MACHINE LEARNING METHODS INSPIRED BY CHALLENGES IN TOTAL SYNTHESIS

Thesis by

Michael Robert Maser

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2022

(Defended Jan. 26, 2022)

© 2022

Michael Robert Maser

ORCID: 0000-0001-7895-7804

All Rights Reserved

To my wife, Victoria

ACKNOWLEDGEMENTS

I am very grateful for my time at the California Institute of Technology. I'd like to thank Caltech for serving as the grounds of my most significant personal and professional development.

Of the people most influential to my career path, I'd like first and foremost to thank my advisor, Prof. Sarah Reisman. Her offering of practical and theoretical insight was certainly of great support in my days of synthesis research. I am especially thankful for Sarah's willingness to allow my transition out of the lab and into the computer science realm. I ultimately felt large control over the directions of our ML research, and this was paramount to my development as an ML Scientist, which is now my career role. I will remember Sarah's lab as the space within which this path began, and I will remain thankful to her for it.

Second, I'd like to thank Prof. Brian Stoltz, my PhD Committee Chair and someone I consider a remarkably reliable mentor and friend. Brian is an incredible teacher; he is a true expert, yet connects intellectually with students of all levels laterally. From my local perspective, this makes Brian an irreplaceable pillar of the Division. Beyond academics, I'll long remember early mornings on the baseball field with Brian and our small group in the Wood Bat Baseball Club and at San Marino Softball.

My other three Committee members are people I learned a great deal from consider a great privilege to call colleagues and friends.

First, Prof. Greg Fu was another remarkable teacher with an amazing walking understanding of molecular structure and reactivity. I consider it a great honor to have conversed with him over topics of our field to know that he influenced my work first hand.

Second, Prof. Jennifer Listgarten of UC Berkeley was my first memorable idol in the ML research world. I was fortunate enough to be in the audience of a research talk she gave at Caltech, and even more fortunate that Jenn responded to a follow-up email from me with interest in collaboration on ML + Chemistry research. I learned a great deal from our conversations on constructing ML research problems and I'm very grateful to her and her group for their time and teachings.

Finally, Prof. Yisong Yue of Computing and Mathematical Sciences has served as something of a second mentor to me at Caltech. He was very influential in the directions of our early ML research out of CMS 273 and he remains a most valuable collaborator on our current work in computer vision and active learning. Like the Listgarten group, I learned a great deal just by observing him and his group process research problems from ML angles; it was invaluable to have them as models, colleagues, and now friends.

Beyond my Committee, I'd like to thank Prof. Pietro Perona and his group for the invaluable insight and discussions on our ML collaborations, Profs. Max Robb, Bob Grubbs, Tom Miller, and their groups for being great friends and resources in CCE, Dr. Scott Virgil for extensive time and effort in helping get our robotic chemistry work up and running, as well as for general effort in maintaining the Catalysis Center, Dr. Dave VanderVelde for efforts in running the NMR lab and valuable insight on various NMR experiments over the years, Dr. Mona Shahgholi for all mass spectrometry experiments and for going above and beyond for time sensitive data as well as precious ritterazine compounds, Beth Marshall and all previous Group Administrators for their tireless and extremely valuable work, all of the staff in the CCE Division Office—particularly Alison

Ross—as well as Caltech Administration in general, and the NSF, NIH, Caltech CCE, Research Corp., and Google, LLC for fellowship and project funding.

At the student/postdoc level, there are countless people to thank. First, thank you to all the people I had the privilege of being project partners with; I enjoyed working with you all immensely. Specifically, I'd like to name Dr. Serim Ryou, Alex Cui, Travis DeLano, Dr. Yasuaki Nakayama, Sara Dibrell, Dr. Caitlin Lacker, Emily Chen, Dr. Uriah Israel, Jennifer Sun, Megan Santhumayor, Julen Lujambio, Arya Mevada, and Yuanzhe Xie. I learned a lot from all of these people, which certainly shaped my understanding of our fields and research processes.

Working with our CMS 273 team (Serim, Alex, and Travis) was perhaps my favorite research experience of my career thus far. I will look back very fondly on our weekly meetings outside of Moore Lab, in the Schlinger conference room, or eventually on Zoom. The three of them were the most delightful project partners; I hope to have the privilege of enjoying research experiences like this one for the remainder of my career.

Beyond research, thank you to all the amazing colleagues and friends I was fortunate enough to interact with over the years. Specifically, I'd like to thank Drs. Luke Hanna, Caitlin Lacker, Suzie Stevenson, Eric Welin, Willie Wolf, Marco Brandstätter, Kelsey Poremba, and Jordan Beck, as well as Nick Fastuca, Nick Hafeman, Ray Turro, Travis DeLano, and Karli Holman for your friendship and for being positive presences in the group(s). I can recall many delightful memories of coffee breaks, lunches, ping pong matches, softball games, Dodger games, and other moments that were much-needed reminders that life outside of lab still existed. Thank you all.

Outside of Caltech, there a few people whom, again, I consider a great privilege to be colleagues and friends with. First, I'd like to thank Dr. Steven Kearnes, formerly of Google Research and now of Relay Therapeutics, and Prof. Connor Coley of MIT. Working with Steve and Connor on the Open Reaction Database was probably the real start of my career in the greater software/ML field. I learned an incredible amount from observing their thought and software practices, which is how I ultimately became able to create functional and sustainable software. ORD meetings with Connor and Steve will last as some of my favorite memories of grad school.

This experience as well as Steve's invitation led me to work with Google Research for a few months in my fourth year, for which I am extremely grateful. I'd specifically like to thank my mentors Dr. James Thompson, Dr. Jin Xu, Brian Lee, and Dr. Arun Narayanaswamy, as well as team members Dr. Wen Torng, Jess Xu, Dr. Jennifer Wei, Dr. Alex Wiltschko, Dr. Patrick Riley, and my supervisor Dr. JW Feng. This experience was a great jumpstart to my career in ML industry, where I hope to spend the foreseeable future.

Before grad school, a number of people and places had great influence on my career. First, I'd like to thank the University of California, Santa Barbara, like Caltech, for serving as the grounds of major personal and professional development for me. I hold very fond memories of my time there as a student/researcher in chemistry. I'd specifically like to thank Prof. Bruce Lipshutz for his extensive mentorship and devotion to undergraduate research, education, and development. It was in Bruce's lab that I first learned to conduct research, and I'm very grateful for his commitment and guidance.

In the Lipshutz Lab, I was mentored by Drs. Dan Lippincott and Roscoe Linstadt. I don't believe I would have pursued a PhD in Chemistry if not for these two. I learned

every foundational skill in synthesis from them, and I am most thankful to them for the years of friendship and time spent mentoring me for nearly zero personal gain on their part. While working with Dan and Roscoe I also met Dr. Carina Jette in the Lipshutz Lab, who went on to complete her PhD with Prof. Brian Stoltz at Caltech. Carina has also remained a great friend and was very supportive in my navigation of the graduate school process. I am extremely thankful to all three of you.

During my time at UCSB, I was fortunate enough to spend a summer doing medicinal chemistry research at Novartis in Emeryville. I am very grateful to my mentors Drs. Ben Taft and Dennis Koester, who have remained great friends and career resources over the years. Before this, my first research experience at UCSB and thus the beginning of my research career was in a biophysics lab that I was invited to by Dr. Rob Geller, my undergraduate physics instructor, whom I'd like to thank for his mentorship and friendship.

In life outside of research/academia, I have been very privileged to know the support, friendship, and family relations of many people. I'd like to thank my parents Shannon and Rob Maser for their support over the years. I'd like to thank my grandparents Phil, Jeanne, Cathy, Wayne, Marilyn, and Martin for the relationships they have maintained. I'd like to thank my sister Lauren, her husband Drew, and their son Avery Helenihi for their warmth and familial nature. Finally, I'd like to thank the extended Perry, Moore, Mitchell, and Maser families for their love and kindness.

Beyond family, I want to thank my best friends from childhood, Nick Truex and Nathan D'Emilio. Despite my general absence in the last several years, they remain unwavering friends and I am very grateful for their relationships and memories. I'd like to thank Scott Yehl, my best friend from UCSB, for many memories as well.

Finally, the very most important person I'd like to thank is my wife, Dr. Victoria McHugh. To be clear, Victoria is *singularly* responsible for whatever is perceived as "my" career trajectory.

We're particularly private, so I won't say much more here. I would, however, like to thank her family—the Mchughs/Salangsangs—for their unfailing support of her and us and everyone else over the years. Thank you, Josephine and Pete, Harrison (and Natalie and Charlotte), Dani, Elliott, Bella, and Nia. I'd also like to thank Drs. Hannah Cone and Jess Everista, Victoria's best friends from grad school; some of our favorite memories from this time in our lives was with you.

One more time, if there's anything to take away from these acknowledgements, it's that there would be no contents in this book if not for Victoria.

ABSTRACT

Synthetic organic chemists face a dearth of challenges in the efficient construction of functional molecules, particularly bioactive compounds. Predictive approaches offer reductions to research timelines and resource costs and allow chemists to devote their expertise where it is most valuable. Promising machine learning (ML) methods have evolved for uncovering patterns in chemical data that are beyond the grasp of expert humans, but a number of grand challenges in molecular ML remain. First, the learning of chemical structure representations rooted in physical first principles has yet to be robustly demonstrated. Second, the practical task of predicting successful "over-the-arrow" reaction conditions remains elusive. Finally, the demonstration of such solutions in the context of complex synthesis has yet to be realized.

Herein, approaches to these grand challenges are developed and described. Inspiration is derived from the successful synthesis of the anticancer marine natural product ritterazine B. Reaction condition prediction is approached first, where a novel graph neural network architecture is developed under a multilabel classification framework. The resulting model is successfully demonstrated on datasets of four high-value reaction types in modern synthesis. Next, 3D-to-1D representation learning is approached by development of a volumetric neural architecture based on inception networks. Such voxel models are demonstrated for the prediction of expensive quantum mechanical properties from space-filled data alone.

The merging of these approaches for reaction condition optimization and utility in complex settings is discussed and forecasted for future works, which are currently underway.

PUBLISHED CONTENT AND CONTRIBUTIONS

Portions of the work described herein were disclosed in the following communications:

1. Dibrell*, S. E.; Maser*, M. R.; Reisman, S. E. SeO_2-Mediated Oxidative Transposition of Pauson–Khand Products. *J. Am. Chem. Soc.* **2020**, *142*, 6483–6487. DOI: 10.1021/jacs.9b13818. Copyright © 2020 American Chemical Society

   M.R.M. participated in the conception of the project, conducted experiments, and participated in the preparation of the supporting data and writing of the manuscript.

2. Ryou*, S.; Maser*, M. R.; Cui*, A. Y.; DeLano, T. J.; Yue, Y.; Reisman, S. E. Graph Neural Networks for the Prediction of Substrate-Specific Organic Reaction Conditions. *arXiv:2007.04275* **2020**, *Preprint*. International Conference on Machine Learning 2020 Workshop on Graph Representation Learning and Beyond. Copyright © 2020 The Authors

   M.R.M. participated in the conception of the project, conducted experiments, and participated in the preparation of the supporting data and writing of the manuscript.

3. Maser*, M. R.; Ryou*, S.; Cui*, A. Y.; DeLano, T. J.; Yue, Y.; Reisman, S. E. Multilabel Classification Models for the Prediction of Cross-Coupling Reaction Conditions. *J. Chem. Inf. Model.* **2021**, *61*, 156–166. DOI: 10.1021/acs.jcim.0c01234. Copyright © 2021 American Chemical Society

   M.R.M. participated in the conception of the project, conducted experiments, and participated in the preparation of the supporting data and writing of the manuscript.

4. Nakayama*, Y.; Maser*, M. R.; Okita, T.; Dubrovskiy, A. V.; Campbell, T. L.; Reisman, S. E. Total synthesis of Ritterazine B. *J. Am. Chem. Soc.* **2021**, *143*, 4187–4192. DOI: 10.1021/jacs.1c01372. Copyright © 2021 American Chemical Society

M.R.M. contributed to the synthesis development, conducted experiments, and participated in the preparation of the supporting data and writing of the manuscript.

5. Reisman, S. E.; Maser, M. R.; Nakayama, Y.; Okita, T.; Dubrovskiy, A. V.; Campbell, T. L. Synthesis of Ritterazine B and Analogs for the Treatment of Cancer. *U.S. Provisional Patent Application*. Copyright © 2021 The Authors

   M.R.M. contributed to the synthesis development, conducted experiments, and participated in the preparation of the supporting data and patent documents.

6. Maser, M. R.; Reisman, S. E. 3D Computer Vision Models Predict DFT-Level HOMO-LUMO Gap Energies from Force-Field-Optimized Molecular Geometries. *ChemRxiv* **2021**, *Preprint*. DOI: 10.33774/chemrxiv-2021-11r61. Copyright © 2021 The Authors

   M.R.M. participated in the conception of the project, conducted experiments, and participated in the preparation of the supporting data and writing of the manuscript.

7. Kearnes, S.; Maser, M.; Wleklinski, M.; Kast, A.; Doyle, A.; Dreher, S.; Hawkins, J.; Jensen, K.; Coley, C. The Open Reaction Database. *J. Am. Chem. Soc.* **2021**, *143*, 18820. DOI: 10.1021/jacs.1c09820. Copyright © 2021 American Chemical Society

   M.R.M. contributed to the software development and participated in the writing of the manuscript.

TABLE OF CONTENTS

CHAPTER 1 ***1***

SeO_2-Mediated Oxidative Transposition of Pauson–Khand Products

1.1 INTRODUCTION....1

1.2 CONDITION OPTIMIZATION....2

1.3 INVESTIGATION OF SCOPE....4

1.4 MECHANISTIC STUDIES....6

1.5 SYNTHETIC UTILITY DEMONSTRATION....11

1.6 INVESTIGATION OF ADDITIONAL RING SYSTEMS....12

1.7 CONCLUSION....14

1.8 EXPERIMENTAL SECTION....15

1.8.1 Materials and Methods....15

1.8.2 General Synthetic Schemes and Procedures....17

1.8.3 Substrate Preparation....22

1.8.4 SeO_2-Mediated Oxidative Transpositions....55

1.8.5 Stereochemical Analysis....77

1.8.6 Mechanistic Investigation....93

1.8.7 Product Diversification....108

1.9 REFERENCES....117

Appendix 1 ***122***

Spectra Relevant to Chapter 1

CHAPTER 2 ***274***

Total Synthesis of Ritterazine B and Analogs for the Treatment of Cancer

2.1 INTRODUCTION.....274

2.2 RETROSYNTHETIC ANALYSIS 275

2.3 FORWARD SYNTHESIS.....277

2.3.1 Accessing the Divergent Intermediate 277

2.3.2 Initial Attempts for Alkyne Installation.....278

2.3.3 Second Generation Approach 279

2.3.3.1 Preparing the Propargyl Bromides 279

2.3.3.2 Preparing the Western Coupling Partner.....280

2.3.3.3 Preparing the Eastern Coupling Partner.....284

2.3.3.4 Completion of the Synthesis 289

2.4 ADDITIONAL RITTERAZINES.....291

2.4.1 Identifying Promising Targets 291

2.4.2 Synthesis of Ritterazine K 292

2.4.3 Synthesis of 7′-Desoxy-Ritterazine B 294

2.5 BIOLOGICAL EVALUATION 293

2.6 CONCLUSION.....296

2.7 EXPERIMENTAL SECTION.....297

2.7.1 Materials and Methods.....297

2.7.2 Synthetic Procedures.....299

2.7.2.1 Ritterazine B 299

2.7.2.2 Ritterazine K 397

2.7.2.3 7′-Desoxy-Ritterazine B.....409

2.8 REFERENCES 416

APPENDIX 2 ***421***

Spectra Relevant to Chapter 2

CHAPTER 3 ***592***

Multilabel Classification Models for the Prediction of Substrate-Specific Organic Reaction Conditions

3.1 INTRODUCTION 592

3.2 INITIAL STUDIES 594

3.3 REVISED APPROACH AND METHODS 599

3.3.1 Data Acquisition and Pre-Processing 599

3.3.2 Problem Setup 603

3.3.3 Model Construction 604

3.3.3.1 Gradient-Boosting Machines (GBMs) 605

3.3.3.2 Graph Convolutional Networks (GCNs) 608

3.4 RESULTS AND DISCUSSION 611

3.4.1 Model Performance 611

3.4.2 Interpretability 622

3.4.2.1 Tree Methods 622

3.4.2.2 Deep Learning Methods 624

3.4.3 Yield Analysis 628

3.5 CONCLUSION AND OUTLOOK 630

3.6 EXPERIMENTAL SECTION 632

3.6.1 Data Preparation and Reaction Dictionaries 632

3.6.2 Computational Details and Hyperparameters 651

3.6.2.1 Gradient-Boosting Machines (GBMs) 651

3.6.2.1.1 Binary Relevance Method (BM) 652

3.6.2.1.2 Classifier Trellises (CT) 652

3.6.2.2 Graph Convolutional Networks (GCNs) 657

3.6.3 Adversarial Controls 659

3.6.3.1 Binary Relevance GBMs (BM-GBMs) 660

3.6.3.2 Classifier Trellis GBMs (CT-GBMs) 664
3.6.4 Interpretability 667
3.6.4.1 GBM Feature Importances (FIs).....667
3.6.4.2 AR-GCN Attention Visualizations.....684
3.6.5 Exploratory Data Analysis (EDA) 687
3.6.5.1 Suzuki Dataset 689
3.6.5.2 C–N Dataset.....692
3.6.5.3 Negishi Dataset.....695
3.6.5.4 PKR Dataset 698

3.7 REFERENCES 701

APPENDIX 3 ***709***

Robotic High-Throughput Experimentation and Active Learning for Automated Reaction Optimization

A3.1 INTRODUCTION 709

A3.2 INITIAL STUDIES.....710

A3.3 TROUBLESHOOTING AND HTE OPTIMIZATION.....712

A3.4 FOLLOW-UP CYCLOBUTANE STUDIES.....716

A3.5 CONCLUSION AND FUTURE OUTLOOK 717

A3.6 REFERENCES 719

A3.7 NMR SPECTRA.....721

APPENDIX 4 ***744***

The Open Reaction Database

A4.1 BACKGROUND.....744

A4.2 INTRODUCTION 745

A4.3 SCHEMA 747

A4.4 INTERFACES 751

A4.5 DATA SET EXAMPLES 751

A4.6 DOWNSTREAM USE CASES 754

A4.7 DISCUSSION 755

A4.8 OUTLOOK 757

A4.9 REFERENCES 759

CHAPTER 4 ***763***

3D Computer Vision Architectures for Molecular Shape Recognition

4.1 INTRODUCTION 763

4.2 APPROACH AND METHODS 765

4.2.1 Data 765

4.2.2 Augmentation 767

4.2.3 Deep Learning Models 767

4.3 RESULTS AND ANALYSIS 771

4.3.1 Architecture Screen – 100k Examples 771

4.3.2 Best Architecture Scaleup – 1M Examples 774

4.3.3 Baseline Methods 775

4.4 DEMONSTRATION OF 3D SHAPE LEARNING 778

4.5 DISCUSSION AND OUTLOOK 781

4.6 CONCLUSION 784

4.7 EXPERIMENTAL SECTION 785

4.7.1 Computational Details and Hyperparameters 785

4.7.2 Model Architectures and Implementations 785

4.7.2.1 VoxNet 786
4.7.2.2 V-CNN I.....787
4.7.2.3 V-CNN II.....789
4.7.2.4 ShapeEncoder-d 792
4.7.2.5 QMCVNet.....794
4.7.3 Voxelization Parameter Screen 797
4.7.4 Modeling Results: Scaleup and Extended Training 803
4.7.5 Additional Sample Code 805

4.8 REFERENCES 815

ABOUT THE AUTHOR.....820

LIST OF ABBREVIATIONS

$[\alpha]_D$	angle of optical rotation of plane-polarized light
Å	angstrom(s)
Ac	acetyl
acac	acetylacetonate
alk	alkyl
^{t}Am	*tert*-amyl
APCI	atmospheric pressure chemical ionization
app	apparent
aq	aqueous
Ar	aryl
atm	atmosphere(s)
BHT	2,6-di-*tert*-butyl-4-methylphenol ("butylated hydroxytoluene")
BiOX	bi(oxazoline)
BINAP	2,2'-bis(diphenylphosphino)-1,1'-binaphthyl
BINOL	1,1'-bi(2-naphthol)
Bn	benzyl
Boc	*tert*-butoxycarbonyl
BOX	bis(oxazoline)
bp	boiling point
br	broad
Bu	butyl
^{i}Bu	*iso*-butyl

^{n}Bu	butyl or *norm*-butyl
^{s}Bu	*sec*-butyl
^{t}Bu	*tert*-butyl
Bz	benzoyl
c	concentration of sample for measurement of optical rotation
^{13}C	carbon-13 isotope
/C	supported on activated carbon charcoal
°C	degrees Celsius
calc'd	calculated
CAM	cerium ammonium molybdate
cat.	catalyst
Cbz	benzyloxycarbonyl
cf.	consult or compare to (Latin: *confer*)
cis	on the same side
cm^{-1}	wavenumber(s)
cod	1,5-cyclooctadiene
conc.	concentrated
conv.	conversion
Cp	cyclopentadienyl
Cy	cyclohexyl
Cyp	cyclopentyl
Δ	heat or difference
δ	chemical shift in ppm

d	doublet or dropout
d	deutero or dextrorotatory
D	deuterium
dba	dibenzylideneacetone
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCE	1,2-dichloroethane
DIPEA	*N*,*N*-diisopropylethylamine
DIBAL	diisobutylaluminum hydride
DFT	density functional theory
DMA	*N*,*N*-dimethylacetamide
DMAP	4-(dimethylamino)pyridine
DME	1,2-dimethoxyethane
DMF	*N*,*N*-dimethylformamide
DMPU	*N*,*N*'-dimethylpropylene urea
DMSO	dimethylsulfoxide
dppb	1,4-bis(diphenylphosphino)butane
dppbz	1,2-bis(diphenylphosphino)benzene
dppf	1,1'-bis(diphenylphosphino)ferrocene
dppe	1,2-bis(diphenylphosphino)ethane
dr	diastereomeric ratio
dtbpy	4,4'-di-*tert*-butyl-2,2'-bipyridine
E	methyl carboxylate (CO_2CH_3)
E^+	electrophile

E	trans (entgegen) olefin geometry
EDCI	*N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride
ee	enantiomeric excess
e.g.	for example (Latin: *exempli gratia*)
EI	electron impact
epi	epimeric
equiv	equivalent(s)
ESI	electrospray ionization
Et	ethyl
et al.	and others (Latin: *et alii*)
FAB	fast atom bombardment
g	gram(s)
GC	gas chromatography
h	hour(s)
^{1}H	proton
[H]	reduction
hex	hexyl
HMDS	hexamethyldisilazane
HMPA	hexamethylphosphoramide
$h\nu$	light
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
Hz	hertz

IC_{50}	half maximal inhibitory concentration (50%)
i.e.	that is (Latin: *id est*)
in situ	in the reaction mixture
IPA	isopropanol
IR	infrared spectroscopy
J	coupling constant
k	rate constant
kcal	kilocalorie(s)
kg	kilogram(s)
L	liter or neutral ligand
l	levorotatory
LA	Lewis acid
LC_{50}	half maximal lethal concentration (50%)
LC/MS	liquid chromatography–mass spectrometry
LDA	lithium diisopropylamide
LED	light-emitting diode
m	multiplet or meter(s)
M	molar or molecular ion
m	meta
μ	micro
m-CPBA	*meta*-chloroperbenzoic acid
Me	methyl
mg	milligram(s)

MHz	megahertz
min	minute(s)
mL	milliliter(s)
MM	mixed method
mol	mole(s)
MOM	methoxymethyl
mp	melting point
Ms	methanesulfonyl (mesyl)
MS	molecular sieves or mass spectrometry
m/*z*	mass-to-charge ratio
naphth	naphthyl
NBS	*N*-bromosuccinimide
ND	not determined
NHC	*N*-heterocyclic carbene
NHP	*N*-hydroxyphthalimide
nm	nanometer(s)
NMP	*N*-methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
o	ortho
[O]	oxidation
p	para
Ph	phenyl
pH	hydrogen ion concentration in aqueous solution

phen	1,10-phenanthroline
PHOX	phosphinooxazoline
pin	pinacol
Piv	pivaloyl
pK_a	acid dissociation constant
Pr	propyl
^{i}Pr	isopropyl
^{n}Pr	propyl or *norm*-propyl
py	pyridine
PyBOX	pyridine-bis(oxazoline)
PyOx	pyridine-oxazoline
q	quartet
quant.	quantitative
R	alkyl group
R_L	large group
R	rectus
RCM	ring-closing metathesis
recry.	recrystallization
ref	reference
R_f	retention factor
rt	room temperature
s	singlet or seconds
S	sinister

sat.	saturated
SET	single-electron transfer
SFC	supercritical fluid chromatography
SM	starting material
t	triplet
TBAB	tetra-*n*-butylammonium bromide
TBAF	tetra-*n*-butylammonium fluoride
TBAI	tetra-*n*-butylammonium iodide
TBAT	tetra-*n*-butylammonium difluorotriphenylsilicate
TBS	*tert*-butyldimethylsilyl
TDAE	tetrakis(dimethylamino)ethylene
Tf	trifluoromethanesulfonyl
TFA	trifluoroacetic acid
temp	temperature
terpy	2,2′:6′,2″-terpyridine
THF	tetrahydrofuran
TIPS	triisopropylsilyl
TLC	thin layer chromatography
TMEDA	*N*,*N*,*N*’,*N*’-tetramethylethylenediamine
TMS	trimethylsilyl
TOF	time-of-flight
tol	toluene
trans	on the opposite side

Ts	*para*-toluenesulfonyl (tosyl)
UV	ultraviolet
vide infra	see below
vide supra	see above
v/v	volume per volume
w/v	weight per volume
X	anionic ligand or halide
xs	excess
Y	anionic ligand or halide
Z	cis (zusammen) olefin geometry

Chapter 1

SeO_2-Mediated Oxidative Transposition of Pauson–Khand Products[1]

1.1 INTRODUCTION

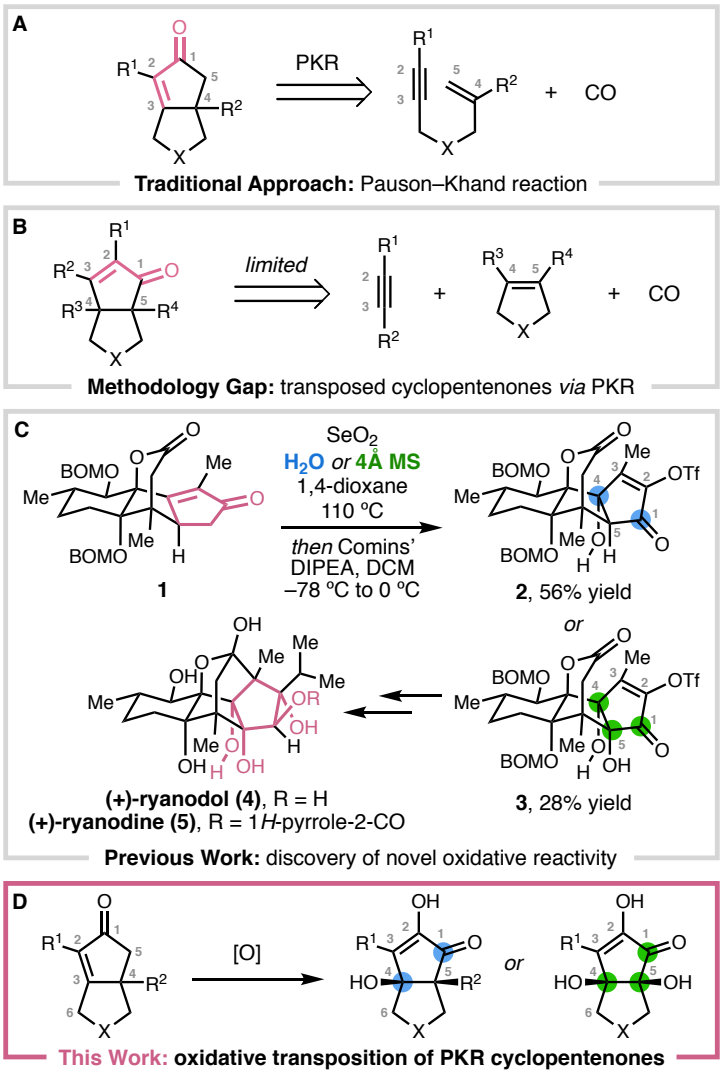
The Pauson–Khand reaction is a powerful method to prepare bicyclic cyclopentenones,[1] structural features found in many natural products. The intramolecular PKR is well-developed and affords 3,4-fused bicycles from the corresponding tethered enynes (**Figure 1.1A**).[2a] In contrast, the intermolecular PKR, which is required to access the isomeric 4,5-fused bicycles, is less developed and has severe limitations in scope (**Figure 1.1B**).[2b,c] A method to rapidly access 4,5-fused bicyclic enones, which are inaccessible directly by the PKR, would provide valuable synthetic building blocks.

As part of our recent syntheses of the complex diterpenes (+)-ryanodol (**4**) and (+)-ryanodine (**5**), our group discovered an unusual oxidative transformation mediated by selenium dioxide (SeO_2).[3] Treatment of PKR product **1** with SeO_2 under either aqueous or anhydrous conditions provided highly oxidized compounds **2** and **3**, respectively (**Figure 1.1C**). This serendipitous discovery streamlined access to 4,5-fused cyclopentenone **3**, a key intermediate in the synthesis of **4** and **5**. Although these di- and trioxidations were modest yielding, the ability to construct functional group relationships elusive via canonical PKRs led us to consider their wider application in synthesis (**Figure 1.1D**). We

[1] Portions of this chapter were adapted from the following communications: Dibrell*, S. E.; Maser*, M. R.; Reisman, S. E. SeO_2-Mediated Oxidative Transposition of Pauson–Khand Products. *J. Am. Chem. Soc.* **2020**, *142*, 6483–6487. DOI: 10.1021/jacs.9b13818. Copyright © 2020 American Chemical Society. The research discussed in this chapter was completed in collaboration with Sara E. Dibrell, a current graduate student in the Reisman Lab.

thus initiated an investigation of the SeO_2-mediated oxidative transposition of 3,4-fused PKR products to demonstrate the ability of this method to prepare highly substituted bicyclopentenones.

Figure 1.1 *Motivation for reaction development*

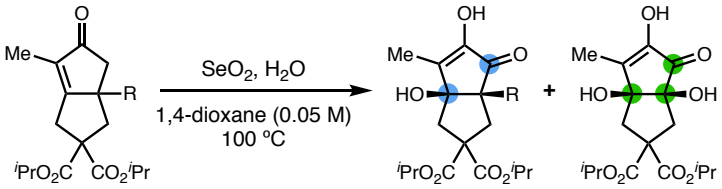


1.2 CONDITION OPTIMIZATION

Our studies began with 3,4-fused bicycle **6a**, which is readily accessible from commercially available diisopropyl malonate (see Experimental Section). Substrate **6a** was selected for these initial studies because the C4-γ-quaternary substitution prevents

trioxidation, simplifying reaction analysis for the dioxidation process. Although **6a** successfully underwent the desired dioxidation under our previously reported conditions,[3] the yield of **8a** was significantly lower than that observed in the ryanodane system (**Table 1.1**, entry 1). Attempts to push formation of **8a** by increasing the temperature and/or equivs of SeO_2 led to messy reaction profiles and overoxidation (see Experimental Section). Fortunately, increasing the amount of water to 100 equiv led to a significant boost in conversion, providing **8a** in 71% yield (entry 3). Under these optimized conditions, the primary mass balance was unreacted starting material (**6a**) and minor amounts of over-oxidation to the C3-tertiary alcohol (not shown).

Table 1.1 *Optimization of conditions*



6a, R = Me
7a, R = H

8a, R = Me
9a, R = H

10a

Entry	R	SeO_2 (equiv)	H_2O (equiv)	Time (h)	**8a** *or* **9a** (%)[a]	**10a** (%)[a]
1	Me	10	10	24	13	—
2	Me	10	0	3	11	—
3	Me	10	100	38	71[b]	—
4	Me	3	30	24	32	—
5	Me	1.5	15	24	29	—
6	Me	3	100	48	62	—
7	Me	1.5	100	120	42	—
8	H	10	100	4	64[b]	0
9	H	10	100	46	0	14
10	H	10	10	1	0	15
11	H	10	0	2	0	21[b]
12	H	3	0	3	0	10

[a]Yield determined by ^{1}H NMR versus an internal standard. See Experimental Section for additional information. [b]Isolated yield.

In addition to improving the yield of **8a**, increasing the [H_2O] also drastically improved the reaction profile. When the equivalents of SeO_2 were reduced but a 10:1 ratio of water to SeO_2 maintained, the yield of **8a** dropped (entries 4 and 5). However, increasing the equivalents of water with reduced amounts of SeO_2 improved the yield, affording **8a** in 42% yield with just 1.5 equiv SeO_2 (entry 7).[4] When substrate **7a**, which lacks the C4-γ-quaternary center, was subjected to the optimized conditions, dioxidation product **9a** was formed in comparable yield to that of **8a** (entry 8). Interestingly, subjection of **7a** to these same conditions for extended reaction times led to the formation of trioxidation product **10a**, which had previously only been observed under rigorously anhydrous conditions (entry 9).[3a] Motivated by this observation, we turned our attention to investigating the formation of **10a**.

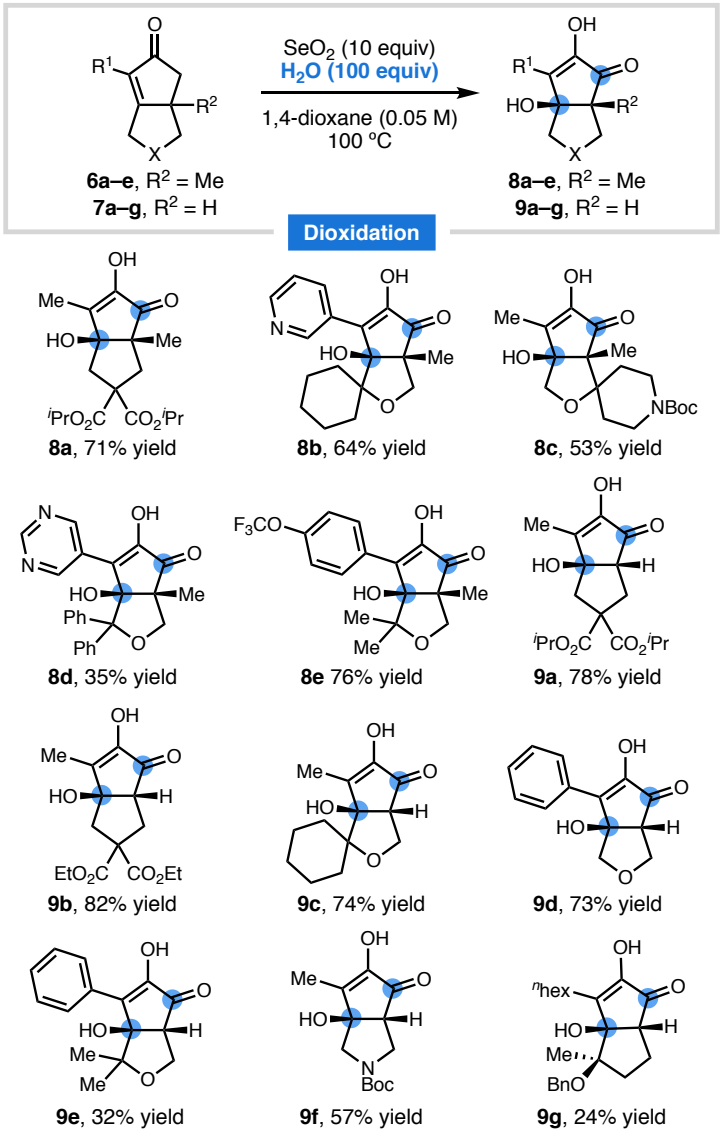
Consistent with our prior studies,[3a] reducing the equivalents of water in the SeO_2-mediated reaction provided trioxidation product **10a** in faster rates with cleaner reaction profiles (entry 10 vs. entry 9). The complete exclusion of water provided **10a** in 21% yield with no trace of undesired **9a** (entry 11). Reducing the amount of SeO_2 decreased the yield of **10a** (entry 12), and thus the conditions of entry 11 were selected as optimal for the trioxidation reaction. Under these optimized conditions, the mass balance consisted of unreacted starting material (**7a**) and a complex mixture of minor reaction byproducts.

1.3 INVESTIGATION OF SCOPE

With optimized conditions in hand, we investigated the scope of both oxidative transformations. A variety of carbo- and heterocyclic 3,4-fused cyclopentenones were prepared via the PKR and subjected to the dioxidation conditions (**Table 1.2**). Most substrates underwent the oxidative transposition in good yields, providing the oxidized 4,5-

fused bicycles as single diastereomers.[5] The reaction is tolerant of heterocycles (**8b–e** and **9c–f**), C6-γ-quaternary centers (**8b**, **8d–e**, **9c**, **9e**, and **9g**), and protected amines (**8c** and **9f**). Triflation of **8e** afforded the enol triflate as a crystalline solid, and single crystal X-ray diffraction confirmed the *syn* ring fusion (see Experimental Section).[6]

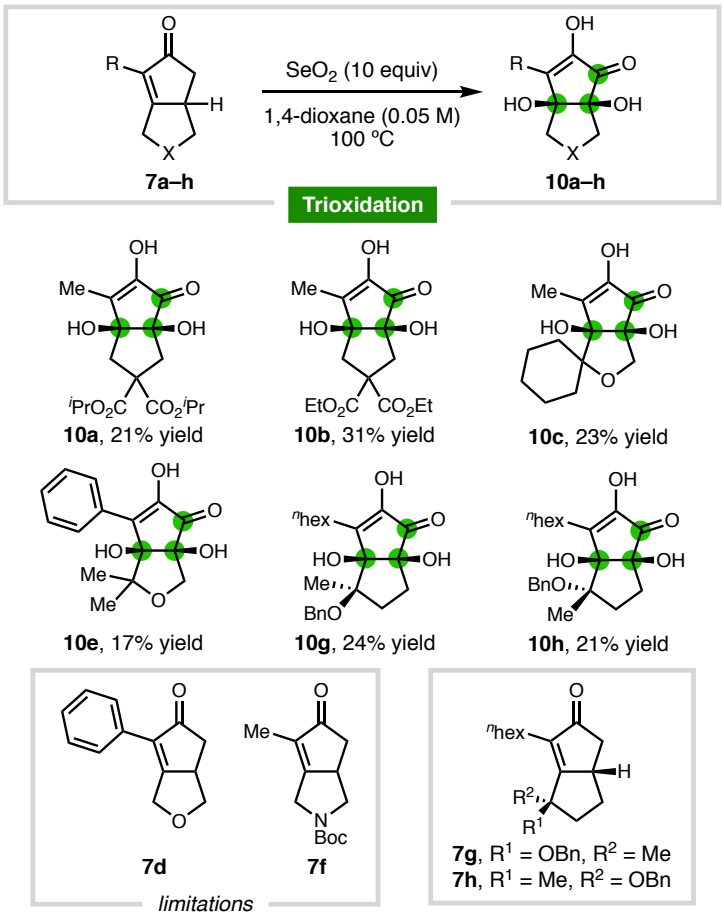
Table 1.2 *Scope of dioxidation*



We next explored the scope of the trioxidation reaction (**Table 1.3**). While competitive decomposition of the trioxidation products (**10a–h**) generally resulted in

diminished yields relative to those of the dioxidations, the desired products could be isolated in approximately 20–30% yield, similar to that observed in the trioxidation of **1**. Substrates containing unsubstituted, saturated heterocyclic rings, such as **7d** and **7f**, marked a limitation of this method. Trioxidation products **10g** and **10h** were formed as single diastereomers from cyclopentenones **7g** and **7h**, respectively.

Table 1.3 *Scope of trioxidation*



1.4 MECHANISTIC STUDIES

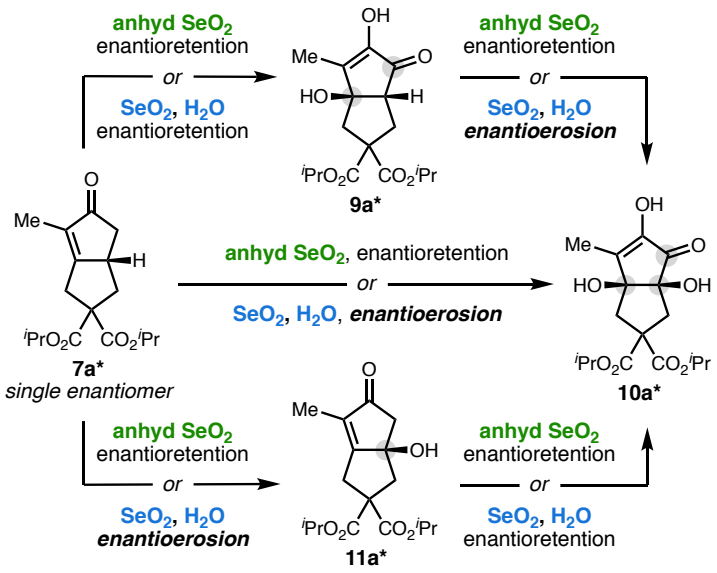
These results led us to investigate potentially operative reaction pathways. To this end, we separated enantiomers of model substrate **7a** via chiral HPLC (**7a***, **Table 1.4**).[7] Treatment of **7a*** with SeO_2 in the presence of water furnished dioxidation product **9a***

with retention of enantiomeric excess (ee, entry 1). Under the same conditions, trace yield of trioxidation product **10a*** was formed after extended reaction times; however, formation of **10a*** under these conditions proceeded with significant enantioerosion (entry 3). In contrast, treatment of **7a*** with anhydrous SeO_2 provided **10a*** with enantioretention (entry 4). Treatment of dioxidation product **9a*** with H_2O/SeO_2 resulted in the slow formation of **10a*** with substantial loss of ee (entry 5), whereas subjection of **9a*** to anhydrous SeO_2 resulted in the fast conversion to **10a*** with complete retention of ee (entry 6).

Close monitoring of the above reactions revealed the presence of a fleeting species with a mass consistent with mono-oxidation. This species was isolated and characterized as tertiary alcohol **11a***, which presumably arises via an allylic C–H oxidation pathway (**Table 1.4**, entries 7 and 8).[8] Subjection of **11a*** to both standard anhydrous and aqueous conditions resulted in conversion to trioxidation product **10a*** with retention of enantiopurity (entries 9 and 10). Taken together, these results demonstrate that under optimal conditions both **9a*** and **10a*** can be formed with enantioretention and that both **9a*** and **11a*** are viable intermediates to **10a***.

While the exact mechanism for enantioerosion is currently unclear, control experiments have confirmed that **7a*** and **10a*** are configurationally stable under both the dioxidation and trioxidation reaction conditions. On the other hand, **9a*** undergoes slow enantioerosion in the presence of both SeO_2 and H_2O (see Experimental Section). Whereas the racemization of **9a*** could lead to **10a*** with enantioerosion, a second possibility is that in the presence of SeO_2 and H_2O, there is a competing oxidation pathway for **7a*** or **9a*** that involves an achiral cyclopentadienone intermediate.

Table 1.4 *Stereochemical analysis of oxidations*[a]



Entry	Reaction	Conditions	ee (%)	Yield (%)	Time (h)
1	**7a*** → **9a***	SeO_2, H_2O	>99 → >99	59	24
2	**7a*** → **9a***	anhyd SeO_2	>99 → >99	4	1
3	**7a*** → **10a***	SeO_2, H_2O	>99 → 60	3	24
4	**7a*** → **10a***	anhyd SeO_2	>99 → >99	15	1
5	**9a*** → **10a***	SeO_2, H_2O	>99 → 46	12	46
6	**9a*** → **10a***	anhyd SeO_2	>99 → >99	24	2
7	**7a*** → **11a***	SeO_2, H_2O	>99 → 36	2	1
8	**7a*** → **11a***	anhyd SeO_2	>99 → 92	10	1
9	**11a*** → **10a***	SeO_2, H_2O	92 → 92	5	60
10	**11a*** → **10a***	anhyd SeO_2	92 → 92	1	3

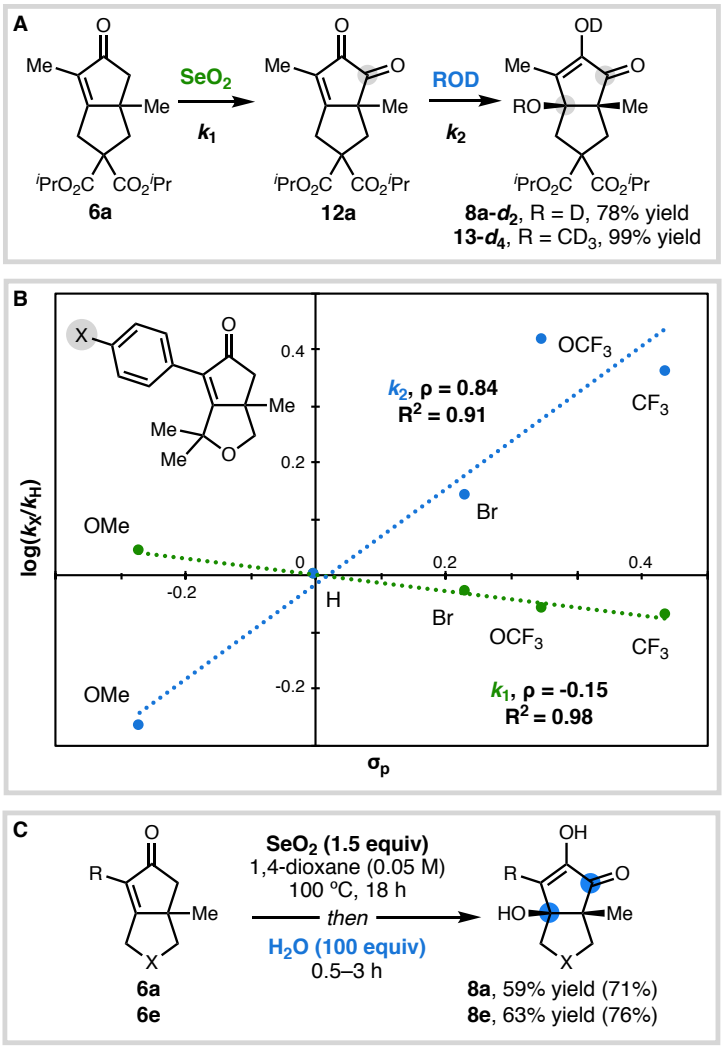
[a]Reactions performed under anhydrous [SeO_2 (10 equiv), 1,4-dioxane (0.05 M), 100 °C] or aqueous [with H_2O (100 equiv)] conditions.

In order to better understand the dioxidation reaction, we searched for potential intermediates in the oxidation of C4-γ-quaternary substituted substrate **6a**. Under the standard dioxidation conditions (H_2O/SeO_2), no species *en route* to **8a** could be observed. However, with the exclusion of water, **6a** was observed by ^1H and ^{13}C NMR to convert to the product of a Riley oxidation (**12a**, **Figure 1.2A**).[9] Following *in situ* formation of ene-diketone **12a**, D_2O was added to determine if **12a** was a viable intermediate to **8a**. Indeed,

deuterated dioxidation product **8a-*d*$_2$** was formed. Addition of deuterated methanol instead of D_2O provided the analogous compound **13-*d*$_4$** in excellent yield.[10]

To further support the hypothesis that conjugate addition to an ene-diketone intermediate leads to the dioxidation products, we studied each step individually (see Experimental Section). Correlation of the measured reaction rates with Hammett parameters[11] (**Figure 1.2B**) indicated that the rate of the conjugate addition (k_2, blue) is more dependent on the substrate electronics than the rate of the Riley oxidation (k_1, green). Intuitively, these studies revealed that the oxidation step proceeded faster for electron-rich substrates, whereas conjugate addition proceeded faster for electron-deficient systems. Given that excess water depressed the rate of the Riley oxidation (of which H_2O is a byproduct), we postulated that a protocol separating the two steps could allow faster access to dioxidation products while maintaining clean reaction profiles. Indeed, this alternative protocol proceeds with only 1.5 equiv SeO_2 and can afford dioxidation products in yields comparable to those resulting from the standard aqueous conditions (**Figure 1.2C**).[12]

Figure 1.2 *Additional experiments*



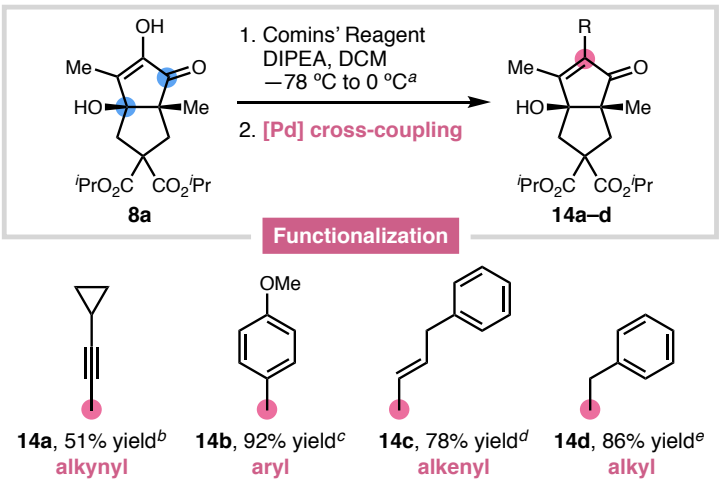
(a) Formation of **12a** under anhydrous conditions [SeO_2 (10 equiv), 1,4-dioxane-d_8 (0.1 M), 100 °C] followed by trapping with deuterated water or methanol (100 equiv). Yield of **8a-*d*$_2$** and **13-*d*$_4$** from **12a** determined by ^{1}H NMR versus an internal standard. (b, green) Hammett plot for the oxidation of enones [X = OMe (**6f**), H (**6g**), Br (**6h**), OCF_3 (**6e**), CF_3 (**6i**)] under anhydrous conditions [SeO_2 (10 equiv), 1,4-dioxane-d_8 (0.1 M), 100 °C]. (b, blue) Hammett plot for the addition of deuterated water to ene-diketones (**12e–i-*d*$_2$**) [D_2O (100 equiv), 1,4-dioxane-d_8 (0.1 M), 100 °C]. (c) Yields determined by ^{1}H

NMR versus an internal standard. In parentheses are isolated yields obtained under standard dioxidation conditions.

1.5 SYNTHETIC UTILITY DEMONSTRATION

At this stage, we turned our attention to investigating the synthetic utility of the multioxidation products. Modern cross-coupling methods offer many opportunities for further elaboration of the α-hydroxyenone functionality. To this end, the enol triflate of **8a** was prepared in 90% yield and treated with a variety of palladium catalysts and coupling partners (**Table 1.5**). Although **8a-OTf** is a challenging cross-coupling substrate, Sonogashira and Suzuki reactions produced highly functionalized products **14a-d** in good to excellent yields.[13] Reaction with alkynyl, aryl, alkenyl and alkyl partners demonstrates the ability to access diverse, fully substituted cyclopentenones from simple PKR products.

Table 1.5 *Functionalization of dioxidation product*



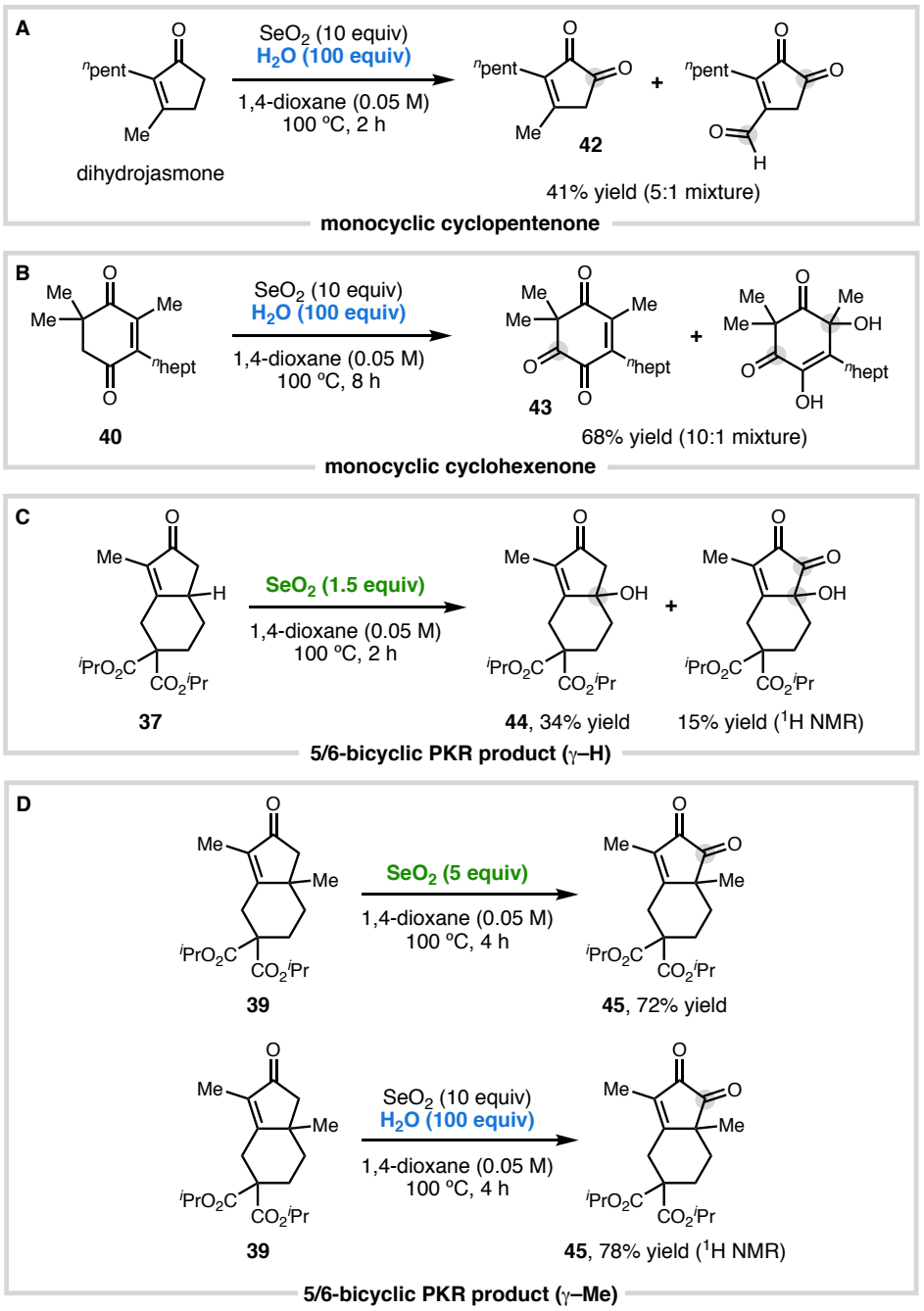
[a]90% isolated yield. [b]ethynylcyclopropane, $Pd(PPh_3)_2Cl_2$, CuI, Et_3N, 50 °C. [c](4-methoxyphenyl)boronic acid, $Pd(OAc)_2$, PCy_3, KF, THF. [d]*(E)*-(3-phenylprop-1-en-1-yl)boronic acid, $Pd(OAc)_2$, PCy_3, KF, THF [e]potassium benzyltrifluoroborate, $Pd(dppf)Cl_2$, Cs_2CO_3, THF/H_2O, 80 °C.

1.6 INVESTIGATION OF ADDITIONAL RING SYSTEMS

Preliminary efforts to expand the scope of the SeO_2-mediated multioxidations to other, non-5/5-fused enones have not been successful (**Figure 1.3**). While the initial Riley oxidation to the ene-diketone is observed, these systems do not undergo spontaneous addition of water. The monocyclic cyclopentenone natural product dihydrojasmone afforded a 5:1 mixture of Riley oxidation product **42** and a byproduct of further oxidation in 41% yield; no hydration products were observed (**Figure 1.3A**). Ketoisophorone derivative **39,** a monocyclic cyclohexanone, underwent oxidation to the trione **43** in 68% yield, with a minor amount of oxy-Michael product analogous to the dioxidations being observed (**Figure 1.3B**).

A bicyclic cyclopentenone (**37**) with a 6/5 instead of 5/5 ring system gave a roughly 2:1 mixture of allylic oxidation product **44** and the further oxidized diketol, which was only observable by NMR and could not be isolated by chromatography (**Figure 1.3C**). Finally, the C4-γ-quaternary analog of **37** gave high yields of ene-diketone **45** under both aqueous and anhydrous conditions, demonstrating its reluctance to further oxygen incorporation (**Figure 1.3D**).

Figure 1.3 *Multi-oxidation attempts with non-bicyclopentenone scaffolds*



Altogether, the difference in reactivity between these alternate enone systems and the substrates explored previously indicates a special proclivity for bicyclopentenones to undergo the oxy-Michael addition. Given the observation that the multioxidation reactions

could proceed through ene-diketone intermediates such as **12**, we hypothesize that these compounds are especially strained given the planar, nearly all sp^2-hybridized diketocyclopentene. It appears that this motif is particularly susceptible to conjugate addition compared to lesser strained systems (cf **Figure 1.3C-D**), perhaps due to the resulting relief of such strain by formation of a new sp^3 center at the ring junction.

1.7 CONCLUSION

In conclusion, SeO_2-mediated oxidative transpositions of PKR products have been developed to access highly oxidized 4,5-fused bicyclic enones. The reactions occur under aqueous or anhydrous conditions to afford di- or trioxidized products, respectively. Investigation of the mechanism identified potential intermediates and elucidated the reactivity dependence on water concentration. Notably, the di- and trioxidations enable the construction of transposed and highly functionalized PKR products pertinent to the synthesis of complex targets.

1.8 EXPERIMENTAL SECTION

1.8.1 MATERIALS AND METHODS

Unless otherwise stated, reactions were performed with freshly dried solvents utilizing standard Schlenk techniques. Glassware was oven-dried at 120 °C for a minimum of four hours or flame-dried utilizing a Bunsen burner under high vacuum. THF, DCM, MeCN, PhH, and PhMe were dried by passing through activated alumina columns. MeOH (HPLC grade) was purchased from Fisher Scientific. 1,4-dioxane, anhydrous ≥99.9%, was purchased from Millipore Sigma. DCE, Et_3N, *i*-Pr_2NH, DIPEA, Pyr, and 2,6-lutidine were distilled from calcium hydride prior to use and stored under N_2 or Ar. Commercial reagents were used directly as supplied from commercial sources and without further purification unless otherwise specified. All reactions were monitored by thin layer chromatography using EMD/Merck silica gel 60 F254 pre-coated plates (0.25 mm) and were visualized by UV (254 nm) and $KMnO_4$, *p*-anisaldehyde, iodine, or CAM staining. Flash column chromatography was performed as described by Still et al.[14] using silica gel (SiliaFlash® P60, particle size 40-63 microns [230 to 400 mesh]) purchased from Silicycle. ^1H and ^{13}C NMR spectra were recorded on a Bruker Advance III HD with Prodigy Cryoprobe (at 400 MHz and 101 MHz, respectively) or Varian Inova 500 (at 500 MHz and 126 MHz, respectively) and are reported relative to internal $CDCl_3$ (^1H, δ = 7.26), CD_3OD (^1H, δ = 3.31), 1,4-dioxane-d_8 (^1H, δ = 3.53), $(CH_3)_2SO$ (^1H, δ = 2.50), or pyridine-d_4 (^1H, δ = 8.74) and $CDCl_3$ (^{13}C, δ = 77.16), CD_3OD (^{13}C, δ = 49.0), 1,4-dioxane-d_8 (^{13}C, δ = 66.66), $(CH_3)_2SO$ (^{13}C, δ = 39.51), or pyridine-d_4 (^{13}C, δ = 150.35). Data for ^1H NMR spectra are reported as follows: chemical shift (δ ppm) (multiplicity, coupling constant (Hz), integration). Multiplicity and qualifier abbreviations are as follows: s = singlet, d = doublet,

t = triplet, q = quartet, p = pentet, hept = heptet, m = multiplet. IR spectra were recorded on a Perkin Elmer Paragon 1000 spectrometer and are reported in frequency of absorption (cm^{-1}). Analytical chiral SFC was performed with a Mettler SFC supercritical CO_2 analytical chromatography system (CO_2 = 1450 psi, column temperature = 40 °C) with a Chiralcel OD-H column (4.6 mm x 25 cm). Preparative and analytical chiral HPLC was performed with an Agilent 1100 Series HPLC with a Chiralpak IH column (4.6 mm x 25 cm, Daicel Chemical Industries, Ltd.). HRMS were acquired using an Agilent 6200 Series TOF with an Agilent G1978A Multimode source in electrospray ionization (ESI) mode. Molecular formulas of the compounds [M] are given, with the observed ion fragment in brackets, e.g. $[M+H]^+$.

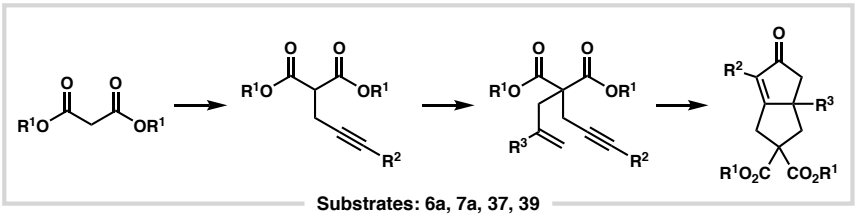
Abbreviations used: Et_2O – diethyl ether; CH_2Cl_2 – methylene chloride; PhMe – toluene; EtOAc – ethyl acetate; THF – tetrahydrofuran; *t*-BuOH – *tert*-butanol; $CHCl_3$ – chloroform; MeCN – acetonitrile; DMF – *N,N*-dimethylformamide; dme – 1,2-dimethoxyethane; TBME – *tert*-butyl methyl ether

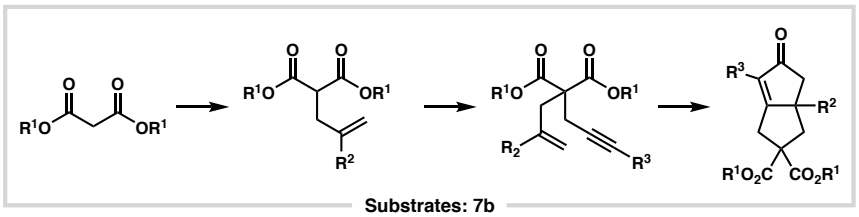
X-Ray Structure Determination

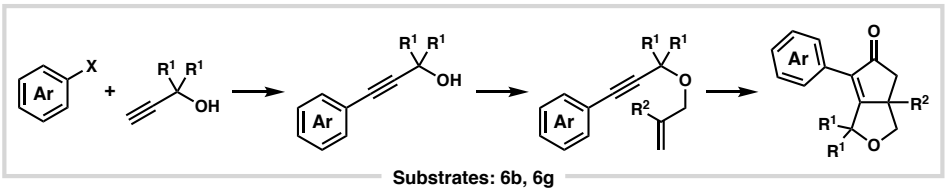
Low-temperature diffraction data (φ- and ω-scans) were collected on a Bruker AXS D8 VENTURE KAPPA diffractometer coupled to a PHOTON II CPAD detector with Mo-K_α radiation (λ = 0.71073 Å) from a $I_\mu S$ HB micro-focus sealed X-ray tube. All diffractometer manipulations, including data collection, integration, and scaling were carried out using the Bruker APEXII software.[15] Absorption corrections were applied using SADABS.[16] The structure was solved by intrinsic phasing using SHELXT[17] and refined against F^2 on all data by full-matrix least squares with SHELXL-2014[16] using established refinement techniques.[18] All non-hydrogen atoms were refined anisotropically. Unless otherwise

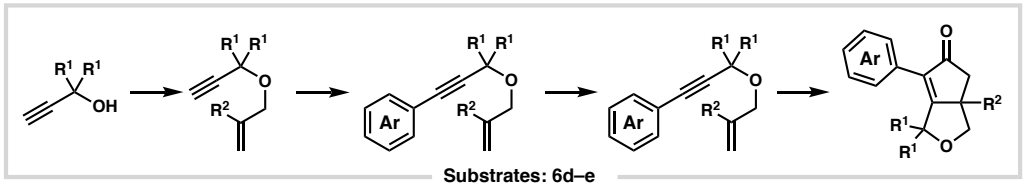
noted, all hydrogen atoms were included into the model at geometrically calculated positions and refined using a riding model. The isotropic displacement parameters of all hydrogen atoms were fixed to 1.2 times the *U* value of the atoms they are linked to (1.5 times for methyl groups and hydroxyl groups). Crystallographic data for **15e** can be obtained free of charge from The Cambridge Crystallographic Data Centre (CCDC) via www.ccdc.cam.ac.uk/data_request/cif under CCDC deposition numbers **1970101**. Graphical representation of the structures with 50% probability thermal ellipsoids was generated using Mercury visualization software.

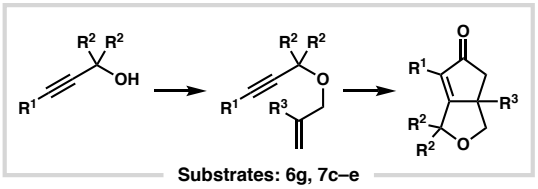
1.8.2 GENERAL SYNTHETIC SCHEMES AND PROCEDURES

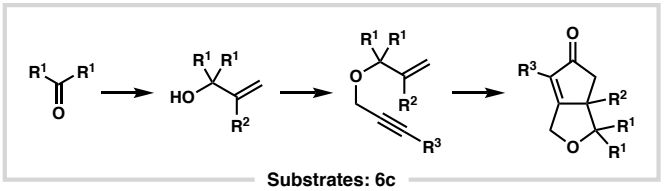


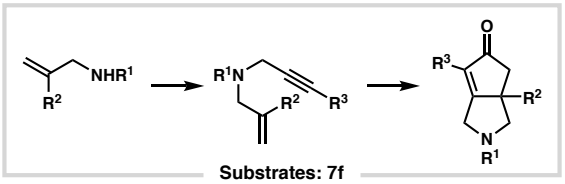


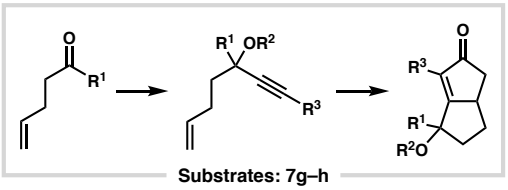


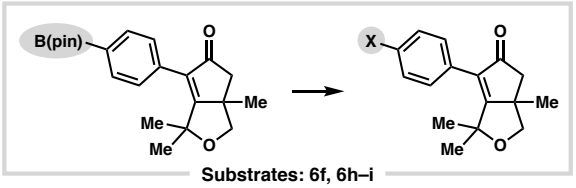

Substrates: 6d–e

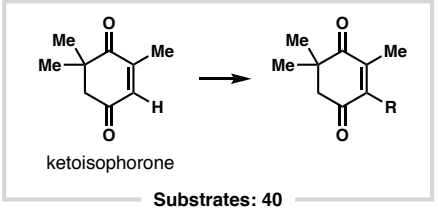

Substrates: 6g, 7c–e


Substrates: 6c


Substrates: 7f


Substrates: 7g–h


Substrates: 6f, 6h–i


ketoisophorone
Substrates: 40

Enyne Preparation: General Procedure A

To a dry round-bottomed flask with a Teflon-coated stir bar was added NaH in a glovebox. The flask was sealed with a rubber septum, removed from the glovebox, and placed under an atmosphere of dry N_2. Dry THF or DMF was then added and the mixture stirred. The resulting slurry was cooled to 0 °C in an ice water bath. The substrate was then added dropwise via syringe, with a vent needle in place to ensure efficient release of hydrogen gas into a well-ventilated fume hood. The resulting mixture was removed from the ice bath and allowed to stir at room temperature for 30 minutes or until hydrogen evolution ceased. At this time, the reaction was returned to the ice bath, and the alkyl bromide was added dropwise. The resulting mixture was allowed to reach room temperature while stirring and monitored by TLC. Upon completion, the reaction was quenched by dropwise addition of sat. aq. NH_4Cl, being careful not to cause vigorous hydrogen gas evolution or exotherm. The reaction was then diluted with water and Et_2O. The layers were separated and the aqueous layer extracted twice with Et_2O. Combined organics were washed with water and brine, dried over Na_2SO_4, filtered, and concentrated. The crude residue was purified by column chromatography to afford the desired product.

Enyne Preparation: General Procedure B

To a dry round-bottomed flask with a Teflon-coated stir bar was added $Pd(PPh_3)_2Cl_2$ and CuI in a glovebox. The flask was sealed with a rubber septum, removed from the glovebox, and placed under an atmosphere of dry N_2. The alkyne and aryl halide were then added. The combined solids were taken up in freshly distilled amine and dry solvent then stirred. The resulting yellow solution was stirred at various temperatures and monitored by TLC.

Upon complete consumption of the aryl halide (typically accompanied by rapid gelation of the reaction mixture), the reaction was allowed to reach to room temperature then diluted with EtOAc and water. The layers were separated and the aqueous layer extracted twice with EtOAc. Combined organics were washed with water and brine, dried over Na_2SO_4, filtered, and concentrated. The crude residue was purified by column chromatography to afford the desired product.

PKR: General Procedure C[19]

To a dry round-bottomed flask with a Teflon-coated stir bar was added the enyne and placed under an atmosphere of dry N_2. Anhydrous solvent was added and the solution stirred at room temperature. $Co_2(CO)_8$ was added in one portion, and the reaction mixture was stirred for 1 h with a vent needle to allow efficient release of CO. CAUTION: all manipulations with CO should be performed in a well-ventilated fume hood. Upon complete consumption of the enyne as judged by TLC, anhydrous DMSO was added dropwise via syringe. The reaction mixture was then placed in a preheated oil bath and monitored by TLC. Upon complete consumption of the intermediate Co-alkyne complex, the reaction was allowed to reach room temperature then diluted with EtOAc. Celite was added and the reaction mixture stirred overnight open to the air to sequester any Co species. The slurry was filtered over a pad of silica gel, washing with additional EtOAc. The filtrate was concentrated, and the crude residue was purified by column chromatography to afford the desired product. Note: initial elution with nonpolar solvent system is particularly important, as it washes away any remaining Co impurities that otherwise hamper purification.

PKR: General Procedure D[20]

To a dry round-bottomed flask with a Teflon-coated stir bar was added $[Rh(CO)_2Cl]_2$ in a glovebox. The flask was sealed with a rubber septum, removed from the glovebox, and placed under an atmosphere of dry N_2. The solid was taken up in *m*-xylene or PhMe and stirred, and the enyne was added via syringe. The resulting solution was stirred at room temperature, sparged with dry N_2 then dry CO successively for five minutes each, and was then kept under an atmosphere of CO via balloon. CAUTION: all manipulations with CO should be performed in a well-ventilated fume hood. At this time, the reaction was placed in a preheated oil bath at 120 °C and was stirred and monitored by TLC. Upon completion, the reaction was allowed to reach room temperature, and the CO was cautiously released into the fume hood. The solvent was evaporated, and the crude residue was purified by column chromatography to afford the desired product. Note: additional chromatography may be required to remove trace rhodium (i.e. passing a concentrated solution of the desired product in DCM through a pre-packed plug of SiliCycle® thiol-functionalized silica gel).

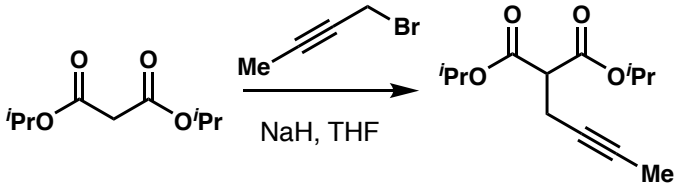
PKR: General Procedure E[21]

To a dry round-bottomed flask with a Teflon-coated stir bar was added $Mo(CO)_3(DMF)_3$ in a glovebox. The flask was sealed with a rubber septum, removed from the glovebox, and placed under an atmosphere of dry N_2. The enyne was added as a solution in DCM, and the reaction was stirred at room temperature with a vent needle to allow efficient release of CO. Note: in the event of solvent evaporation, fresh DCM was added periodically.

CAUTION: all manipulations with CO should be performed in a well-ventilated fume hood. Upon completion as judged by TLC, the reaction mixture was filtered over a pad of silica gel and celite, eluting with DCM. The filtrate was concentrated, and the crude residue was purified by column chromatography to afford the desired product.

1.8.3 SUBSTRATE PREPARATION

diisopropyl 2-(but-2-yn-1-yl)malonate (16)



Prepared from diisopropyl malonate (11.4 mL, 60.0 mmol, 3.0 equiv), 1-bromobut-2-yne (1.75 mL, 20.0 mmol, 1.0 equiv), NaH (95%, 720 mg, 30.0 mmol, 1.5 equiv), and THF (60 mL, 0.33 M) following General Procedure A. The crude residue was purified by column chromatography (silica, 8 to 15% Et_2O/hexanes) to afford **16** (2.99 g, 63% yield) as a colorless oil.

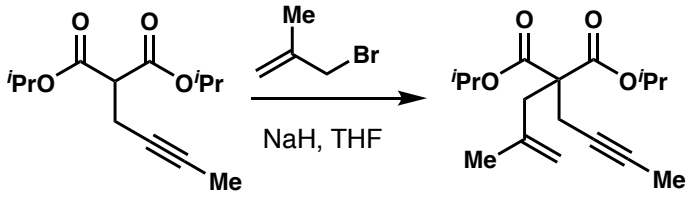
^{1}H NMR (500 MHz, CD_3Cl) δ 5.08 (hept, *J* = 6.2 Hz, 2H), 3.44 (t, *J* = 7.8 Hz, 1H), 2.69 (dq, *J* = 7.7, 2.5 Hz, 2H), 1.75 (t, *J* = 2.5 Hz, 3H), 1.26 (dd, *J* = 6.3, 4.3 Hz, 12H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 167.98, 77.77, 74.96, 69.20, 52.15, 21.79, 21.69, 18.80, 3.60.

FTIR (NaCl, thin film, cm^{-1}): 3055, 2978, 2850, 1706, 1655, 1301, 1120, 1026, 907, 766, 696.

HRMS (TOF-ESI, m/z): calc'd for $C_{13}H_{20}O_4$: 241.1434 $[M+H]^+$; found: 241.1439.

diisopropyl 2-(but-2-yn-1-yl)-2-(2-methylallyl)malonate (17)



Prepared from **16** (721 mg, 3.0 mmol, 1.0 equiv), methallyl bromide (0.34 mL, 3.3 mmol, 1.1 equiv), NaH (95%, 84.0 mg, 3.3 mmol, 1.1 equiv), and THF (6 mL, 0.50 M) following General Procedure A. The crude, clear colorless oil of **17** was used without further purification (821 mg, 93% yield).

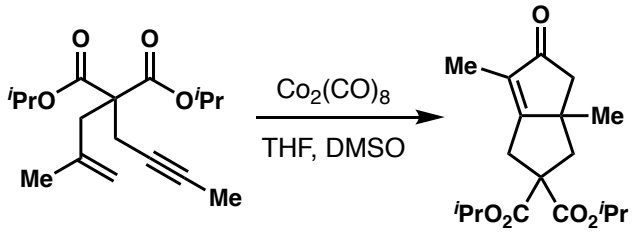
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.05 (hept, *J* = 6.3 Hz, 2H), 4.88 (dt, *J* = 2.0, 1.5 Hz, 1H), 4.84 (dq, *J* = 1.8, 0.9 Hz, 1H), 2.79 (d, *J* = 0.9 Hz, 2H), 2.74 (q, *J* = 2.5 Hz, 2H), 1.78 – 1.72 (m, 3H), 1.69 (dd, *J* = 1.5, 0.8 Hz, 3H), 1.24 (dd, *J* = 6.3, 1.4 Hz, 12H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 170.17, 140.58, 115.99, 78.87, 74.05, 69.06, 56.67, 39.39, 23.57, 23.03, 21.68, 3.59.

FTIR (NaCl, thin film, cm^{-1}): 3465, 3080, 2980, 2938, 1732, 1645, 1456, 1374, 1277.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{26}O_4$: 295.1904 $[M+H]^+$; found: 295.1904.

diisopropyl 3a,6-dimethyl-5-oxo-3,3a,4,5-tetrahydropentalene-2,2(1H)-dicarboxylate (6a)



Prepared from **17** (2.95 g, 10.0 mmol, 1.0 equiv), $Co_2(CO)_8$ (3.77 g, 11.0 mmol, 1.1 equiv), DMSO (7.1 mL, 100.0 mmol, 10.0 equiv), and THF (100 mL, 0.10 M) following General Procedure C. The crude residue was purified by column chromatography (silica, 40 to 50% Et_2O/hexanes) to afford **6a** (3.06 g, 95% yield) as a white, glassy solid.

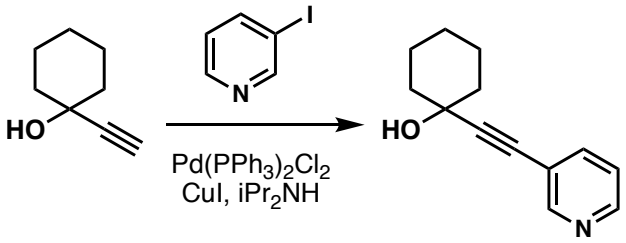
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.10 (hept, *J* = 6.3 Hz, 1H), 5.00 (hept, *J* = 6.3 Hz, 1H), 3.34 (dq, *J* = 17.4, 1.9 Hz, 1H), 3.12 (d, *J* = 17.4 Hz, 1H), 2.55 (d, *J* = 13.6 Hz, 1H), 2.40 (d, *J* = 17.5 Hz, 1H), 2.34 (d, *J* = 17.5 Hz, 1H), 2.11 (d, *J* = 13.5 Hz, 1H), 1.70 (d, *J* = 1.7 Hz, 3H), 1.30 – 1.19 (m, 12H), 1.12 (s, 3H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 209.55, 181.23, 171.41, 171.17, 131.63, 69.72, 69.68, 60.51, 51.23, 47.67, 44.74, 33.14, 26.76, 21.66, 21.64, 21.60, 21.57, 8.63.

FTIR (NaCl, thin film, cm^{-1}): 2981, 2934, 2874, 1727, 1714, 1677, 1375, 1262, 1183, 1104, 1064, 912.

HRMS (TOF-ESI, m/z): calc'd for $C_{18}H_{26}O_5$: 323.1853 $[M+H]^+$; found: 323.1858.

3-((1-((2-methylallyl)oxy)cyclohexyl)ethynyl)pyridine (18)

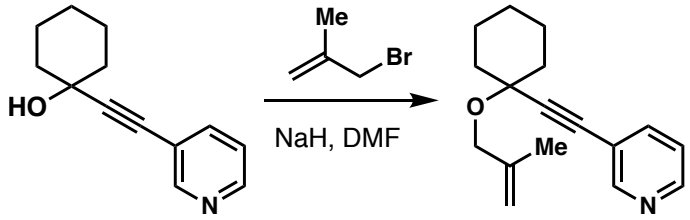


Prepared from 1-ethynylcyclohexan-1-ol (1.49 g, 12.0 mmol, 1.2 equiv), 3-iodopyridine (2.05 g, 10.0 mmol, 1.0 equiv), $Pd(PPh_3)_2Cl_2$ (211 mg, 0.3 mmol, 0.03 equiv), CuI (58 mg, 0.30 mmol, 0.03 equiv), and *i*Pr_2NH (20 mL, 0.50 M) at 80 ºC following General Procedure

B. The crude residue was purified by column chromatography (silica, 55 to 65% EtOAc/hexanes) to afford **18** (2.00 g, >99% yield) as a tan, amorphous solid.

Spectral data matched those reported in the literature.[22]

3-((1-((2-methylallyl)oxy)cyclohexyl)ethynyl)pyridine (19)



Prepared from **18** (1.01 g, 5.0 mmol, 1.0 equiv), methallyl bromide (0.56 mL, 5.5 mmol, 1.1 equiv), NaH (95%, 140 mg, 5.5 mmol, 1.1 equiv), and DMF (20 mL, 0.25 M) following General Procedure A. The crude, clear amber oil of **19** was used without further purification (1.23 g, 97% yield).

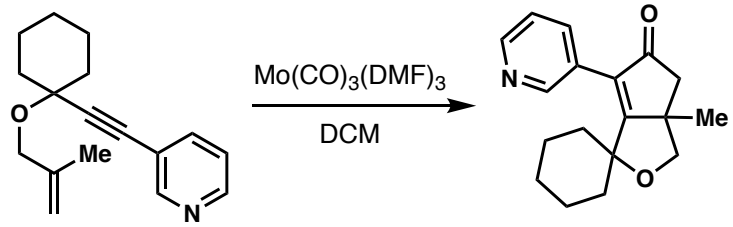
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.61 (d, *J* = 55.6 Hz, 2H), 7.71 (dt, *J* = 7.9, 1.8 Hz, 1H), 7.24 (d, *J* = 7.9 Hz, 1H), 5.04 (dp, *J* = 1.9, 1.0 Hz, 1H), 4.88 (dp, *J* = 2.4, 1.1 Hz, 1H), 4.07 (d, *J* = 1.1 Hz, 2H), 2.00 (ddd, *J* = 12.0, 7.2, 4.5 Hz, 2H), 1.79 (t, *J* = 1.2 Hz, 3H), 1.72 (qt, *J* = 9.5, 2.5 Hz, 3H), 1.57 (dddt, *J* = 17.8, 11.5, 8.6, 3.5 Hz, 3H), 1.45 – 1.30 (m, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 152.51, 148.66, 143.07, 138.68, 123.11, 111.66, 94.60, 82.66, 74.23, 67.62, 37.28, 25.57, 22.96, 20.02.

FTIR (NaCl, thin film, cm^{-1}): 3077, 3029, 2935, 2857, 1656, 1560, 1475, 1448, 1406, 1086.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{21}NO$: 256.1696 $[M+H]^+$; found: 256.1697.

3a'-methyl-6'-(pyridin-3-yl)-3a',4'-dihydrospiro[cyclohexane-1,1'-cyclopenta[*c*]furan]-5'(3'*H*)-one (6b)



Prepared from **19** (1.22 g, 4.8 mmol, 1.0 equiv) and $Mo(CO)_3(DMF)_3$ (2.10 g, 5.3 mmol, 1.1 equiv) following General Procedure E. The crude residue was purified by column chromatography (silica, 30 to 50% EtOAc/hexanes) to afford **6b** (241 mg, 18% yield) as a tan, amorphous solid.

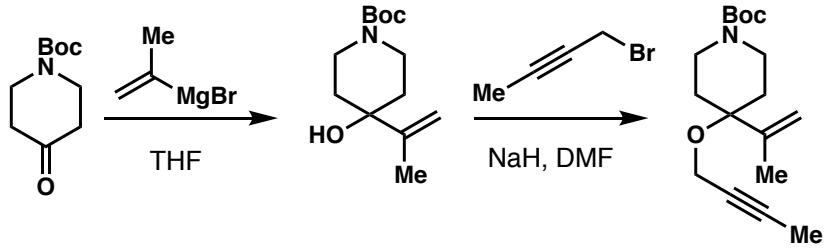
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.53 (dd, *J* = 4.9, 1.7 Hz, 1H), 8.45 (dd, *J* = 2.3, 0.9 Hz, 1H), 7.56 (dt, *J* = 7.8, 2.0 Hz, 1H), 7.28 (ddd, *J* = 7.8, 4.9, 0.9 Hz, 1H), 3.96 (d, *J* = 8.5 Hz, 1H), 3.48 (dd, *J* = 8.5, 0.9 Hz, 1H), 2.46 (d, *J* = 1.4 Hz, 2H), 2.17 (dp, *J* = 14.2, 2.7 Hz, 1H), 1.79 (td, *J* = 12.9, 4.8 Hz, 1H), 1.71 – 1.45 (m, 4H), 1.41 (s, 3H), 1.40 – 1.25 (m, 2H), 1.07 – 0.90 (m, 2H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 207.28, 191.05, 149.86, 149.59, 136.91, 131.83, 127.70, 123.46, 81.04, 74.72, 50.50, 48.92, 38.56, 32.03, 26.20, 24.90, 22.02, 21.79.

FTIR (NaCl, thin film, cm^{-1}): 2934, 2855, 1710, 1652, 1447, 1411, 1117, 1015, 919, 734, 713.

HRMS (TOF-ESI, m/z): calc'd for $C_{18}H_{21}NO_2$: 284.1645 $[M+H]^+$; found: 284.1642.

***tert*-butyl 4-(but-2-yn-1-yloxy)-4-(prop-1-en-2-yl)piperidine-1-carboxylate (20)**



To a dry round-bottomed flask with a Teflon-coated stir bar was added isopropenylmagnesium bromide (0.5 M in THF, 40.0 mL, 20.0 mmol, 2.0 equiv) and stirred at room temperature under an atmosphere of dry N_2. *N*-boc-piperidone (2.00 g, 10.0 mmol, 1.0 equiv) was then added as a solution in THF (27 mL, 0.15 M final), and the solution was stirred at room temperature and monitored by TLC. Upon complete consumption of the ketone, the reaction was quenched by slow dropwise addition of sat. aq. NH_4Cl, being careful not to cause vigorous propene gas evolution or exotherm. When propene evolution ceased, the reaction was diluted with water and EtOAc. The layers were separated and the aqueous extracted twice with EtOAc. Combined organics were washed with water and brine, dried over Na_2SO_4, filtered, and concentrated. The crude oil (10.0 mmol, 1.0 equiv) was used directly in the next step, following General Procedure A with 1-bromobut-2-yne (1.1 mL, 12.5 mmol, 1.25 equiv), NaH (95%, 316 mg, 12.5 mmol, 1.25 equiv), and DMF (40 mL, 0.25 M). The crude oil was purified by column chromatography (silica, 15 to 25% Et_2O/hexanes) to afford **20** (1.64 g, 56% yield over 2 steps) as a colorless oil.

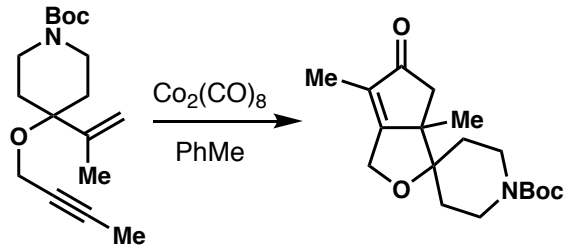
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.04 (p, *J* = 1.4 Hz, 1H), 4.91 (t, *J* = 1.0 Hz, 1H), 3.99 – 3.72 (m, 4H), 3.18 (s, 2H), 1.93 – 1.84 (m, 2H), 1.84 (t, *J* = 2.4 Hz, 3H), 1.74 (dd, *J* = 1.4, 0.7 Hz, 3H), 1.63 – 1.53 (m, 2H), 1.45 (s, 9H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 154.99, 146.69, 114.02, 81.64, 79.44, 77.03, 76.03, 50.94, 28.59, 18.23, 3.89.

FTIR (NaCl, thin film, cm^{-1}): 3495, 3364, 3091, 2927, 2318, 2241, 1958, 1822, 1689, 1454.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{27}NO_3$: 294.2064 $[M+H]^+$; found: 294.2078.

***tert*-butyl 4,6a-dimethyl-5-oxo-3,5,6,6a-tetrahydrospiro[cyclopenta[*c*]furan-1,4'-piperidine]-1'-carboxylate (6c)**



Prepared from **20** (733 mg, 2.5 mmol, 1.0 equiv), $Co_2(CO)_8$ (941 mg, 2.75 mmol, 1.1 equiv), and PhMe (25 mL, 0.10 M) following modified General Procedure C without addition of DMSO. Upon complete conversion of the enyne to the Co-alkyne complex, the reaction was placed in a preheated oil bath at 110 °C. Upon complete consumption of the Co-alkyne complex, the reaction was allowed to reach room temperature, filtered over a pad of SiO_2 and Celite, and the filtrate concentrated. The crude residue was purified by column chromatography (silica, 5 to 40% EtOAc/hexanes) to afford **6c** (494 mg, 62% yield) as an off-white, amorphous solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 4.64 (dq, *J* = 16.0, 1.7 Hz, 1H), 4.43 (d, *J* = 16.0 Hz, 1H), 4.14 – 3.82 (m, 2H), 3.01 (s, 2H), 2.34 (d, *J* = 17.7 Hz, 1H), 2.18 (d, *J* = 17.7 Hz, 1H), 1.85

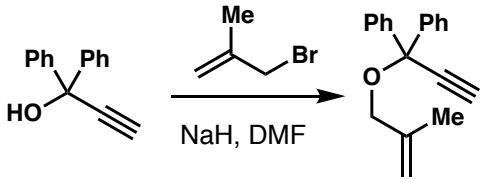
– 1.74 (m, 1H), 1.70 (dd, *J* = 1.6, 1.0 Hz, 3H), 1.54 (td, *J* = 13.2, 5.3 Hz, 1H), 1.44 (s, 9H), 1.36 (dd, *J* = 12.7, 4.5 Hz, 1H), 1.29 – 1.19 (m, 1H), 1.18 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 209.39, 180.68, 154.83, 130.62, 81.26, 79.66, 62.56, 52.62, 45.28, 31.68, 29.78, 28.56, 22.72, 8.74.

FTIR (NaCl, thin film, cm^{-1}): 3415, 2974, 2935, 1769, 1715, 1693, 1427, 1366, 1247, 1163, 1052, 737.

HRMS (TOF-ESI, m/z): calc'd for $C_{18}H_{27}NO_4$: 322.2013 $[M+H]^+$; found: 322.2000.

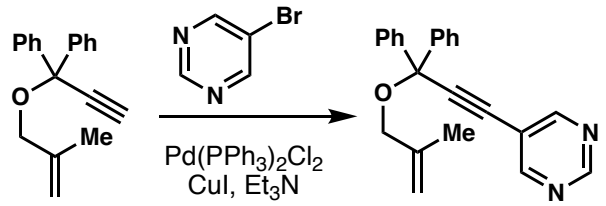
(1-((2-methylallyl)oxy)prop-2-yne-1,1-diyl)dibenzene (21)



Prepared from 1,1-diphenylprop-2-yn-1-ol (4.17 g, 20.0 mmol, 1.0 equiv), methallyl bromide (2.22 mL, 22.0 mmol, 1.1 equiv), NaH (95%, 556 mg, 22.0 mmol, 1.1 equiv), and DMF (80 mL, 0.25 M) following General Procedure A. The crude oil was purified by column chromatography (silica, 1 to 3% Et_2O/hexanes) to afford **21** (4.69 g, 90% yield) as a yellow oil.

Spectral data matched those reported in the literature.[23]

5-(3-((2-methylallyl)oxy)-3,3-diphenylprop-1-yn-1-yl)pyrimidine (22)



Prepared from **21** (374 mg, 1.42 mmol, 1.0 equiv), 5-bromopyrimidine (240 mg, 1.5 mmol, 1.05 equiv), $Pd(PPh_3)_2Cl_2$ (43 mg, 0.06 mmol, 0.04 equiv), CuI (11.5 mg, 0.06 mmol, 0.04 equiv), and Et_3N (6 mL, 0.25 M) at 50 ºC following General Procedure B. The crude residue was purified by column chromatography (silica, 25 to 30% Et_2O/hexanes) to afford **22** (365 mg, 75% yield) as a yellow, amorphous solid.

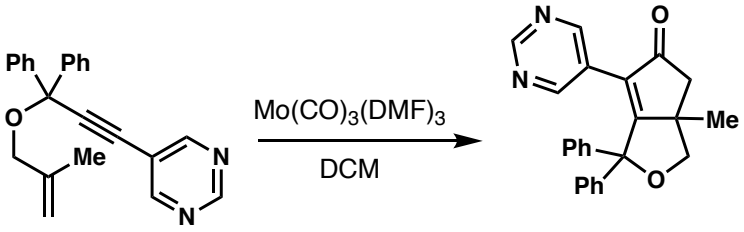
^{1}H NMR (400 MHz, $CDCl_3$) δ 9.17 (s, 1H), 8.85 (s, 2H), 7.64 – 7.56 (m, 4H), 7.40 – 7.32 (m, 4H), 7.34 – 7.25 (m, 2H), 5.12 (tp, *J* = 1.9, 0.9 Hz, 1H), 4.93 (dh, *J* = 2.6, 1.3 Hz, 1H), 3.97 (d, *J* = 1.5 Hz, 2H), 1.81 (dd, *J* = 1.5, 0.8 Hz, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 159.05, 157.30, 143.00, 142.24, 128.52, 128.13, 126.72, 119.29, 111.65, 96.49, 82.48, 80.69, 68.99, 20.13.

FTIR (NaCl, thin film, cm^{-1}): 3356, 3312, 2953, 2922, 2851, 2352, 2333, 1633, 1540, 1447, 1415, 1267.

HRMS (TOF-ESI, m/z): calc'd for $C_{23}H_{20}N_2O$: 341.1648 $[M+H]^+$; found: 341.1653.

3a-methyl-1,1-diphenyl-6-(pyrimidin-5-yl)-3a,4-dihydro-1*H*-cyclopenta[*c*]furan-5(3*H*)-one (6d)



Prepared from **22** (340 mg, 1.0 mmol, 1.0 equiv) and $Mo(CO)_3(DMF)_3$ (401 mg, 1.0 mmol, 1.0 equiv) following General Procedure E. The crude residue was purified by column chromatography (silica, 30 to 60% EtOAc/hexanes) to afford **6d** (101 mg, 28% yield) as a light yellow, amorphous solid.

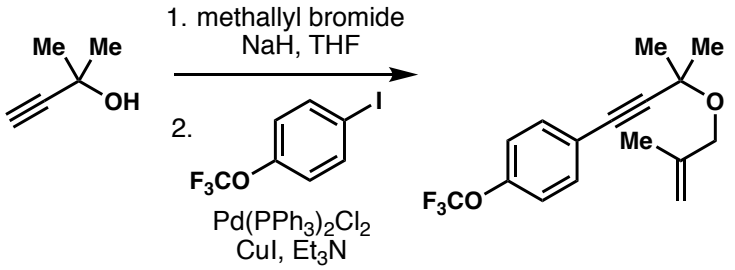
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.92 (s, 1H), 8.22 (s, 2H), 7.50 – 7.37 (m, 5H), 7.07 – 6.93 (m, 3H), 6.84 – 6.76 (m, 2H), 4.39 (d, *J* = 8.2 Hz, 1H), 4.20 (dd, *J* = 8.1, 0.8 Hz, 1H), 2.81 (d, *J* = 17.7 Hz, 1H), 2.71 (d, *J* = 17.7 Hz, 1H), 1.31 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 206.12, 187.62, 157.69, 156.30, 142.88, 141.70, 133.17, 128.77, 128.70, 128.27, 127.92, 127.76, 127.56, 124.55, 87.88, 78.83, 50.50, 49.64, 27.33.

FTIR (NaCl, thin film, cm^{-1}): 3055, 2969, 2926, 2858, 2362, 1715, 1548, 1447, 1413.

HRMS (TOF-ESI, m/z): calc'd for $C_{24}H_{20}N_2O_2$: 369.1598 $[M+H]^+$; found: 369.1612.

1-(3-methyl-3-((2-methylallyl)oxy)but-1-yn-1-yl)-4-(trifluoromethoxy)benzene (23)



Prepared in two steps. Intermediate prepared from 2-methyl-3-butyn-2-ol (1.94 mL, 20.0 mmol, 1.0 equiv), methallyl bromide (2.22 mL, 22.0 mmol, 1.1 equiv), NaH (95%, 555 mg, 22.0 mmol, 1.1 equiv), and THF (80 mL, 0.25 M) following General Procedure A. The crude residue was purified by column chromatography (silica, 2 to 4% Et_2O/hexanes) to afford the desired intermediate as a clear colorless oil, which was left as a solution in hexanes to avoid product evaporation upon concentration. The intermediate (~690 mg in hexanes, ~5.0 mmol, 1.0 equiv) was then used directly in the next step, following General Procedure B with *p*-trifluoromethoxy-iodobenzene (0.79 mL, 5.0 mmol, 1.0 equiv), $Pd(PPh_3)_2Cl_2$ (141 mg, 0.2 mmol, 0.04 equiv), CuI (38.1 mg, 0.2 mmol, 0.04 equiv), and

Et_3N (20 mL, 0.25 M). The crude oil was purified by column chromatography (silica, 2 to 3% Et_2O/hexanes) to afford **23** (799 mg, 54% yield over 2 steps) as a colorless oil.

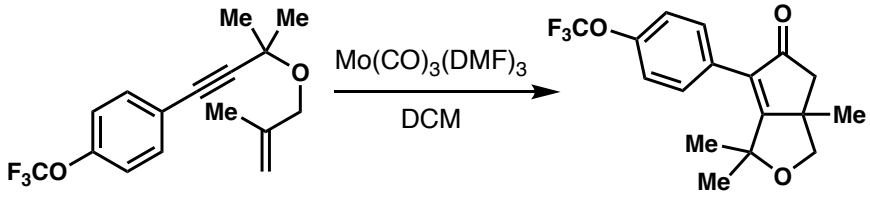
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.48 – 7.40 (m, 2H), 7.15 (ddt, *J* = 7.7, 2.0, 1.0 Hz, 2H), 5.03 (dh, *J* = 2.2, 1.1 Hz, 1H), 4.88 (dqd, *J* = 2.4, 1.5, 0.8 Hz, 1H), 4.08 – 4.03 (m, 2H), 1.81 – 1.74 (m, 3H), 1.57 (s, 6H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 148.98, 148.96, 143.00, 133.31, 120.50 (q, $^1J_{CF}$ = 160 Hz), 121.87, 120.96, 111.74, 92.63, 82.78, 70.90, 68.60, 28.96, 19.92.

FTIR (NaCl, thin film, cm^{-1}): 3437, 2984, 2935, 1726, 1608, 1507, 1264, 1219, 1162.

HRMS (TOF-ESI, m/z): calc'd for $C_{12}H_{11}F_3O_2$: 227.0678 $[M–C_4H_7O+H]^+$; found: 227.0688.

1,1,3a-trimethyl-6-(4-(trifluoromethoxy)phenyl)-3a,4-dihydro-1*H*-cyclopenta[*c*]furan-5(3*H*)-one (6e)



Prepared from **23** (745 mg, 2.5 mmol, 1.0 equiv) and $Mo(CO)_3(DMF)_3$ (1.00 g, 2.5 mmol, 1.0 equiv) following General Procedure E. The crude residue was purified by column chromatography (silica, 30 to 35% EtOAc/hexanes). To remove residual [Mo] byproducts, the sample was taken up in 5 mL DCM, then 450 mg (~0.2 mmol) TAAcONa capped silica gel (SiliCycle®) was added and the mixture stirred overnight. The mixture was then filtered and concentrated to afford **6e** (330 mg, 41% yield) as an off-white amorphous solid.

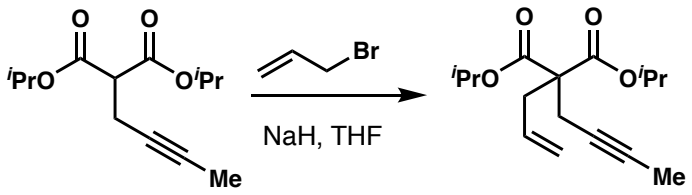
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.31 – 7.23 (m, 2H), 7.23 – 7.15 (m, 2H), 3.95 (d, *J* = 8.4 Hz, 1H), 3.51 (dd, *J* = 8.5, 0.9 Hz, 1H), 2.48 (d, *J* = 1.0 Hz, 2H), 1.67 (s, 3H), 1.44 (s, 3H), 1.02 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ ^{13}C NMR (101 MHz, $CDCl_3$) δ 207.17, 189.24, 149.40, 149.38, 149.36, 133.97, 130.81, 129.80, 120.54 (q, $^1J_{CF}$ = 258 Hz), 120.99, 79.23, 75.06, 50.63, 49.04, 29.60, 26.04, 25.13.

FTIR (NaCl, thin film, cm^{-1}): 2977, 2933, 2848, 1712, 1652, 1507, 1263, 1224, 1162, 1019.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{17}F_3O_3$: 327.1203 $[M+H]^+$; found: 327.1203.

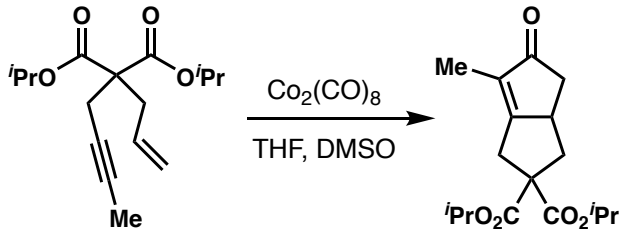
diisopropyl 2-allyl-2-(but-2-yn-1-yl)malonate (24)



Prepared from **16** (1.44 g, 6.0 mmol, 1.0 equiv), allyl bromide (0.55 mL, 6.3 mmol, 1.05 equiv), NaH (95%, 173 mg, 7.2 mmol, 1.2 equiv), and THF (15 mL, 0.40 M) following General Procedure A. The crude oil was purified by column chromatography (silica, 7 to 10% Et_2O/hexanes) to afford **24** (1.63 g, 97% yield) as a colorless oil.

Spectral data matched those reported in the literature.[24]

Diisopropyl 6-methyl-5-oxo-3,3a,4,5-tetrahydropentalene-2,2(1H)-dicarboxylate (7a)



Prepared from **24** (1.40 g, 5.0 mmol, 1.0 equiv), $Co_2(CO)_8$ (2.06 g, 6.0 mmol, 1.2 equiv), and THF (50 mL, 0.10 M) following General Procedure C. The crude residue was purified by column chromatography (silica, 10 to 60% Et_2O/hexanes) to afford **7a** (1.30 g, 85% yield) as a white powder.

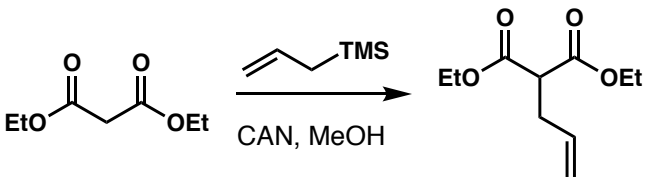
^{1}H NMR (400 MHz, CD_3OD) δ 5.04 (dhept, *J* = 20.8, 6.3 Hz, 2H), 3.19 (dt, *J* = 2.5, 1.5 Hz, 2H), 3.05 – 2.95 (m, 1H), 2.72 (dd, *J* = 12.7, 7.6 Hz, 1H), 2.61 (dd, *J* = 18.0, 6.2 Hz, 1H), 2.16 – 2.06 (m, 1H), 1.69 (dt, *J* = 2.5, 1.3 Hz, 3H), 1.64 (t, *J* = 12.5 Hz, 1H), 1.27 (d, *J* = 6.3 Hz, 6H), 1.24 (dd, *J* = 6.3, 3.7 Hz, 6H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 209.80, 178.32, 171.41, 170.79, 133.14, 69.77, 69.68, 61.28, 42.90, 41.66, 39.32, 34.19, 21.77, 8.81.

FTIR (NaCl, thin film, cm^{-1}): 2981, 2924, 1728, 1714, 1678, 1455, 1375, 1271, 1192, 1102, 911.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{24}O_5$: 309.1697 $[M+H]^+$; found: 309.1684.

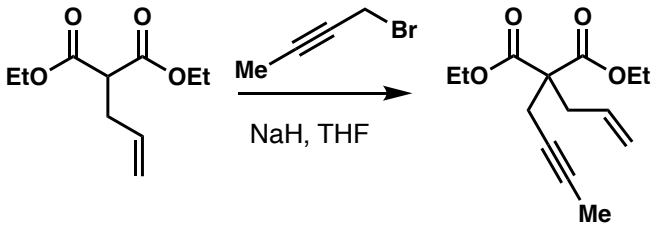
diethyl 2-allylmalonate (25)



Prepared from diethylmalonate (2.29 mL, 15.0 mmol, 1.0 equiv) following a procedure by Shiao and coworkers.[25] The crude oil was purified by column chromatography (silica, 15% Et_2O/hexanes) to afford **25** (1.70 g, 57% yield) as a colorless oil.

Spectral data matched those reported in the literature.[25]

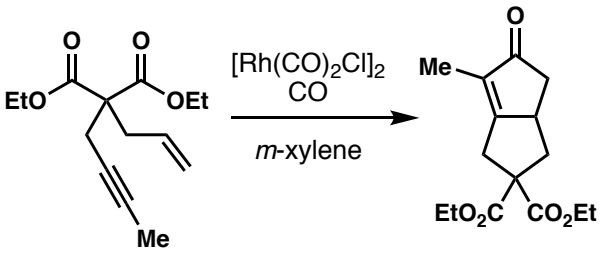
diethyl 2-allyl-2-(but-2-yn-1-yl)malonate (26)



Prepared from **25** (1.00 g, 5.0 mmol, 1.0 equiv), 1-bromobut-2-yne (0.44 mL, 5.0 mmol, 1.0 equiv), NaH (60% in mineral oil, 240 mg, 6.0 mmol, 1.2 equiv), and THF (10 mL, 0.50 M) following General Procedure A. The crude residue was purified by column chromatography (12% Et_2O/hexanes) to afford **26** (1.12 g, 89% yield) as a colorless oil.

Spectral data matched those reported in the literature.[26]

diethyl 6-methyl-5-oxo-3,3a,4,5-tetrahydropentalene-2,2(1*H*)-dicarboxylate (7b)

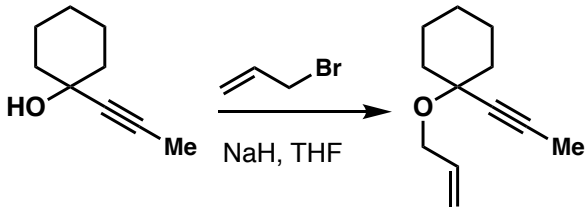


Prepared from **26** (504 mg, 2.0 mmol, 1.0 equiv), $[Rh(CO)_2Cl]_2$ (78.0 mg, 0.2 mmol, 0.1 equiv), and *m*-xylene (20 mL, 0.10 M) following General Procedure D. The crude residue

was purified by column chromatography (silica, 30% EtOAc/hexanes) to afford **7b** (438 mg, 79% yield) as an off-white powder.

Spectral data matched those reported in the literature.[27]

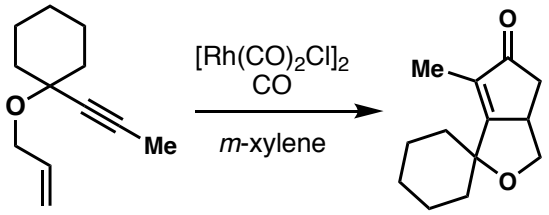
1-(allyloxy)-1-(prop-1-yn-1-yl)cyclohexane (27)



Prepared from 1-(prop-1-yn-1-yl)cyclohexan-1-ol (1.39 g, 10.0 mmol, 1.0 equiv), allyl bromide (1.73 mL, 20.0 mmol, 2.0 equiv), NaH (95%, 360 mg, 15.0 mmol, 1.5 equiv), and THF (20 mL, 0.50 M) following General Procedure A. The crude oil was purified by column chromatography (silica, 2 to 5% Et_2O/hexanes) to afford **27** (914 mg, 52% yield) as a colorless oil.

Spectral data matched those reported in the literature.[28]

6'-methyl-3a',4'-dihydrospiro[cyclohexane-1,1'-cyclopenta[*c*]furan]-5'(3'*H*)-one (7c)



Prepared from **27** (446 mg, 2.5 mmol, 1.0 equiv), $[Rh(CO)_2Cl]_2$ (73 mg, 0.19 mmol, 0.75 equiv), and *m*-xylene (12.5 mL, 0.20 M) following General Procedure D. The crude oil was

purified by column chromatography (silica, 20 to 30% EtOAc/hexanes) to afford **7c** (27 mg, 68% yield) as a clear yellow oil.

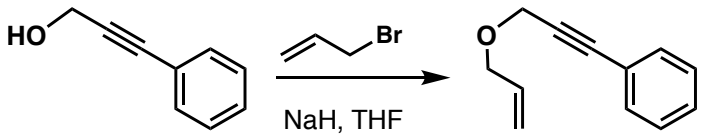
^{1}H NMR (500 MHz, $CDCl_3$) δ 4.29 – 4.17 (m, 1H), 3.28 – 3.17 (m, 3H), 2.58 (dd, *J* = 17.6, 5.9 Hz, 1H), 2.09 (dd, *J* = 17.5, 3.5 Hz, 1H), 1.79 (d, *J* = 2.2 Hz, 4H), 1.76 (ddt, *J* = 11.5, 6.5, 2.5 Hz, 2H), 1.72 – 1.59 (m, 6H), 1.33 – 1.19 (m, 1H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 210.00, 182.70, 130.57, 79.72, 69.77, 44.10, 38.61, 35.28, 32.22, 25.40, 22.19, 21.58, 8.36.

FTIR (NaCl, thin film, cm^{-1}): 3412, 2924, 2856, 1698, 1682, 1446, 1269, 1058, 1024, 917, 840, 738, 665.

HRMS (TOF-ESI, m/z): calc'd for $C_{13}H_{18}O_2$: 207.1380 $[M+H]^+$; found: 207.1376.

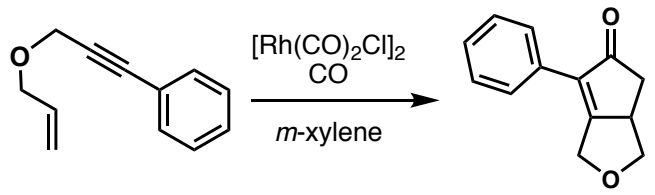
(3-(allyloxy)prop-1-yn-1-yl)benzene (28)



Prepared from 3-phenylprop-2-yn-1-ol (1.25 mL, 10.0 mmol, 1.0 equiv), allyl bromide (1.73 mL, 20.0 mmol, 2.0 equiv), NaH (95%, 300 mg, 12.5 mmol, 1.25 equiv), and THF (20 mL, 0.50 M) following General Procedure A. The resulting crude, colorless oil of **28** (1.68 g, 98% yield) was used without further purification.

Spectral data matched those reported in the literature.[29]

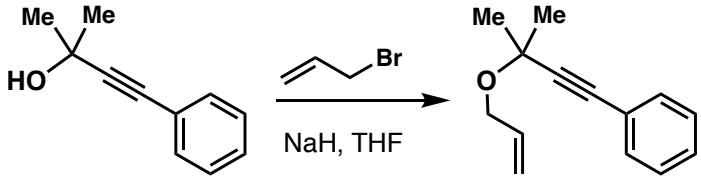
6-phenyl-3a,4-dihydro-1*H*-cyclopenta[*c*]furan-5(3*H*)-one (7d)



Prepared from **28** (345 mg, 2.0 mmol, 1.0 equiv), $[Rh(CO)_2Cl]_2$ (59 mg, 0.15 mmol, 0.75 equiv), and *m*-xylene (20 mL, 0.10 M) following General Procedure D. The crude residue was purified by column chromatography (silica, 30 to 40% EtOAc/hexanes) to afford **7d** (336 mg, 84% yield) as a white powder.

Spectral data matched those reported in the literature.[30]

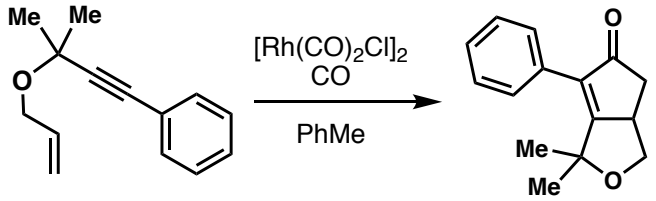
(3-(allyloxy)-3-methylbut-1-yn-1-yl)benzene (29)



Prepared from 2-methyl-4-phenylbut-3-yn-2-ol (801 mg, 5.0 mmol, 1.0 equiv), allyl bromide (0.87 mL, 10.0 mmol, 2.0 equiv), NaH (95%, 158 mg, 6.25 mmol, 1.25 equiv), and THF (10 mL, 0.50 M) following General Procedure A. The resulting yellow oil of **29** (283 mg, 48% yield) was used without further purification.

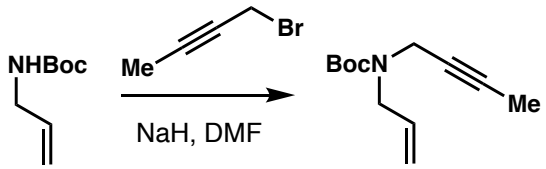
Spectral data matched those reported in the literature.[31]

1,1-dimethyl-6-phenyl-3a,4-dihydro-1*H*-cyclopenta[*c*]furan-5(3*H*)-one (7e)



Prepared from **29** (20.0 mg, 0.1 mmol, 1.0 equiv), $[Rh(CO)_2Cl]_2$ (2.9 mg, 0.0075 mmol, 0.075 equiv), and PhMe (1 mL, 0.10 M) following General Procedure D. The crude residue was purified by column chromatography (silica, 15 to 25% EtOAc/hexanes) to yield **7e** (15.5 mg, 68% yield) as a white powder. Spectral data matched those reported in the literature.[31]

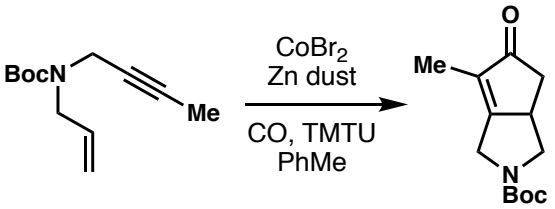
***tert*-butyl allyl(but-2-yn-1-yl)carbamate (30)**



Prepared from *tert*-butyl allylcarbamate (786 mg, 5.0 mmol, 1.0 equiv), 1-bromobut-2-yne (0.44 mL, 5.0 mmol, 1.0 equiv), NaH (95%, 168 mg, 7.0 mmol, 1.4 equiv), and DMF (20 mL, 0.25 M) following General Procedure A. The crude residue of **30** was used in the next step without further purification (881 mg, 85% yield).

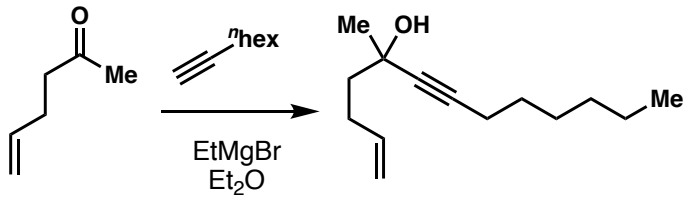
Spectral data matched those reported in the literature.[30]

tert-butyl 6-methyl-5-oxo-3,3a,4,5-tetrahydrocyclopenta[_c_]pyrrole-2(1_H_)-carboxylate (7f)



To a dry vial with a Teflon-coated stir bar was added $CoBr_2$ (22 mg, 0.1 mmol, 0.1 equiv) in a glovebox. The flask was sealed with a rubber septum, removed from the glovebox, and placed under an atmosphere of dry CO via balloon. CAUTION: all manipulations with CO should be performed in a well-ventilated fume hood. Zinc dust (131 mg, 2.0 mmol, 2.0 equiv) and tetramethyltiourea (80 mg, 0.6 mmol, 0.6 equiv) were then added and the solids taken up in PhMe (20 mL, 0.50 M) and stirred. The solution was degassed with CO for 5 min and the reaction kept under an atmosphere of CO following addition of **30** (295 mg, 1.0 mmol, 1.0 equiv). The reaction was brought to 70 °C in a preheated oil bath and monitored by TLC. Upon complete consumption of the enyne (typically indicated by a color change from deep green to light turquoise green or colorless), the reaction mixture was filtered over a pad of silica gel and celite, eluting with EtOAc. The filtrate was concentrated, and the crude residue was purified by column chromatography (silica, 30 to 40% EtOAc/hexanes) to afford **7f** (165 mg, 70% yield) as an off-white, amorphous solid. Spectral data matched those reported in the literature.[30]

5-methyltridec-1-en-6-yn-5-ol (31)



To a dry round-bottom flask with a Teflon-coated stir bar was added oct-1-yne (8.9 mL, 60.0 mmol, 3.0 equiv) and Et_2O (100 mL, 0.1 M). 3 M EtMgBr in Et_2O (20.0 mL, 60.0 mmol, 3.0 equiv) was then added dropwise. The solution was then heated to 30 ºC for 1 h then allowed to reach room temperature. To a separate dry round-bottom flask with a Teflon-coated stir bar was added hex-5-en-2-one (2.3 mL, 20.0 mmol, 1.0 equiv) and Et_2O (80 mL). This solution was added to the first solution slowly via canula. The reaction was allowed to stir at room temperature and monitored by TLC. Upon completion, the reaction was quenched by dropwise addition of sat. aq. NH_4Cl. The resulting mixture was diluted with Et_2O and water and the layers separated. The aqueous layer was extracted twice with Et_2O. Combined organics were washed with water and brine, dried over Na_2SO_4, filtered, and concentrated. The crude residue was purified by column chromatography (silica, 15 to 20% Et_2O/hexanes) to afford **31** (2.89 g, 70% yield) as a colorless oil.

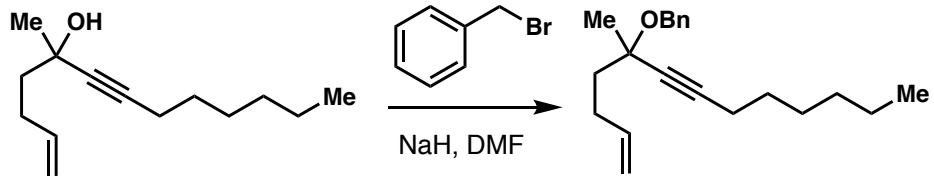
^{1}H NMR (500 MHz, $CDCl_3$) δ 5.88 (ddt, *J* = 16.9, 10.1, 6.6 Hz, 1H), 5.07 (dq, *J* = 17.1, 1.7 Hz, 1H), 4.98 (ddt, *J* = 10.2, 1.8, 1.3 Hz, 1H), 2.37 – 2.22 (m, 2H), 2.19 (t, *J* = 7.1 Hz, 2H), 1.96 – 1.91 (m, 1H), 1.80 – 1.66 (m, 2H), 1.55 – 1.48 (m, 2H), 1.47 (s, 3H), 1.43 – 1.35 (m, 2H), 1.35 – 1.24 (m, 3H), 0.90 (t, *J* = 7.0 Hz, 3H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 138.74, 114.81, 84.41, 83.80, 68.39, 43.03, 31.44, 30.46, 29.53, 28.81, 28.64, 22.69, 18.73, 14.18.

FTIR (NaCl, thin film, cm^{-1}): 3364, 3078, 2930, 2859, 2239, 1642, 1455, 1371, 1127.

HRMS (TOF-ESI, m/z): calc'd for $C_{14}H_{24}O$: 191.1800 $[M–H_2O+H]^+$; found: 191.1794.

(((5-methyltridec-1-en-6-yn-5-yl)oxy)methyl)benzene (32)



Prepared from **31** (1.04 g, 5.0 mmol, 1.0 equiv), benzyl bromide (0.66 mL, 5.5 mmol, 1.1 equiv), NaH (95%, 139 mg, 5.5 mmol, 1.1 equiv), and DMF (20 mL, 0.25 M) following General Procedure A. The resulting crude residue of **32** was used in the next step without further purification (1.46 g, 98% yield).

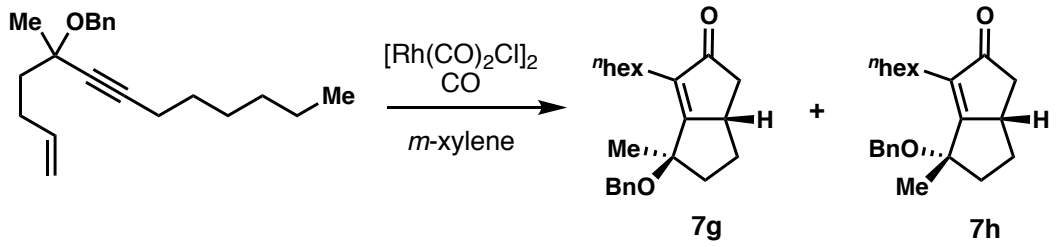
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.40 – 7.33 (m, 3H), 7.35 – 7.21 (m, 2H), 5.87 (ddt, *J* = 16.8, 10.2, 6.5 Hz, 1H), 5.04 (dq, *J* = 17.2, 1.7 Hz, 1H), 4.95 (ddt, *J* = 10.2, 2.3, 1.3 Hz, 1H), 4.67 (d, *J* = 11.2 Hz, 1H), 4.57 (d, *J* = 11.2 Hz, 1H), 2.42 – 2.25 (m, 1H), 2.23 (t, *J* = 7.0 Hz, 2H), 1.92 – 1.71 (m, 2H), 1.60 – 1.48 (m, 2H), 1.48 (s, 3H), 1.45 – 1.35 (m, 2H), 1.29 (ddtt, *J* = 10.0, 7.1, 5.1, 2.5 Hz, 3H), 0.96 – 0.82 (m, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 139.64, 138.90, 128.37, 127.73, 127.32, 114.37, 86.46, 81.41, 73.80, 66.17, 41.38, 31.45, 29.08, 28.91, 28.64, 26.97, 22.71, 18.79, 14.19.

FTIR (NaCl, thin film, cm^{-1}): 2931, 2860, 2237, 1455, 1088, 1062, 911, 731, 697.

HRMS (TOF-ESI, m/z): calc'd for $C_{21}H_{30}O$: 299.2369 $[M+H]^+$; found: 299.2378.

(4*R*,6a*S*)-4-(benzyloxy)-3-hexyl-4-methyl-4,5,6,6a-tetrahydropentalen-2(1*H*)-one (7g) and **(4*S*,6a*S*)-4-(benzyloxy)-3-hexyl-4-methyl-4,5,6,6a-tetrahydropentalen-2(1*H*)-one (7h)**



Prepared from **32** (1.19 g, 4.0 mmol, 1.0 equiv), $[Rh(CO)_2Cl]_2$ (77.7 mg, 0.2 mmol, 0.05 equiv), and *m*-xylene (40 mL, 0.10 M) following General Procedure D. The crude residue was purified by column chromatography (silica, 15 to 30% EtOAc/hexanes) to afford **7g** (693 mg, 53% yield) and **7h** (312 mg, 24% yield) as off-white oils.

(7g):

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 – 7.25 (m, 5H), 4.45 (d, *J* = 10.8 Hz, 1H), 4.36 (d, *J* = 10.7 Hz, 1H), 3.11 (d, *J* = 8.8 Hz, 1H), 2.68 (dd, *J* = 17.9, 6.4 Hz, 1H), 2.42 – 2.26 (m, 3H), 2.22 (dtd, *J* = 12.3, 8.1, 2.8 Hz, 1H), 2.08 (dd, *J* = 17.9, 3.3 Hz, 1H), 1.99 (ddd, *J* = 13.6, 8.9, 2.8 Hz, 1H), 1.66 (s, 3H), 1.56 (d, *J* = 1.1 Hz, 1H), 1.58 – 1.15 (m, 9H), 0.92 – 0.85 (m, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 211.84, 179.39, 139.80, 139.20, 129.01, 128.16, 128.07, 81.63, 66.98, 43.78, 43.18, 42.03, 32.21, 30.37, 29.45, 28.99, 24.37, 23.20, 23.13, 14.68.

FTIR (NaCl, thin film, cm^{-1}): 2930, 1713, 1652, 1462, 1086, 736, 697.

HRMS (TOF-ESI, m/z): calc'd for $C_{22}H_{30}O_2$: 327.2319 $[M+H]^+$; found: 327.2326.

(7h):

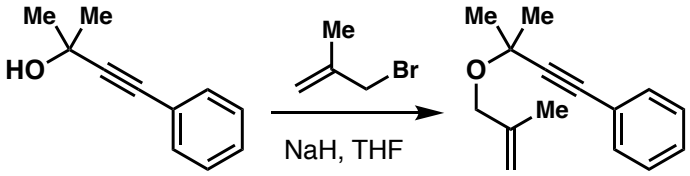
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.36 – 7.30 (m, 4H), 7.32 – 7.19 (m, 1H), 4.45 (d, *J* = 10.9 Hz, 1H), 4.26 (d, *J* = 11.2 Hz, 1H), 2.93 – 2.81 (m, 1H), 2.65 (dd, *J* = 17.9, 6.4 Hz, 1H), 2.47 – 2.39 (m, 1H), 2.39 – 2.25 (m, 2H), 2.19 – 2.09 (m, 3H), 1.95 (ddd, *J* = 14.5, 12.0, 7.4 Hz, 1H), 1.62 (s, 3H), 1.49 – 1.13 (m, 12H), 0.89 – 0.80 (m, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 210.71, 181.73, 138.82, 138.11, 128.50, 127.55, 127.34, 81.23, 65.78, 42.91, 42.04, 38.46, 31.80, 31.20, 29.82, 28.79, 27.48, 23.53, 23.37, 22.74, 14.23.

FTIR (NaCl, thin film, cm^{-1}): 2928, 2858, 1698, 1660, 1455, 1062, 733, 696.

HRMS (TOF-ESI, m/z): calc'd for $C_{22}H_{30}O_2$: 327.2319 $[M+H]^+$; found: 327.2314.

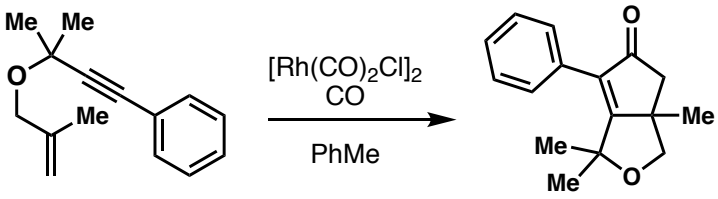
(3-methyl-3-((2-methylallyl)oxy)but-1-yn-1-yl)benzene (33)



Prepared from 2-methyl-4-phenylbut-3yn-2-ol (270 mg, 1.7 mmol, 1.0 equiv), methallyl bromide (0.3 mL, 3.4 mmol, 2.0 equiv), NaH (95%, 53.2 mg, 2.1 mmol, 1.25 equiv), and THF (3.4 mL, 0.50 M) following General Procedure A. The crude residue was purified by column chromatography (silica, 0 to 10% EtOAc/hexanes) to afford **33** (127 mg, 35% yield) as a white powder.

Spectral data matched those reported in the literature.[32]

1,1,3a-trimethyl-6-phenyl-3a,4-dihydro-1*H*-cyclopenta[*c*]furan-5(3*H*)-one (6f)



Prepared from **33** (60 mg, 0.28 mmol, 1.0 equiv), $[Rh(CO)_2Cl]_2$ (10.9 mg, 0.03 mmol, 0.1 equiv), and PhMe (1.4 mL, 0.20 M) following General Procedure D. The crude residue was purified by column chromatography (silica, 7% EtOAc/hexanes) to afford **6f** (21.7 mg, 32% yield) as a white powder.

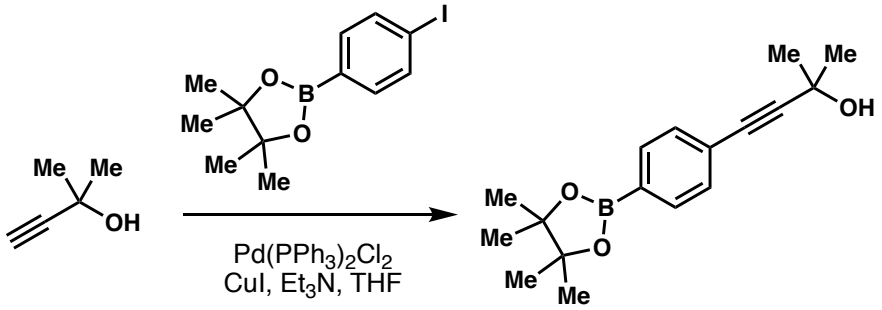
^{1}H NMR (400 MHz, 1,4-dioxane-*d*$_8$) δ 7.23 – 7.00 (m, 5H), 3.66 (d, *J* = 8.3 Hz, 1H), 3.25 (d, *J* = 8.3 Hz, 1H), 2.23 (q, *J* = 16.9 Hz, 2H), 1.46 (s, 3H), 1.21 (s, 3H), 0.77 (s, 3H).

^{13}C NMR (101 MHz, 1,4-dioxane-*d*$_8$) δ 206.54, 188.15, 135.69, 132.59, 130.11, 129.02, 128.94, 79.53, 75.42, 50.96, 49.25, 29.93, 26.00, 25.20.

FTIR (NaCl, thin film, cm^{-1}): 2974, 2928, 2852, 1713, 1652, 1235, 1137, 1018, 699.

HRMS (TOF-ESI, m/z): calc'd for $C_{16}H_{18}O_2$: 243.1380 $[M+H]^+$; found: 243.1385.

2-methyl-4-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)but-3-yn-2-ol (34)



Prepared from 2-methylbut-3-yn-2-ol (0.12 mL, 1.2 mmol, 1.0 equiv), 4-iodophenylboronic acid pinacol ester (450 mg, 1.4 mmol, 1.1 equiv), $Pd(PPh_3)_2Cl_2$ (8.7

mg, 0.012 mmol, 0.01 equiv), CuI (4.7 mg, 0.025 mmol, 0.02 equiv), and Et_3N (0.52 mL, 3.7 mmol, 3.0 equiv), and THF (1.2 mL, 1.00 M) following General Procedure B. The crude residue was purified by column chromatography (silica, 0 to 20% Et_2O/hexanes) to afford **34** (355 mg, 73% yield) as a pale yellow powder.

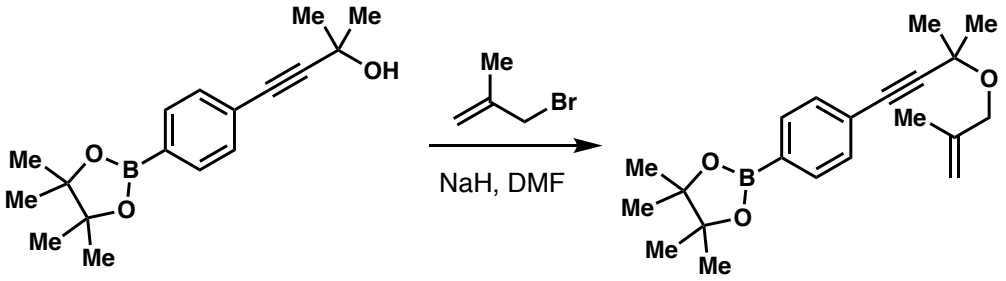
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.77 – 7.70 (m, 2H), 7.44 – 7.37 (m, 2H), 1.62 (s, 6H), 1.34 (s, 12H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 134.64, 130.96, 125.54, 114.80, 95.13, 84.11, 82.45, 65.81, 31.60, 25.03.

FTIR (NaCl, thin film, cm^{-1}): 3390, 2981, 2933, 1608, 1393, 1364, 1146, 1088, 963, 858, 658.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{23}BO_3$: 268.1744 $[M–H_2O+H]^+$; found: 268.1743.

4,4,5,5-tetramethyl-2-(4-(3-methyl-3-((2-methylallyl)oxy)but-1-yn-1-yl)phenyl)-1,3,2-dioxaborolane (35)



Prepared **34** (242 mg, 0.7 mmol, 1.0 equiv), methallyl bromide (0.7 mL, 1.7 mmol, 2.5 equiv), NaH (95%, 32 mg, 1.3 mmol, 1.9 equiv), and DMF (3.4 mL, 0.20 M) following General Procedure A. The crude residue was purified by column chromatography (silica, 0 to 20% EtOAc/hexanes) to afford **35** (75 mg, 33% yield) as a yellow powder.

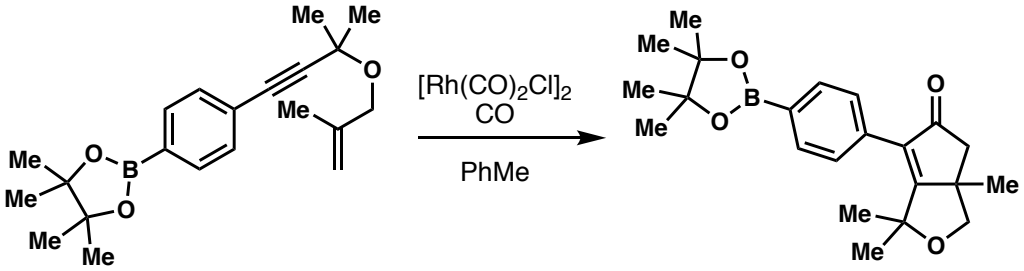
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.77 – 7.70 (m, 2H), 7.44 – 7.37 (m, 2H), 5.05 – 5.00 (m, 1H), 4.90 – 4.84 (m, 1H), 4.07 (s, 2H), 1.81 – 1.76 (m, 3H), 1.57 (s, 6H), 1.34 (s, 12H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 143.14, 134.65, 130.96, 125.77, 111.72, 93.02, 84.34, 84.10, 70.98, 68.61, 29.02, 25.02, 19.94.

FTIR (NaCl, thin film, cm^{-1}): 2981, 2931, 2856, 1608, 1398, 1360, 1323, 1144, 1089.

HRMS (TOF-ESI, m/z): calc'd for $C_{21}H_{29}BO_3$: 268.1744 $[M–C_4H_7O+H]^+$; found: 268.1741.

1,1,3a-trimethyl-6-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)-3a,4-dihydro-1*H*-cyclopenta[*c*]furan-5(3*H*)-one (6g)



Prepared from **35** (75.4 mg, 0.22 mmol), $[Rh(CO)_2Cl]_2$ (8.6 mg, 0.022 mmol, 0.1 equiv), and PhMe (1.1 mL, 0.20 M), following General Procedure D. The crude residue was purified by column chromatography (silica, 7% EtOAc/hexanes) to yield **6g** (25.1 mg, 31% yield) as an off-white powder.

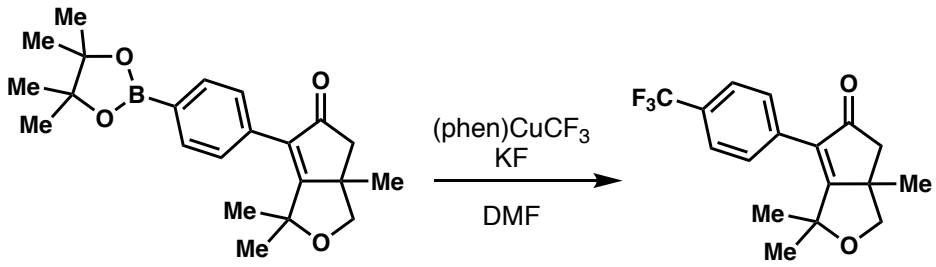
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.83 – 7.77 (m, 2H), 7.28 – 7.22 (m, 2H), 3.97 (d, J = 8.4 Hz, 1H), 3.53 (d, J = 8.4 Hz, 1H), 2.50 (s, 2H), 1.69 (s, 3H), 1.46 (s, 3H), 1.32 (s, 12H), 1.01 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 207.10, 188.65, 135.30, 134.74, 133.86, 128.44, 83.93, 79.26, 75.01, 50.40, 49.09, 29.47, 25.90, 24.99, 24.93, 24.88, 24.84, 14.22 (C_{Ar}–B not observed).

FTIR (NaCl, thin film, cm^{-1}): 2978, 2927, 2850, 1711, 1610, 1399, 1360, 1144, 1088, 1021.

HRMS (TOF-ESI, m/z): calc'd for $C_{22}H_{29}BO_4$: 368.2268 $[M+H]^+$; found: 368.2278.

1,1,3a-trimethyl-6-(4-(trifluoromethyl)phenyl)-3a,4-dihydro-1*H*-cyclopenta[*c*]furan-5(3*H*)-one (6h)



Prepared from **6g** (60.0 mg, 0.16 mmol, 1.0 equiv), (phen)$CuCF_3$ (61.1 mg, 0.2 mmol, 1.25 equiv), KF (9.46 mg, 0.16 mmol, 1.0 equiv), and DMF (1.63 mL, 0.10 M) at 100 ºC following a literature procedure.[33] The crude residue was purified by column chromatography (silica, 0 to 20% EtOAc/hex) to yield **6h** (24.5 mg, 48% yield) as an off-white powder.

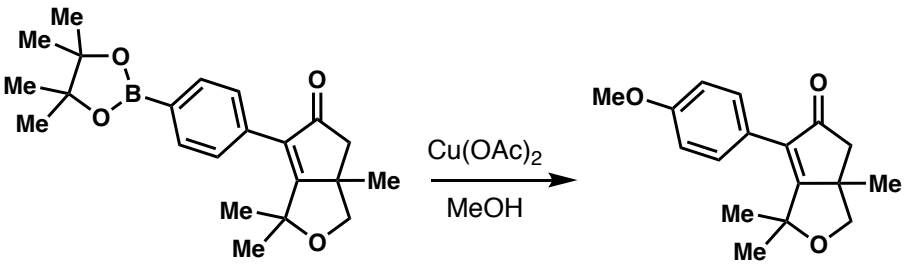
^{1}H NMR (400 MHz, 1,4-dioxane-*d*$_8$) δ 7.50 (d, *J* = 8.1 Hz, 2H), 7.24 (d, *J* = 8.1 Hz, 2H), 3.69 (d, *J* = 8.3 Hz, 1H), 3.28 (d, *J* = 8.4 Hz, 1H), 2.26 (q, *J* = 17.0 Hz, 2H), 1.47 (s, 3H), 1.22 (s, 3H), 0.78 (s, 3H).

^{13}C NMR (101 MHz, 1,4-dioxane-*d*$_8$) δ 206.01, 190.02, 136.45, 134.54, 130.69 (q, $^2J_{CF}$ = 32 Hz), 130.58, 126.05 (q, $^3J_{CF}$ = 4 Hz), 125.24 (q, $^1J_{CF}$ = 272 Hz), 79.53, 75.32, 51.30, 49.20, 29.75, 25.92, 25.36.

FTIR (NaCl, thin film, cm^{-1}): 2977, 2933, 2851, 1712, 1325, 1163, 1127, 1067, 1018.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{18}F_3O_2$: 311.1253 $[M+H]^+$; found: 311.1258.

6-(4-methoxyphenyl)-1,1,3a-trimethyl-3a,4-dihydro-1*H*-cyclopenta[*c*]furan-5(3*H*)-one (6i)



Prepared from **6g** (60 mg, 0.16 mmol, 1.0 equiv), $Cu(OAc)_2$ (3.0 mg, 0.016 mmol, 0.1 equiv), MeOH (0.8 mL, 0.20 M), and a balloon of O_2 following a literature procedure.[34] The crude residue was purified by column chromatography (silica, 0 to 20% EtOAc/hexanes) to yield **6i** (17 mg, 39% yield) as an off-white powder.

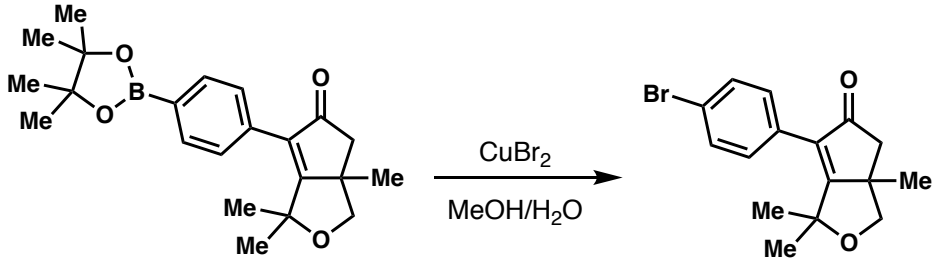
^{1}H NMR (400 MHz, 1,4-dioxane-*d*$_8$) δ 7.05 – 6.97 (m, 2H), 6.75 – 6.66 (m, 2H), 3.65 (d, *J* = 8.2 Hz, 1H), 3.23 (d, *J* = 8.1 Hz, 1H), 2.24 – 2.11 (m, 2H), 1.46 (s, 3H), 1.19 (s, 3H), 0.80 (s, 3H).

^{13}C NMR (101 MHz, 1,4-dioxane-*d*$_8$) δ 207.02, 186.72, 160.67, 135.20, 131.32, 124.68, 114.38, 79.53, 75.48, 55.34, 50.79, 49.16, 29.97, 26.03, 25.04.

FTIR (NaCl, thin film, cm^{-1}): 2974, 2934, 2852, 2342, 1706, 1512, 1249, 1018, 1031.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{20}O_3$: 273.1485 $[M+H]^+$; found: 273.1498.

6-(4-bromophenyl)-1,1,3a-trimethyl-3a,4-dihydro-1*H*-cyclopenta[*c*]furan-5(3*H*)-one (6j)



Prepared from **6g** (54 mg, 0.15 mmol, 1.0 equiv), $CuBr_2$ (32.7 mg, 0.15 mmol, 1.0 equiv), MeOH (1.8 mL, 0.04 M), and H_2O (1.8 mL, 0.04 M) at 100 ºC following a literature procedure.[35] The crude residue was purified by column chromatography (silica, 0 to 20% EtOAc/hexanes) to yield **6j** (26.8 mg, 57% yield) as an off-white powder.

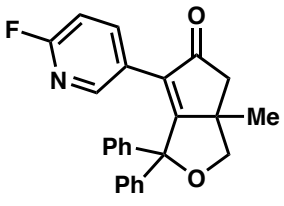
^{1}H NMR (400 MHz, 1,4-dioxane-*d*$_8$) δ 7.40 – 7.30 (m, 2H), 7.05 – 6.95 (m, 2H), 3.67 (d, *J* = 8.3 Hz, 1H), 3.25 (d, *J* = 8.3 Hz, 1H), 2.30 – 2.14 (m, 2H), 1.45 (s, 3H), 1.20 (s, 3H), 0.79 (s, 3H).

^{13}C NMR (101 MHz, 1,4-dioxane-*d*$_8$) δ 206.21, 188.84, 134.61, 132.28, 131.83, 131.52, 123.11, 79.52, 75.35, 51.12, 49.15, 29.82, 25.93, 25.22.

FTIR (NaCl, thin film, cm^{-1}): 2974, 2928, 2849, 1708, 1487, 1234, 1013.

HRMS (TOF-ESI, m/z): calc'd for $C_{16}H_{17}BrO_2$: 321.0485 $[M+H]^+$; found: 321.0475.

6-(6-fluoropyridin-3-yl)-3a-methyl-1,1-diphenyl-3a,4-dihydro-1*H*-cyclopenta[*c*]furan-5(3*H*)-one (6k)

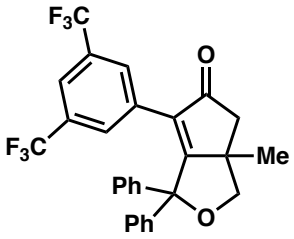


^{1}H NMR (400 MHz, $CDCl_3$) δ 7.81 (dt, *J* = 2.5, 0.8 Hz, 1H), 7.46 – 7.35 (m, 5H), 7.17 (ddd, *J* = 8.5, 7.6, 2.5 Hz, 1H), 7.06 – 6.93

(m, 3H), 6.84 – 6.76 (m, 2H), 6.59 (ddd, *J* = 8.4, 3.0, 0.7 Hz, 1H), 4.37 (d, *J* = 8.1 Hz, 1H), 4.17 (dd, *J* = 8.2, 0.8 Hz, 1H), 2.77 (d, *J* = 17.7 Hz, 1H), 2.68 (d, *J* = 17.6 Hz, 1H), 1.31 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 206.84, 186.03, 164.27, 161.87, 147.82, 147.67, 143.10, 142.03, 141.57, 141.49, 135.07, 128.66, 128.57, 127.93, 127.80, 127.71, 127.68, 124.09, 124.04, 108.79, 108.42, 87.95, 78.79, 50.28, 49.55, 27.34.

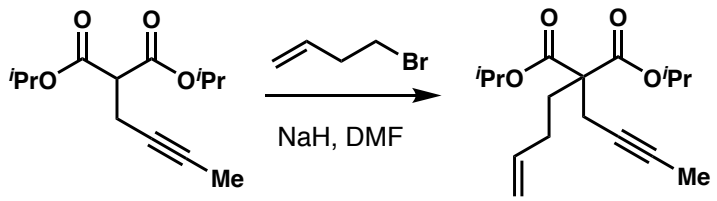
6-(3,5-bis(trifluoromethyl)phenyl)-3a-methyl-1,1-diphenyl-3a,4-dihydro-1*H*-cyclopenta[*c*]furan-5(3*H*)-one (6l)



^{1}H NMR (400 MHz, $CDCl_3$) δ 7.61 (tt, *J* = 1.6, 0.8 Hz, 1H), 7.50 – 7.37 (m, 5H), 7.30 (d, *J* = 1.8 Hz, 2H), 7.00 – 6.93 (m, 1H), 6.93 – 6.84 (m, 2H), 6.75 – 6.67 (m, 2H), 4.39 (d, *J* = 8.2 Hz, 1H), 4.23 – 4.16 (m, 1H), 2.81 (d, *J* = 17.6 Hz, 1H), 2.71 (d, *J* = 17.7 Hz, 1H), 1.32 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 206.16, 186.80, 142.83, 141.75, 136.76, 132.24, 131.50, 131.17, 130.84, 129.29, 128.66, 128.08, 127.68, 127.60, 127.52, 124.44, 121.90, 121.72, 87.88, 78.76, 50.27, 49.67, 27.29.

diisopropyl 2-(but-2-yn-1-yl)-2-(but-3-en-1-yl)malonate (36)



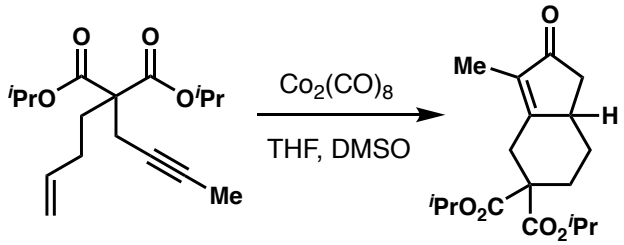
Prepared from **16** (361 mg, 1.5 mmol, 1.0 equiv), homoallyl bromide (0.23 mL, 2.25 mmol, 1.5 equiv), NaH (95%, 57.0 mg, 2.25 mmol, 1.5 equiv), and DMF (3 mL, 0.50 M) following

General Procedure A. The crude, clear colorless oil of **36** was used without further purification (417 mg, 93% yield).

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.81 (ddt, *J* = 16.7, 10.2, 6.4 Hz, 1H), 5.11 – 5.00 (m, 3H), 4.97 (dq, *J* = 10.1, 1.3 Hz, 1H), 2.73 (q, *J* = 2.6 Hz, 2H), 2.14 – 2.05 (m, 2H), 2.01 – 1.91 (m, 2H), 1.74 (t, *J* = 2.6 Hz, 3H), 1.22 (dd, *J* = 6.3, 1.7 Hz, 13H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 170.15, 137.89, 115.10, 78.60, 77.48, 77.16, 76.84, 73.63, 68.92, 56.91, 31.19, 28.48, 23.14, 21.71, 21.67, 3.61.

diisopropyl 3-methyl-2-oxo-1,2,4,6,7,7a-hexahydro-5*H*-indene-5,5-dicarboxylate (37)



Prepared from **36** (368 mg, 1.25 mmol, 1.0 equiv), $Co_2(CO)_8$ (513 mg, 1.5 mmol, 1.2 equiv), DMSO (0.89 mL, 12.5 mmol, 10.0 equiv), and THF (12.5 mL, 0.10 M) following General Procedure C. The crude residue was purified by column chromatography (silica, 55 to 65% Et_2O/hexanes) to afford **37** (165 mg, 41% yield) as a white, amorphous solid.

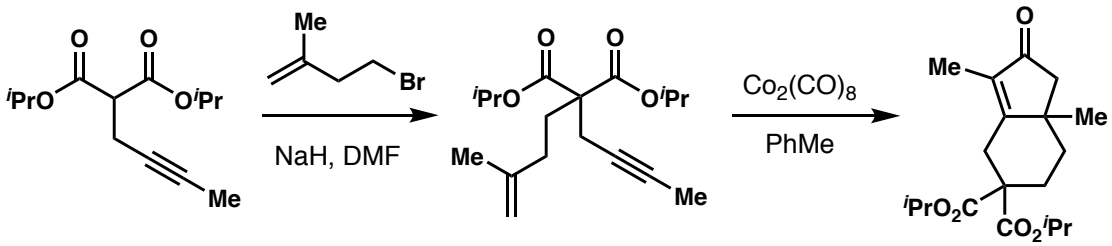
^{1}H NMR (400 MHz, $CDCl_3$) δ 5.04 (dhept, *J* = 22.4, 6.3 Hz, 2H), 3.49 (dd, *J* = 14.0, 2.3 Hz, 1H), 2.64 – 2.42 (m, 4H), 2.12 (ddt, *J* = 13.5, 5.4, 3.4 Hz, 1H), 1.99 – 1.87 (m, 2H), 1.75 (t, *J* = 1.6 Hz, 3H), 1.29 – 1.17 (m, 13H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 208.62, 170.97, 170.32, 169.41, 136.06, 77.48, 77.16, 76.84, 69.63, 69.18, 56.46, 41.00, 39.31, 33.25, 31.00, 30.70, 21.71, 21.70, 21.69, 21.67, 8.00.

FTIR (NaCl, thin film, cm^{-1}): 3446, 2980, 2932, 2873, 1723, 1714, 1696, 1453, 1373, 1254, 1107.

HRMS (GC-EI+, m/z): calc'd for $C_{18}H_{26}O_5$: 322.1780 $[M]^+$; found: 322.1782.

diisopropyl 3,7a-dimethyl-2-oxo-1,2,4,6,7,7a-hexahydro-5*H*-indene-5,5-dicarboxylate (38)



Prepared in two steps. Intermediate prepared from **16** (800 mg, 3.3 mmol, 1.0 equiv), 4-bromo-2-methylbut-1-ene (0.44 mL, 3.6 mmol, 1.1 equiv), NaH (95%, 151 mg, 6.0 mmol, 1.8 equiv), and DMF (4.15 mL, 0.80 M) following General Procedure A. The crude residue was purified by column chromatography (silica, 2% EtOAc/hexanes) to afford the crude intermediate (433 mg, 42% yield) as a colorless oil. The intermediate (308 mg, 1.0 mmol, 1.0 equiv) was then used directly in the next step with $Co_2(CO)_8$ (410 mg, 1.2 mmol, 1.2 equiv) and PhMe (10 mL, 0.10 M), following modified General Procedure C without addition of DMSO. Upon complete conversion of the enyne to the Co-alkyne complex, the reaction was placed in a preheated oil bath at 110 °C. Upon complete consumption of the Co-alkyne complex, the reaction was allowed to reach room temperature, filtered over a pad of SiO_2 and Celite, and the filtrate concentrated. The crude residue was purified by column chromatography (silica, 10 to 20 to 35% EtOAc/hexanes) to afford **38** (134 mg, 40% yield) as a clear oil.

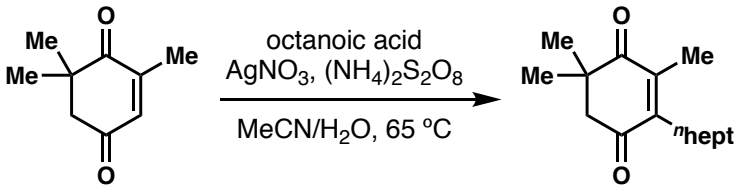
^{1}H NMR (400 MHz, $CDCl_3$) δ 5.03 (dp, *J* = 26.7, 6.3 Hz, 2H), 3.35 (dd, *J* = 14.0, 2.3 Hz, 1H), 2.67 (dd, *J* = 14.0, 1.7 Hz, 1H), 2.39 – 2.31 (m, 1H), 2.31 – 2.12 (m, 2H), 2.06 (td, *J* = 14.1, 4.0 Hz, 1H), 1.87 (ddd, *J* = 13.6, 3.9, 2.8 Hz, 1H), 1.72 (d, *J* = 1.4 Hz, 3H), 1.51 (td, *J* = 13.8, 4.0 Hz, 1H), 1.29 – 1.17 (m, 16H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 207.88, 173.62, 170.97, 169.30, 135.44, 77.48, 77.16, 76.84, 69.63, 69.21, 56.86, 51.02, 40.64, 36.46, 29.74, 27.66, 24.56, 21.71, 21.68, 8.16.

FTIR (NaCl, thin film, cm^{-1}): 3456, 3394, 2980, 2924, 2870, 2049, 2021, 1731, 1704, 1659, 1456.

HRMS (GC-EI+, m/z): calc'd for $C_{19}H_{28}O_5$: 336.1937 $[M]^+$; found: 336.1919.

2-heptyl-3,5,5-trimethylcyclohex-2-ene-1,4-dione (39)



To a dry round-bottomed flask with a Teflon-coated stir bar was added ketoisophorone (0.74 mL, 5.0 mmol, 1.0 equiv), MeCN (15 mL), and H_2O (3 mL). Octanoic acid (1.98 mL, 12.5 mmol, 2.5 equiv) was added, and the mixture was stirred placed in a preheated oil bath at 65 °C. Silver nitrate (170 mg, 1.00 mmol, 0.20 equiv) was added in a single portion, and the flask was sealed with a rubber septum and flushed with N_2. Ammonium persulfate (1.49 g, 6.50 mmol, 1.3 equiv) was added as a solution in MeCN (15 mL) and H_2O (12 mL) (0.1 M final concentration) over 1.5 h via syringe pump. Upon complete consumption of the ketoisophorone as judged by TLC, the reaction was allowed to reach room temperature and concentrated to approximately ½ volume. The resulting liquid was

extracted three times with EtOAc. Combined organics were washed with sat. aq. $NaHCO_3$ and brine, dried over Na_2SO_4, filtered, and concentrated. The crude residue was purified by column chromatography (silica, 5 to 10% Et_2O/hexanes) to afford **39** (380 mg, 31% yield) as a clear oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ 2.70 (s, 2H), 2.41 (dd, *J* = 8.3, 6.4 Hz, 2H), 1.99 (s, 3H), 1.39 – 1.24 (m, 10H), 1.21 (s, 6H), 0.90 – 0.84 (m, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 203.74, 197.83, 149.20, 143.45, 77.48, 77.16, 76.84, 52.01, 45.23, 31.87, 30.01, 29.22, 28.46, 27.05, 26.38, 22.76, 14.22, 13.19.

FTIR (NaCl, thin film, cm^{-1}): 3340, 2957, 2923, 2856, 1682, 1614, 1469, 1378, 1276.

HRMS (GC-FAB+, m/z): calc'd for $C_{16}H_{27}O_2$: 251.2011 $[M+H]^+$; found: 251.2001.

1.8.4 SEO_2-MEDIATED OXIDATIVE TRANSPOSITIONS

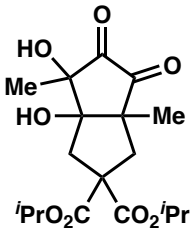
Reaction Optimization: General Procedure F

To a flame-dried vial with a Teflon-coated stir bar was added **6a** or **7a** (1.0 equiv), SeO_2, and 4 Å molecular sieves (activated by flame-drying under high vacuum for 10 minutes), if applicable. To this mixture was added 1,4-dioxane (0.05 M) then H_2O, if applicable. The reaction was capped and sealed with Teflon tape, stirred, and brought to 100 °C in a preheated aluminum reaction block. The reaction was monitored by LCMS. When the reaction was judged to be complete, it was allowed to reach room temperature then concentrated. To the crude residue was added pyrazine as an internal standard, then the mixture was taken up in CD_3OD and filtered over a cotton plug into an NMR tube. Yield determined by ^{1}H NMR versus the internal standard.

Note: *The reactions were judged to be complete when the greatest amount of starting material had been consumed and the least amount of the desired product was degraded by further oxidation. This is particularly important for the trioxidation; usually, after two-thirds of the starting material converted, rapid degradation of the desired product was observed.*

diisopropyl 3a,4-dihydroxy-4,6a-dimethyl-5,6-dioxohexahydropentalene-2,2(1*H*)-dicarboxylate (41)

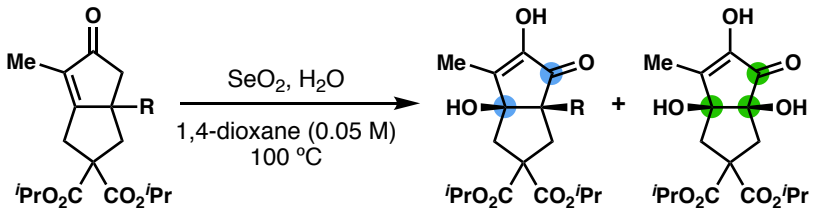
Observed to be a major byproduct during optimization of **6a** subjected to dioxidation conditions, generally in 10–30% yield (by ^{1}H NMR versus an internal standard).



^{1}H NMR (400 MHz, CD_3OD) δ 5.08 (dp, *J* = 12.5, 6.2 Hz, 1H), 4.87 (p, *J* = 6.3 Hz, 1H), 3.01 (dd, *J* = 15.6, 1.2 Hz, 1H), 2.78 – 2.69 (m, 1H), 2.67 – 2.55 (m, 2H), 1.97 (s, 3H), 1.29 – 1.20 (m, 9H), 1.15 (dd, *J* = 7.2, 6.3 Hz, 5H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 202.54, 190.10, 173.51, 168.62, 102.83, 86.53, 77.36, 71.47, 69.94, 57.86, 55.94, 45.12, 42.98, 21.69, 21.53, 21.40, 18.75, 14.34.

Table 1.6. *Optimization of the di- and trioxidation reactions*



6a, R = Me
7a, R = H

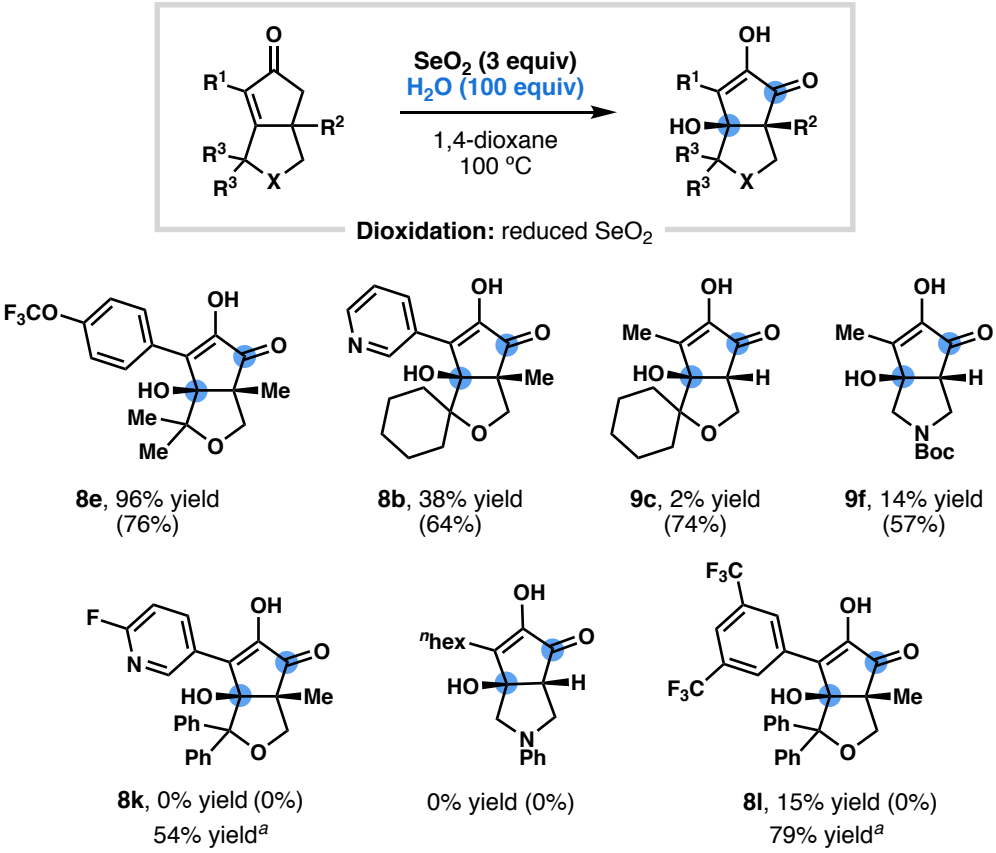
8a, R = Me
9a, R = H

10a

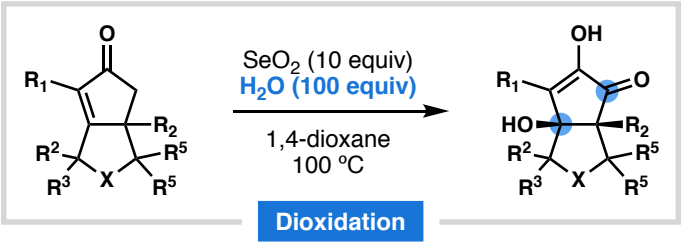
Entry	R	SeO_2 (equiv)	H_2O (equiv)	Time (h)	8a *or* 9a (%)[a]	10a (%)[a]	RSM (%)[a]	Scale (mmol)
1	Me	10	0	3	11	—	39	0.10
2[b]	Me	10	0	35	0	—	100	0.05
3	Me	10	10	24	13	—	2	0.03
4	Me	10	100	38	71[c]	—	0	1.00
5	Me	3	30	12	58	—	28	0.03
6	Me	3	30	24	32	—	0	0.03
7	Me	3	60	24	54	—	20	0.03
8	Me	3	100	48	62	—	24	0.03
9	Me	3	100	120	52	—	0	0.03
10	Me	1.5	15	24	29	—	25	0.03
11	Me	1.5	100	48	37	—	51	0.03
12	Me	1.5	100	120	42	—	39	0.03
13	H	10	0	2	0	21[c]	ND	0.02
14	H	10	10	1	0	15	ND	0.02
15	H	10	100	4	64[c]	0	ND	0.03
16	H	10	100	46	0	14	ND	0.07
17	H	3	0	3	0	10	ND	0.05
18	H	3	30	6	18	17	ND	0.02
19	H	3	100	14	7	15	ND	0.02

[a]Yield determined by ^{1}H NMR versus an internal standard. [b]Reaction performed with 4 Å molecular sieves. [c]Isolated yield.

Table 1.7. *Exploration of the generality of the dioxidation with reduced SeO_2 (3.0 equiv)*



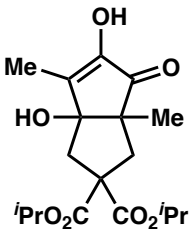
Yields determined by ^{1}H NMR versus an internal standard. Yield in parentheses obtained under standard dioxidation conditions (SeO_2, 10.0 equiv). [a]Yield obtained under alternative dioxidation protocol conditions: SeO_2 (1.5 equiv), 1,4-dioxane (0.05 M), 100 °C, 18 h; *then* H_2O (100.0 equiv), 0.5–3 h.



Dioxidation: General Procedure G

To a round-bottomed flask with a Teflon-coated stir bar was added the enone (**6a–e** or **7a–g**) (1.0 equiv), SeO_2 (10.0 equiv), 1,4-dioxane (0.05 M), and water (100.0 equiv). The reaction was sealed with a rubber septum, stirred, and brought to 100 °C in a preheated oil bath and monitored by TLC and/or LCMS. After the reaction was judged to be complete, typically 24–48 h (see note in General Procedure F), the reaction was allowed to reach room temperature and filtered over a pad of celite, washing with EtOAc. The filtrate was washed with sat. aq. $NaHCO_3$ twice then with water. The combined aqueous washings were extracted with EtOAc. The combined organics were dried over Na_2SO_4, filtered, and concentrated. The crude residue was purified by column chromatography to afford the desired product. Alternatively, some compounds could be purified by trituration.

diisopropyl 5,6a-dihydroxy-3a,6-dimethyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (8a)



Prepared from **6a** (323 mg, 1.0 mmol) following General Procedure G. The crude residue was purified by column chromatography (silica, 30 to 35% EtOAc/hexanes) to yield **8a** (251 mg, 71% yield) as an off-white powder.

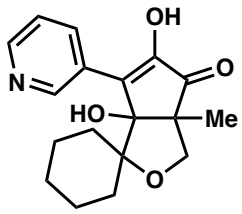
^{1}H NMR (400 MHz, CD_3OD) δ 5.08 (hept, *J* = 6.3 Hz, 1H), 4.97 (hept, *J* = 6.3 Hz, 1H), 3.36 (dq, *J* = 17.6, 1.9 Hz, 1H), 3.17 (d, *J* = 17.5 Hz, 1H), 2.55 (d, *J* = 13.6 Hz, 1H), 2.37 (s, 2H), 2.13 (d, *J* = 13.7 Hz, 1H), 1.67 (d, *J* = 1.7 Hz, 3H), 1.28 (dd, *J* = 6.2, 1.9 Hz, 6H), 1.21 (dd, *J* = 6.2, 2.9 Hz, 6H), 1.12 (s, 3H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 203.71, 171.01, 169.52, 148.40, 141.75, 85.25, 69.93, 69.61, 58.23, 53.88, 42.74, 42.42, 21.61, 21.60, 21.57, 21.47, 19.31, 9.09.

FTIR (NaCl, thin film, cm^{-1}): 3444, 2981, 2937, 1731, 1455, 1372, 1266, 1106, 912, 823.

HRMS (TOF-ESI, m/z): calc'd for $C_{18}H_{26}O_7$: 337.1646 $[M–H_2O+H]^+$; found: 337.1658.

5',6a'-dihydroxy-3a'-methyl-6'-(pyridin-3-yl)-3a',6a'-dihydrospiro[cyclohexane-1,1'-cyclopenta[*c*]furan]-4'(3'*H*)-one (8b)



Prepared from **6b** (14.4 mg, 0.05 mmol) following General Procedure G. The crude residue was purified by column chromatography (silica, 5% MeOH/DCM) to yield **8b** (10 mg, 64% yield) as an off-white powder.

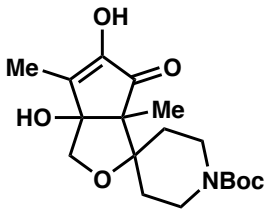
^{1}H NMR (400 MHz, pyridine-d_5) δ 10.25 (dd, *J* = 2.3, 0.9 Hz, 1H), 9.11 (dt, *J* = 8.1, 1.9 Hz, 1H), 8.70 (dd, *J* = 4.7, 1.7 Hz, 1H), 7.47 (s, 1H), 7.37 (ddd, *J* = 8.2, 4.7, 0.9 Hz, 1H), 3.76 (d, *J* = 9.4 Hz, 1H), 2.44 – 2.34 (m, 1H), 1.95 (td, *J* = 13.8, 4.1 Hz, 1H), 1.79 – 1.52 (m, 5H), 1.48 (s, 3H), 1.08 (qt, *J* = 13.1, 4.1 Hz, 1H).

^{13}C NMR (101 MHz, pyridine-d_5) δ 206.56, 153.09, 151.95, 150.63, 149.38, 137.32, 136.22, 135.72, 135.47, 133.14, 124.21, 123.72, 89.73, 86.24, 72.20, 58.90, 36.94, 29.16, 26.38, 23.93, 22.64, 17.80.

FTIR (NaCl, thin film, cm^{-1}): 3335, 2911, 2851, 2639, 2544, 2355, 2336, 2004, 1664, 1539.

HRMS (TOF-ESI, m/z): calc'd for $C_{18}H_{21}NO_4$: 316.1543 $[M+H]^+$; found: 316.1552.

***tert*-butyl 3a,5-dihydroxy-4,6a-dimethyl-6-oxo-3,3a,6,6a-tetrahydrospiro[cyclopenta[*c*]furan-1,4'-piperidine]-1'-carboxylate (8c)**



Prepared from **6c** (161 mg, 0.5 mmol) following General Procedure G. The crude residue was purified by column chromatography (silica, 5 to 7% MeOH/DCM) to yield **8c** (94 mg, 53% yield) as off-white powder.

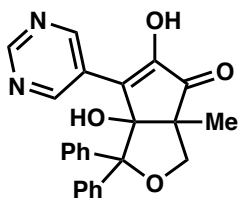
^{1}H NMR (400 MHz, $CDCl_3$) δ 3.95 (d, *J* = 10.0 Hz, 1H), 3.64 (d, *J* = 10.1 Hz, 1H), 2.94 (s, 2H), 1.99 (s, 3H), 1.92 – 1.74 (m, 2H), 1.74 – 1.58 (m, 2H), 1.45 (s, 10H), 1.38 – 1.27 (m, 1H), 1.02 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 200.79, 153.90, 147.50, 139.31, 85.58, 85.35, 81.65, 78.73, 69.57, 58.29, 32.10, 28.84, 27.58, 27.15, 13.06, 8.20.

FTIR (NaCl, thin film, cm^{-1}): 3367, 2975, 2933, 2874, 1682, 1428, 1367, 1289, 1251.

HRMS (TOF-ESI, m/z): calc'd for $C_{18}H_{27}NO_6$: 354.1911 $[M+H]^+$; found: 354.1926.

5,6a-dihydroxy-3a-methyl-1,1-diphenyl-6-(pyrimidin-5-yl)-1,3,3a,6a-tetrahydro-4*H*-cyclopenta[*c*]furan-4-one (8d)



Prepared from **6d** (18.4 mg, 0.05 mmol) following General Procedure G. The crude residue was purified by trituration from DCM and pentanes to yield **8d** (7 mg, 35% yield) as an off-white powder.

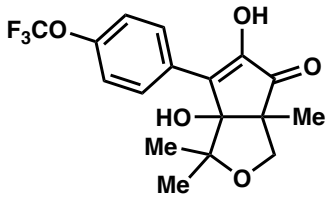
^{1}H NMR (400 MHz, pyridine-d_4) δ 9.12 (s, 1H), 8.54 (s, 2H), 7.78 (dd, *J* = 6.7, 3.0 Hz, 2H), 7.57 – 7.50 (m, 2H), 7.39 – 7.30 (m, 3H), 7.30 – 7.23 (m, 3H), 4.48 (d, *J* = 8.8 Hz, 1H), 3.87 (d, *J* = 8.8 Hz, 1H), 2.25 (s, 3H).

^{13}C NMR (101 MHz, pyridine-d_4) δ 202.69, 195.78, 186.85, 183.09, 182.92, 176.74, 176.47, 176.20, 169.54, 169.07, 162.36, 162.12, 161.87, 159.35, 157.26, 156.51, 155.15, 154.99, 154.86, 154.46, 154.26, 149.86, 121.19, 103.12, 83.26, 49.23, 26.56.

FTIR (NaCl, thin film, cm^{-1}): 3062, 2924, 2870, 2537, 2249, 1714, 1562, 1446, 1413, 1230, 1060.

HRMS (TOF-ESI, m/z): calc'd for $C_{24}H_{20}N_2O_4$: 401.1496 $[M+H]^+$; found: 401.1503.

5,6a-dihydroxy-1,1,3a-trimethyl-6-(4-(trifluoromethoxy)phenyl)-1,3,3a,6a-tetrahydro-4*H*-cyclopenta[*c*]furan-4-one (8e)



Prepared from **6e** (164 mg, 0.5 mmol) following General Procedure G. The crude residue was purified by trituration from DCM and pentanes to yield **8e** (135 mg, 76% yield) as an off-white powder.

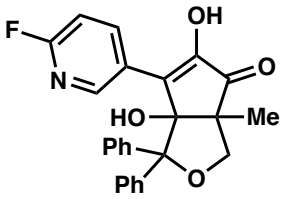
^{1}H NMR (500 MHz, $CDCl_3$) δ 7.27 (d, *J* = 9.2 Hz, 10H), 6.25 (dt, *J* = 8.1, 1.0 Hz, 2H), 5.53 (s, 1H), 3.06 (d, *J* = 9.7 Hz, 1H), 2.71 – 2.62 (m, 1H), 1.39 (s, 1H), 0.39 (s, 3H), 0.23 (s, 3H), 0.02 (s, 3H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 204.08, 155.58, 149.35, 148.19, 135.62, 131.93, 131.67, 120.43, 89.18, 84.53, 71.45, 58.21, 26.87, 21.58, 16.51.

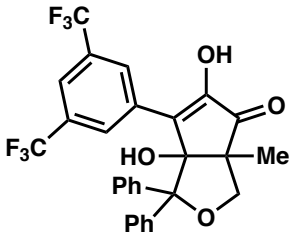
FTIR (NaCl, thin film, cm^{-1}): 3310, 2979, 2371, 2348, 1719, 1701, 1388, 1260, 1198.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{17}F_3O_5$: 359.1101 $[M+H]^+$; found: 359.1114.

6-(6-fluoropyridin-3-yl)-5,6a-dihydroxy-3a-methyl-1,1-diphenyl-1,3,3a,6a-tetrahydro-4*H*-cyclopenta[*c*]furan-4-one (8k)

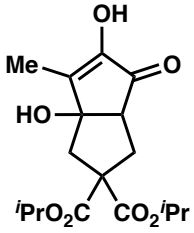


6-(3,5-bis(trifluoromethyl)phenyl)-5,6a-dihydroxy-3a-methyl-1,1-diphenyl-1,3,3a,6a-tetrahydro-4*H*-cyclopenta[*c*]furan-4-one (8l)



^{1}H NMR (400 MHz, $CDCl_3$) δ 7.31 (s, 1H), 7.27 (d, *J* = 4.5 Hz, 4H), 7.25 – 7.18 (m, 1H), 7.17 – 7.11 (m, 3H), 6.65 – 6.51 (m, 2H), 6.47 – 6.39 (m, 2H), 4.13 (d, *J* = 9.1 Hz, 1H), 3.77 (d, *J* = 9.1 Hz, 1H), 1.07 (s, 3H).

Diisopropyl 5,6a-dihydroxy-6-methyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (9a)



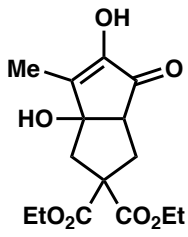
Prepared from **7a** (154.1 mg, 0.5 mmol) following General Procedure G. The crude residue was purified by column chromatography (silica, 60% EtOAc/hexanes) to yield **9a** (132 mg, 78% yield) as an off-white powder. **^{1}H NMR (400 MHz, CD_3OD) δ** 4.98 (p, J = 6.3 Hz, 1H), 4.83 (dt, *J* = 12.6, 6.3 Hz, 1H), 2.72 (dd, *J* = 13.6, 1.6 Hz, 1H), 2.56 – 2.39 (m, 3H), 2.31 (d, *J* = 13.5 Hz, 1H), 1.88 (d, *J* = 0.6 Hz, 3H), 1.27 – 1.11 (m, 12H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 200.52, 171.93, 169.63, 147.90, 141.77, 85.26, 70.26, 69.70, 61.98, 55.66, 43.27, 34.53, 21.74, 21.73, 21.69, 21.63, 9.29.

FTIR (NaCl, thin film, cm^{-1}): 3431, 2988, 2936, 2742, 1732, 1455, 1377, 1168, 1022, 985, 901, 832, 758.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{24}O_7$: 323.1489 $[M–H_2O+H]^+$; found: 323.1480.

diethyl 5,6a-dihydroxy-6-methyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (9b)



Prepared from **7b** (14.1 mg, 0.05 mmol) following General Procedure G. The crude residue was purified by column chromatography (silica, 60% EtOAc/hexanes) to yield **9b** (12.7 mg, 82% yield) as an off-white powder.

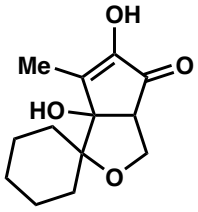
^{1}H NMR (500 MHz, CD_3OD) δ 5.73 (s, 1H), 4.25 (qd, *J* = 7.1, 2.6 Hz, 2H), 4.13 (q, *J* = 7.1 Hz, 2H), 2.98 (s, 1H), 2.79 (dd, *J* = 10.5, 5.3 Hz, 1H), 2.64 (ddd, *J* = 13.7, 10.4, 0.9 Hz, 1H), 2.61 – 2.50 (m, 2H), 2.43 (ddd, *J* = 13.8, 5.3, 0.8 Hz, 1H), 2.01 (d, *J* = 0.6 Hz, 3H), 1.30 (t, *J* = 7.1 Hz, 4H), 1.24 (t, *J* = 7.1 Hz, 3H).

^{13}C NMR (101 MHz, CD_3OD) δ 202.81, 171.92, 171.73, 151.49, 144.09, 84.55, 63.03, 62.97, 61.63, 55.97, 43.22, 35.59, 14.27, 14.22, 8.93.

FTIR (NaCl, thin film, cm^{-1}): 3381, 2984, 2940, 1727, 1714, 1672, 1437, 1405, 1370, 1262.

HRMS (TOF-ESI, m/z): calc'd for $C_{15}H_{20}O_7$: 311.1136 $[M–H]^-$; found: 311.1122.

5',6a'-dihydroxy-6'-methyl-3a',6a'-dihydrospiro[cyclohexane-1,1'-cyclopenta[*c*]furan]-4'(3'*H*)-one (9c)



Prepared from **7c** (104 mg, 0.5 mmol) following General Procedure G. The crude residue was purified by column chromatography (silica, 50% EtOAc/hexanes) to yield **9c** (87 mg, 74% yield) as an off-white powder.

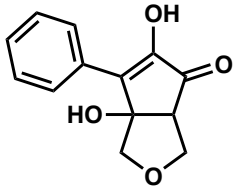
^{1}H NMR (400 MHz, $CDCl_3$) δ 4.00 (dd, *J* = 9.8, 8.6 Hz, 1H), 3.79 (dd, *J* = 9.8, 3.9 Hz, 1H), 2.89 – 2.81 (m, 1H), 2.07 – 1.94 (m, 4H), 1.73 – 1.60 (m, 2H), 1.60 – 1.38 (m, 4H), 1.34 – 1.09 (m, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 199.38, 148.53, 140.21, 87.16, 82.89, 62.25, 58.42, 31.52, 28.00, 24.58, 21.43, 20.97, 9.74.

FTIR (NaCl, thin film, cm^{-1}): 3342, 2928, 2858, 1713, 1698, 1668, 1402, 1361, 1133. 1060, 1022.

HRMS (TOF-ESI, m/z): calc'd for $C_{13}H_{18}O_4$: 239.1278 $[M+H]^+$; found: 239.1282.

5,6a-dihydroxy-6-phenyl-1,3,3a,6a-tetrahydro-4*H*-cyclopenta[*c*]furan-4-one (9d)



Prepared from **7d** (500 mg, 2.5 mmol) following General Procedure G. The crude residue was purified by trituration from DCM and pentanes to yield **9d** (397.2 mg, 73% yield) as an off-white powder.

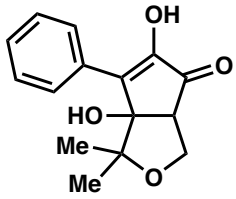
^{1}H NMR (400 MHz, CD_3OD) δ 8.18 – 8.11 (m, 2H), 7.46 – 7.38 (m, 2H), 7.38 – 7.30 (m, 1H), 4.12 (d, *J* = 9.2 Hz, 1H), 4.00 – 3.90 (m, 2H), 3.75 (d, *J* = 9.2 Hz, 1H), 2.74 (dd, *J* = 5.6, 3.5 Hz, 1H).

^{13}C NMR (101 MHz, CD_3OD) δ 202.00, 151.69, 139.33, 133.66, 130.60, 129.82, 129.29, 85.13, 76.95, 69.52, 58.97.

FTIR (NaCl, thin film, cm^{-1}): 3339, 2880, 2753, 1680, 1501, 1369, 1268, 1016, 698.

HRMS (TOF-ESI, m/z): calc'd for $C_{13}H_{12}O_4$: 233.0808 $[M+H]^+$; found: 233.0802.

5,6a-dihydroxy-1,1-dimethyl-6-phenyl-1,3,3a,6a-tetrahydro-4*H*-cyclopenta[*c*]furan-4-one (9e)



Prepared from **7e** (22.8 mg, 0.11 mmol) following General Procedure G. The crude residue was purified by column chromatography (silica, 20% EtOAc/hexanes) to yield **9e** (8.0 mg, 32% yield) as an off-white powder.

^{1}H NMR (400 MHz, CD_3OD) δ 8.27 – 8.19 (m, 2H), 7.45 – 7.35 (m, 2H), 4.17 – 4.08 (m, 1H), 3.82 (dd, *J* = 9.4, 3.6 Hz, 1H), 2.94 (dd, *J* = 8.2, 3.6 Hz, 1H), 1.40 (d, *J* = 4.0 Hz, 3H), 0.92 (s, 5H).

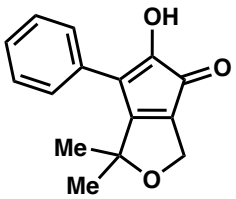
^{13}C NMR (101 MHz, CD_3OD) δ 202.61, 150.95, 139.36, 135.53, 130.86, 129.53, 128.92, 88.78, 85.74, 64.87, 61.51, 26.00, 22.64.

FTIR (NaCl, thin film, cm^{-1}): 3388, 2925, 2855, 1694, 1385, 1204, 1062.

HRMS (TOF-ESI, m/z): calc'd for $C_{15}H_{16}O_4$: 261.1121 $[M+H]^+$; found: 261.1122.

5-hydroxy-1,1-dimethyl-6-phenyl-1,3-dihydro-4*H*-cyclopenta[*c*]furan-4-one (40)

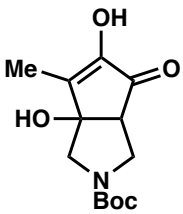
Prepared from **7e** (30.6 mg, 0.1 mmol) following General Procedure G. The crude residue was purified by column chromatography (silica, 30% EtOAc/hexanes) then preparative thin layer chromatography (silica, 10% MeOH/DCM) to yield **40** (7.5 mg, 30% yield) as an off-white powder.



^{1}H NMR (400 MHz, $CDCl_3$) δ 8.03 – 7.86 (m, 2H), 7.60 (ddt, *J* = 7.9, 6.9, 1.3 Hz, 1H), 7.57 – 7.42 (m, 2H), 4.89 (s, 2H), 1.40 (s, 6H).

^{13}C NMR (101 MHz, $CDCl_3$) δ ^{13}C NMR (101 MHz, $CDCl_3$) δ 193.41, 165.32, 157.15, 135.83, 134.29, 129.33, 128.91, 128.47, 91.27, 71.82, 29.86, 26.75.

tert-butyl 5,6a-dihydroxy-6-methyl-4-oxo-3,3a,4,6a-tetrahydrocyclopenta[_c_]pyrrole-2(1_H_)-carboxylate (9f)



Prepared from **7f** (24 mg, 0.1 mmol) following General Procedure G. The crude residue was purified by column chromatography (silica, 75% EtOAc/hexanes) to yield **9f** (15 mg, 57% yield) as an off-white powder.

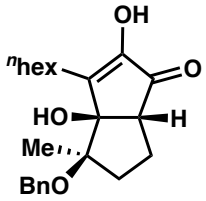
^{1}H NMR (400 MHz, CD_3OD) δ 3.81 (d, *J* = 11.9 Hz, 1H), 3.73 (dd, *J* = 11.7, 1.9 Hz, 1H), 3.49 (t, *J* = 10.2 Hz, 1H), 3.16 (d, *J* = 11.9 Hz, 1H), 2.55 (dd, *J* = 8.5, 1.9 Hz, 1H), 1.98 – 1.93 (m, 3H), 1.41 (s, 9H).

^{13}C NMR (101 MHz, CD_3OD, asterisk denotes minor rotamer) δ 201.58, 156.39, 151.57, 144.65*, 144.40*, 83.96*, 83.31*, 81.61, 56.88*, 56.22*, 55.31*, 54.75*, 47.87*, 47.42*, 28.65, 9.22*.

FTIR (NaCl, thin film, cm^{-1}): 3446, 2862, 2352, 1698, 1668, 1660, 1634, 1436, 1224.

HRMS (TOF-ESI, m/z): calc'd for $C_{13}H_{19}O_5$: 270.1336 $[M+H]^+$; found: 270.1341.

(3a*R*,4*R*,6a*R*)-4-(benzyloxy)-3-hexyl-2,3a-dihydroxy-4-methyl-4,5,6,6a-tetrahydropentalen-1(3a*H*)-one (9g)



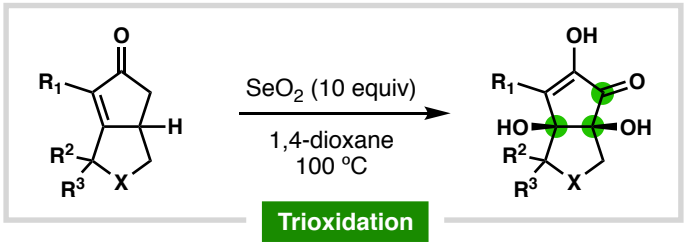
Prepared from **7g** (6.5 mg, 0.02 mmol) following General Procedure G. The crude residue was purified by preparative thin layer chromatography (silica, 40% EtOAc/hexanes) to yield **9g** (1.7 mg, 24% yield) as a white amorphous solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.44 – 7.28 (m, 5H), 5.63 (s, 1H), 4.56 (d, *J* = 10.9 Hz, 1H), 4.44 (d, *J* = 10.9 Hz, 1H), 4.23 (d, *J* = 0.8 Hz, 1H), 2.76 – 2.68 (m, 1H), 2.52 – 2.38 (m, 1H), 2.38 – 2.23 (m, 1H), 2.23 – 2.11 (m, 1H), 2.11 – 1.98 (m, 1H), 1.77 – 1.59 (m, 5H), 1.42 – 1.25 (m, 5H), 1.24 (s, 3H), 0.93 – 0.81 (m, 4H).

^{13}C NMR (101 MHz, $CDCl_3$) δ ^{13}C NMR (101 MHz, $CDCl_3$) δ 202.81, 149.56, 145.27, 138.17, 128.69, 127.98, 127.70, 87.38, 84.14, 65.43, 56.15, 35.06, 31.73, 30.07, 27.66, 26.78, 22.75, 22.69, 19.59, 14.25.

FTIR (NaCl, thin film, cm^{-1}): 3388, 2925, 2855, 1694, 1385, 104, 1062.

HRMS (TOF-ESI, m/z): calc'd for $C_{22}H_{30}O_4$: 359.2217 $[M+H]^+$; found: 359.2205.

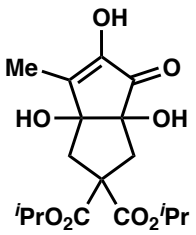


Trioxidation: General Procedure H

To a flame-dried vial with a Teflon-coated stir bar was added the enone (**7a–h**) (1.0 equiv), SeO_2 (10.0 equiv), and 1,4-dioxane (0.05 M). The reaction was capped and sealed with

Teflon tape, stirred, and brought to 100 °C in a preheated aluminum reaction block and monitored by TLC and/or LCMS. After the reaction was judged to be complete, typically 1–4 h (see note in General Procedure F), the reaction was allowed to reach room temperature then concentrated. The crude residue was purified by chromatography to afford the desired product.

Diisopropyl 3a,5,6a-trihydroxy-6-methyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (10a)



Prepared from **7a** (7.0 mg, 0.02 mmol) following General Procedure H. The crude residue was purified by preparative thin layer chromatography (silica, 70% EtOAc/hexanes) to yield **10a** (1.7 mg, 21% yield) as a white amorphous solid.

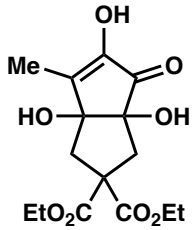
^{1}H NMR (400 MHz, CD_3OD) δ 4.97 (hept, *J* = 6.3 Hz, 1H), 4.87 – 4.78 (m, 1H), 2.86 (dd, *J* = 13.8, 1.9 Hz, 1H), 2.80 (dd, *J* = 13.3, 1.9 Hz, 1H), 2.33 (d, *J* = 13.4 Hz, 1H), 2.14 (d, *J* = 13.7 Hz, 1H), 1.86 (s, 3H), 1.30 – 1.11 (m, 12H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 198.08, 172.12, 169.27, 147.56, 142.77, 83.23, 79.99, 70.64, 69.95, 59.43, 42.99, 42.03, 21.74, 21.68, 9.58.

FTIR (NaCl, thin film, cm^{-1}): 3438, 2984, 2931, 1727, 1714, 1660, 1263, 1100, 902.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{24}O_8$: 374.1809 $[M+NH_4]^+$; found: 374.1813.

diethyl 3a,5,6a-trihydroxy-6-methyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (10b)



Prepared from **7b** (14.1 mg, 0.05 mmol) following General Procedure H. The crude residue was purified by column chromatography (silica, 5% MeOH/DCM) to yield **10b** (5.1 mg, 31% yield) as a white amorphous solid.

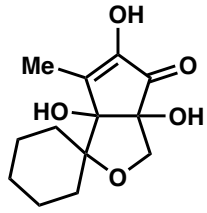
^{1}H NMR (400 MHz, CD_3OD) δ 4.15 (qd, *J* = 7.1, 1.3 Hz, 2H), 4.03 (qd, *J* = 7.2, 2.2 Hz, 2H), 2.86 (ddd, *J* = 20.9, 13.5, 2.0 Hz, 2H), 2.35 (d, *J* = 13.3 Hz, 1H), 2.17 (d, *J* = 13.7 Hz, 1H), 1.87 (s, 3H), 1.21 (dt, *J* = 8.2, 7.1 Hz, 6H).

^{13}C NMR (101 MHz, CD_3OD) δ 201.23, 171.88, 171.37, 152.07, 143.21, 82.47, 79.98, 63.17, 63.08, 57.10, 43.88, 41.87, 14.25, 14.22, 9.06.

FTIR (NaCl, thin film, cm^{-1}): 3402, 2984, 2940, 2852, 1728, 1667, 1404, 1366, 1252.

HRMS (TOF-ESI, m/z): calc'd for $C_{15}H_{20}O_8$: 327.1085 $[M–H]^-$; found: 327.1077.

3a',5',6a'-trihydroxy-6'-methyl-3a',6a'-dihydrospiro[cyclohexane-1,1'-cyclopenta[*c*]furan]-4'(3'*H*)-one (10c)



Prepared from **7c** (20.6 mg, 0.1 mmol) following General Procedure H. The crude residue was purified by column chromatography (silica, 40 to 50% EtOAc/hexanes) to yield **10c** (6.2 mg, 23% yield) as a white amorphous solid.

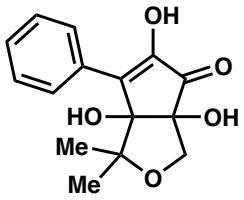
^{1}H NMR (400 MHz, CD_3OD) δ 5.49 (s, 1H), 3.89 (d, *J* = 10.0 Hz, 1H), 3.55 (d, *J* = 10.1 Hz, 1H), 1.97 (s, 3H), 1.77 – 1.35 (m, 8H), 1.35 – 1.12 (m, 2H).

^{13}C NMR (101 MHz, CD_3OD) δ 201.20, 152.33, 145.38, 85.90, 85.24, 81.39, 70.51, 54.81, 35.25, 30.76, 28.45, 26.69, 23.64, 22.39, 12.11.

FTIR (NaCl, thin film, cm^{-1}): 3394, 2925, 2855, 1714, 1650, 1460, 1361, 1109, 1072.

HRMS (TOF-ESI, m/z): calc'd for $C_{13}H_{18}O_5$: 237.1121 $[M–H_2O+H]^+$; found: 237.1111.

3a,5,6a-trihydroxy-1,1-dimethyl-6-phenyl-1,3,3a,6a-tetrahydro-4*H*-cyclopenta[*c*]furan-4-one (10e)



Prepared from **7e** (15.0 mg, 0.07 mmol) following General Procedure H. The crude residue was purified by preparative thin layer chromatography (silica, 40 to 55% EtOAc/hexanes then 10% MeOH/DCM) to yield **10e** (3.0 mg, 17% yield) as a white amorphous solid.

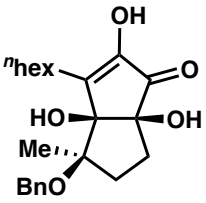
^{1}H NMR (400 MHz, CD_3OD) δ 8.24 – 8.17 (m, 2H), 7.43 – 7.28 (m, 3H), 3.98 (d, *J* = 10.1 Hz, 1H), 3.71 (d, *J* = 10.1 Hz, 1H), 1.35 (s, 3H), 0.90 (s, 3H).

^{13}C NMR (101 MHz, $(CD_3)_2SO$) δ 201.01, 150.42, 138.58, 135.22, 129.95, 128.81, 128.28, 84.91, 84.44, 80.63, 69.83, 27.14, 22.36.

FTIR (NaCl, thin film, cm^{-1}): 3341, 2920, 1694, 1463, 1393, 1049, 730.

HRMS (TOF-ESI, m/z): calc'd for $C_{15}H_{16}O_5$: 259.0965 $[M–H_2O+H]^+$; found: 259.0963.

(3a*S*,4*R*,6a*S*)-4-(benzyloxy)-3-hexyl-2,3a,6a-trihydroxy-4-methyl-4,5,6,6a-tetrahydropentalen-1(3a*H*)-one (10g)



Prepared from **7g** (32.6 mg, 0.1 mmol) following General Procedure H. The crude residue was purified by column chromatography (silica, 20%

EtOAc/hexanes then 2% MeOH/DCM) to yield **10g** (9 mg, 24% yield) as a pale orange amorphous solid.

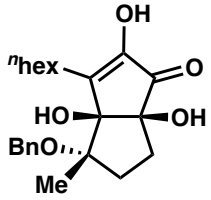
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.43 – 7.30 (m, 5H), 4.55 – 4.43 (m, 2H), 4.41 (s, 1H), 3.32 (s, 1H), 2.51 (ddd, *J* = 13.9, 9.3, 6.7 Hz, 1H), 2.32 (ddd, *J* = 13.9, 9.1, 7.2 Hz, 1H), 2.13 – 2.02 (m, 2H), 1.94 – 1.80 (m, 1H), 1.77 – 1.60 (m, 1H), 1.39 (s, 3H), 1.37 – 1.18 (m, 6H), 0.92 – 0.83 (m, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 200.48, 150.53, 144.83, 137.59, 128.78, 128.14, 127.78, 85.29, 83.68, 79.42, 64.78, 31.69, 30.50, 30.26, 29.93, 27.83, 27.48, 22.73, 19.88, 14.23.

FTIR (NaCl, thin film, cm^{-1}): 3444, 2926, 2856, 1714, 1660, 1393, 1104, 1040, 738.

HRMS (TOF-ESI, m/z): calc'd for $C_{22}H_{30}O_5$: 375.2166 $[M+H]^+$; found: 375.2174.

(3a*S*,4*S*,6a*S*)-4-(benzyloxy)-3-hexyl-2,3a,6a-trihydroxy-4-methyl-4,5,6,6a-tetrahydropentalen-1(3a*H*)-one (10h)



Prepared from **7h** (32.6 mg, 0.1 mmol) following General Procedure H. The crude residue was purified by column chromatography (silica, 20% EtOAc/hexanes) to yield **10h** (7.7 mg, 21% yield) as a pale orange amorphous solid.

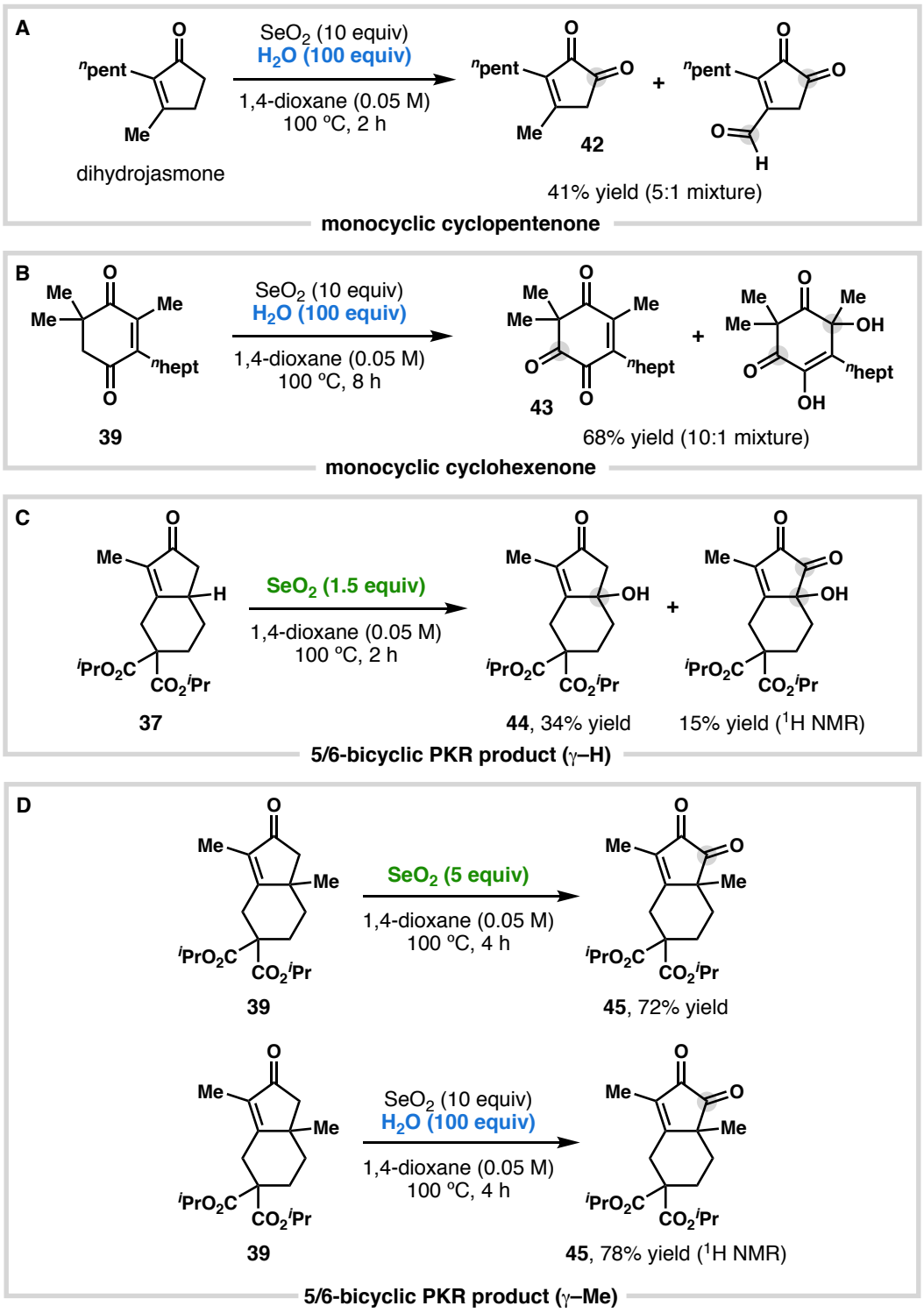
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.42 – 7.17 (m, 5H), 5.44 (s, 1H), 4.50 (d, *J* = 11.4 Hz, 1H), 4.33 (d, *J* = 11.4 Hz, 1H), 3.21 (s, 1H), 2.98 (s, 1H), 2.54 (ddd, *J* = 14.3, 10.4, 5.8 Hz, 1H), 2.41 (ddd, *J* = 14.3, 10.4, 6.0 Hz, 2H), 2.08 – 1.90 (m, 2H), 1.83 (q, *J* = 6.4, 5.8 Hz, 1H), 1.77 – 1.67 (m, 1H), 1.68 – 1.51 (m, 1H), 1.49 (s, 3H), 1.37 – 1.11 (m, 6H), 0.84 (t, *J* = 7.0 Hz, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ δ 199.61, 149.37, 148.48, 138.98, 128.37, 127.31, 126.98, 86.24, 85.37, 80.20, 64.78, 33.08, 31.74, 31.59, 30.04, 27.67, 27.42, 22.75, 18.07, 14.24.

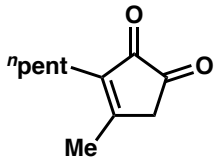
FTIR (NaCl, thin film, cm^{-1}): 3388, 2927, 2855, 1698, 1651, 1454, 1394, 1106, 1052, 734.

HRMS (TOF-ESI, m/z): calc'd for $C_{22}H_{30}O_5$: 375.2166 $[M+H]^+$; found: 375.2164.

Scheme 1.1. *Investigation of oxidative transformations in other ring systems*



4-methyl-3-pentylcyclopent-3-ene-1,2-dione (42)



See Scheme S1a. Prepared from dihydrojasmone (41.6 mg, 0.25 mmol) following General Procedure G. The crude residue was purified by column chromatography (silica, 50% EtOAc/hexanes) to yield **42** as a 5:1 mixture with an aldehyde byproduct (17.9 mg, 41% combined) as a pale yellow oil.

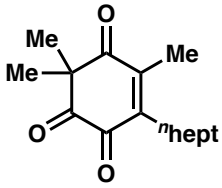
^{1}H NMR (400 MHz, $CDCl_3$) δ 2.94 (h, *J* = 1.2 Hz, 2H), 2.37 (t, *J* = 7.7 Hz, 2H), 2.23 (d, *J* = 1.2 Hz, 3H), 1.47 – 1.43 (m, 2H), 1.36 – 1.28 (m, 6H), 0.90 – 0.86 (m, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 198.55, 188.85, 166.73, 146.73, 77.48, 77.16, 76.84, 40.04, 31.84, 27.83, 23.45, 22.56, 17.82, 14.12.

FTIR (NaCl, thin film, cm^{-1}) of mixture: 3387, 2952, 2924, 2855, 1714, 1455, 1385, 1261, 1094.

HRMS (GC-EI+, m/z): calc'd for $C_{11}H_{16}O_2$: 180.1150 $[M]^+$; found: 180.1141.

6-heptyl-3,3,5-trimethylcyclohex-5-ene-1,2,4-trione (43)



See Scheme S1b. Prepared from **39** (47.3 mg, 0.20 mmol) following General Procedure G. The crude residue was purified by column chromatography (silica, 20% EtOAc/hexanes) to yield **43** as a 10:1 mixture with a byproduct of oxy-Michael addition (35.6 mg, 68% yield combined) as a pale yellow oil.

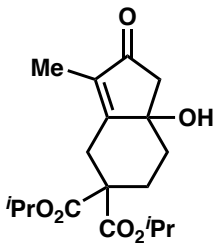
^{1}H NMR (400 MHz, $CDCl_3$) δ 2.62 – 2.55 (m, 2H), 2.15 (s, 3H), 1.42 (s, 6H), 1.30 (m, 10H), 0.90 – 0.85 (m, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 197.37, 194.47, 184.30, 150.74, 148.59, 77.48, 77.16, 76.84, 61.84, 31.79, 29.96, 29.11, 28.20, 26.91, 22.73, 21.88, 14.45, 14.20.

FTIR (NaCl, thin film, cm^{-1}) of mixture: 3442, 2953, 2924, 2854, 1728, 1667, 1463, 1380, 1312, 1156, 1034.

HRMS (GC-EI+, m/z): calc'd for $C_{16}H_{24}O_3$: 264.1726 $[M]^+$; found: 264.1749.

diisopropyl 7a-hydroxy-3-methyl-2-oxo-1,2,4,6,7,7a-hexahydro-5*H*-indene-5,5-dicarboxylate (44)



See Scheme S1c. Prepared from **37** (16.1 mg, 0.05 mmol) following General Procedure H with SeO_2 (8.3 mg, 0.075 mmol, 1.5 equiv). The crude residue was purified by column chromatography (silica, 45% EtOAc/hexanes) to yield **44** (5.7 mg, 34% yield) as a pale yellow oil.

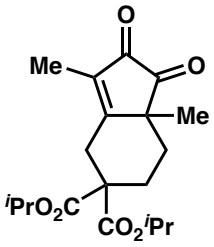
^{1}H NMR (400 MHz, $CDCl_3$) δ 5.03 (dp, *J* = 29.1, 6.3 Hz, 2H), 3.31 (dd, *J* = 13.9, 1.5 Hz, 1H), 2.89 (dq, *J* = 13.8, 1.6 Hz, 1H), 2.55 – 2.40 (m, 2H), 2.35 – 2.28 (m, 2H), 2.19 (dt, *J* = 14.3, 3.3 Hz, 1H), 1.75 (d, *J* = 1.5 Hz, 3H), 1.73 – 1.71 (m, 1H), 1.69 – 1.61 (m, 1H), 1.25 (dd, *J* = 6.3, 1.7 Hz, 7H), 1.20 (dd, *J* = 6.3, 4.0 Hz, 6H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 205.34, 170.48, 169.23, 167.35, 137.18, 77.35, 77.03, 76.71, 73.98, 69.55, 69.16, 56.23, 49.77, 35.64, 29.27, 26.74, 21.58, 21.55, 7.90.

FTIR (NaCl, thin film, cm^{-1}): 3443, 2981, 2856, 1810, 1714, 1667, 1454, 1376, 1303, 1251, 1104.

HRMS (GC-FAB+, m/z): calc'd for $C_{18}H_{27}O_6$: 339.1808 $[M+H]^+$; found: 339.1797.

diisopropyl 3,7a-dimethyl-1,2-dioxo-1,2,4,6,7,7a-hexahydro-5*H*-indene-5,5-dicarboxylate (45)



See Scheme S1d. Prepared from **39** (16.8 mg, 0.05 mmol) following General Procedure H with SeO_2 (27.7 mg, 0.25 mmol, 5.0 equiv). The crude residue was purified by column chromatography (silica, 25% EtOAc/hexanes) to yield **45** (12.6 mg, 72% yield) as a pale yellow oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.08 (hept, *J* = 6.3 Hz, 1H), 4.98 (hept, *J* = 6.3 Hz, 1H), 3.49 (dd, *J* = 14.1, 2.1 Hz, 1H), 2.74 (dq, *J* = 14.1, 1.4 Hz, 1H), 2.42 (ddt, *J* = 14.4, 4.1, 2.6 Hz, 1H), 2.15 (td, *J* = 14.2, 4.0 Hz, 1H), 1.95 (d, *J* = 1.3 Hz, 3H), 1.89 (ddd, *J* = 13.9, 4.0, 2.8 Hz, 1H), 1.51 (td, *J* = 13.9, 4.1 Hz, 1H), 1.31 (s, 3H), 1.26 (dd, *J* = 6.3, 3.5 Hz, 7H), 1.19 (dd, *J* = 6.3, 1.5 Hz, 6H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 202.05, 188.28, 172.37, 170.29, 168.81, 140.59, 77.48, 77.16, 76.84, 70.09, 69.70, 56.53, 43.83, 29.93, 29.27, 26.38, 21.68, 21.66, 19.94, 8.55.

FTIR (NaCl, thin film, cm^{-1}): 3505, 3408, 2980, 2939, 2872, 1732, 1633, 1456, 1386.

HRMS (TOF-ES+, m/z): calc'd for $C_{19}H_{27}O_6$: 351.1808 $[M+H]^+$; found: 351.1784.

1.8.5 STEREOCHEMICAL ANALYSIS

Stereochemical Analysis: General Procedure I

To a flame-dried vial with a Teflon-coated stir bar was added the substrate (1.0 equiv), SeO_2 (10.0 equiv), 1,4-dioxane (0.05 M), and water (100.0 equiv). The reaction was capped and sealed with Teflon tape, stirred, and brought to 100 °C in a preheated aluminum reaction block and monitored by TLC and/or LCMS. At specified time points, the reaction was allowed to reach room temperature and an aliquot was removed and concentrated. To

the residue was added an internal standard (pyrazine, ~1.0 equiv), then the mixture was taken up in CD_3OD and filtered over a cotton plug into an NMR tube. Yield determined by ^1H NMR versus the internal standard. Alternatively, isolated yields were determined by concentration of the entire reaction and purification of the crude residue by chromatography to afford the desired product. Enantiomeric excess was determined by analytical chiral SFC or HPLC.

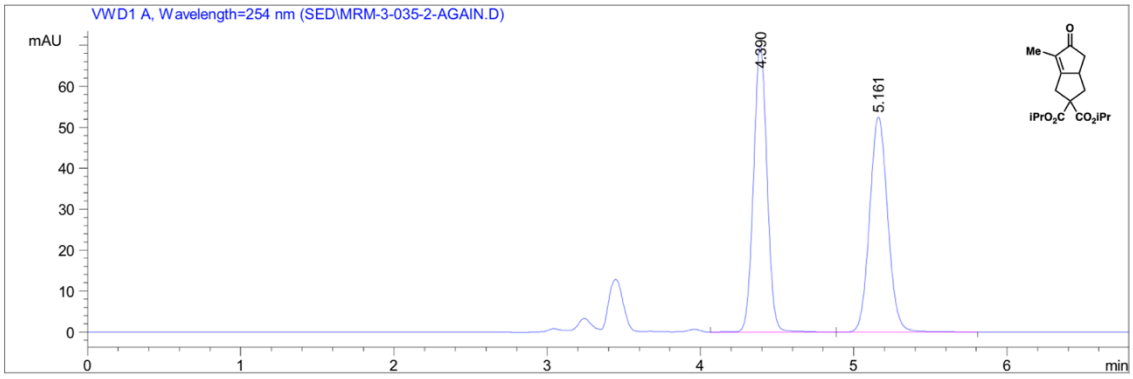
Stereochemical Analysis: General Procedure J

General Procedure I was followed except that the addition of water to the reaction mixture was omitted.

Diisopropyl 6-methyl-5-oxo-3,3a,4,5-tetrahydropentalene-2,2(1H)-dicarboxylate (7a)

Chiral HPLC: (IH, 1.0 mL/min, 10% IPA/hexanes, λ = 254 nm): t_R = 4.4 min, t_R = 5.2 min.

7a: racemic

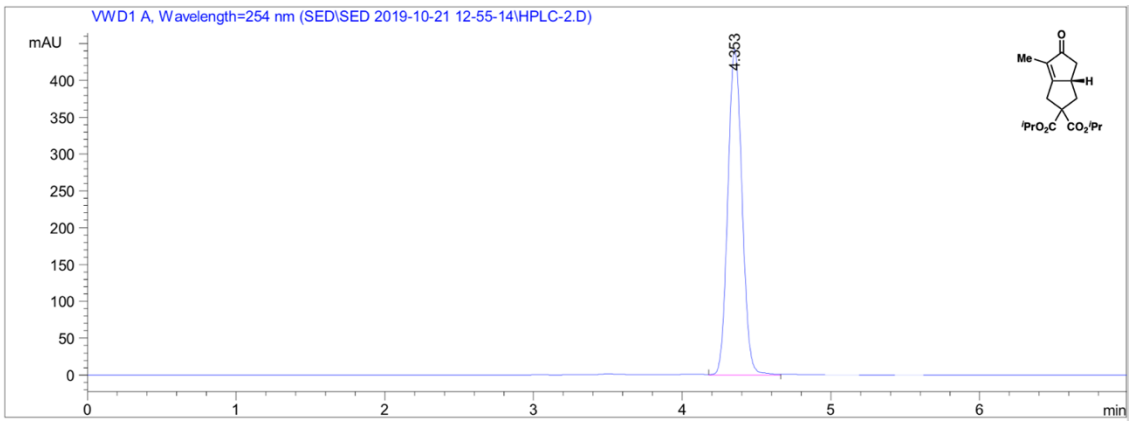


Peak #	RetTime [min]	Type	Width [min]	Area mAU *s	Height [mAU]	Area %
1	4.390	VV	0.0954	433.76349	70.21899	50.0120
2	5.161	VV	0.1281	433.55533	52.62845	49.9880

Prepared by purification of **7a** by chiral HPLC (IH, 1.0 mL/min, 10% IPA/hexanes, λ = 254 nm) to yield **(+)-7a*** as a single enantiomer.

$[\alpha]_D^{22} =$ +126 (c = 0.2, $CHCl_3$).

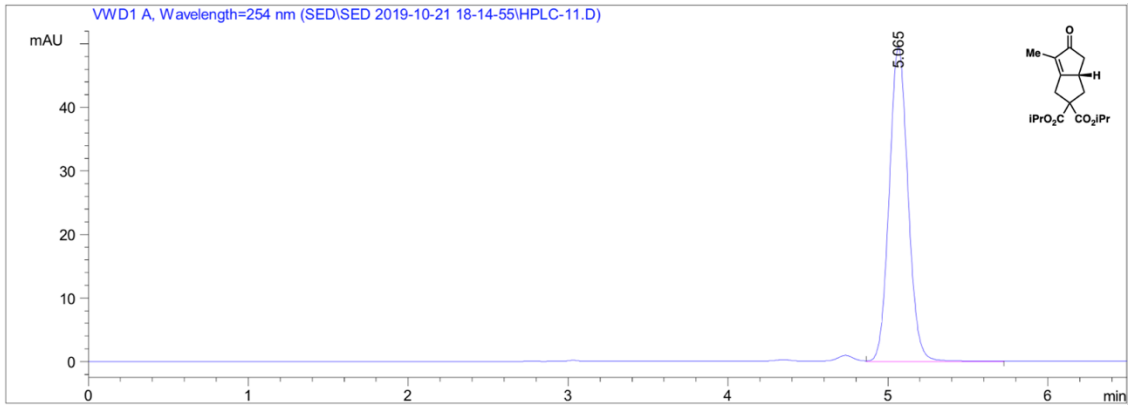
(+)-7a*: enantioenriched, >99% ee



Prepared by purification of **7a** by chiral HPLC (IH, 1.0 mL/min, 10% IPA/hexanes, λ = 254 nm) to yield **(–)-7a*** as a single enantiomer.

$[\alpha]_D^{22} =$ –95 (c = 0.3, $CHCl_3$).

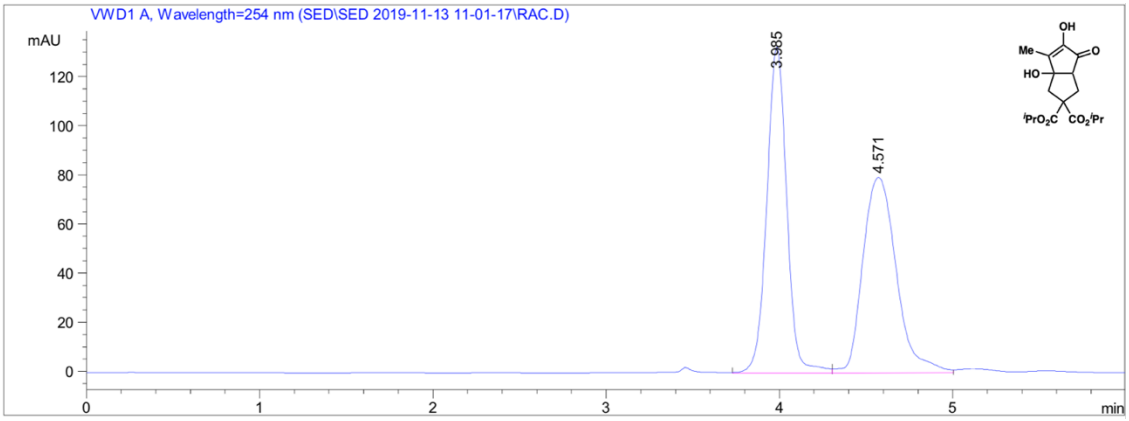
(–)-7a*: enantioenriched, >99% ee



Diisopropyl 5,6a-dihydroxy-6-methyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (9a)

Chiral HPLC: (IH, 1.0 mL/min, 65% IPA/hexanes, λ = 254 nm): t_R = 3.9 min, t_R = 4.6 min.

9a: racemic

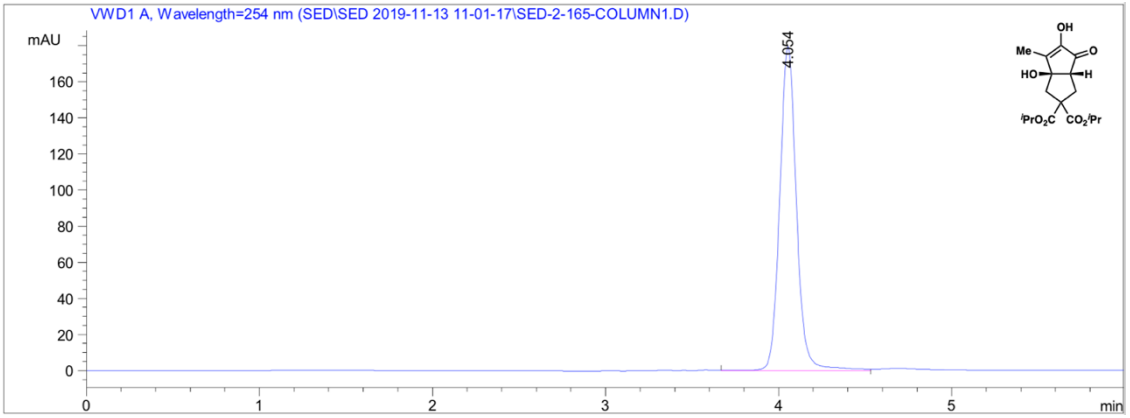


Peak #	RetTime [min]	Type	Width [min]	Area mAU *s	Height [mAU]	Area %
1	3.985	BV	0.1241	1063.69312	132.66316	49.8308
2	4.571	VV	0.2153	1070.91724	79.57430	50.1692

Prepared from **(–)-7a*** (176 mg, 0.57 mmol, >99% ee) following General Procedure I. The crude residue was purified by column chromatography (silica, 0 to 40% EtOAc/hexanes) to yield **(–)-9a*** (114 mg, 59% yield) in >99% ee.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{22}} =$ –67 (c = 0.5, $CHCl_3$).

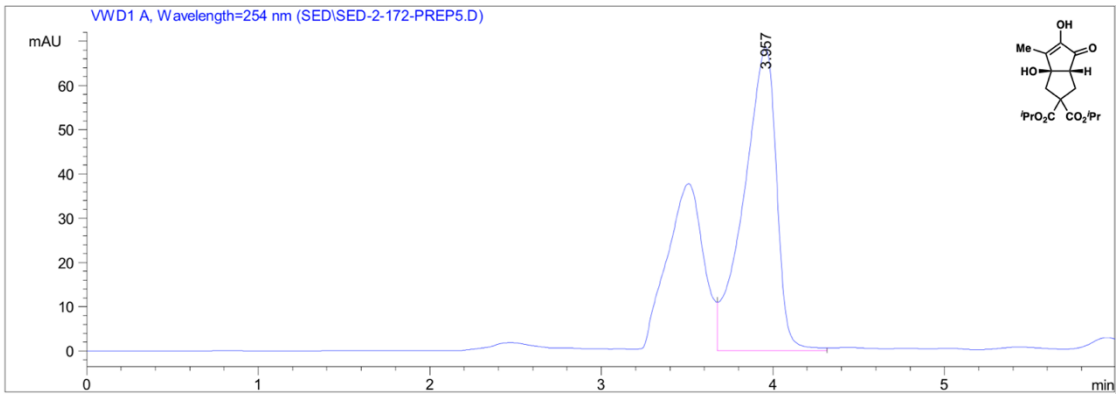
(–)-9a*: enantioenriched, >99% ee



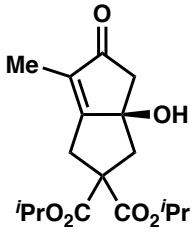
Prepared from **(–)-7a*** (22 mg, 0.07 mmol, >99% ee) following General Procedure J. The crude residue was purified by preparative thin layer chromatography (silica, 7% MeOH/DCM) to yield **(–)-9a*** (1.0 mg, 4% yield) in >99% ee.

$[\alpha]_D^{22} = -15$ (c = 0.1, $CHCl_3$).

(–)-9a*: enantioenriched, >99% ee



Diisopropyl 3a-hydroxy-6-methyl-5-oxo-3,3a,4,5-tetrahydropentalene-2,2(1*H*)-dicarboxylate (11a)



^{1}H NMR (500 MHz, $CDCl_3$) δ 5.09 (dp, *J* = 37.2, 6.2 Hz, 2H), 3.33 (dd, *J* = 18.5, 2.0 Hz, 1H), 3.26 (d, *J* = 18.5 Hz, 1H), 3.00 (s, 1H), 2.89 (d, *J* = 14.3 Hz, 1H), 2.63 (d, *J* = 17.9 Hz, 1H), 2.51 (d, *J* = 18.0 Hz, 1H), 2.11 (d, *J* = 14.4 Hz, 1H), 1.74 (dd, *J* = 1.9, 1.0 Hz, 3H), 1.37 – 1.14 (m, 12H).

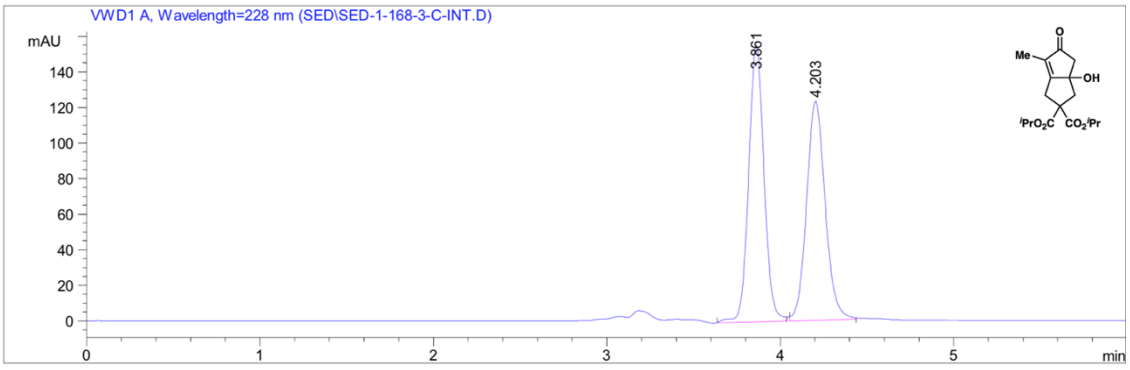
^{13}C NMR (101 MHz, $CDCl_3$) δ 207.52, 174.39, 173.24, 170.51, 134.39, 82.47, 70.68, 70.01, 61.00, 48.11, 45.22, 32.91, 21.74, 21.70, 8.78.

FTIR (NaCl, thin film, cm^{-1}): 3472, 2927, 2855, 1716, 1682, 1456, 1375, 1267, 1188, 1109, 1027, 913.

HRMS (TOF-ESI, m/z): calc'd for $C_{17}H_{24}O_6$: 307.1540 $[M–H_2O+H]^+$; found: 307.1549.

Chiral HPLC: (IH, 1.0 mL/min, 60% IPA/hexanes, λ = 228 nm): t_R = 3.9 min, t_R = 4.2 min.

11a: racemic

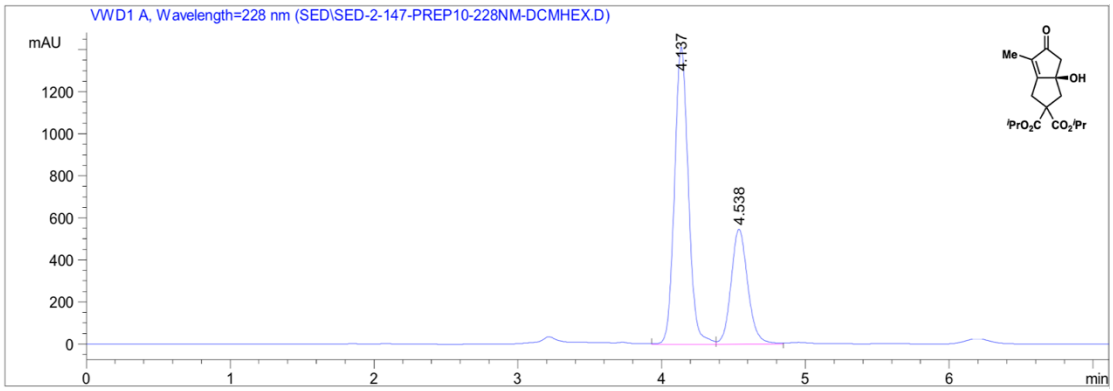


Peak #	RetTime [min]	Type	Width [min]	Area mAU *s	Height [mAU]	Area %
1	3.861	BB	0.0953	958.98309	155.45844	51.0419
2	4.203	BB	0.1149	919.83228	123.07118	48.9581

Prepared from **(+)-7a*** (200 mg, 0.65 mmol, >99% ee) following General Procedure I. The crude residue was purified by column chromatography (silica, 0 to 30% EtOAc/hexanes) followed by preparative thin layer chromatography (silica, 30 to 50% EtOAc/hexanes) to yield **(+)-11a*** (3.5 mg, 2% yield) in 36% ee.

$[\alpha]_D^{22} =$ +6 (c = 0.2, $CHCl_3$).

(+)-11a*: enantioenriched, 36% ee

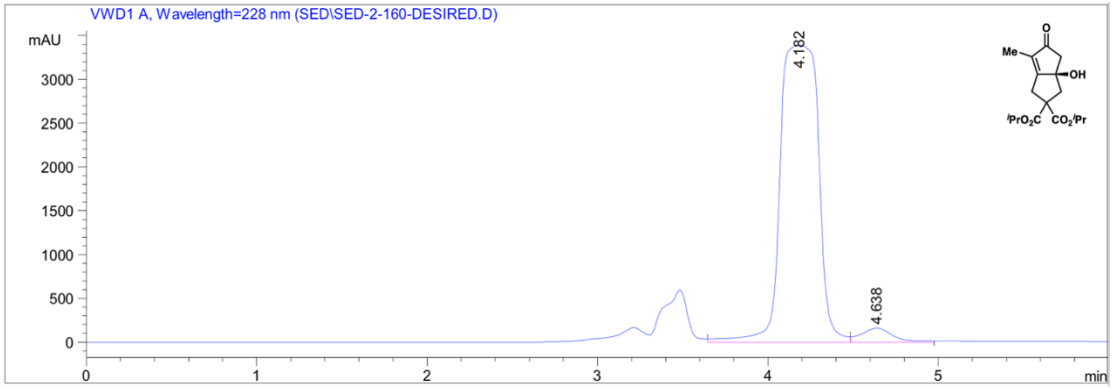


Peak #	RetTime [min]	Type	Width [min]	Area mAU *s	Height [mAU]	Area %
1	4.137	VV	0.1047	9508.88184	1414.22021	68.1725
2	4.538	VV	0.1250	4439.37842	548.41638	31.8275

Prepared from **(+)-7a*** (150 mg, 0.49 mmol) following General Procedure J. The crude residue was purified by column chromatography (silica, 0 to 30% EtOAc/hexanes). This procedure was repeated seven times, and the combined material yielded yield **(+)-11a*** (41 mg, 26% yield) in 92% ee.

$[\alpha]_D^{22} =$ +12 (c = 1.0, $CHCl_3$).

(+)-11a*: enantioenriched, 92% ee

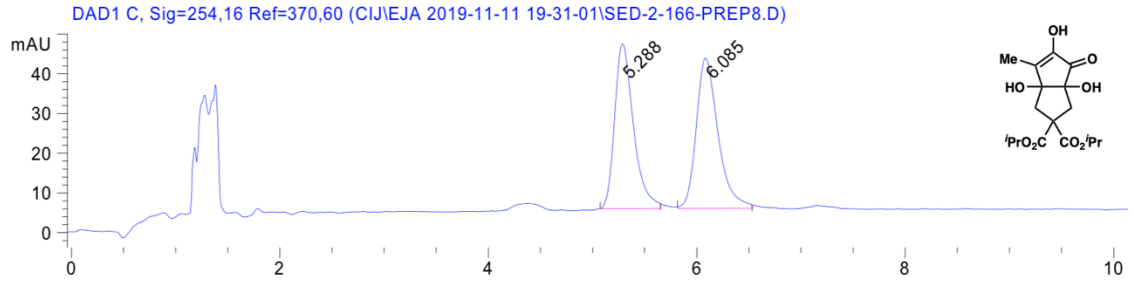


Peak #	RetTime [min]	Type	Width [min]	Area mAU *s	Height [mAU]	Area %
1	4.182	VV	0.2509	5.25926e4	3389.88306	96.1357
2	4.638	VV	0.1925	2114.04004	160.96941	3.8643

Diisopropyl 3a,5,6a-trihydroxy-6-methyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (10a)

Chiral SFC: (OD-H, 1.0 mL/min, 15% IPA in CO_2, λ = 254 nm): t_R = 5.3 min, t_R = 6.0 min.

10a: racemic

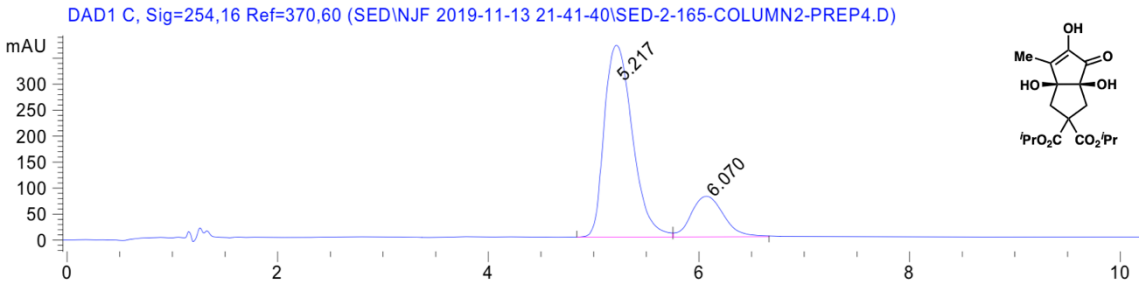


Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	5.288	BB	0.1903	519.55811	41.44424	48.6659
2	6.085	BB	0.2214	548.04315	37.73867	51.3341

Prepared from **(–)-7a*** (176 mg, 0.57 mmol, >99% ee) following General Procedure I. The crude residue was purified by column chromatography (silica, 0 to 40% EtOAc/hexanes) followed by preparative thin layer chromatography (7% MeOH/DCM) to yield **(–)-10a*** (6.9 mg, 3% yield) in 60% ee.

$[\alpha]_D^{22} = -49$ (c = 0.5, $CHCl_3$).

(–)-10a*: enantioenriched, 60% ee

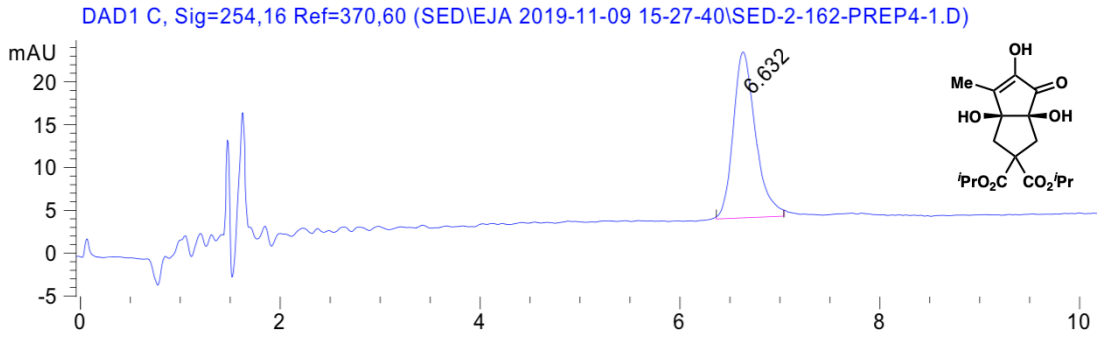


Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	5.217	BV	0.2970	6858.47607	368.79010	80.3824
2	6.070	VB	0.3388	1673.83008	77.87675	19.6176

Prepared from **(–)-7a*** (31 mg, 0.1 mmol, >99% ee) following General Procedure J. The crude residue was purified by preparative thin layer chromatography (silica, 70% EtOAc/hexanes then 7% MeOH/DCM) to yield **(–)-10a*** (5.2 mg, 15% yield) in >99% ee.

$[\alpha]_D^{22} = -59$ (c = 0.3, $CHCl_3$).

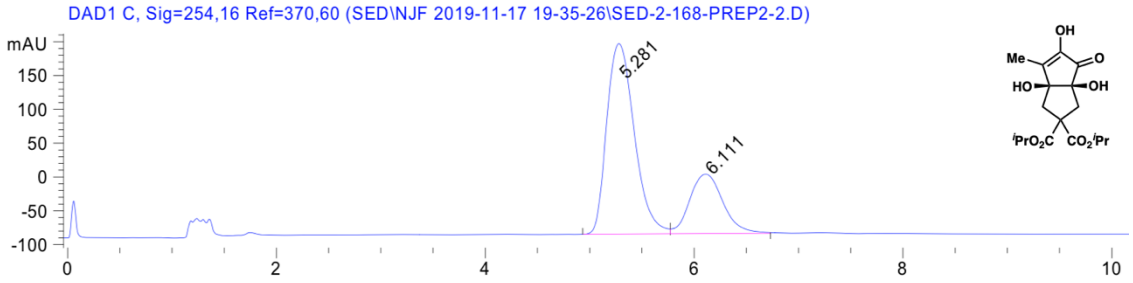
(–)-10a*: enantioenriched, >99% ee



Prepared from **(–)-9a*** (60 mg, 0.18 mmol, >99% ee) following General Procedure I. The crude residue was purified by preparative thin layer chromatography (silica, 10% MeOH/DCM) to yield **(–)-10a*** (7.5 mg, 12% yield) in 46% ee.

$[\alpha]_D^{22} =$ –29 (c = 0.3, $CHCl_3$).

(–)-10a*: enantioenriched, 46% ee

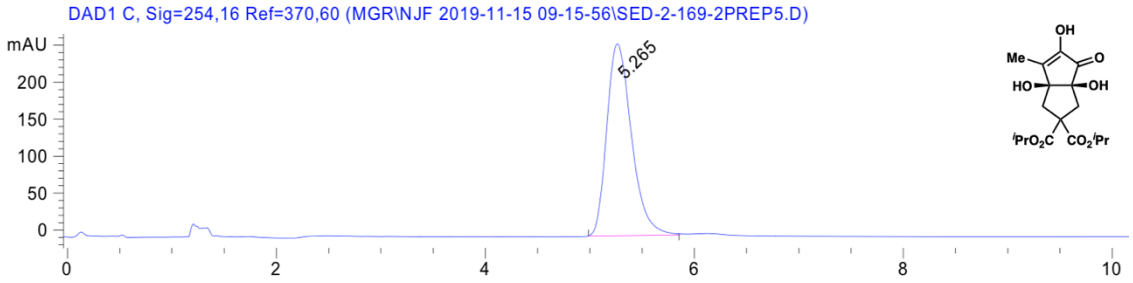


Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	5.281	BV	0.2901	5172.07080	281.81711	72.6377
2	6.111	VB	0.3501	1948.29785	88.05092	27.3623

Prepared from **(–)-9a*** (40 mg, 0.12 mmol, >99% ee) following General Procedure J. The crude residue was purified by preparative thin layer chromatography (silica, 10% MeOH/DCM) to yield **(–)-10a*** (10.1 mg, 24% yield) in >99% ee.

$[\alpha]_D^{22} =$ –90 (c = 0.7, $CHCl_3$).

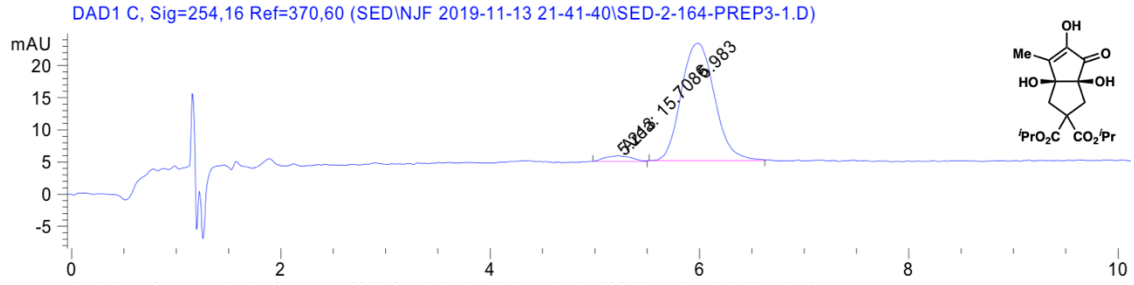
(–)-10a*: enantioenriched, >99% ee



Prepared from **(+)-11a*** (20 mg, 0.06 mmol, 92% ee) following General Procedure I. The crude residue was purified by column chromatography (silica, 7% MeOH/DCM) to yield **(+)-10a*** (1.1 mg, 5% yield) in 92% ee.

$[\alpha]_D^{22} =$ +14 (c = 0.1, $CHCl_3$).

(+)-10a*: enantioenriched, 92% ee

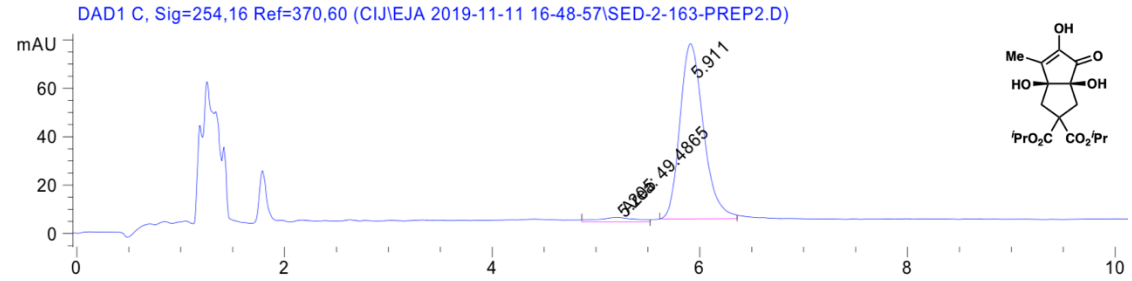


Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	5.213	MM	0.2931	15.70860	8.93133e-1	3.7445
2	5.983	BB	0.3545	403.80246	18.22254	96.2555

Prepared from **(+)-11a*** (20 mg, 0.06 mmol, 92% ee) following General Procedure J. The crude residue was purified by column chromatography (silica, 7% MeOH/DCM) to yield **(+)-10a*** (0.2 mg, 1% yield) in 92% ee.

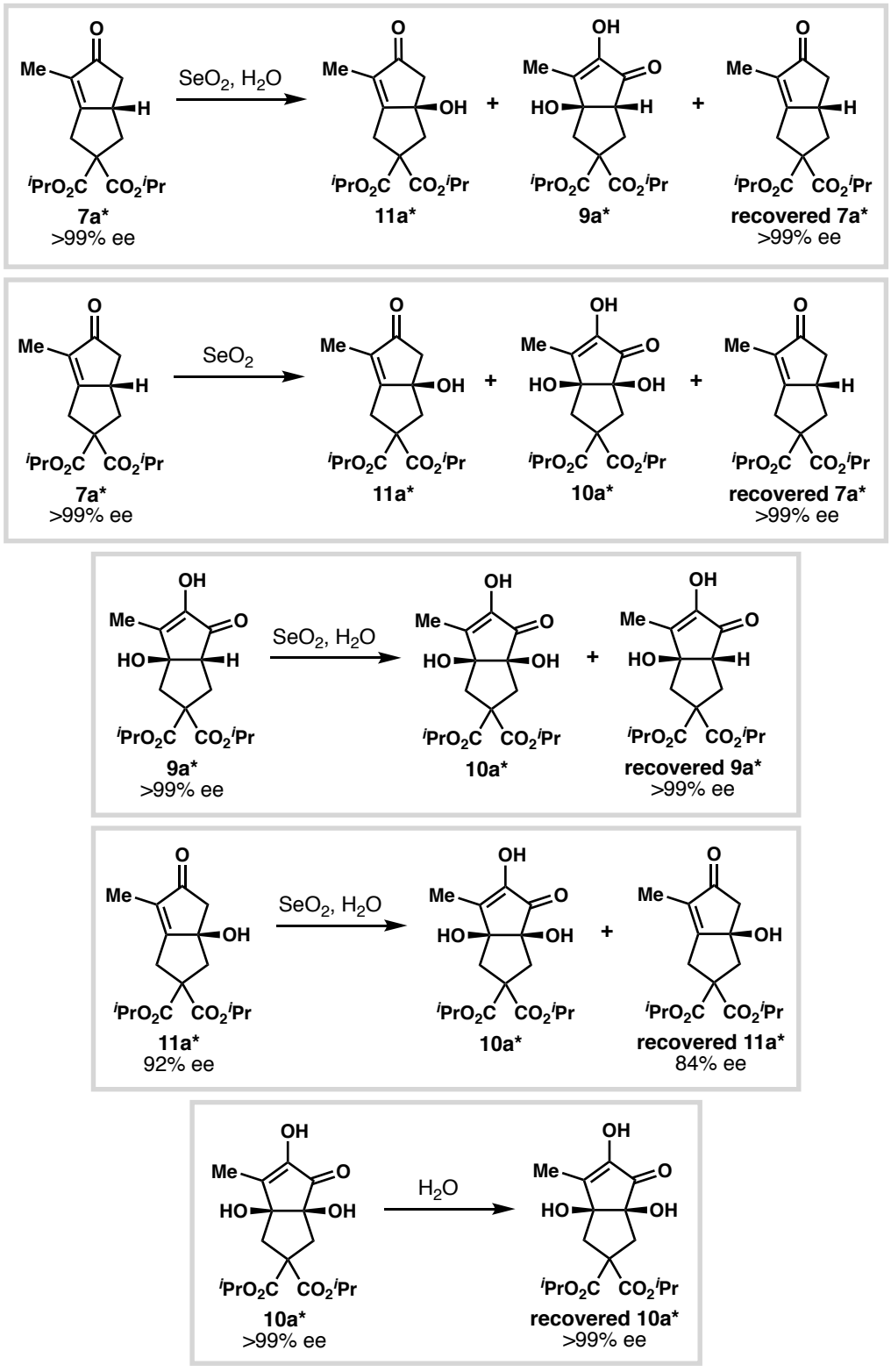
$[\alpha]_D^{22} =$ +1 (c = 0.1, $CHCl_3$).

(+)-10a*: enantioenriched, 92% ee



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	5.205	MM	0.4254	49.48651	1.93883	4.1664
2	5.911	BB	0.2437	1138.27368	72.27718	95.8336

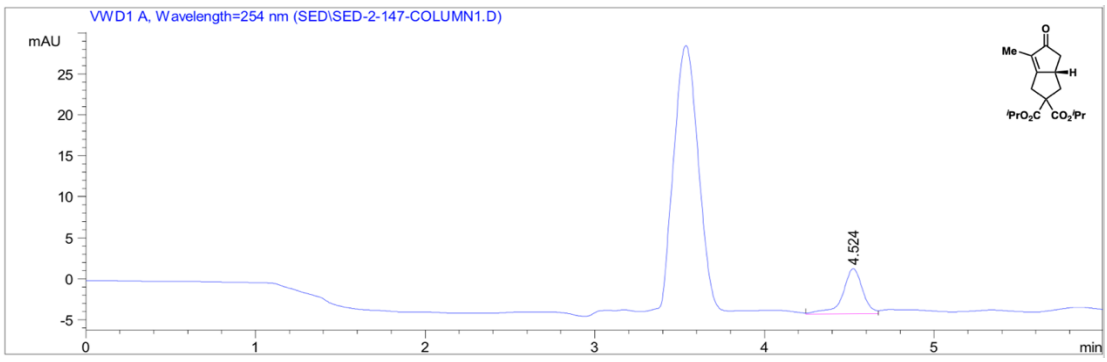
Scheme 1.2. *Control experiments*



Recovered from **(+)-7a*** (200 mg, 0.65 mmol, >99% ee) following General Procedure I. The crude residue was purified by column chromatography (silica, 0 to 30% EtOAc/hexanes) to yield **(+)-7a*** (150 mg, 75% recovery) in >99% ee.

$[\alpha]_D^{22} =$ +126 (c = 0.2, $CHCl_3$).

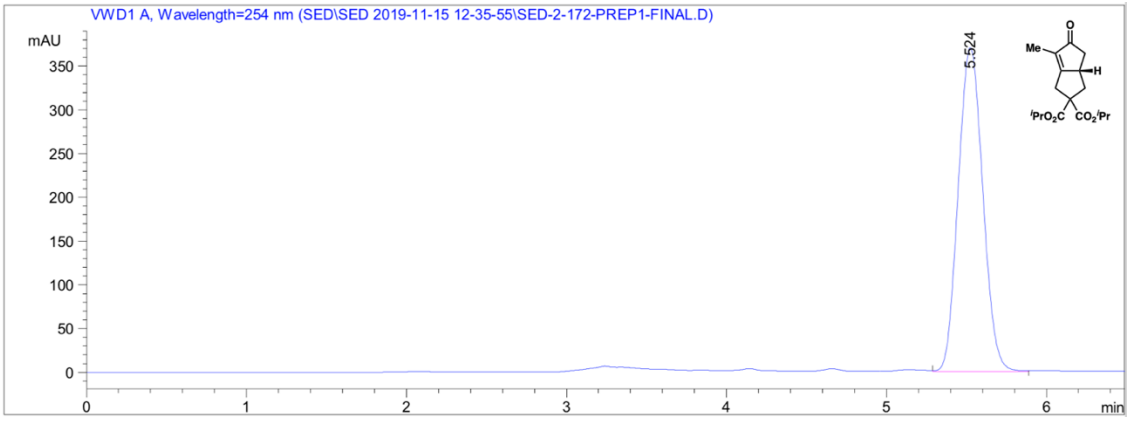
(+)-7a*: enantioenriched, >99% ee



Recovered from **(–)-7a*** (22 mg, 0.07 mmol, >99% ee) following General Procedure J. The crude residue was purified by preparative thin layer chromatography (silica, 7% MeOH/DCM) to yield **(–)-7a*** (1 mg, 4% recovery) in >99% ee.

$[\alpha]_D^{22} =$ –94 (c = 0.7, $CHCl_3$).

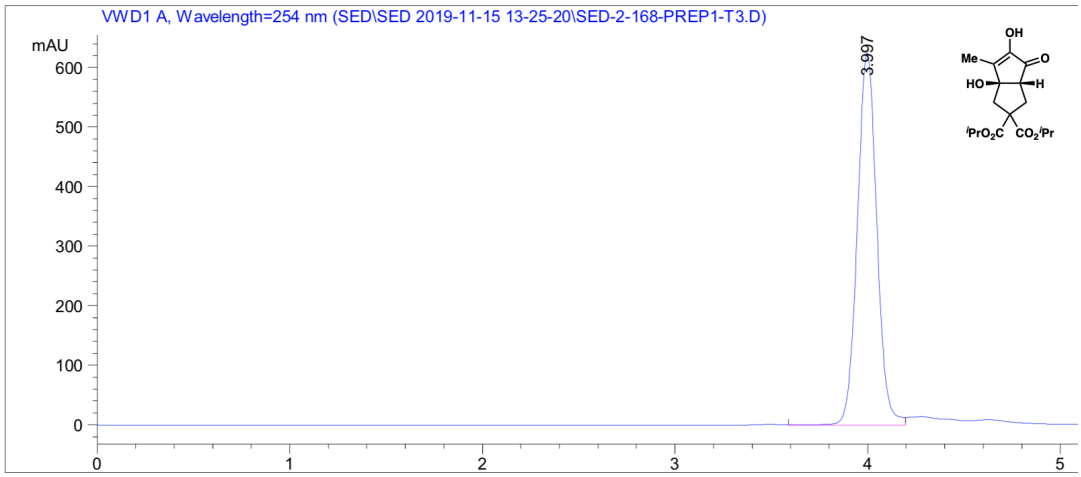
(–)-7a*: enantioenriched, >99% ee



Recovered from **(–)-9a*** (60 mg, 0.18 mmol, >99% ee) following General Procedure I. The crude residue was purified by preparative thin layer chromatography (silica, 10% MeOH/DCM) to yield **(–)-9a*** (2.2 mg, 4% recovery) in >99% ee.

$[\alpha]_D^{22} =$ –36 (c = 0.1, $CHCl_3$).

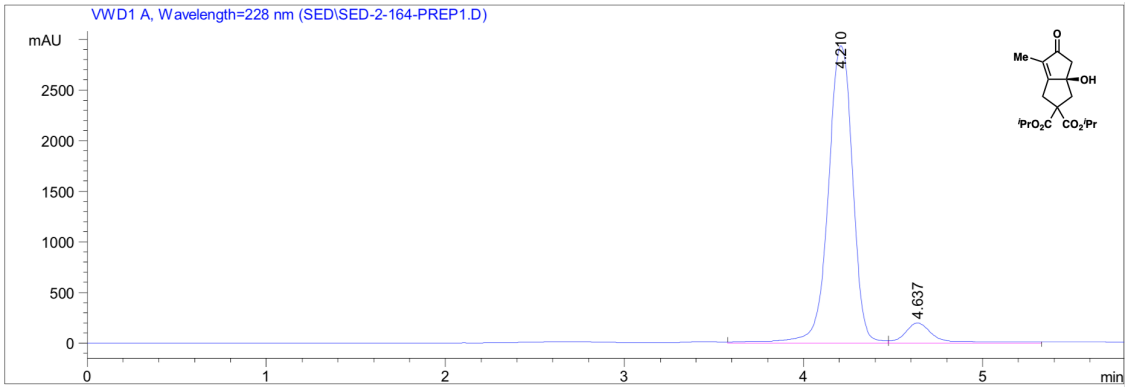
(–)-9a*: enantioenriched, >99% ee



Recovered from **(+)-11a*** (20 mg, 0.65 mmol, 92% ee) following General Procedure I. The crude residue was purified by column chromatography (silica, 0 to 30% EtOAc/hexanes) followed by preparative thin layer chromatography (silica, 7% MeOH/DCM) to yield **(+)-11a*** (2.2 mg, 11% recovery) in 84% ee.

$[\alpha]_D^{22} =$ +30 (c = 0.1, $CHCl_3$).

(+)-11a*: enantioenriched, 84% ee

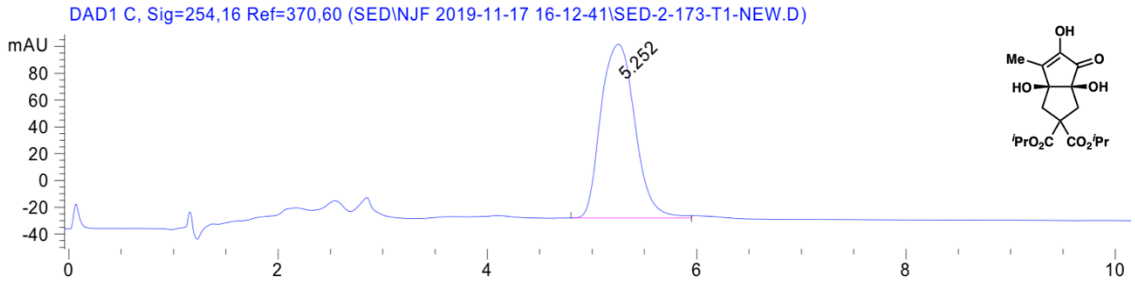


Peak #	RetTime [min]	Type	Width [min]	Area mAU *s	Height [mAU]	Area %
1	4.210	VV	0.1488	2.79395e4	2938.34180	92.1256
2	4.637	VB	0.1743	2388.11646	199.86485	7.8744

Recovered from **(–)-10a*** (10 mg, 0.03 mmol, >99% ee) following General Procedure I except that the addition of SeO_2 to the reaction mixture was omitted. The crude residue afforded **(–)-10a*** (10 mg, 99% recovery) in >99% ee.

$[\alpha]_D^{22} =$ –90 (c = 0.7, $CHCl_3$).

(–)-10a*: enantioenriched, >99% ee

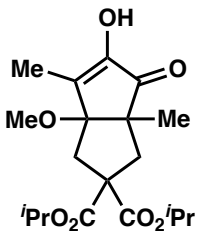


1.8.6 MECHANISTIC INVESTIGATION

Mechanistic Investigation: General Procedure K

All synthetic manipulations were performed in a glovebox. To a flame-dried vial with a Teflon-coated stir bar was added the enone (**6a** or **6e–i**) (0.1 mmol, 1.0 equiv), SeO_2 (111 mg, 1.0 mmol, 10.0 equiv), internal standard 1,2,4,5-tetrachloro-3-nitrobenzene (26 mg, 0.1 mmol, 1.0 equiv), and 1,4-dioxane-d_8 (0.8 mL, 0.125 M). The reaction was capped and sealed with Teflon tape, stirred, and brought to 100 °C in a preheated aluminum reaction block. A t specified time points, the reaction was allowed to reach room temperature and an aliquot (0.1–0.2 mL) was removed and filtered over a cotton plug directly into a dry NMR tube, washing with additional 1,4-dioxane-d_8 (0.3–0.4 mL). The tube was sealed and analyzed by NMR. If applicable, D_2O (180 μL, 10.0 mmol, 100.0 equiv) or CD_3OD (406 μL, 10.0 mmol, 100.0 equiv) was added to the appropriate aliquot, which was then brought to 100 ºC in a preheated oil bath (outside of the glovebox); at specified time points, the aliquots were allowed to reach room temperature, analyzed by NMR, then returned to the heat. Yield determined by ^{1}H NMR versus the internal standard.

diisopropyl 5-hydroxy-6a-methoxy-3a,6-dimethyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (13)



^{1}H NMR (400 MHz, $CDCl_3$) δ 4.98 (hept, *J* = 6.2 Hz, 1H), 4.85 (hept, *J* = 6.3 Hz, 1H), 3.15 (s, 3H), 2.78 (ddd, *J* = 13.8, 5.8, 2.2 Hz, 2H), 2.43 (d, *J* = 13.7 Hz, 1H), 1.97 (d, *J* = 13.9 Hz, 1H), 1.92 (s, 3H), 1.24 – 1.12 (m, 16H).

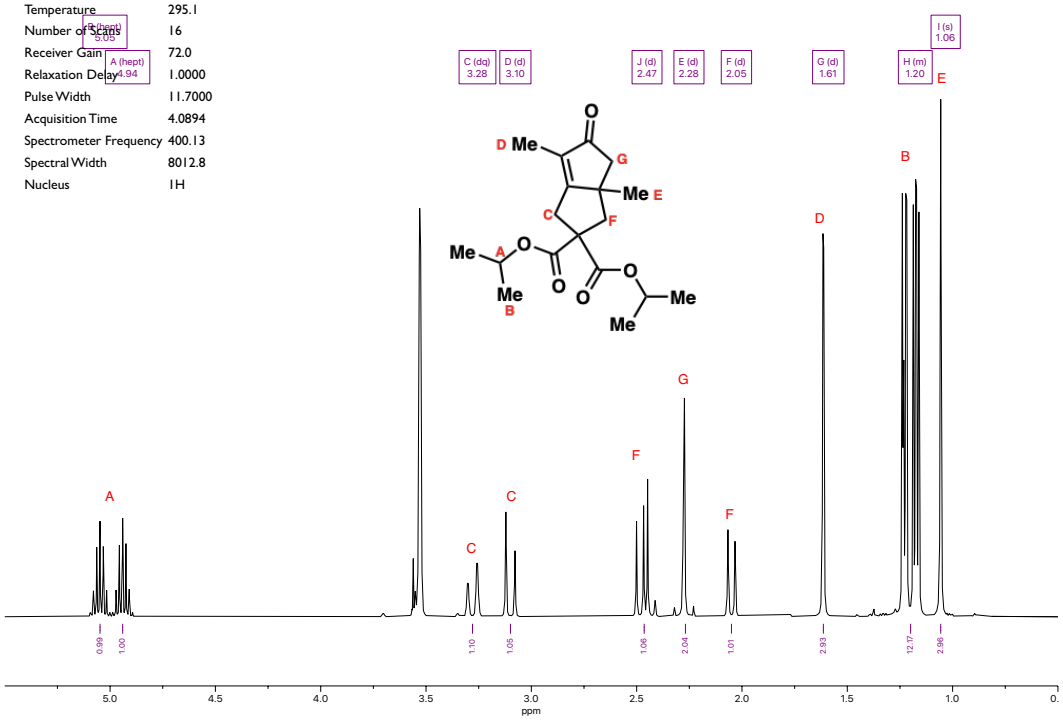
^{13}C NMR (101 MHz, $CDCl_3$) δ 203.53, 170.29, 169.71, 150.71, 138.57, 90.56, 77.36, 69.72, 69.64, 57.48, 53.05, 52.79, 44.42, 40.52, 29.86, 21.66, 21.64, 21.62, 21.53, 18.34, 10.29.

FTIR (NaCl, thin film, cm^{-1}): 3453, 2919, 2851, 1731, 1282, 1076.

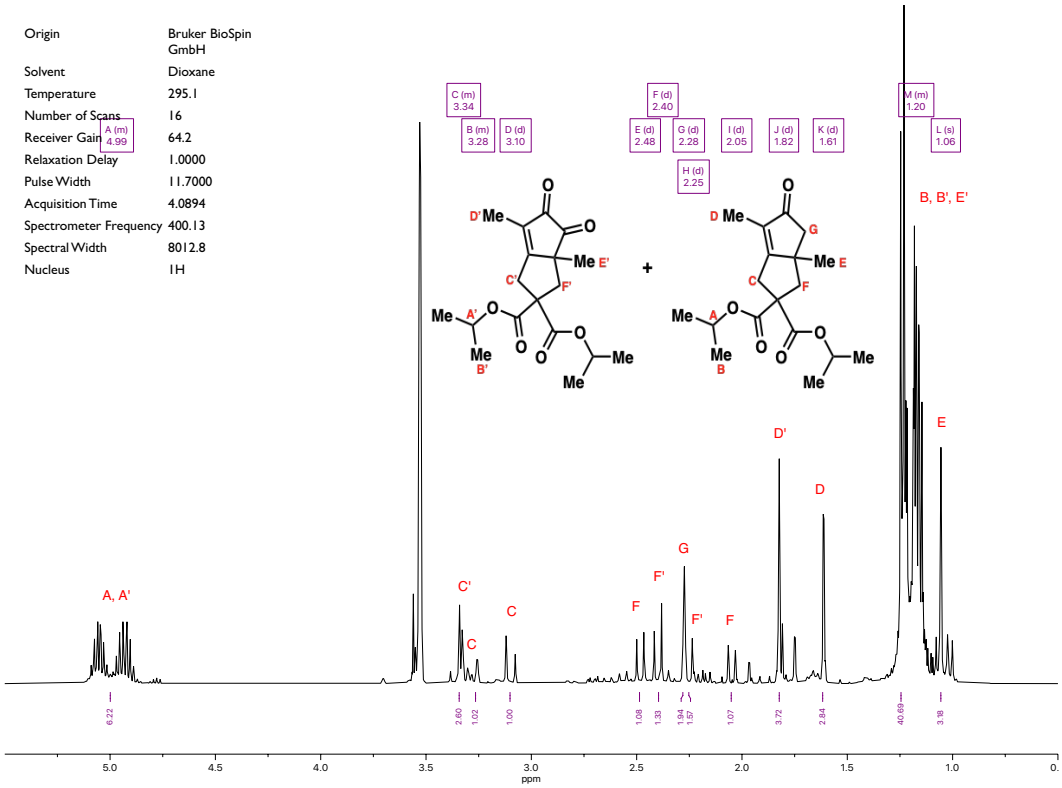
HRMS (TOF-ESI, m/z): calc'd for $C_{19}H_{28}O_7$: 337.1646 $[M–CH_3OH+H]^+$; found: 337.1653.

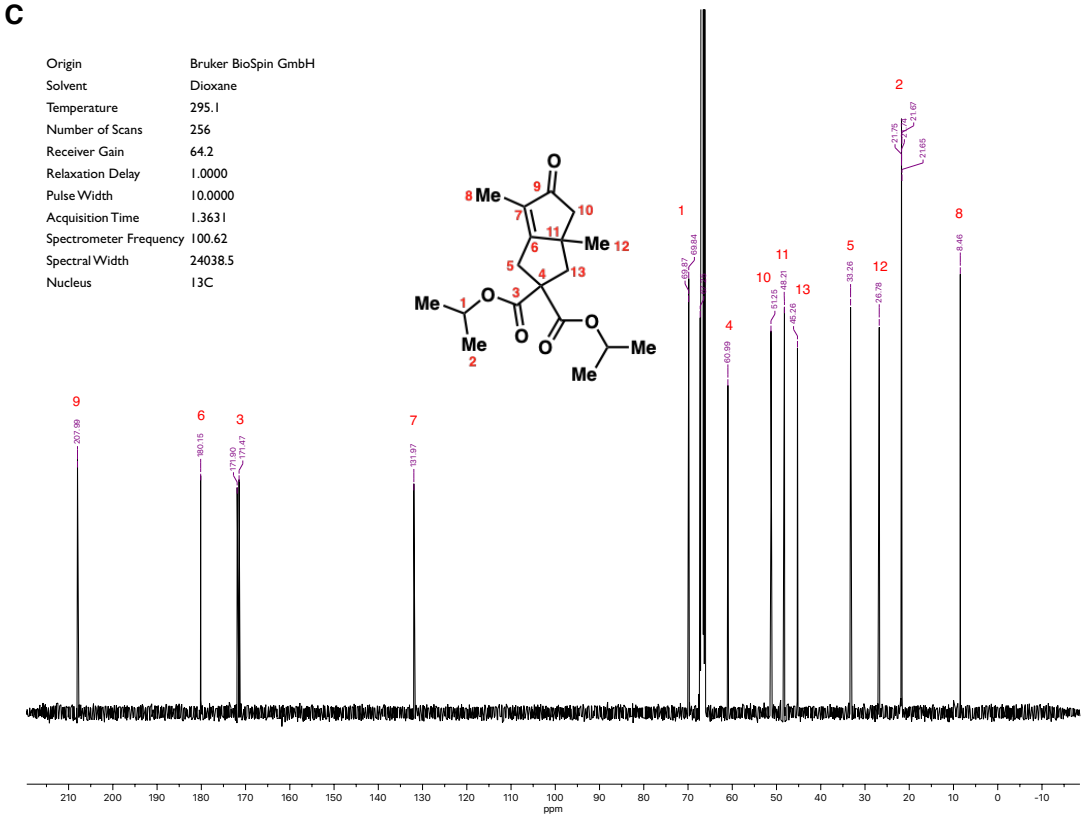
Figure 1.4. *Structural assignment of* ***12a*** *based on ^1H (b), ^{13}C (d), and HSQC (e) NMR spectra of reaction of* ***6a*** *(a,c) under anhydrous conditions [SeO_2 (10.0 equiv), 1,4-dioxane-d_8 (0.125 M), 100 °C] after 6 hours*

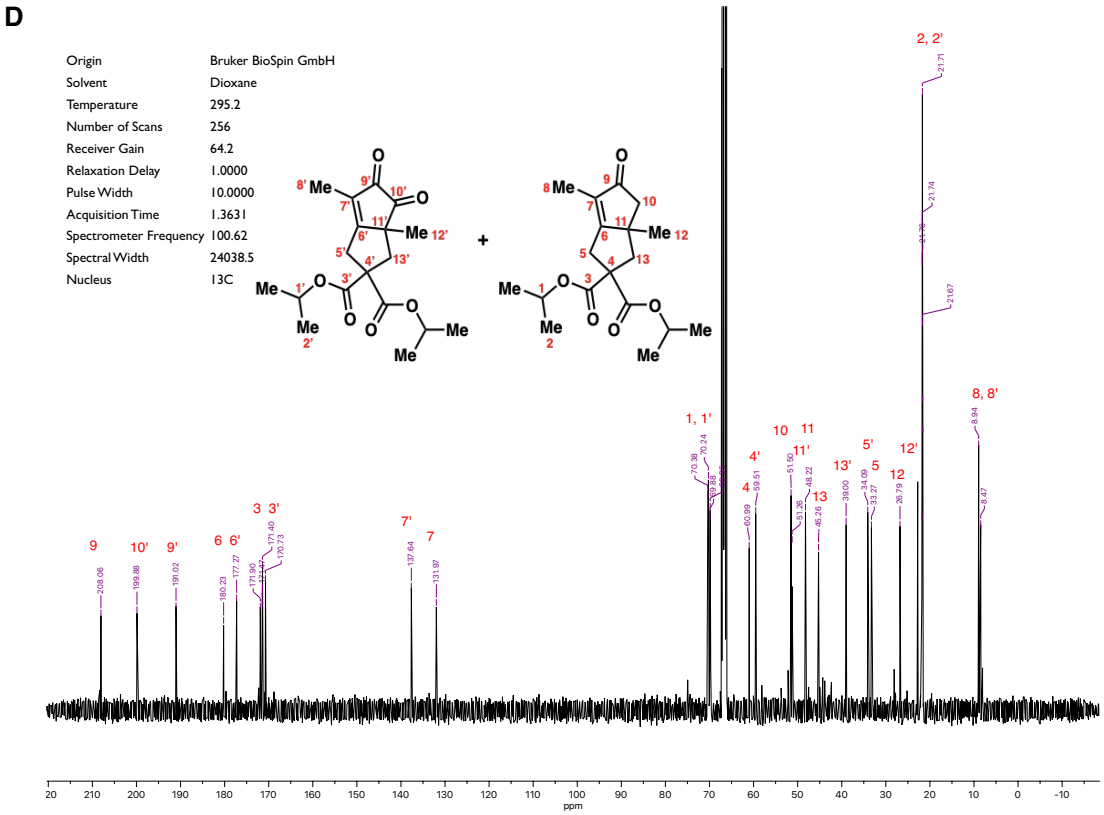
A



B


Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 295.1
Number of Scans 16
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Spectrometer Frequency 400.13
Spectral Width 8012.8
Nucleus 1H
A (m) 4.99
B (m) 3.28
C (m) 3.34
D (d) 3.10
E (d) 2.48
F (d) 2.40
G (d) 2.28
H (d) 2.25
I (d) 2.05
J (d) 1.82
K (d) 1.61
L (s) 1.06
M (m) 1.20
A, A'
B, B', E'
C'
C
D'
D
E
F'
F
G
6.22
2.60
1.02
1.00
1.08
1.33
1.84
1.57
1.07
3.72
2.84
40.69
3.18
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.
ppm


C
Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 295.1
Number of Scans 256
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Spectrometer Frequency 100.62
Spectral Width 24038.5
Nucleus 13C
ppm


D
Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 295.2
Number of Scans 256
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Spectrometer Frequency 100.62
Spectral Width 24038.5
Nucleus 13C
2, 2'
1, 1'
8, 8'
10 11
11'
5'
5
12'
12
4'
4
13'
13
3 3'
7'
7
9
10'
9'
6 6'
ppm

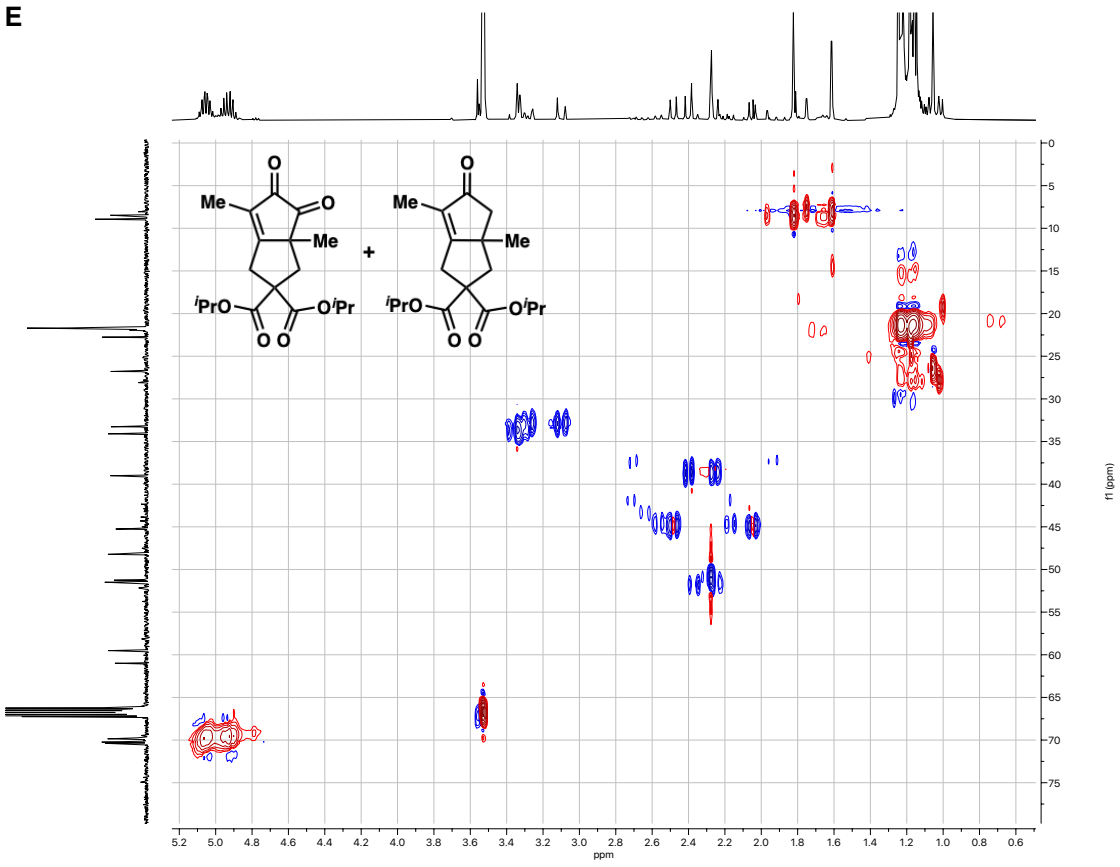

E
Me
O
Me
+
^{i}PrO
O^{i}Pr
ppm
f1 (ppm)

Figure 1.5. *Section of ^1H NMR spectra highlighting in situ formation of* ***12a*** *from reaction of* ***6a*** *under anhydrous conditions [SeO_2 (10.0 equiv), 1,4-dioxane-d_8 (0.125 M), 100 °C] after 3 hours (T_1) followed by in situ formation of* ***8a-d_2*** *from addition of D_2O (100.0 equiv) at 1 h (T_2) and 3 h (T_3)*

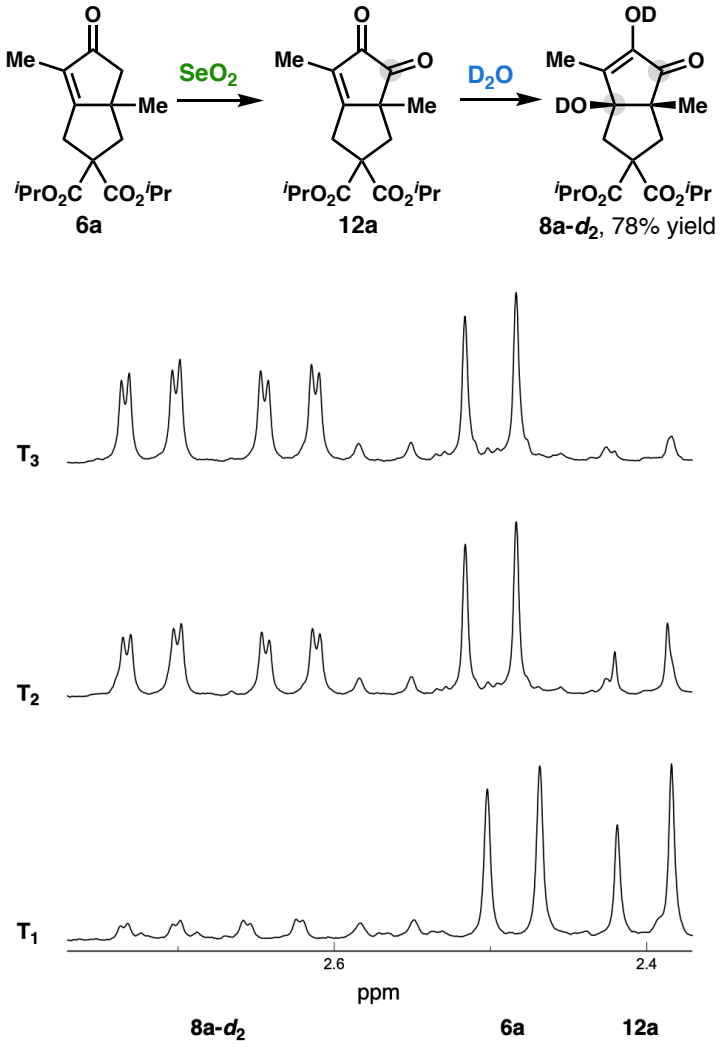


Figure 1.6. *Section of ^1H NMR spectra highlighting in situ formation of* ***12a*** *from reaction of* ***6a*** *under anhydrous conditions [SeO_2 (10.0 equiv), 1,4-dioxane-d_8 (0.125 M), 100 °C] after 3 hours (T_1) followed by in situ formation of* ***13-d_4*** *from addition of CD_3OD (100.0 equiv) at 25 h (T_2)*

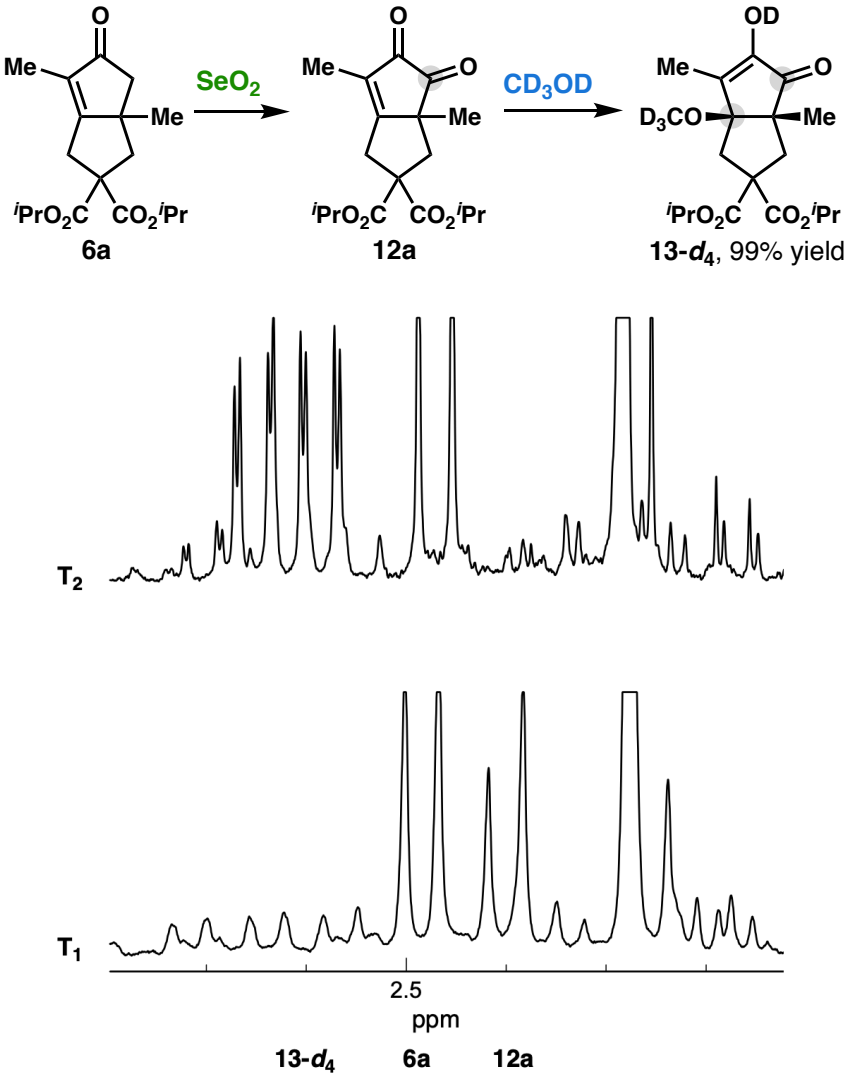
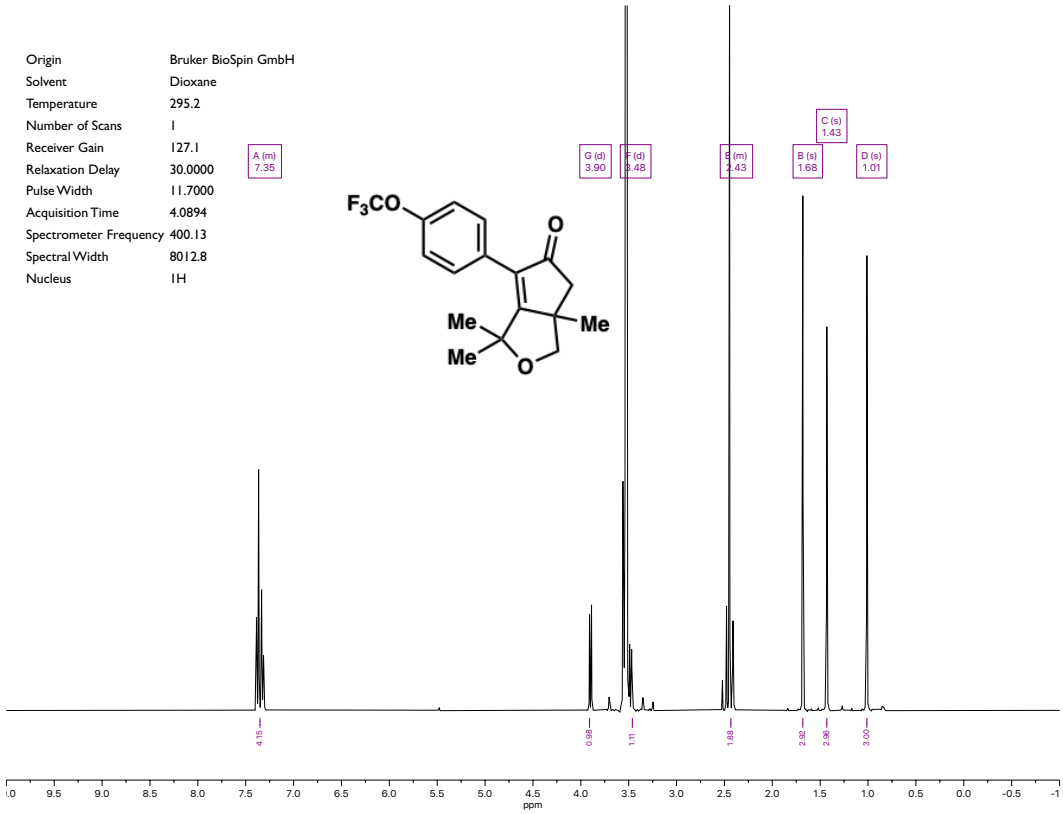
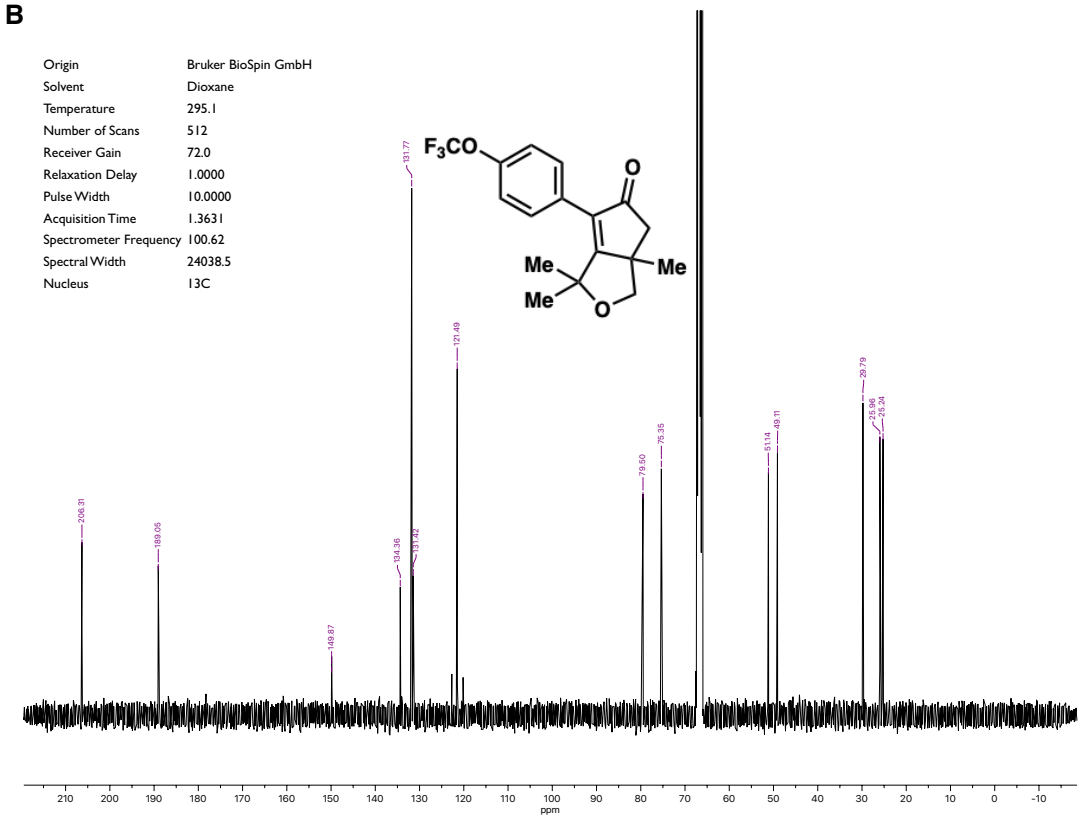
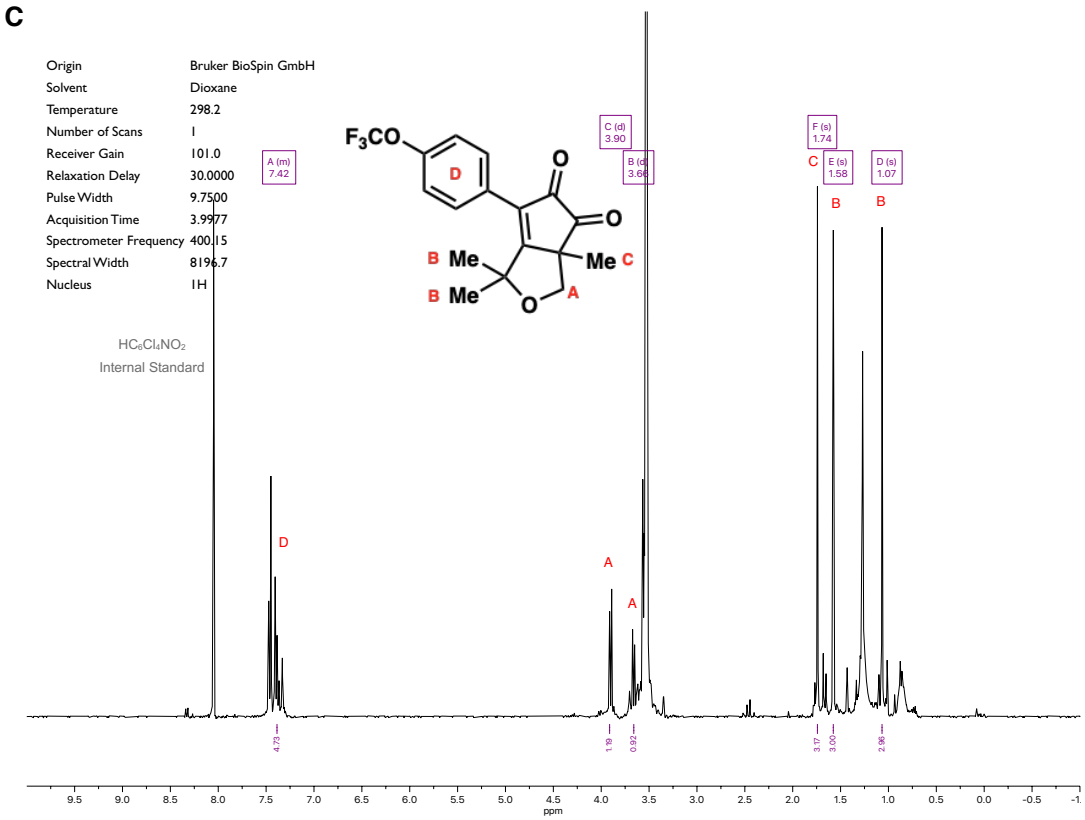


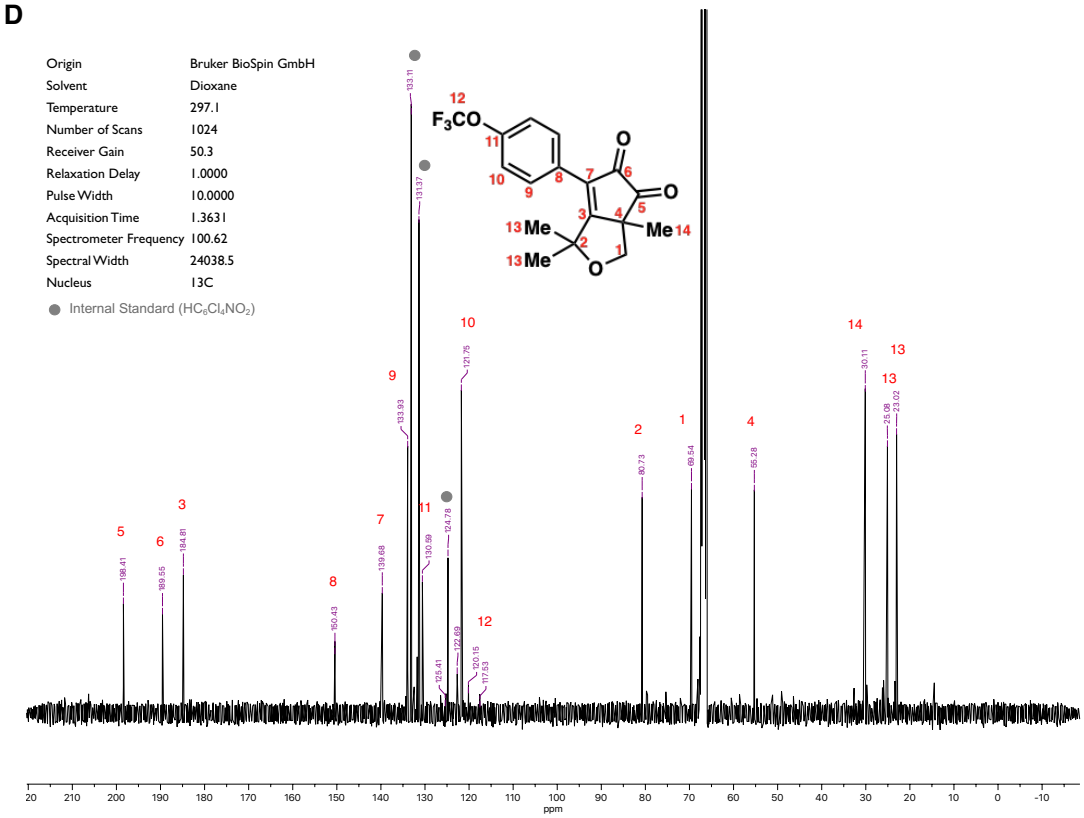
Figure 1.7. *Structural assignment of* ***12e*** *based on* ^{1}H *(c),* ^{13}C *(d), HSQC (e), and HMBC (f) NMR spectra of reaction of* ***6e*** *(a,b) under anhydrous conditions [*SeO_2 **A** *(10.0 equiv), 1,4-dioxane-*d_8 *(0.125 M), 100 °C] after 4 hours (e,f) and 8 hours (c,d).*

Structural assignments of ene-diketones ***12f–i*** *made by analogy to* ***12e***

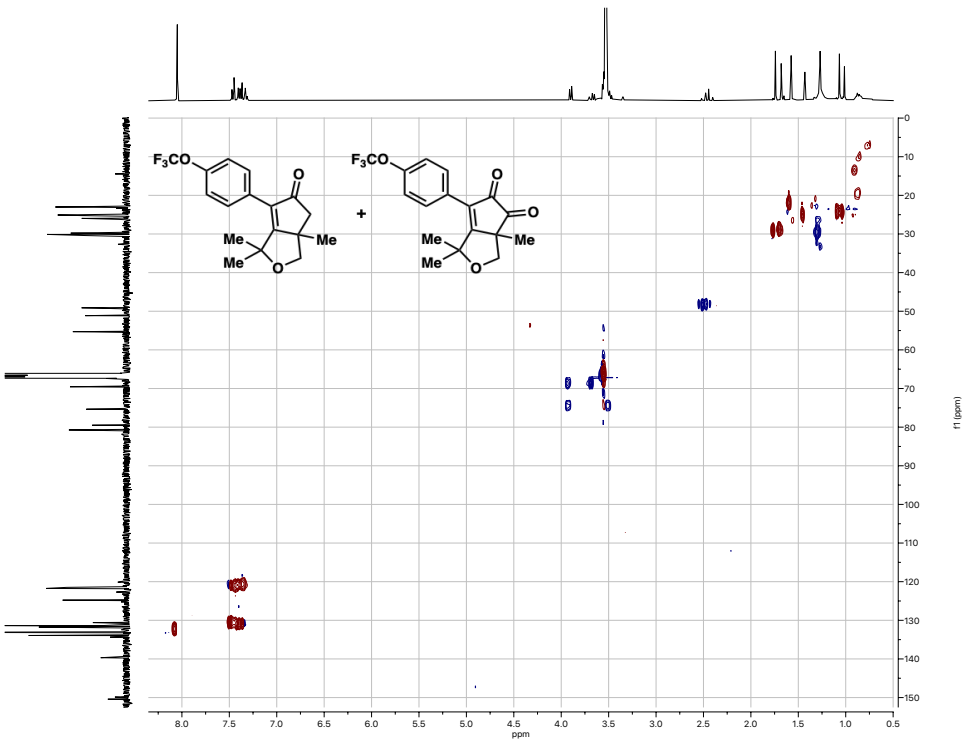



B
Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 295.1
Number of Scans 512
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Spectrometer Frequency 100.62
Spectral Width 24038.5
Nucleus 13C
F3CO
O
Me
Me
Me
O
206.31
189.05
149.87
134.36
131.77
131.42
121.49
79.50
75.35
51.14
49.11
29.79
25.96
25.24
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
ppm

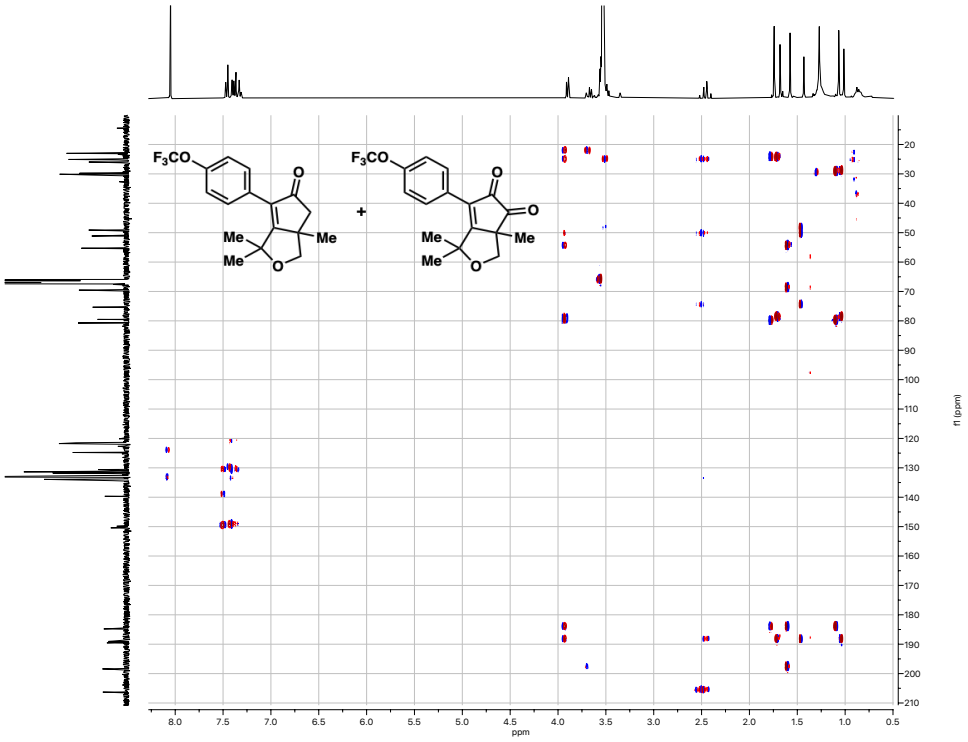

C
Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 298.2
Number of Scans 1
Receiver Gain 101.0
Relaxation Delay 30.0000
Pulse Width 9.7500
Acquisition Time 3.9977
Spectrometer Frequency 400.15
Spectral Width 8196.7
Nucleus 1H
HC6Cl4NO2
Internal Standard
A (m) 7.42
C (d) 3.90
B (d) 3.66
F (s) 1.74
E (s) 1.58
D (s) 1.07
F3CO
D
O
O
B Me
B Me
Me C
O
A
C
B
B
D
A
A
4.23
1.19
0.92
3.17
3.00
2.96
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1.
ppm


D
Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 297.1
Number of Scans 1024
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Spectrometer Frequency 100.62
Spectral Width 24038.5
Nucleus 13C
Internal Standard ($HC_6Cl_4NO_2$)
ppm

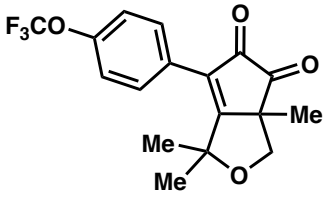
E


F3CO
O
Me
Me
O
Me
+
F3CO
O
O
Me
Me
O
Me
ppm
f1 (ppm)

F


F3CO
O
Me
Me
O
Me
+
F3CO
O
O
Me
Me
O
Me
ppm
f1 (ppm)

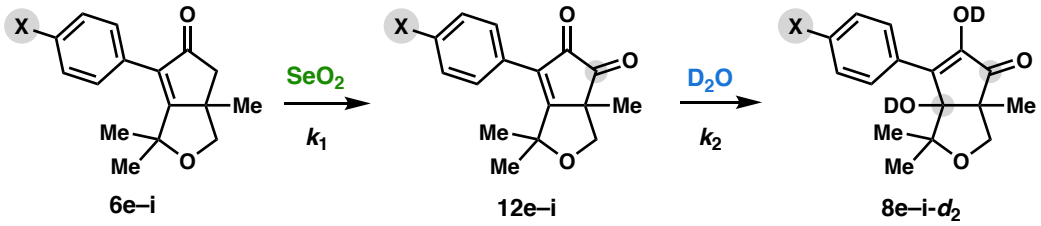
diisopropyl 5-hydroxy-6a-methoxy-3a,6-dimethyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (12e)



^{1}H NMR (400 MHz, 1,4-dioxane-d_8) δ 7.51 – 7.36 (m, 4H), 3.90 (d, J = 8.5 Hz, 1H), 3.66 (d, J = 8.5 Hz, 1H), 1.74 (s, 3H), 1.58 (s, 3H), 1.07 (s, 3H).

^{13}C NMR (101 MHz, 1,4-dioxane-d_8) δ 198.41, 189.55, 184.81, 150.43, 139.68, 133.93, 131.77, 131.50, 130.59, 122.69, 121.75, 121.61, 121.48, 80.73, 69.54, 55.28, 30.41, 30.11, 25.96, 23.02.

Table 1.8. *Tabulated Hammett parameters[36] and rate constants. Initial rates measured for the in situ formation of* ***12e–i*** *from* ***6e–i*** *under anhydrous conditions [SeO_2 (10.0 equiv), 1,4-dioxane-d_8 (0.125 M), 100 ºC] (k_1) and for the in situ formation of* ***8e–i*** *from* ***12e–i*** *[D_2O (100.0 equiv), 1,4-dioxane-d_8 (0.125 M), 100 ºC] (k_2). Structural assignments of dioxidation products* ***8f–i-d₂*** *made by analogy to ^1H and ^{13}C NMR spectra of* ***8e-d₂***

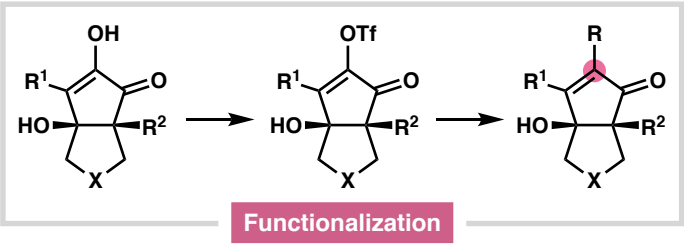


6e–i **12e–i** **8e–i-d_2**

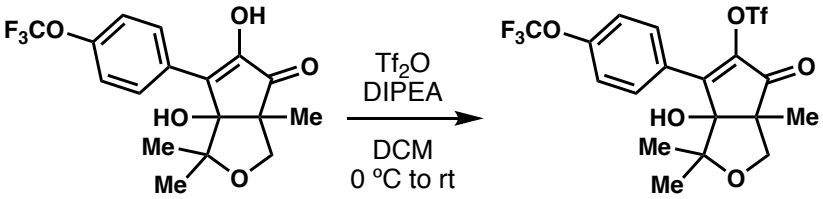
X	σ_p	k_1 (Ms^{-1})	log(k_X/k_H)	k_2 (Ms^{-1})	log(k_X/k_H)
OMe	–0.27	6.5 x10^{-6}	0.042	2.4 x10^{-6}	–0.267
H	0	5.9 x10^{-6}	0	4.4 x10^{-6}	0
Br	0.23	5.5 x10^{-6}	–0.029	6.1 x10^{-6}	0.141
OCF_3	0.35	5.1 x10^{-6}	–0.061	1.1 x10^{-5}	0.415
CF_3	0.54	5.0 x10^{-6}	–0.072	1.0 x10^{-5}	0.357
		ρ	–0.15	**ρ**	0.84
		R^2	0.98	**R^2**	0.91

1.8.7 PRODUCT DIVERSIFICATION

Functionalization of Dioxidation Product



6a-hydroxy-1,1,3a-trimethyl-4-oxo-6-(4-(trifluoromethoxy)phenyl)-3,3a,4,6a-tetrahydro-1*H*-cyclopenta[*c*]furan-5-yl trifluoromethanesulfonate (15e)



To a dry 1-dram vial with a Teflon-coated stir bar was added **8e** (65 mg, 0.2 mmol, 1.0 equiv). The vial was sealed with a rubber septum and placed under an atmosphere of dry N_2. The solid was taken up in DCM (1 mL, 0.20 M) and stirred, iPr_2NEt (53 µL, 0.3 mmol, 1.5 equiv) added, and the solution cooled to 0 °C in an ice water bath. Trifluoromethanesulfonic anhydride (34 µL, 0.2 mmol, 1.0 equiv) was then added dropwise via syringe, with a vent needle to release any triflic acid generated. The resulting mixture was allowed to reach room temperature and monitored by TLC. Upon completion, the reaction was quenched by dropwise addition of water then diluted with water and DCM. The layers were separated, and the aqueous layer was extracted twice with DCM. Combined organics were washed with water and brine, dried over Na_2SO_4, filtered, and

concentrated. The crude residue was purified by trituration from DCM and pentanes (1:4) to yield **15e** (88 mg, 0.20 mmol, 99% yield) as a white solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.00 – 7.92 (m, 2H), 7.32 (ddt, *J* = 8.0, 2.2, 1.1 Hz, 2H), 4.13 (d, *J* = 9.9 Hz, 1H), 3.68 (d, *J* = 9.9 Hz, 1H), 1.31 (d, *J* = 3.0 Hz, 6H), 0.96 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 197.78, 154.23, 150.22, 142.15, 130.58, 127.74, 120.74, 119.77, 118.92, 118.16, 115.73, 87.96, 83.57, 70.39, 57.68, 28.85, 24.99, 20.77, 15.62.

FTIR (NaCl, thin film, cm^{-1}): 3446, 2994, 29256, 2854, 2355, 1738, 1732, 1715, 1632, 1506, 1434, 1212.

HRMS (TOF-ESI, m/z): calc'd for $C_{18}H_{16}F_6O_7S$: 491.0594 $[M+H]^+$; found: 491.0603.

Figure 1.8. *Structure of* ***15e*** *with 50% probability anisotropic displacement ellipsoids. Compound* ***15e*** *crystallizes in the monoclinic space group $P12_1/c1$ with one molecule in the asymmetric unit. The coordinates for the hydrogen atom bound to O3 was located in the difference Fourier synthesis and refined using a riding model*

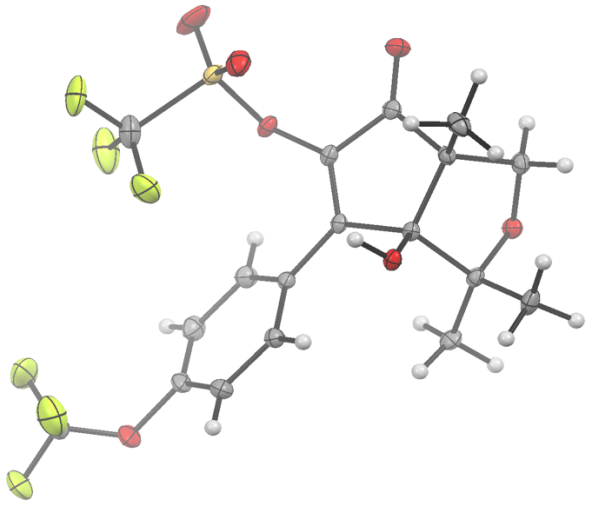
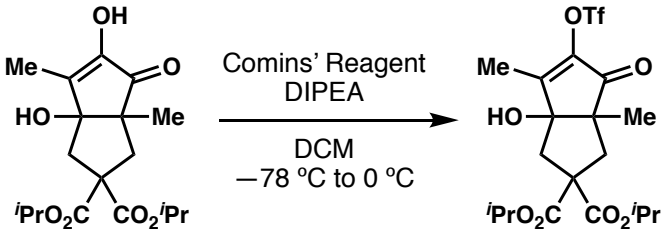


Table 1.9. *Crystal data and structure refinement for v18444_a*

Identification code	V18444_a	
Empirical Formula	C18 H16 F6 O7 S	
Formula weight	490.37	
Temperature	99.99 K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	P 1 21/c 1	
Unit cell dimensions	a = 6.0928(3) Å	α = 90°.
	b = 9.2223(5) Å	β = 91.637(2)°.
	c = 35.161(2) Å	γ = 90°.
Volume	1974.84(18) $Å^3$	
Z	4	
Density (calculated)	1.649 mg/m^3	
Absorption coefficient	0.260 mm^{-1}	
F(000)	1000	
Crystal size	0.26 x 0.23 x 0.15 mm^3	
Theta range for data collection	2.494 to 27.518°.	
Index ranges	$-7\leq h\leq 7$, $-11\leq k\leq 11$, $-45\leq l\leq 45$	
Reflections collected	59485	
Independent reflections	4529 [R(int) = 0.0412]	

Completeness to theta = 26.000°	99.9%	
Absorption correction	Semi-empirical from equivalents	
Max. and min. transmission	0.7456 and 0.7195	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	4529 / 0 / 295	
Goodness-of-fit on F^2	1.090	
Final R indices [I>2sigma(I)]	R1 = 0.0326, wR2 = 0.0807	
R indices (all data)	R1 = 0.0357, wR2 = 0.0820	
Extinction coefficient	n/a	
Largest diff. peak and hole	0.442 and -0.399 e.$Å^{-3}$	

diisopropyl 6a-hydroxy-3a,6-dimethyl-4-oxo-5-(((trifluoromethyl)sulfonyl)oxy)-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (15a)



To a dry round bottom flask with a Teflon-coated stir bar was added **8a** (648 mg, 1.8 mmol, 1.0 equiv). The vial was sealed with a rubber septum and placed under an atmosphere of dry N_2. The solid was taken up in DCM (37 mL, 0.05 M) and stirred, DIPEA (1.6 mL, 1.25 mmol, 5.0 equiv) added, and the solution cooled to –78 °C in a dry ice/acetone bath. Comins' Reagent (719 mg, 0.25 mmol, 1.0 equiv) was then added in one portion. The resulting mixture was allowed to reach 0 ºC following transfer to an ice water bath and

monitored by TLC. Upon completion, the reaction was directly purified by column chromatography (silica, 20% EtOAc/hexanes) to afford **15a** (797 mg, 90% yield) as an orange oil.

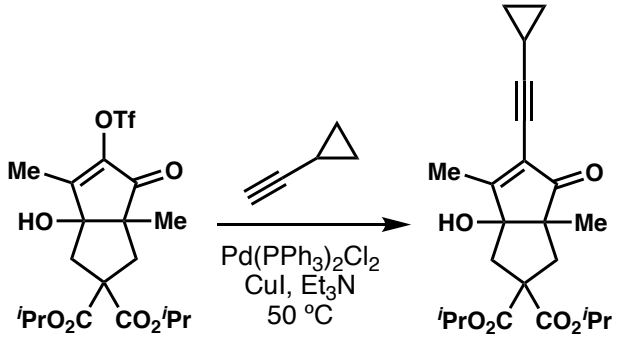
^{1}H NMR (400 MHz, $CDCl_3$) δ 5.06 (heptd, *J* = 6.3, 0.9 Hz, 1H), 4.93 (pd, *J* = 6.3, 0.8 Hz, 1H), 2.72 (d, *J* = 14.5 Hz, 1H), 2.52 (s, 2H), 2.24 (d, *J* = 14.5 Hz, 1H), 2.18 (s, 3H), 1.24 (dd, *J* = 6.3, 3.0 Hz, 6H), 1.21 – 1.09 (m, 9H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 198.44, 171.95, 169.21, 162.78, 142.51, 120.15, 116.96, 85.09, 70.56, 70.02, 59.41, 55.72, 44.20, 41.99, 21.57, 21.55, 21.35, 19.30, 10.82.

FTIR (NaCl, thin film, cm^{-1}): 3477, 2982, 2939, 1738, 1732, 1715, 1428, 1219, 1099, 987, 896.

HRMS (TOF-ESI, m/z): calc'd for $C_{19}H_{25}F_3O_9S$: 504.1510 $[M+NH_4]^+$; found: 504.1521.

diisopropyl 5-(cyclopropylethynyl)-6a-hydroxy-3a,6-dimethyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (14a)



Prepared from **15a** (24.4 mg, 0.05 mmol, 1.0 equiv), cyclopropyl acetylene (22 L, 0.25 mmol, 5.0 equiv), $Pd(PPh_3)_2Cl_2$ (3.5 mg, 0.005 mmol, 0.1 equiv), CuI (1.91 mg, 0.01 mmol, 0.2 equiv), Et_3N (0.20 mL, 0.25 M) at 50 ºC following General Procedure B. The resulting

crude residue was purified by column chromatography (silica, 30% EtOAc/hexanes) to afford **14a** (10.1 mg, 51% yield) as a brown oil.

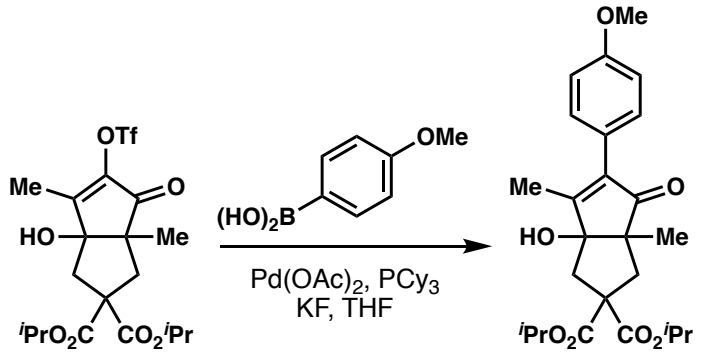
^{1}H NMR (400 MHz, $CDCl_3$) δ 5.02 (hept, *J* = 6.3 Hz, 1H), 4.89 (hept, *J* = 6.3 Hz, 1H), 2.73 (dd, *J* = 14.2, 1.3 Hz, 1H), 2.58 (dd, *J* = 13.9, 1.2 Hz, 1H), 2.56 (s, 1H), 2.44 (dd, *J* = 13.9, 0.8 Hz, 1H), 2.16 (s, 3H), 2.12 (dd, *J* = 14.1, 0.8 Hz, 1H), 1.44 (tt, *J* = 8.2, 5.0 Hz, 1H), 1.22 (dd, *J* = 6.3, 2.2 Hz, 6H), 1.18 (t, *J* = 6.0 Hz, 6H), 1.12 (s, 3H), 0.93 – 0.80 (m, 2H), 0.81 – 0.75 (m, 2H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 206.20, 175.31, 171.87, 169.34, 125.41, 103.47, 87.80, 70.10, 69.76, 65.93, 59.21, 56.14, 44.03, 42.93, 30.10, 29.84, 22.84, 21.58, 21.57, 21.45, 19.47, 13.57, 9.21, 1.16.

FTIR (NaCl, thin film, cm^{-1}): 3480, 2927, 2870, 2231, 1716, 1627, 1456, 1386, 1259, 1098.

HRMS (TOF-ESI, m/z): calc'd for $C_{23}H_{30}O_6$: 403.2155 $[M+H]^+$; found: 403.2135.

diisopropyl 6a-hydroxy-5-(4-methoxyphenyl)-3a,6-dimethyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (14b)



Prepared from **15a** (43 mg, 0.09 mmol, 1.0 equiv), (*p*-methoxyphenyl boronic acid (41 mg, 0.27 mmol, 3.0 equiv), potassium fluoride (30.8 mg, 0.53 mmol, 6.0 equiv), THF (5 mL),

and catalyst stock solution (0.9 mL, 0.10 M, 0.04 equiv [Pd], 0.05 equiv PCy_3), prepared from $Pd(OAc)_2$ (4.5 mg, 0.02 mmol) and tricyclohexyl phosphine (7.1 mg, 0.025 mmol), at 60 ºC following a literature procedure.[37] The crude residue was purified by column chromatography (silica, 30 to 40% EtOAc/hexanes) to afford **14c** (36.4 mg, 92% yield) as a white amorphous solid.

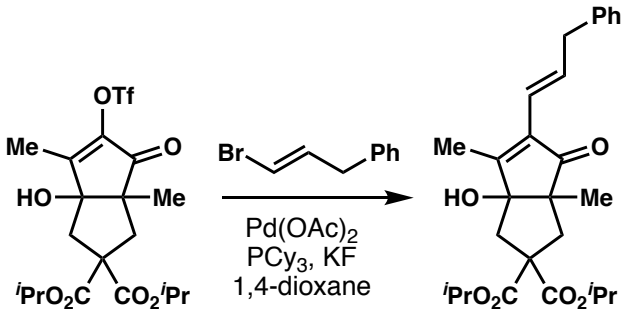
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.32 – 7.22 (m, 2H), 6.96 – 6.88 (m, 2H), 5.04 (hept, *J* = 6.3 Hz, 1H), 4.79 (hept, *J* = 6.3 Hz, 1H), 3.82 (s, 3H), 2.78 (ddd, *J* = 17.5, 13.9, 1.5 Hz, 2H), 2.48 (d, *J* = 12.4 Hz, 1H), 2.19 (d, *J* = 14.0 Hz, 1H), 2.18 (s, 3H), 1.22 (dd, *J* = 6.3, 1.2 Hz, 6H), 1.09 (d, *J* = 6.3 Hz, 3H), 0.99 (d, *J* = 6.3 Hz, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 207.85, 171.68, 169.63, 168.08, 159.46, 138.41, 130.55, 123.30, 113.76, 87.60, 69.99, 69.52, 58.88, 55.67, 55.39, 43.87, 43.14, 21.62, 21.53, 21.23, 19.59, 12.84.

FTIR (NaCl, thin film, cm^{-1}): 3471, 3047, 2843, 2549, 2248, 2056, 1713, 1454, 1249.

HRMS (TOF-ESI, m/z): calc'd for $C_{25}H_{32}O_7$: 445.2221$[M+H]^+$; found: 445.2242.

diisopropyl (*E*)-6a-hydroxy-3a,6-dimethyl-4-oxo-5-(3-phenylprop-1-en-1-yl)-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (14c)



Prepared from **15a** (49 mg, 0.1 mmol, 1.0 equiv), (E)-(3-bromoallyl)benzene (49 mg, 0.3 mmol, 3.0 equiv), potassium fluoride (35 mg, 0.6 mmol, 6.0 equiv), 1,4-dioxane (1 mL), and catalyst stock solution (0.5 mL, 0.10 M, 0.04 equiv [Pd], 0.05 equiv PCy_3), prepared from $Pd(OAc)_2$ (1.8 mg, 0.008 mmol) and tricyclohexyl phosphine (2.8 mg, 0.01 mmol), at 60 ºC following a literature procedure.[37] The crude residue was purified by column chromatography (silica, 30% (Et_2O/EtOAc, 1:1)/hexanes) to afford **14c** (35.3 mg, 78% yield) as a white amorphous solid.

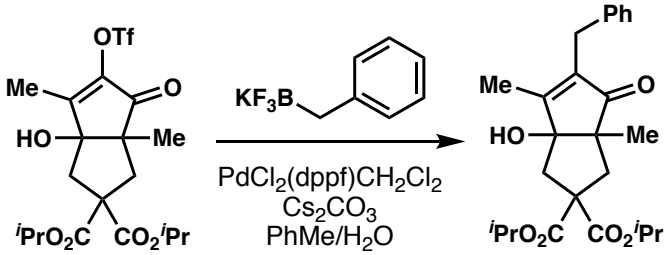
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.33 – 7.23 (m, 2H), 7.19 (tt, *J* = 6.4, 1.2 Hz, 3H), 6.94 (dt, *J* = 15.8, 7.1 Hz, 1H), 6.03 (dt, *J* = 15.8, 1.6 Hz, 1H), 5.01 (hept, *J* = 6.2 Hz, 1H), 4.74 (pd, *J* = 6.3, 1.0 Hz, 1H), 3.51 – 3.44 (m, 2H), 2.74 (dd, *J* = 13.9, 1.4 Hz, 1H), 2.65 (dt, *J* = 13.7, 1.2 Hz, 1H), 2.40 (d, *J* = 13.7 Hz, 1H), 2.12 (d, *J* = 14.2 Hz, 1H), 2.10 (s, 3H), 1.21 (dd, *J* = 6.3, 1.5 Hz, 6H), 1.11 (t, *J* = 3.1 Hz, 6H), 1.05 (d, *J* = 6.3 Hz, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 13C NMR (101 MHz, CDCl3) δ 208.26, 171.72, 169.43, 167.36, 139.74, 136.99, 133.88, 128.78, 128.59, 126.29, 119.62, 87.37, 69.95, 69.56, 58.88, 55.87, 43.83, 43.07, 40.51, 21.59, 21.48, 21.37, 19.59, 12.03.

FTIR (NaCl, thin film, cm^{-1}): 3470, 2981, 2929, 1728, 1715, 1454, 1375, 1258, 1258, 1100.

HRMS (TOF-ESI, m/z): calc'd for $C_{27}H_{34}O_6$: 437.2323 $[M–H_2O+H]^+$; found: 437.2320.

diisopropyl 5-benzyl-6a-hydroxy-3a,6-dimethyl-4-oxo-3,3a,4,6a-tetrahydropentalene-2,2(1*H*)-dicarboxylate (14d)



Prepared from **15a** (36.8 mg, 0.08 mmol, 1.0 equiv), $BnBF_3K$ (41.9 mg, 0.23 mmol, 3.01 equiv), Cs_2CO_3 (70.7 mg, 0.22 mmol, 2.9 equiv), $PdCl_2(dppf)CH_2Cl_2$ (5.9 mg, 0.007 mmol, 0.1 equiv), and PhMe (0.43 mL) and water (0.14 mL, 0.125 M total (PhMe/H_2O, 3:1)) at 80 ºC following a literature procedure.[38] The crude residue was purified by column chromatography (silica, 10 to 20% EtOAc/hexanes) to afford **14d** (27.9 mg, 86% yield) as a yellow gel.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.38 – 7.29 (m, 2H), 7.29 – 7.21 (m, 3H), 5.13 (hept, *J* = 6.2 Hz, 1H), 4.81 (heptd, *J* = 6.3, 1.3 Hz, 1H), 3.63 (d, *J* = 14.5 Hz, 1H), 3.47 (d, *J* = 14.5 Hz, 1H), 2.80 (dd, *J* = 14.2, 1.1 Hz, 1H), 2.68 (s, 1H), 2.65 (dd, *J* = 14.0, 1.1 Hz, 1H), 2.53 (d, *J* = 14.0 Hz, 1H), 2.26 (d, *J* = 14.1 Hz, 1H), 2.16 (s, 3H), 1.32 (dd, *J* = 6.3, 2.8 Hz, 6H), 1.23 (dd, *J* = 6.2, 1.5 Hz, 6H), 1.20 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 208.88, 172.03, 169.47, 169.14, 138.67, 138.36, 128.67, 128.61, 126.25, 88.15, 70.00, 69.35, 59.31, 59.29, 55.95, 43.92, 42.41, 29.14, 21.58, 21.55, 21.45, 19.58, 12.18.

FTIR (NaCl, thin film, cm^{-1}): 3460, 2981, 2933, 1728, 1645, 1376, 1259, 1098.

HRMS (TOF-ESI, m/z): calc'd for $C_{25}H_{32}O_6$: 429.2272 $[M+H]^+$; found: 429.2289.

1.9 REFERENCES

(1) (a) Khand, I. U.; Knox, G. R.; Pauson, P. L.; Watts, W. E.; Foreman, M. I. Organocobalt Complexes. Part II. Reaction of Acetylenehexacarbonyldicobalt Complexes, $(R^1C_2R^2)Co_2(CO)_6$, with Norbornene and its Derivatives. *J. Chem. Soc., Perkin Trans. 1* **1973**, 977. For seminal examples of PKRs in total synthesis, see: (b) Exon, C.; Magnus, P. Stereoselectivity of Intramolecular Dicobalt Octacarbonyl Alkene–Alkyne Cyclizations: Short Synthesis of *dl*-Coriolin. J. Am. Chem. Soc. **1983**, *105*, 2477. (c) Ornum, S. G. V.; Hoerner, S.; Cook, J. M. Recent Adventures with the Pauson-Khand Reaction in Total Synthesis. In *The Pauson-Khand Reaction*; John Wiley & Sons, Ltd, **2012**; pp 211–238.

(2) (a) Shibata, T. Recent Advances in the Catalytic Pauson–Khand-Type Reaction. *Adv. Synth. Catal.* **2006**, *348*, 2328. (b) Gibson, S. E.; Mainolfi, N. The Intermolecular Pauson-Khand Reaction. *Angew. Chem., Int. Ed.* **2005**, *44*, 3022. For a recent approach to the selective formation of intermolecular PKR products *via* siloxy-tethering, see: (c) Gallagher, A. G.; Tian, H.; Torres-Herrera, O. A.; Yin, S.; Xie, A.; Lange, D. M.; Wilson, J. K.; Mueller, L. G.; Gau, M. R.; Carroll, P. J.; Martinez-Solorio, D. Access to Highly Functionalized Cyclopentenones via Diastereoselective Pauson–Khand Reaction of Siloxy-Tethered 1,7-Enynes. *Org. Lett.* **2019**, *21*, 8646.

(3) (a) Chuang, K. V.; Xu, C.; Reisman, S. E. A 15-Step Synthesis of (+)-Ryanodol. *Science* **2016**, *353*, 912. (b) Xu, C.; Han, A.; Virgil, S. C.; Reisman, S. E. Chemical Synthesis of (+)-Ryanodine and (+)-20-Deoxyspiganthine. *ACS Cent. Sci.* **2017**, *3*, 278.

(4) The use of fewer equiv SeO_2 (1.5 or 3.0) proved to be less general (see SI), thus the conditions of entry 3 were selected as optimal for investigating the scope of the dioxidation reaction.

(5) The *trans*-fused isomer of **8a** was calculated to be 18.5 kcal/mol less stable (calculated using Spartan'14, Hartree-Fock model using the 6-311+G** basis set).

(6) Stereochemical assignments of oxidation products were made by analogy to **8e**.

(7) Both enantiomers of **7a***, obtained by chromatographic separation, were used for these studies (see SI). Efforts to prepare **7a*** by asymmetric PKR gave the product

in 59% yield and 72% ee ($[RhCl(CO)_2]_2$, (R)-BINAP, PhMe, CO, 90 °C). See: Furusawa, T.; Morimoto, T.; Ikeda, K.; Tanimoto, H.; Nishiyama, Y.; Kakiuchi, K.; Jeong, N. Asymmetric Pauson–Khand-Type Reactions of 1,6-enynes Using Formaldehyde as a Carbonyl Source by Cooperative Dual Rhodium Catalysis. *Tetrahedron* **2015**, *71*, 875.

(8) (a) Sharpless, K. B.; Lauer, R. F. Selenium Dioxide Oxidation of Olefins. Evidence for the Intermediacy of Allylseleninic Acids. *J. Am. Chem. Soc.* **1972**, *94*, 7154. (b) Arigoni, D.; Vasella, A.; Sharpless, K. B.; Jensen, H. P. Selenium Dioxide Oxidations of Olefins. Trapping of the Allylic Seleninic Acid Intermediate as a Seleninolactone. *J. Am. Chem. Soc.* **1973**, *95*, 7917. (c) Młochowski, J; Wójtowicz-Młochowska, H. Developments in Synthetic Application of Selenium(IV) Oxide and Organoselenium Compounds as Oxygen Donors and Oxygen-Transfer Agents. *Molecules* **2015**, *20*, 10205.

(9) The structure of **12a** was determined by ^{1}H, ^{13}C, and HSQC NMR (see SI). (a) Riley, H. L.; Morley, J. F.; Friend, N. A. C. Selenium Dioxide, a New Oxidising Agent. Part I. Its Reaction with Aldehydes and Ketones. *J. Chem. Soc.* **1932**, 1875. (b) Sharpless, K. B.; Gordon, K. M. Selenium Dioxide Oxidation of Ketones and Aldehydes. Evidence for the Intermediacy of b-Ketoseleninic Acids. *J. Am. Chem. Soc.* **1976**, *98*, 300. For examples of Riley oxidation in synthesis, see: (c) Mehta, G., Shinde, H. M. Enantiospecific total synthesis of 6-epi-(-)-hamigeran B. Intramolecular Heck reaction in a sterically constrained environment. *Tetrahedron Lett.* **2003**, 44, 7049. (d) Trost, B. M.; Pissot-Soldermann, C.; Chen, I.; Schroeder, G. M. An Asymmetric Synthesis of Hamigeran B via a Pd Asymmetric Allylic Alkylation for Enantiodiscrimination. *J. Am. Chem. Soc.* **2004**, *126*, 4480. (e) Condakes, M. L.; Hung, K.; Harwood, S. J.; Maimone, T. J. Total Syntheses of (−)-Majucin and (−)-Jiadifenoxolane A, Complex Majucin-Type Illicium Sesquiterpenes. *J. Am. Chem. Soc.* **2017**, *139*, 17783.

(10) For precedent of an analogous intramolecular oxy-Michael addition to an ene-diketone, see: (a) Kurosu, M.; Marcin, L. R.; Grinsteiner, T. J.; Kishi, Y. Total Synthesis of (±)-Batrachotoxinin A. *J. Am Chem. Soc.* **1998**, *120*, 6627. (b) Xu, C.;

Han, A.; Reisman, S. E. An Oxidative Dearomatization Approach To Prepare the Pentacyclic Core of Ryanodol. *Org. Lett.* **2018**, *20*, 3793.

(11) Hansch, Corwin.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91*, 165.

(12) Because an intermediate analogous to **12a** was not observed during the oxidation of substrates without γ-substitution (**7a-g**), further application of this protocol is limited.

(13) (a) Littke, A. F.; Dai, C.; Fu, G. C. Versatile Catalysts for the Suzuki Cross-Coupling of Arylboronic Acids with Aryl and Vinyl Halides and Triflates under Mild Conditions. *J. Am. Chem. Soc.* **2000**, *122*, 4020. (b) Molander, G. A.; Ito, T. Cross-Coupling Reactions of Potassium Alkyltrifluoroborates with Aryl and 1-Alkenyl Trifluoromethanesulfonates. *Org. Lett.* **2001**, *3*, 393. For discussion of the challenges of cross-coupling of a similar cyclic enol triflate, see: (c) Jolit, A.; Dickinson, C. F.; Kitamura, K.; Walleser, P. M.; Yap, G. P. A.; Tius, M. A. Catalytic Enantioselective Nazarov Cyclization: Catalytic Enantioselective Nazarov Cyclization. *Eur. J. Org. Chem.* **2017**, 6067.

(14) Still, W. C.; Kahn, M.; Mitra, A. Rapid Chromatographic Technique for Preparative Separations with Moderate Resolution. *J. Org. Chem.* **1978**, *43*, 2923–2925.

(15) *APEX2, Verison 2 User Manual, M86-E01078*; Bruker Analytical X-ray Systems: Madison, WI, 2006.

(16) Sheldrick, G. M. *SADABS (Version 2008/1): Program for Absorption Correction for Data from Area Detector Frames*; University of Göttingen, 2008.

(17) Sheldrick, G. M. A Short History of SHELX. *Acta Cryst A* **2008**, *64*, 112–122.

(18) Müller, P. Practical Suggestions for Better Crystal Structures. *Crystallography Reviews* **2009**, *15*, 57–83.

(19) Chung, Y. K.; Lee, B. Y.; Jeong, N.; Hudecek, M.; Pauson, P. L. Promoters for the (Alkyne) Hexacarbonyldicobalt-Based Cyclopentenone Synthesis. *Organometallics* **1993**, *12*, 220–223.

(20) Koga, Y.; Kobayashi, T.; Narasaka, K. Rhodium-Catalyzed Intramolecular Pauson-Khand Reaction. *Chem. Lett.* **1998**, *27*, 249–250.

(21) Adrio, J.; Rivero, M. R.; Carretero, J. C. Mild and Efficient Molybdenum-Mediated Pauson−Khand-Type Reaction. *Org. Lett.* **2005**, *7*, 431–434.

(22) Csékei, M.; Novák, Z.; Kotschy, A. Ethynyl-Cyclohexanol: An Efficient Acetylene Surrogate in Sonogashira Coupling. *Tetrahedron* **2008**, *64*, 975–982.

(23) Clavier, H.; Nolan, S. P. N-Heterocyclic Carbene and Phosphine Ruthenium Indenylidene Precatalysts: A Comparative Study in Olefin Metathesis. *Chem. Eur. J.* **2007**, *13*, 8029–8036.

(24) Ohmura, T.; Sasaki, I.; Suginome, M. Catalytic Generation of Rhodium Silylenoid for Alkene–Alkyne–Silylene [2 + 2 + 1] Cycloaddition. *Org. Lett.* **2019**, *21*, 1649–1653.

(25) Hwu, J. R.; Chen, C. N.; Shiao, S.-S. Silicon-Controlled Allylation of 1,3-Dioxo Compounds by Use of Allyltrimethylsilane and Ceric Ammonium Nitrate. *J. Org. Chem.* **1995**, *60*, 856–862.

(26) Kwong, F. Y.; Li, Y. M.; Lam, W. H.; Qiu, L.; Lee, H. W.; Yeung, C. H.; Chan, K. S.; Chan, A. S. C. Rhodium-Catalyzed Asymmetric Aqueous Pauson-Khand-Type Reaction. *Chem. Eur. J.* **2005**, *11*, 3872–3880.

(27) Morimoto, T.; Fuji, K.; Tsutsumi, K.; Kakiuchi, K. CO-Transfer Carbonylation Reactions. A Catalytic Pauson−Khand-Type Reaction of Enynes with Aldehydes as a Source of Carbon Monoxide. *J. Am. Chem. Soc.* **2002**, *124*, 3806–3807.

(28) Kinder, R. E.; Widenhoefer, R. A. Rhodium-Catalyzed Asymmetric Cyclization/Hydroboration of 1,6-Enynes. *Org. Lett.* **2006**, *8*, 1967–1969.

(29) Josephine Bartlett, A.; Laird, T.; David Ollis, W. Base-Catalysed Intramolecular Cycloadditions of Allyl 3-Phenylprop-2-Ynyl Ethers and 4-Methylpent-4-En-2-Ynyl Prop-2-Ynyl Ethers. *J. Chem. Soc., Perkin Trans. 1* **1975**, *0*, 1315–1320.

(30) Berk, S. C.; Grossman, R. B.; Buchwald, S. L. Development of a Titanocene-Catalyzed Enyne Cyclization/Isocyanide Insertion Reaction. *J. Am. Chem. Soc.* **1994**, *116*, 8593–8601.

(31) Zhao, Z.; Ding, Y.; Zhao, G. Bicyclization of Enynes Using the Cp2TiCl2−Mg−BTC System: A Practical Method to Bicyclic Cyclopentenones. *J. Org. Chem.* **1998**, *63*, 9285–9291.

(32) Tanaka, K.; Okazaki, E.; Shibata, Y. Cationic Rhodium(I)−dppf Complex-Catalyzed Olefin Isomerization/Propargyl Claisen Rearrangement/Carbonyl Migration Cascade. *J. Am. Chem. Soc.* **2009**, *131*, 10822–10823.

(33) Litvinas, N. D.; Fier, P. S.; Hartwig, J. F. A General Strategy for the Perfluoroalkylation of Arenes and Arylbromides by Using Arylboronate Esters and [(Phen)CuRF]. *Angew. Chem. Int. Ed.* **2012**, *51*, 536–539.

(34) King, A. E.; Brunold, T. C.; Stahl, S. S. Mechanistic Study of Copper-Catalyzed Aerobic Oxidative Coupling of Arylboronic Esters and Methanol: Insights into an Organometallic Oxidase Reaction. *J. Am. Chem. Soc.* **2009**, *131*, 5044–5045.

(35) Murphy, J. M.; Liao, X.; Hartwig, J. F. Meta Halogenation of 1,3-Disubstituted Arenes via Iridium-Catalyzed Arene Borylation. *J. Am. Chem. Soc.* **2007**, *129*, 15434–15435.

(36) Hansch, Corwin.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91*, 165–195.

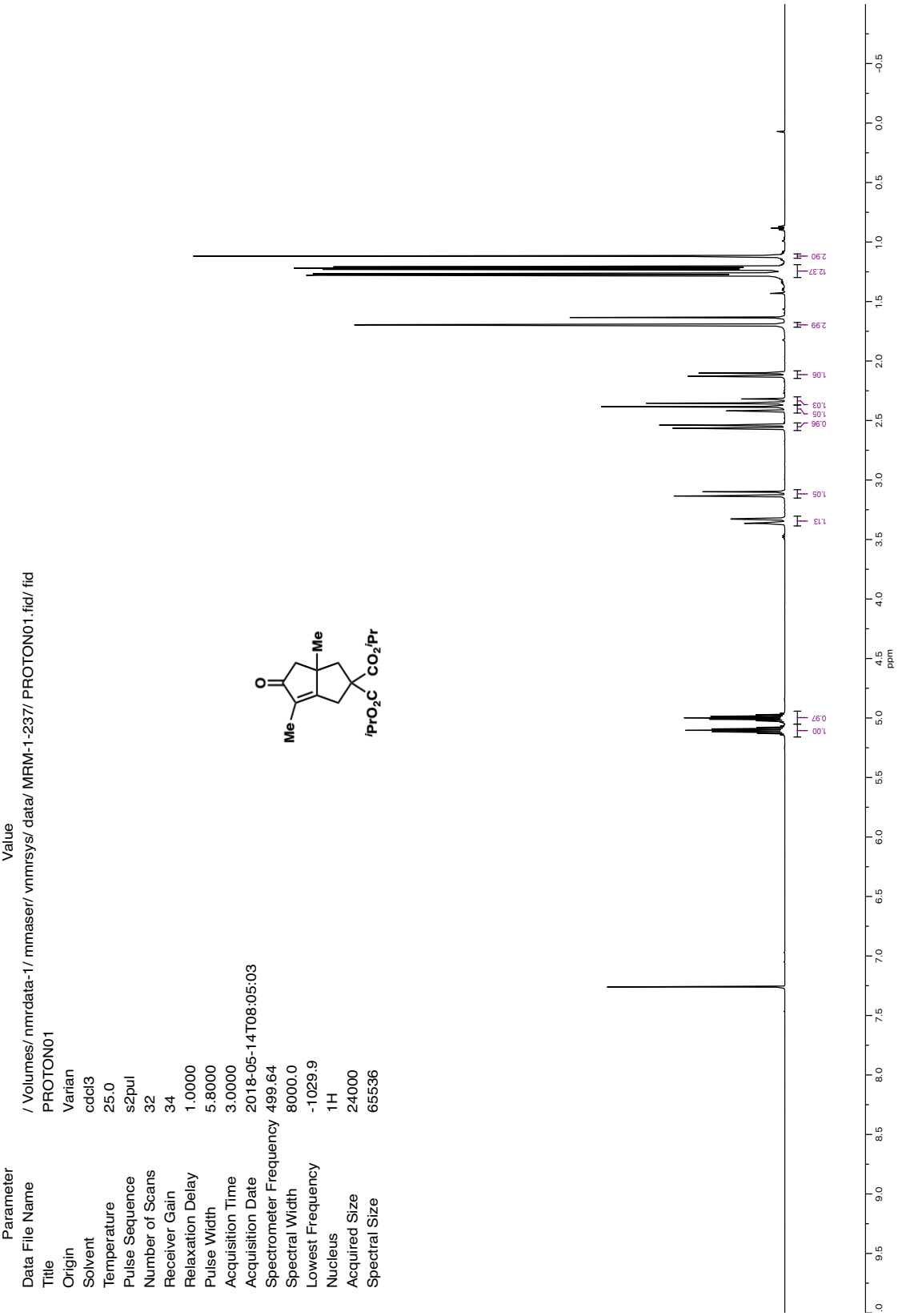
(37) Littke, A. F.; Dai, C.; Fu, G. C. Versatile Catalysts for the Suzuki Cross-Coupling of Arylboronic Acids with Aryl and Vinyl Halides and Triflates under Mild Conditions. *J. Am. Chem. Soc.* **2000**, *122*, 4020–4028.

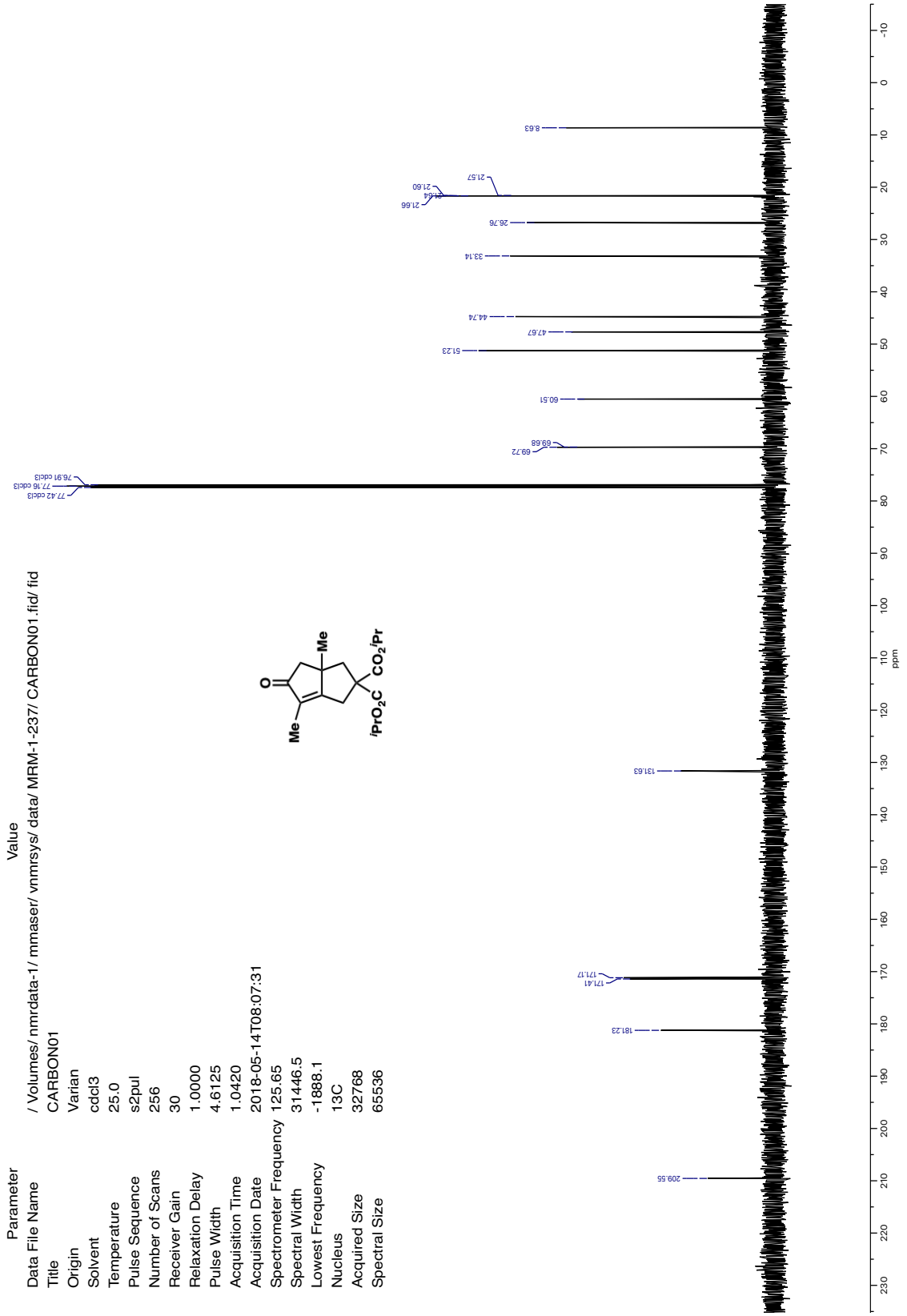
(38) Molander, G. A.; Ito, T. Cross-Coupling Reactions of Potassium Alkyltrifluoroborates with Aryl and 1-Alkenyl Trifluoromethanesulfonates. *Org. Lett.* **2001**, *3*, 393–396.

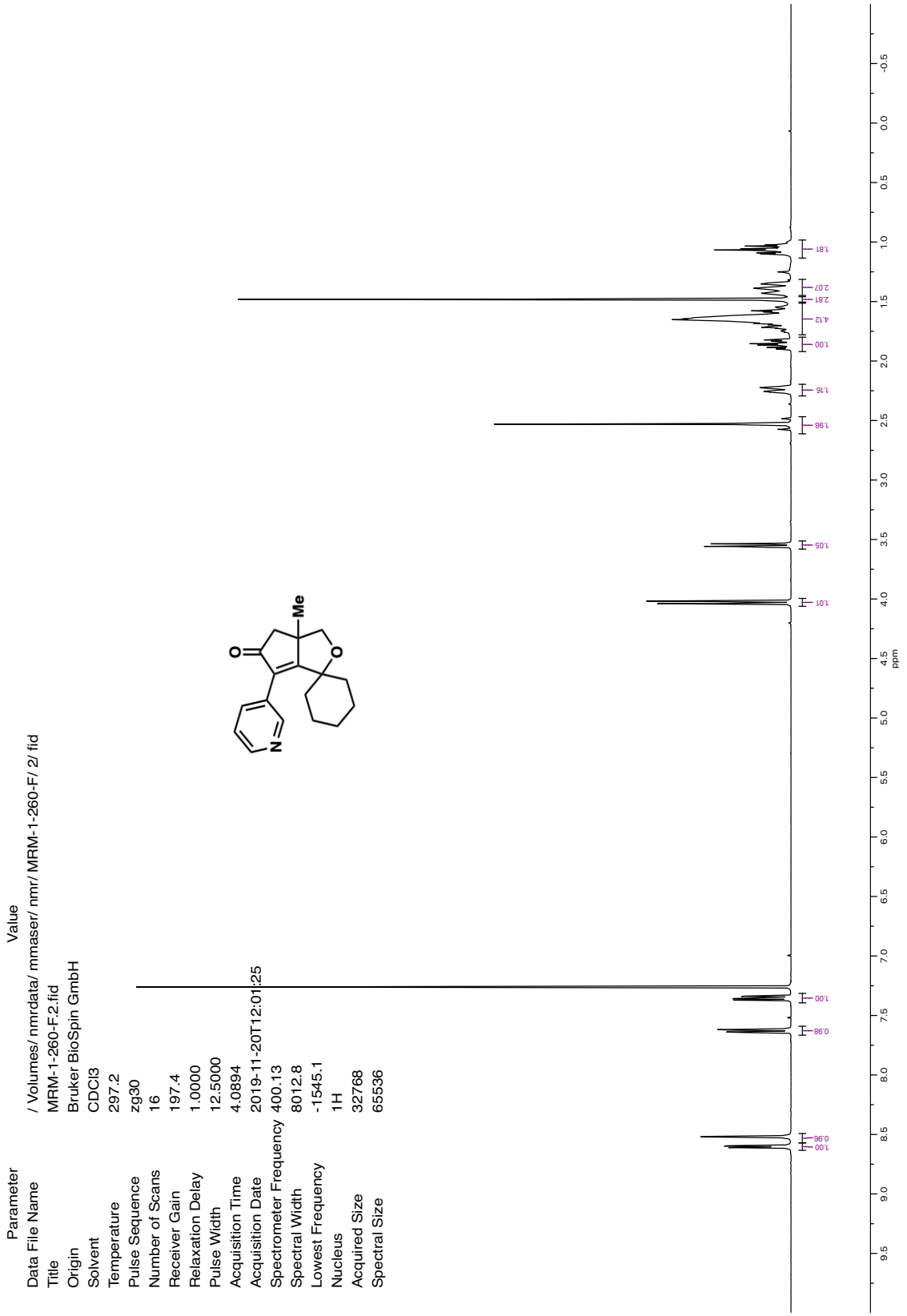
Appendix 1

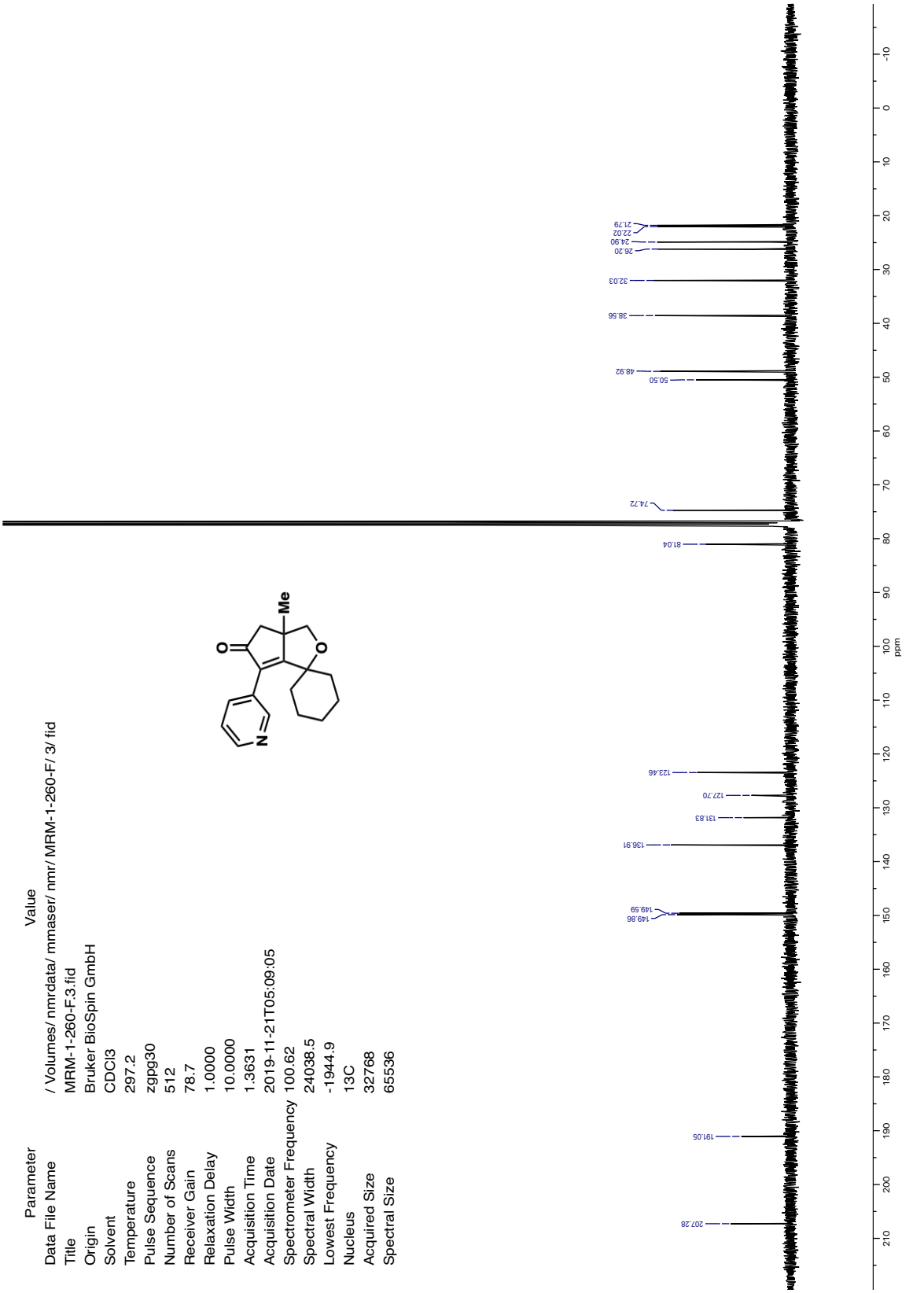
Spectra Relevant to Chapter 1:

SeO2-Mediated Oxidative Transposition of Pauson–Khand Products


Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-237/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 34
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2018-05-14T08:05:03
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1029.9
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
O
Me
Me
$^{i}PrO_2C$ $CO_2{}^{i}Pr$
0.97
1.00
1.05
1.13
0.96
1.03
1.05
1.06
2.99
12.37
2.90
ppm

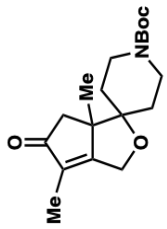

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-237/ CARBON01.fid/ fid
Title CARBON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 256
Receiver Gain 30
Relaxation Delay 1.0000
Pulse Width 4.6125
Acquisition Time 1.0420
Acquisition Date 2018-05-14T08:07:31
Spectrometer Frequency 125.65
Spectral Width 31446.5
Lowest Frequency -1888.1
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
77.42 cdcl3
77.16 cdcl3
76.91 cdcl3
69.72
69.58
60.51
51.23
47.67
44.74
33.14
26.76
21.66
21.64
21.60
21.57
8.63
209.55
181.23
171.41
171.17
131.63
Me
Me
O
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
ppm

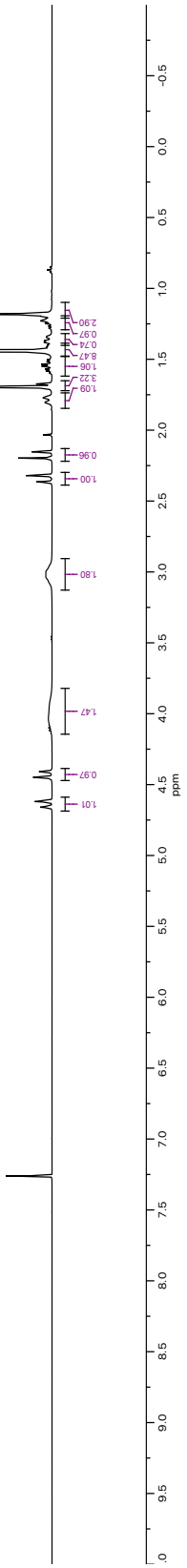

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-260-F/ 2/ fid
Title MRM-1-260-F.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-20T12:01:25
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.1
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
O
O
N
0.96
1.00
0.98
1.00
1.01
1.05
1.98
1.16
1.00
4.12
2.81
2.07
1.81
ppm

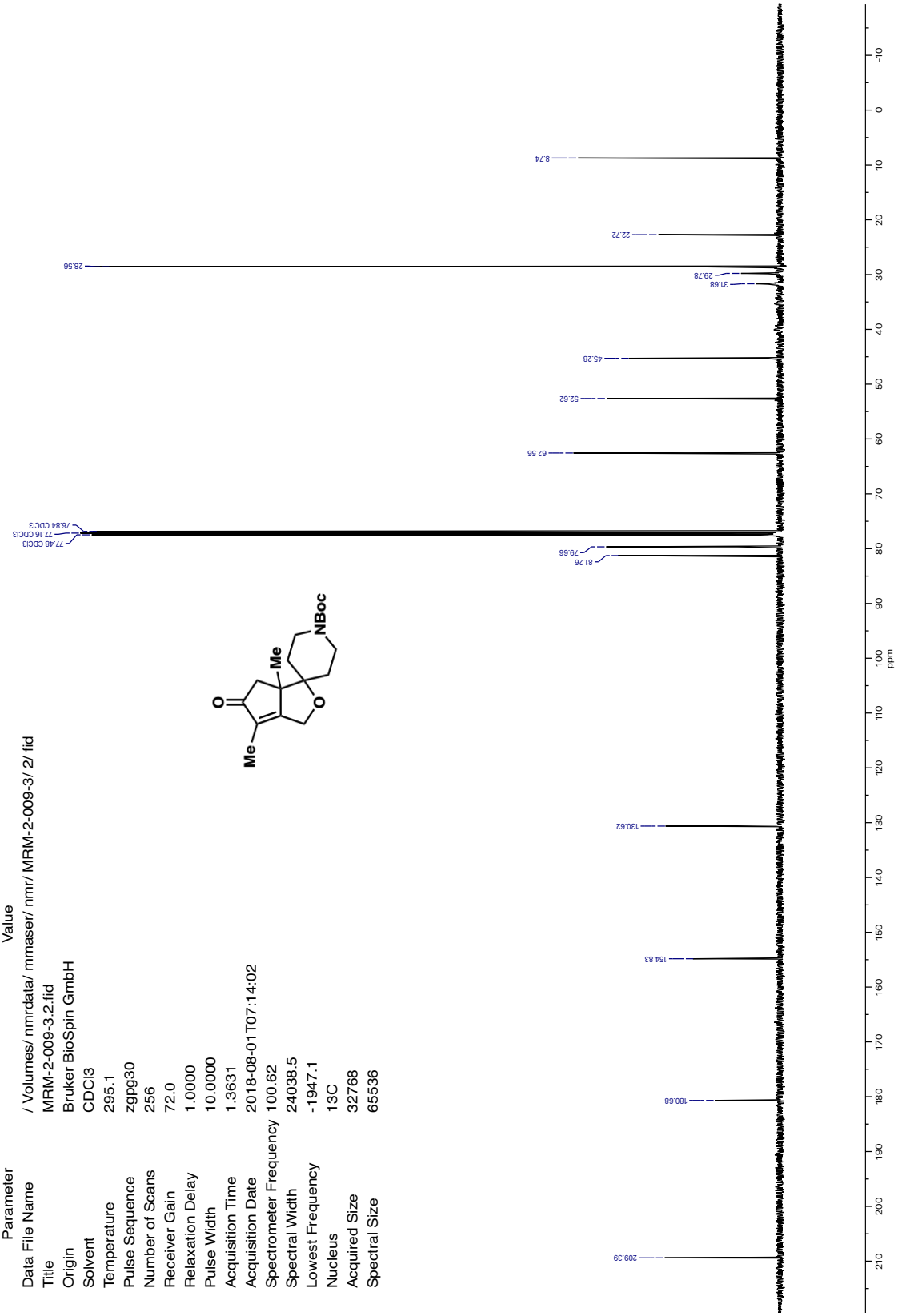


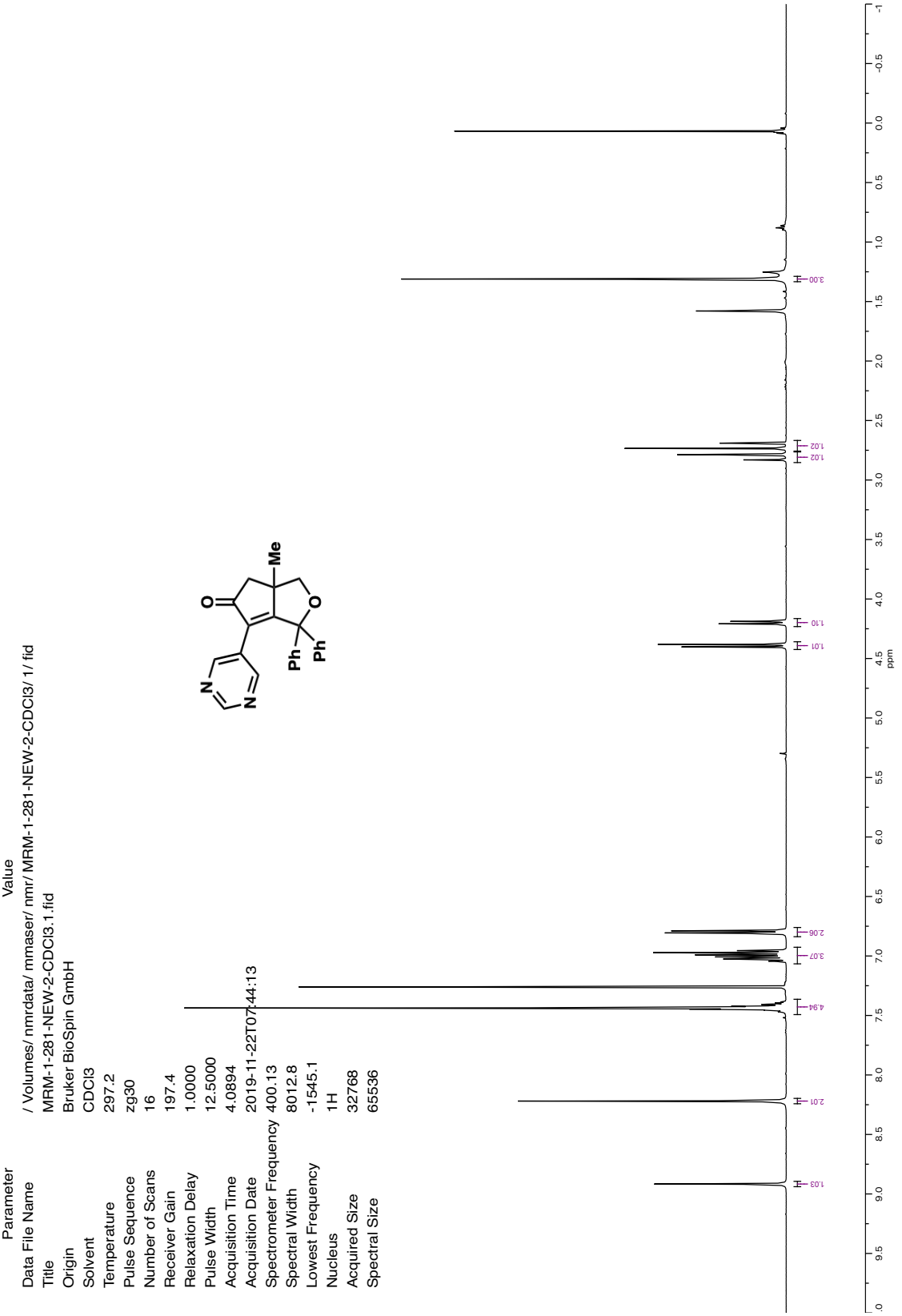
Parameter	Value
Data File Name	/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-260-F/ 3/ fid
Title	MRM-1-260-F.3.fid
Origin	Bruker BioSpin GmbH
Solvent	CDCl3
Temperature	297.2
Pulse Sequence	zgpg30
Number of Scans	512
Receiver Gain	78.7
Relaxation Delay	1.0000
Pulse Width	10.0000
Acquisition Time	1.3631
Acquisition Date	2019-11-21T05:09:05
Spectrometer Frequency	100.62
Spectral Width	24038.5
Lowest Frequency	-1944.9
Nucleus	13C
Acquired Size	32768
Spectral Size	65536

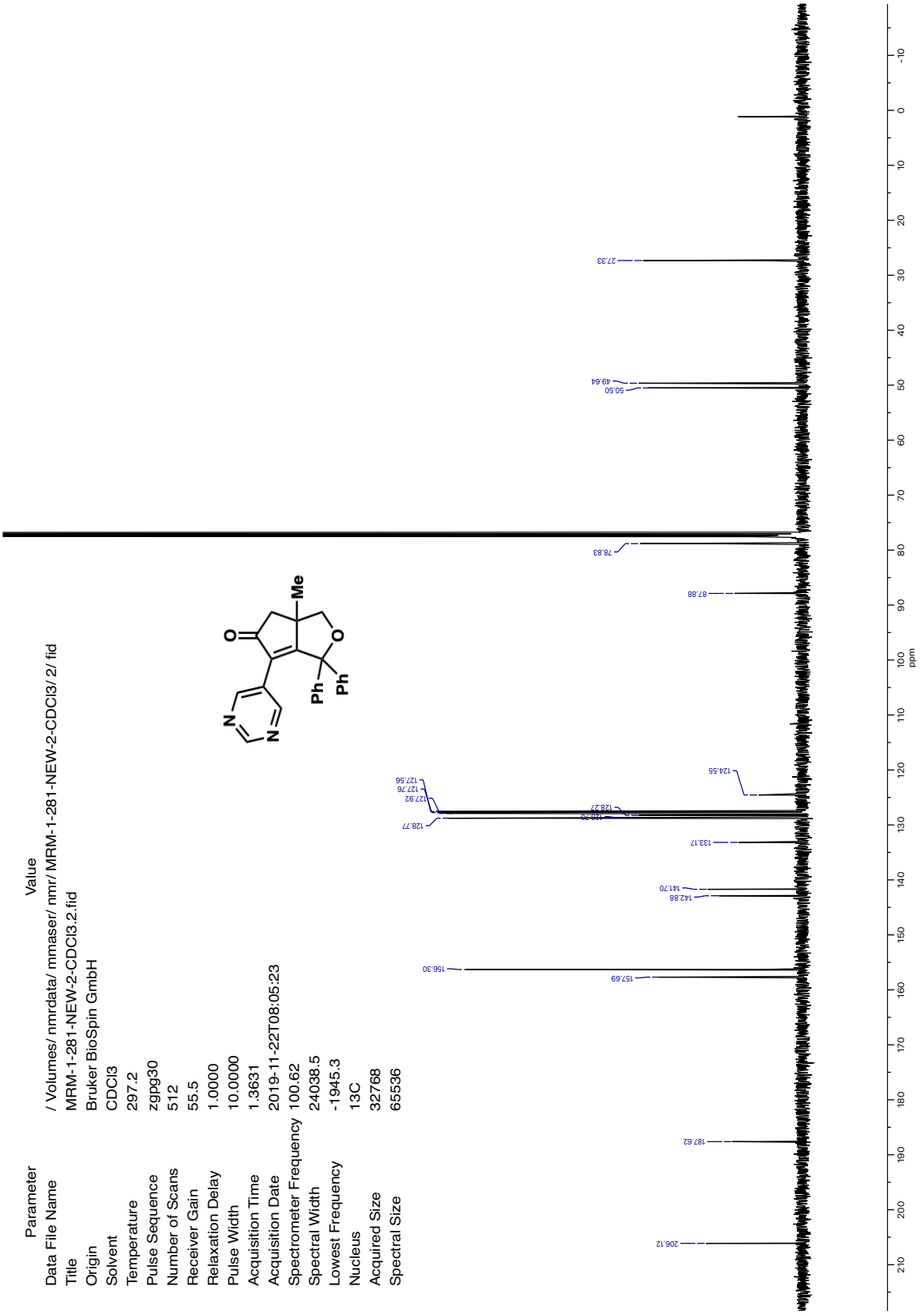
Parameter	Value
Data File Name	/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-009/ 1/ fid
Title	MRM-2-009-3.1.fid
Origin	Bruker BioSpin GmbH
Solvent	CDCl3
Temperature	295.2
Pulse Sequence	zg30
Number of Scans	16
Receiver Gain	78.7
Relaxation Delay	1.0000
Pulse Width	11.7000
Acquisition Time	4.0894
Acquisition Date	2018-08-01T07:03:10
Spectrometer Frequency	400.13
Spectral Width	8012.8
Lowest Frequency	-1545.3
Nucleus	1H
Acquired Size	32768
Spectral Size	65536

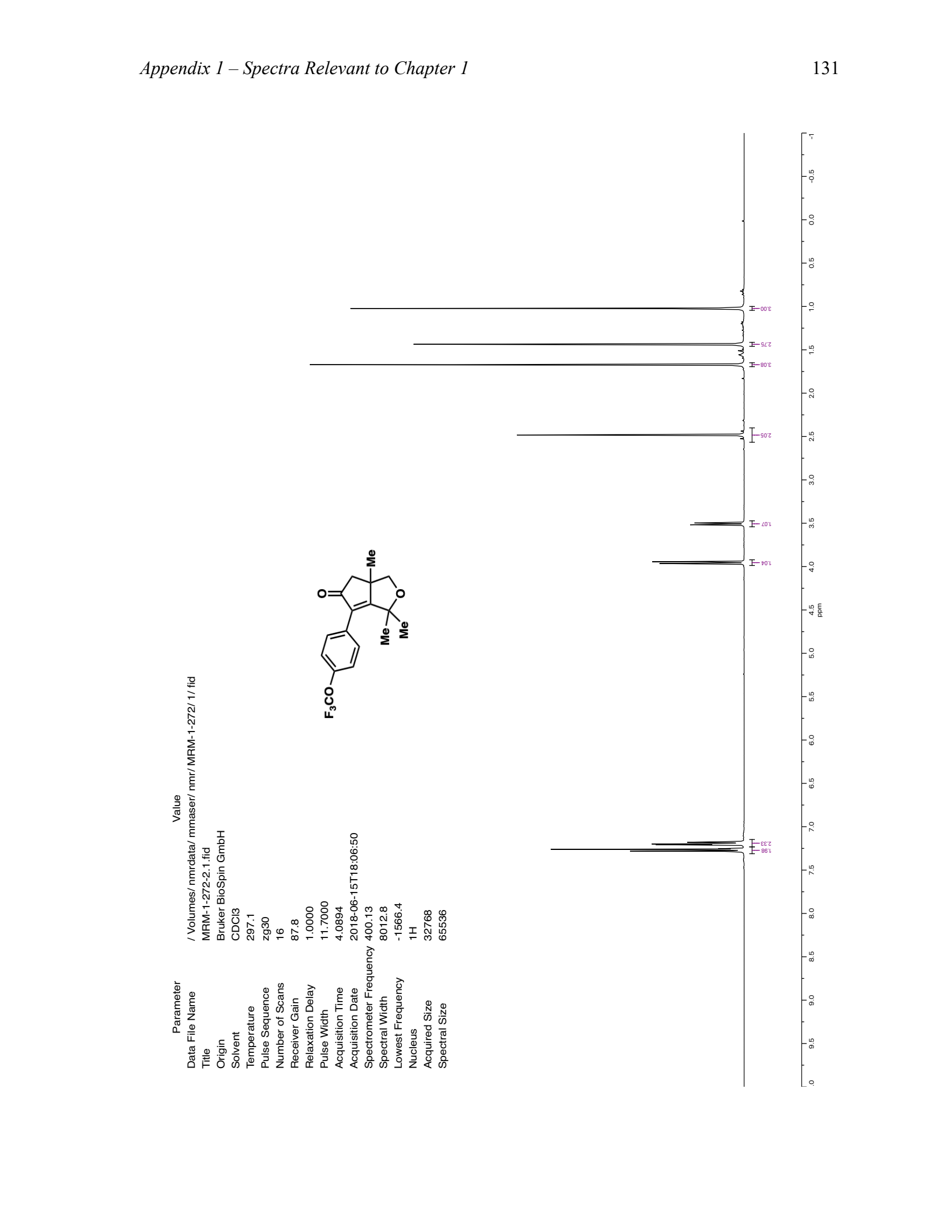


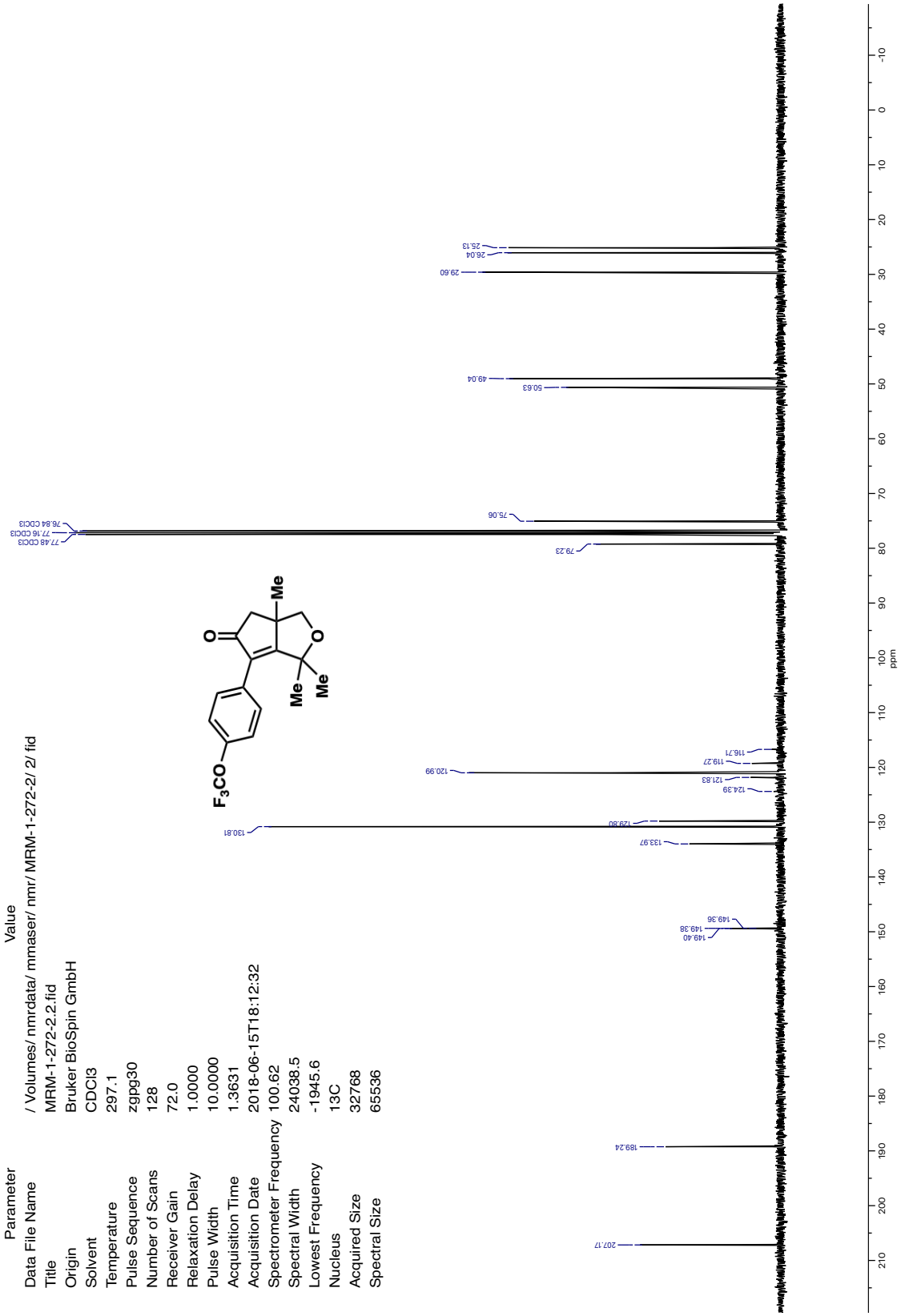


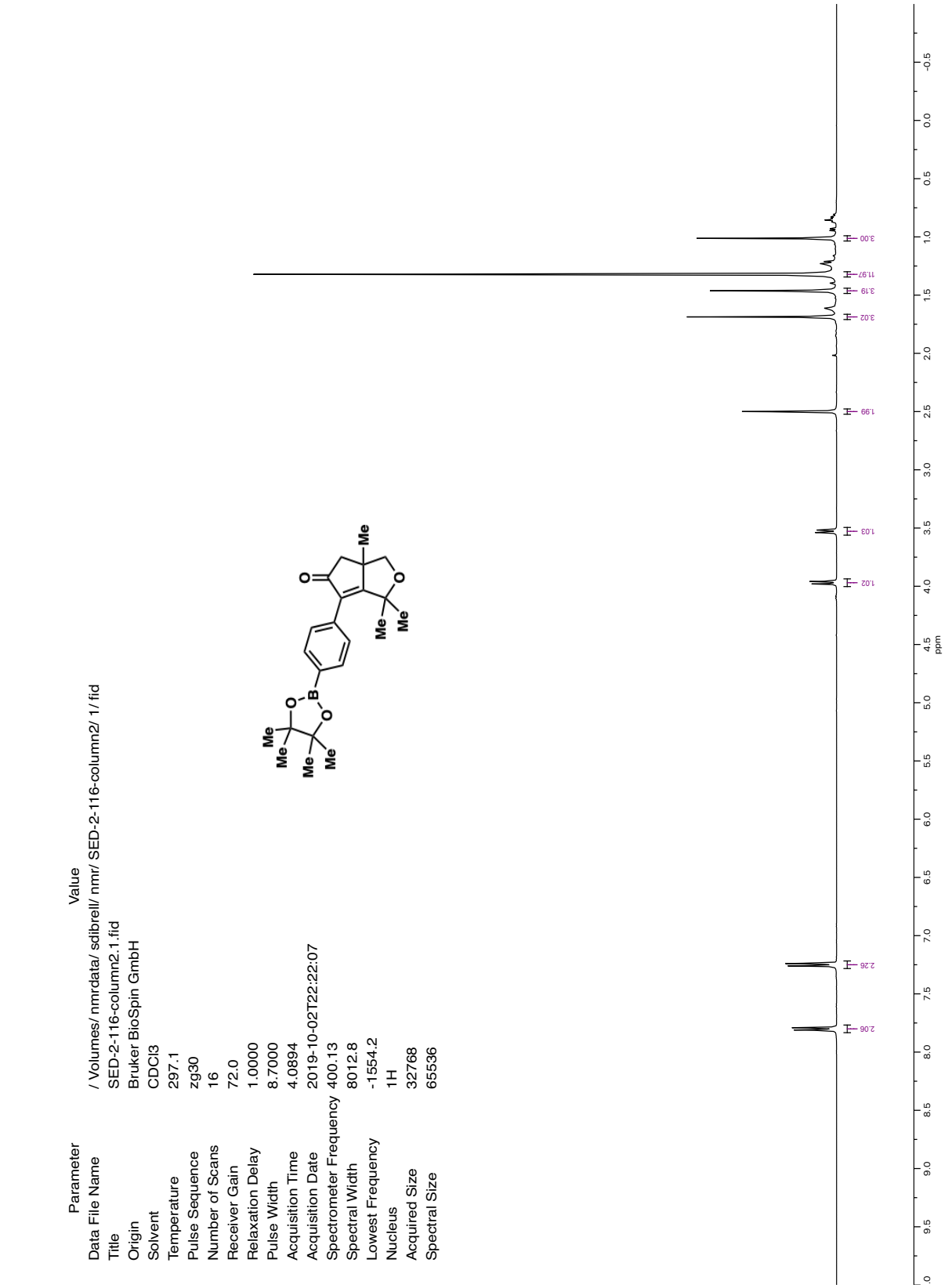

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-009-3/ 2/ fid
Title MRM-2-009-3.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.1
Pulse Sequence zgpg30
Number of Scans 256
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-08-01T07:14:02
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1947.1
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
NBoc
O
O
Me
209.39
180.68
154.83
130.62
81.26
79.66
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
62.56
52.62
45.28
31.68
29.78
28.56
22.72
8.74
ppm

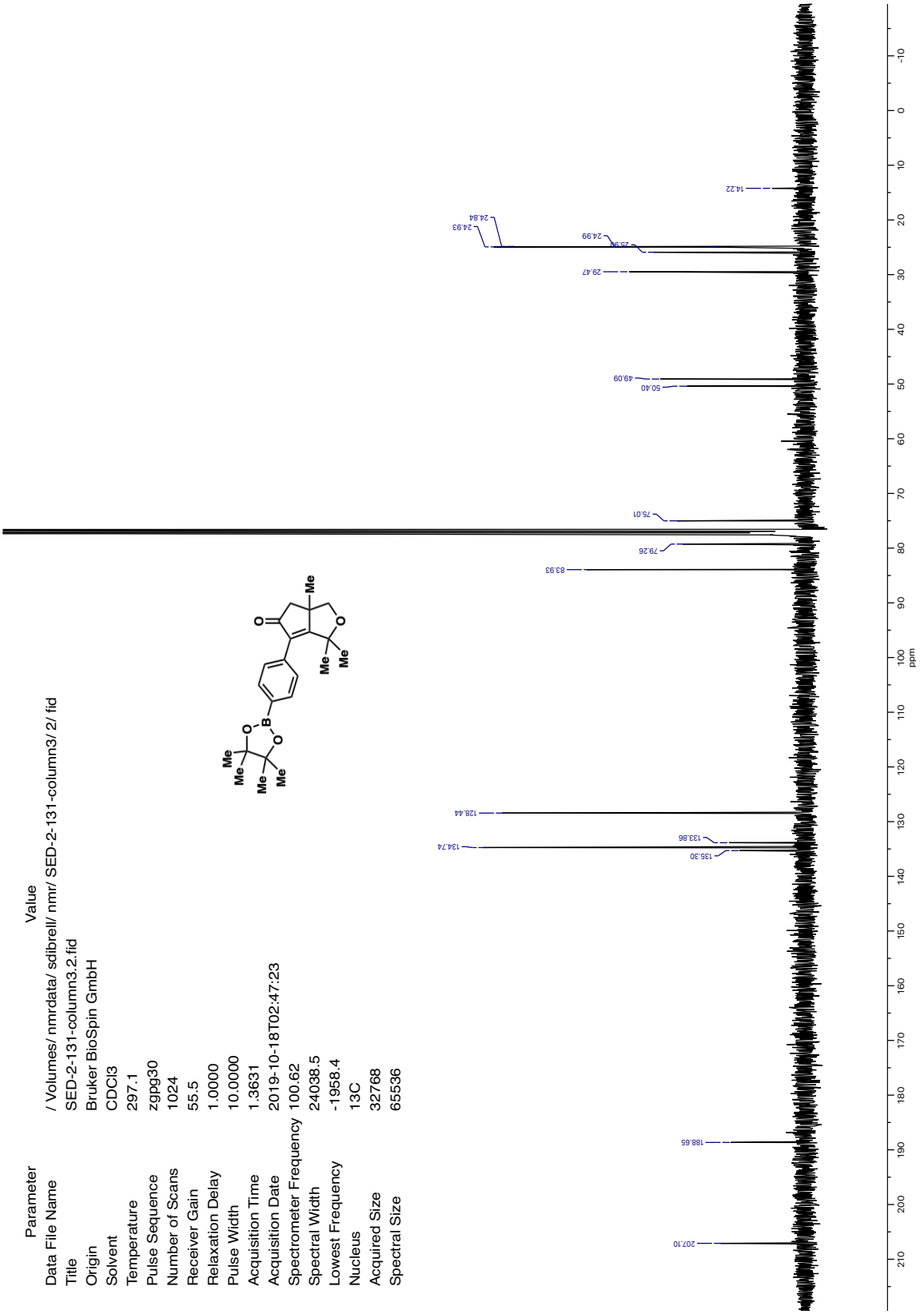

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-281-NEW-2-CDCl3/ 1/ fid
Title MRM-1-281-NEW-2-CDCl3.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-22T07:44:13
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.1
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
O
O
Ph
Ph
N
N
1.03
2.01
4.94
2.06
3.07
1.02
1.02
1.10
1.01
3.00
ppm

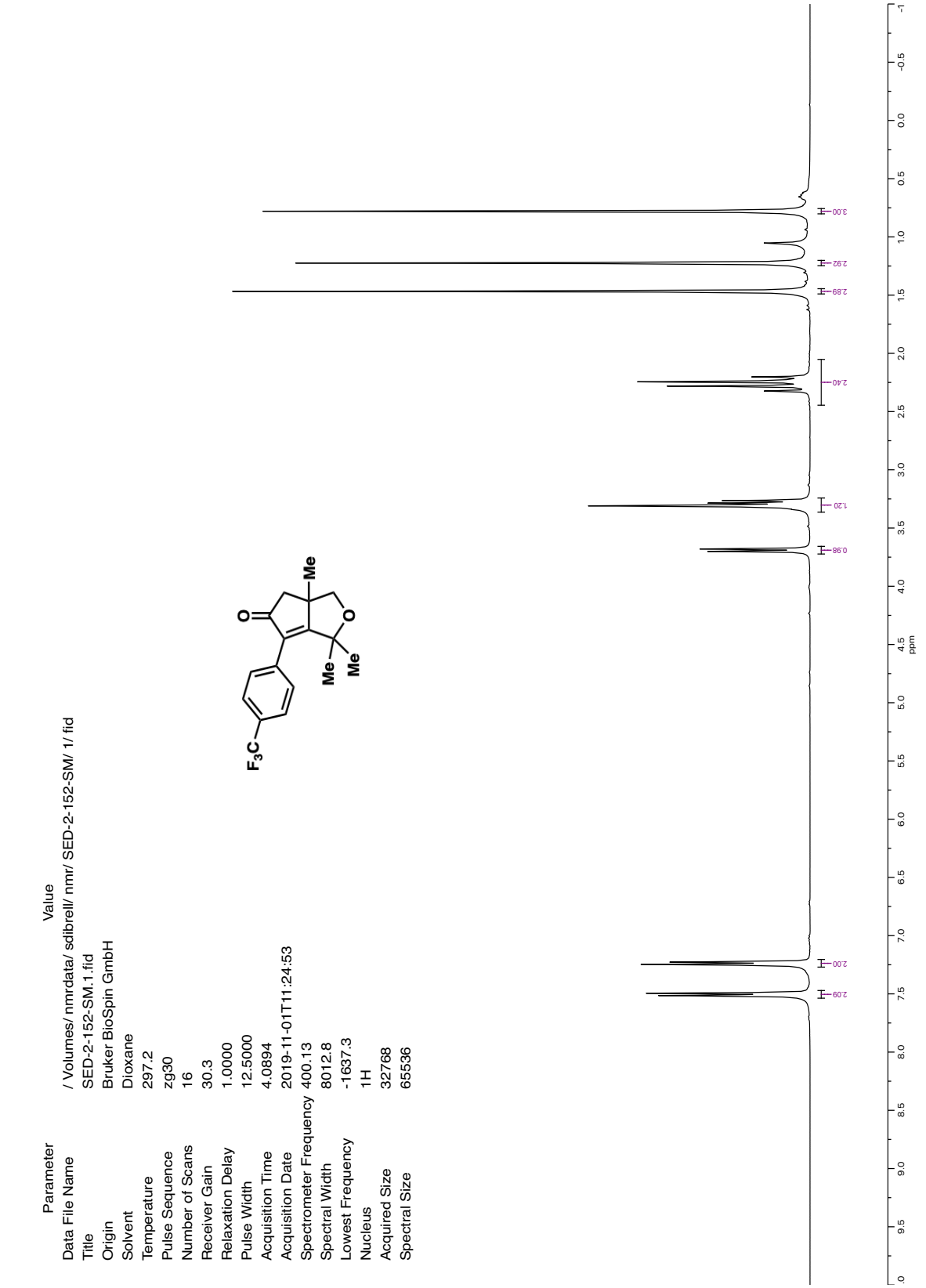

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-281-NEW-2-CDCl3/ 2/ fid
Title MRM-1-281-NEW-2-CDCl3.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-22T08:05:23
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
O
Me
O
Ph
Ph
N
N
206.12
187.62
156.30
157.69
142.88
141.70
133.17
128.77
127.56
127.76
127.92
128.27
128.79
124.55
87.88
78.83
50.50
49.64
27.33
ppm

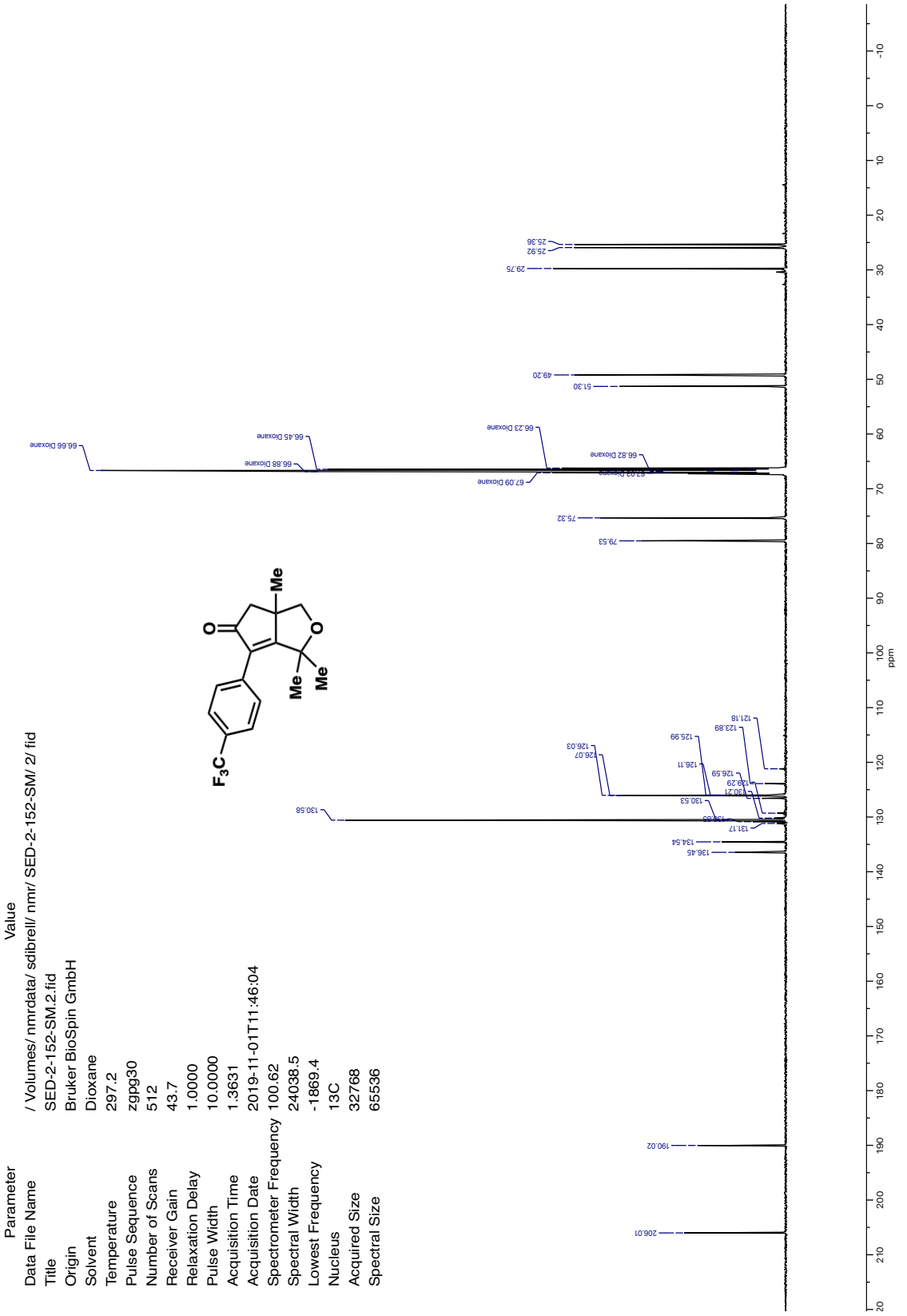

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-272/ 1/ fid
Title MRM-1-272-2.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 87.8
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2018-06-15T18:06:50
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1566.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
Me
O
Me
Me
F3CO
1.98
2.33
1.04
1.07
2.05
3.08
2.75
3.00
ppm

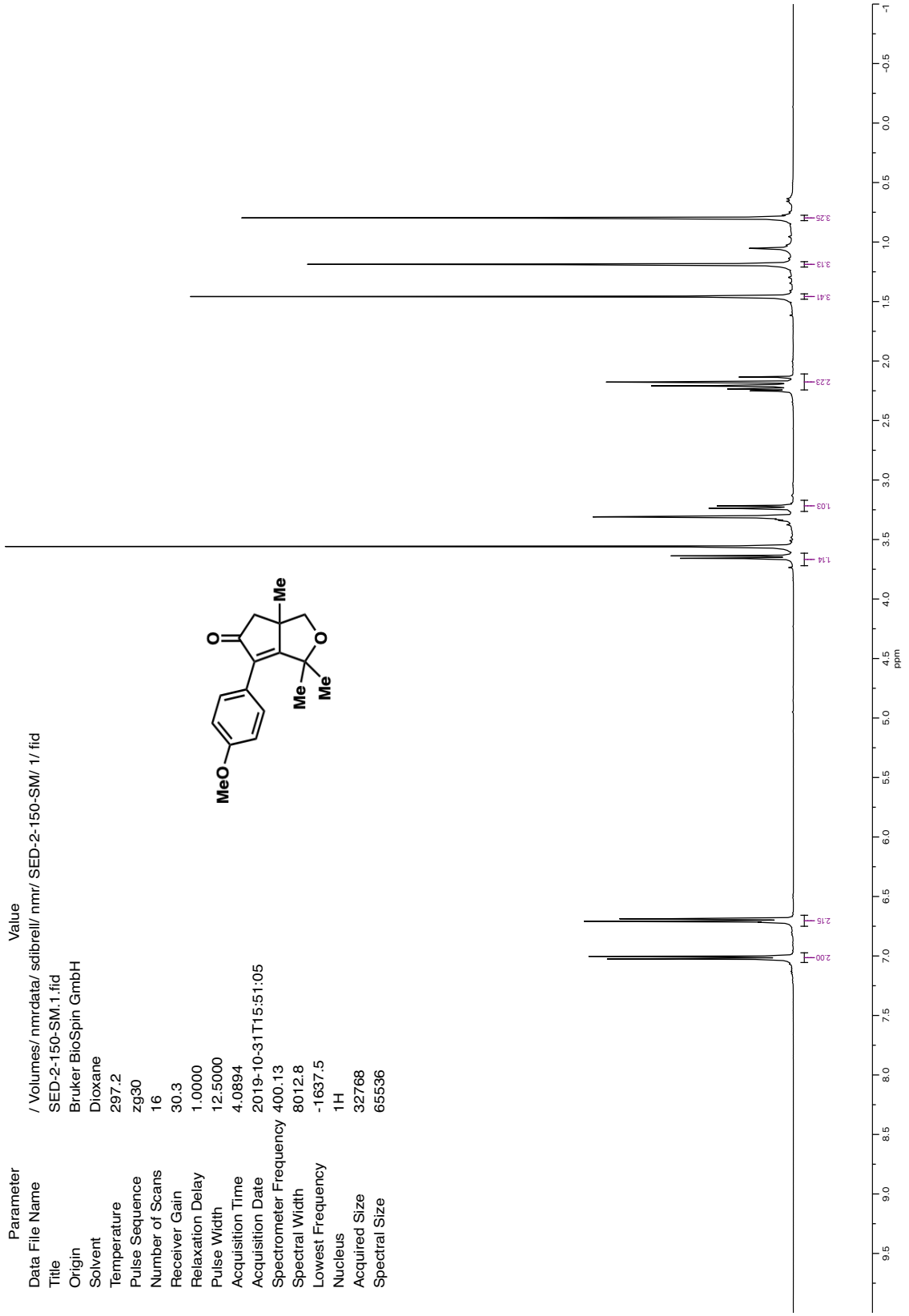

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-272-2/ 2/ fid
Title MRM-1-272-2.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 128
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-06-15T18:12:32
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.6
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
Me
Me
O
O
F3CO
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
207.17
189.24
149.36
149.38
149.40
133.97
129.80
130.81
124.39
121.83
120.99
119.27
116.71
79.23
75.06
50.63
49.04
29.60
26.04
25.13
ppm

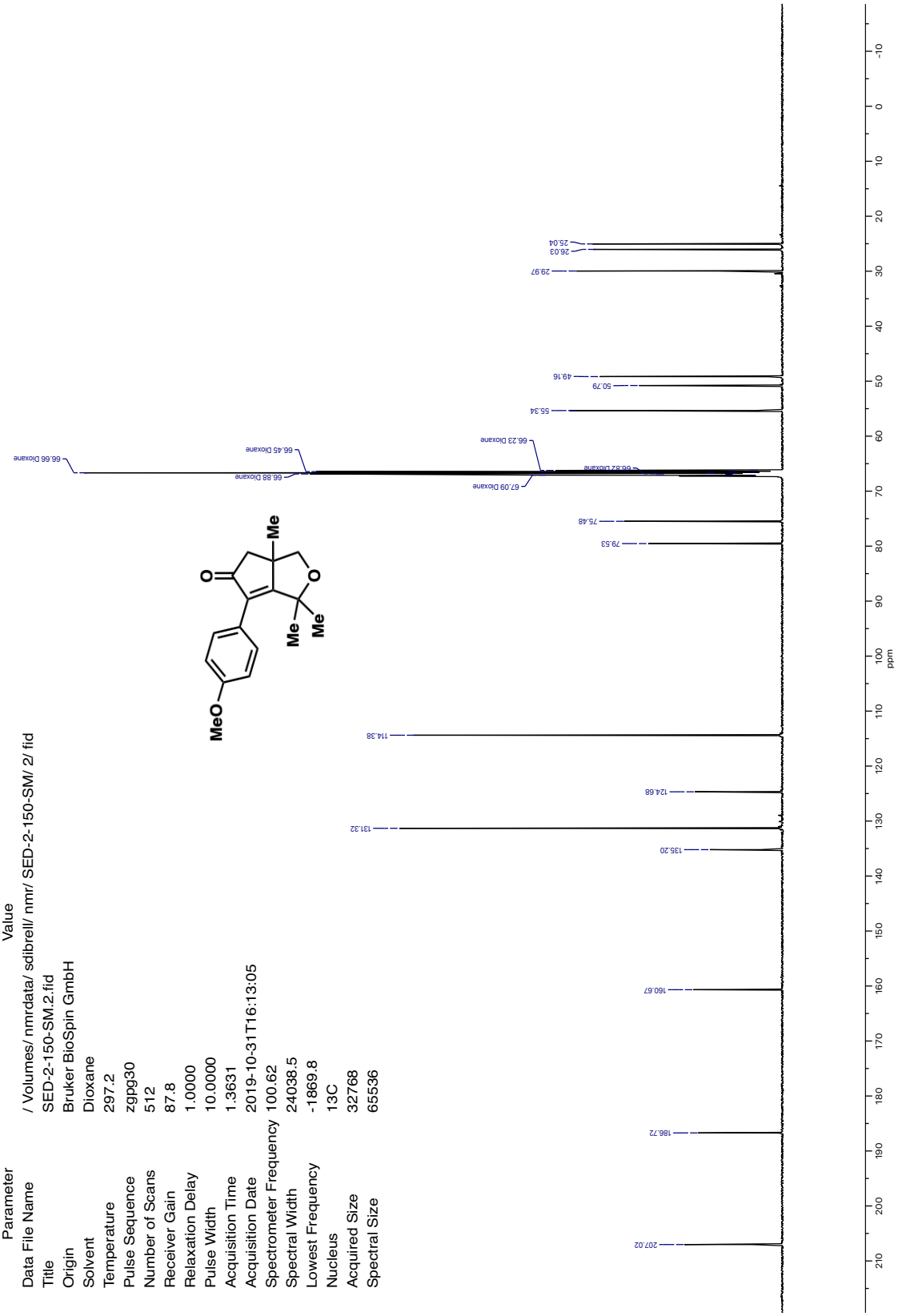

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-116-column2/ 1/ fid
Title SED-2-116-column2.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 8.7000
Acquisition Time 4.0894
Acquisition Date 2019-10-02T22:22:07
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1554.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
Me
Me
Me
Me
Me
O
O
B
O
O
2.06
2.26
1.02
1.03
1.99
3.02
3.19
11.97
3.00
ppm

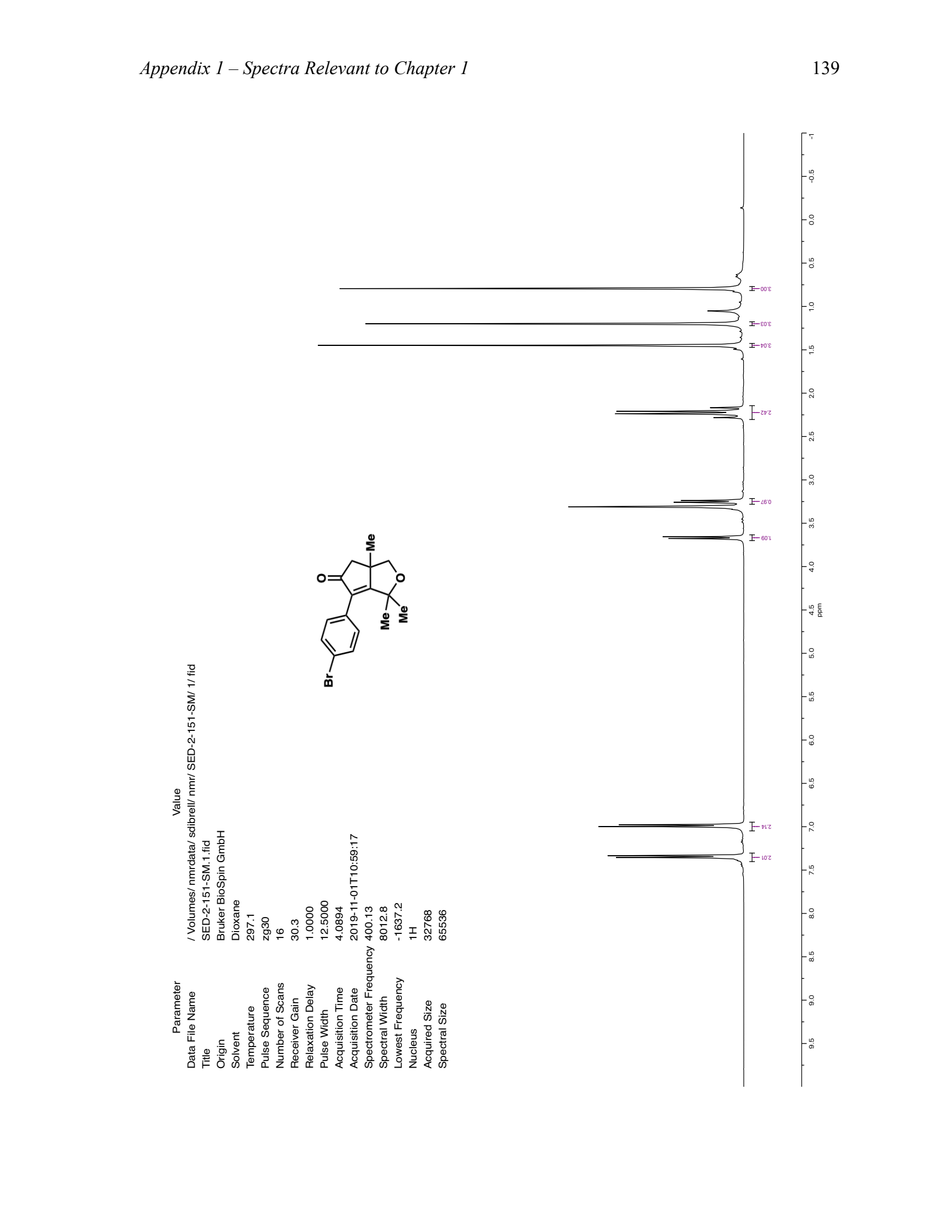

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-131-column3/ 2/ fid
Title
SED-2-131-column3.2.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
zgpg30
Number of Scans
1024
Receiver Gain
55.5
Relaxation Delay
1.0000
Pulse Width
10.0000
Acquisition Time
1.3631
Acquisition Date
2019-10-18T02:47:23
Spectrometer Frequency
100.62
Spectral Width
24038.5
Lowest Frequency
-1958.4
Nucleus
13C
Acquired Size
32768
Spectral Size
65536
Me
Me
Me
Me
Me
Me
O
O
O
O
B
14.22
24.84
24.93
24.99
25.39
29.47
49.09
50.40
75.01
79.26
83.93
128.44
133.86
134.74
135.30
188.65
207.10
ppm

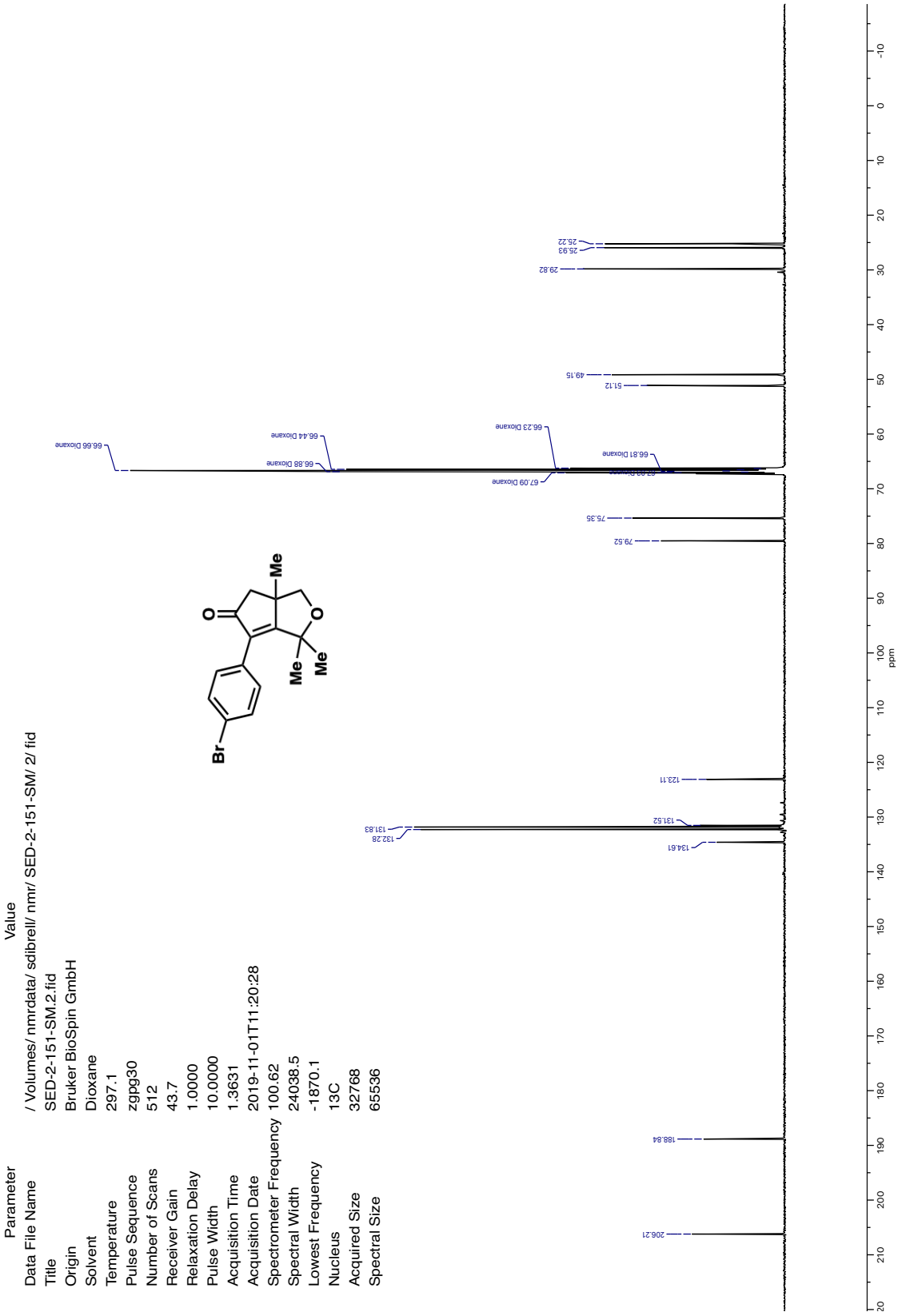

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-152-SM/ 1/ fid
Title SED-2-152-SM.1.fid
Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 30.3
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-01T11:24:53
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1637.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
Me
O
Me
Me
F3C
2.09
2.00
0.98
1.20
2.40
2.89
2.92
3.00
ppm

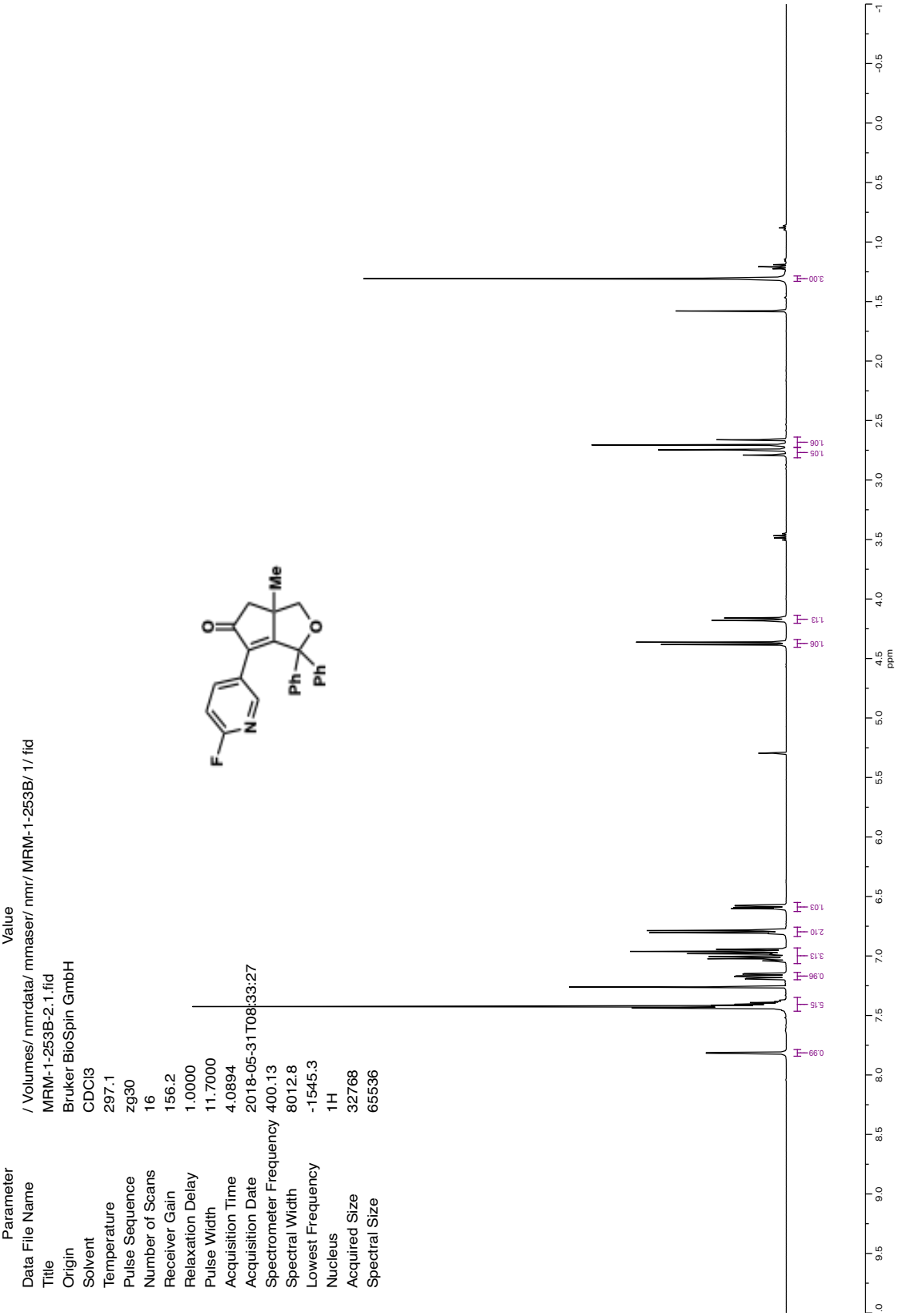

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-152-SM/ 2/ fid
Title SED-2-152-SM.2.fid
Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 43.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-01T11:46:04
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1869.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
O
O
Me
Me
F3C
206.01
190.02
130.58
138.45
134.54
131.72
130.53
130.21
129.29
126.59
126.11
125.99
126.07
126.03
123.89
121.18
79.53
75.32
67.09 Dioxane
66.88 Dioxane
66.45 Dioxane
66.66 Dioxane
66.23 Dioxane
66.82 Dioxane
67.02 Dioxane
49.20
51.30
29.75
25.92
25.36
ppm

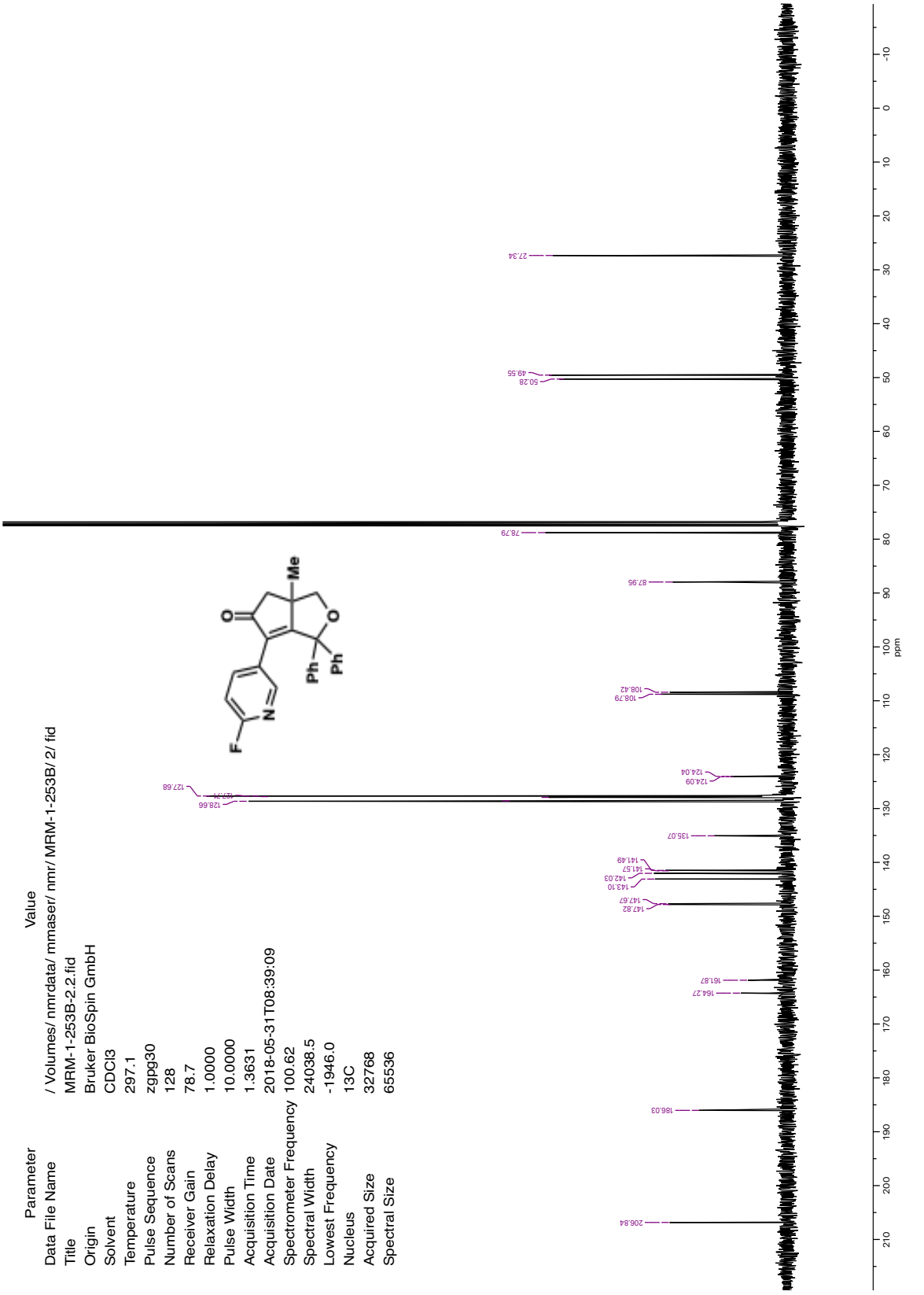

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-150-SM/ 1/ fid
Title SED-2-150-SM.1.fid
Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 30.3
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-10-31T15:51:05
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1637.5
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
MeO
O
O
Me
Me
Me
2.00
2.15
1.14
1.03
2.23
3.41
3.13
3.25
ppm

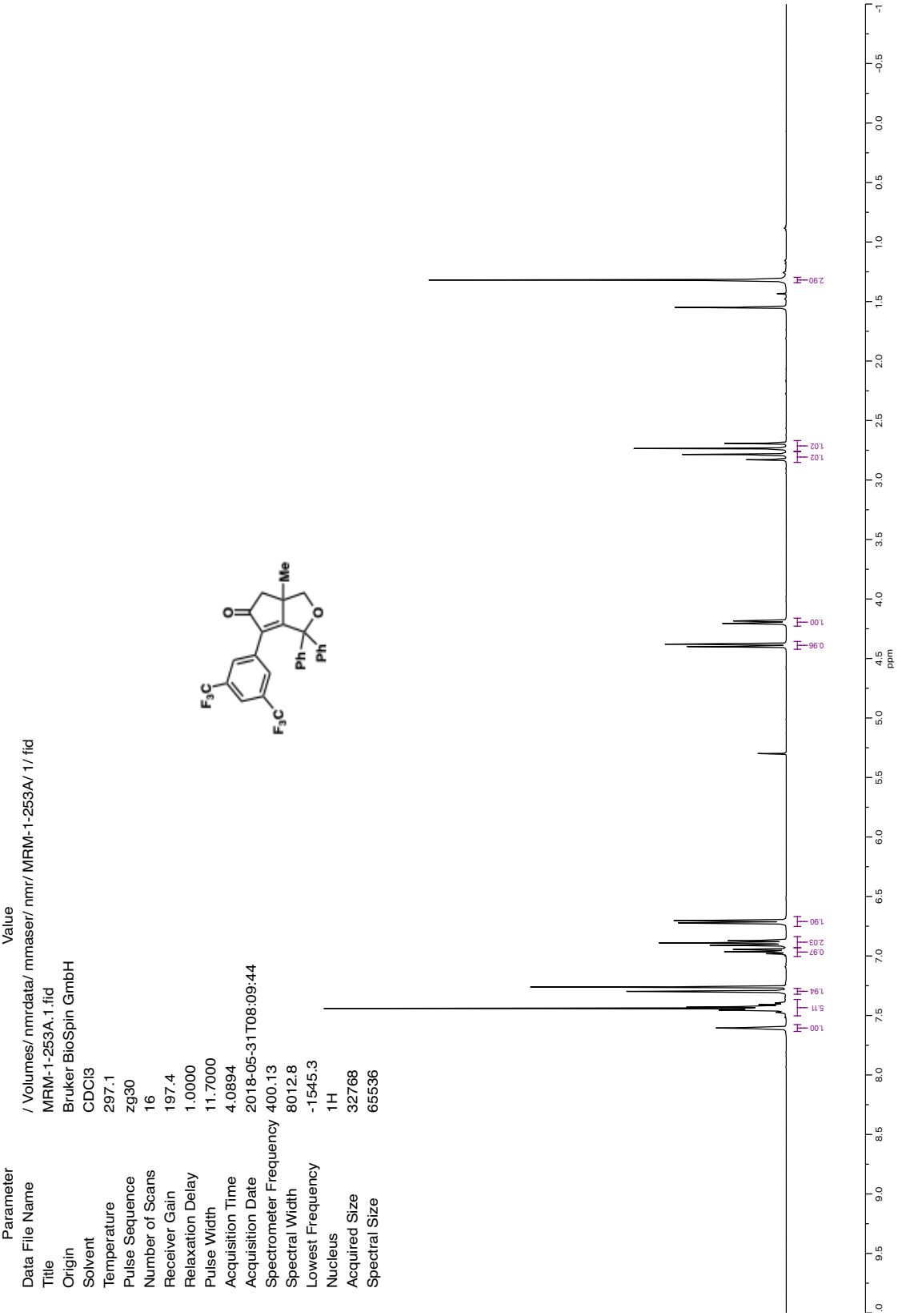

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-150-SM/ 2/ fid
Title SED-2-150-SM.2.fid
Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 87.8
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-10-31T16:13:05
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1869.8
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
MeO
Me
Me
Me
O
O
66.66 Dioxane
66.88 Dioxane
66.45 Dioxane
67.09 Dioxane
66.23 Dioxane
66.32 Dioxane
207.02
186.72
160.67
135.20
131.32
124.68
114.38
79.53
75.48
55.34
50.79
49.16
29.97
26.03
25.04
ppm

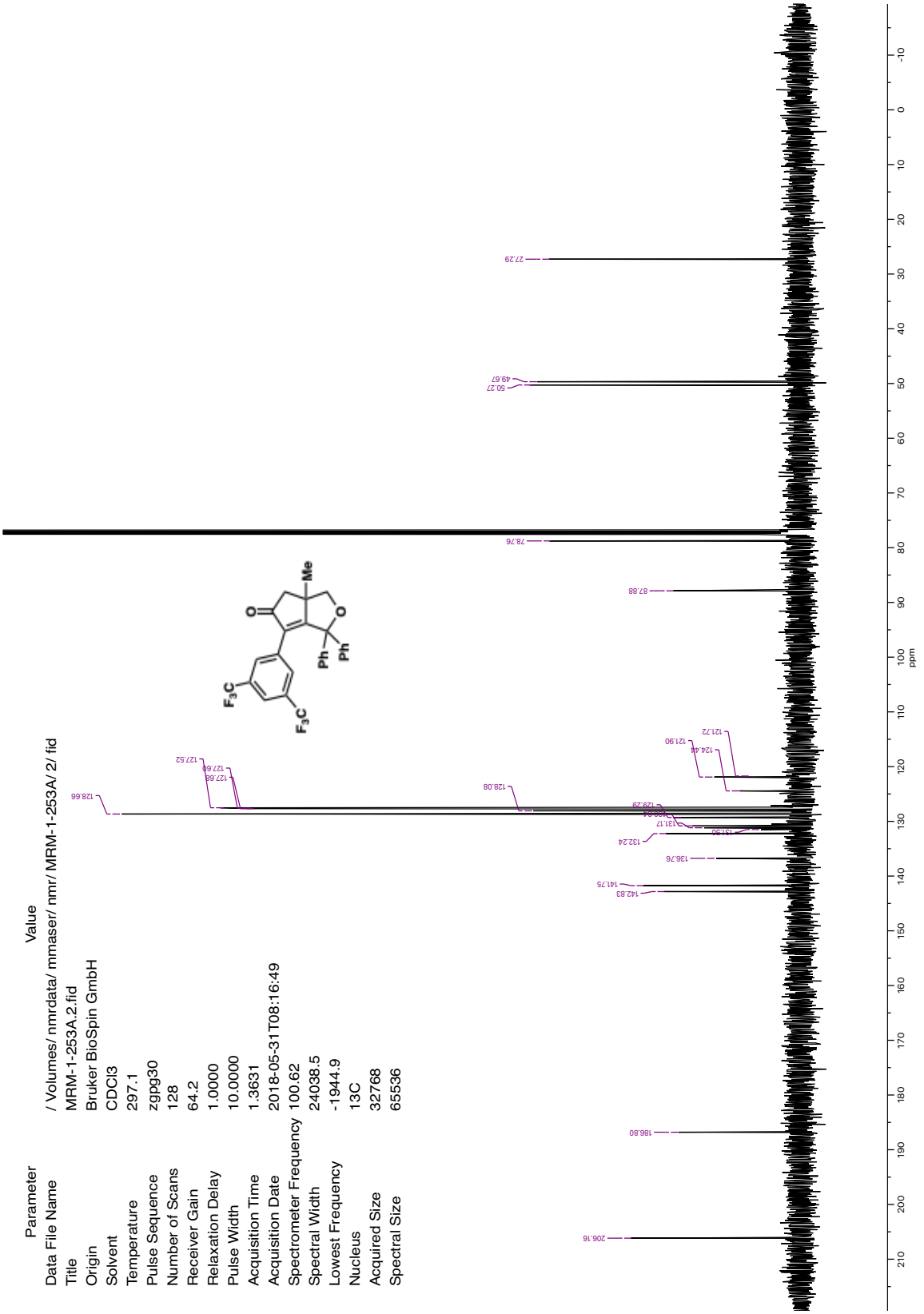

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-151-SM/ 1/ fid
Title SED-2-151-SM.1.fid
Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 30.3
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-01T10:59:17
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1637.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
Me
O
Me
Me
Br
2.01
2.14
1.09
0.97
2.42
3.04
3.03
3.00
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1
ppm

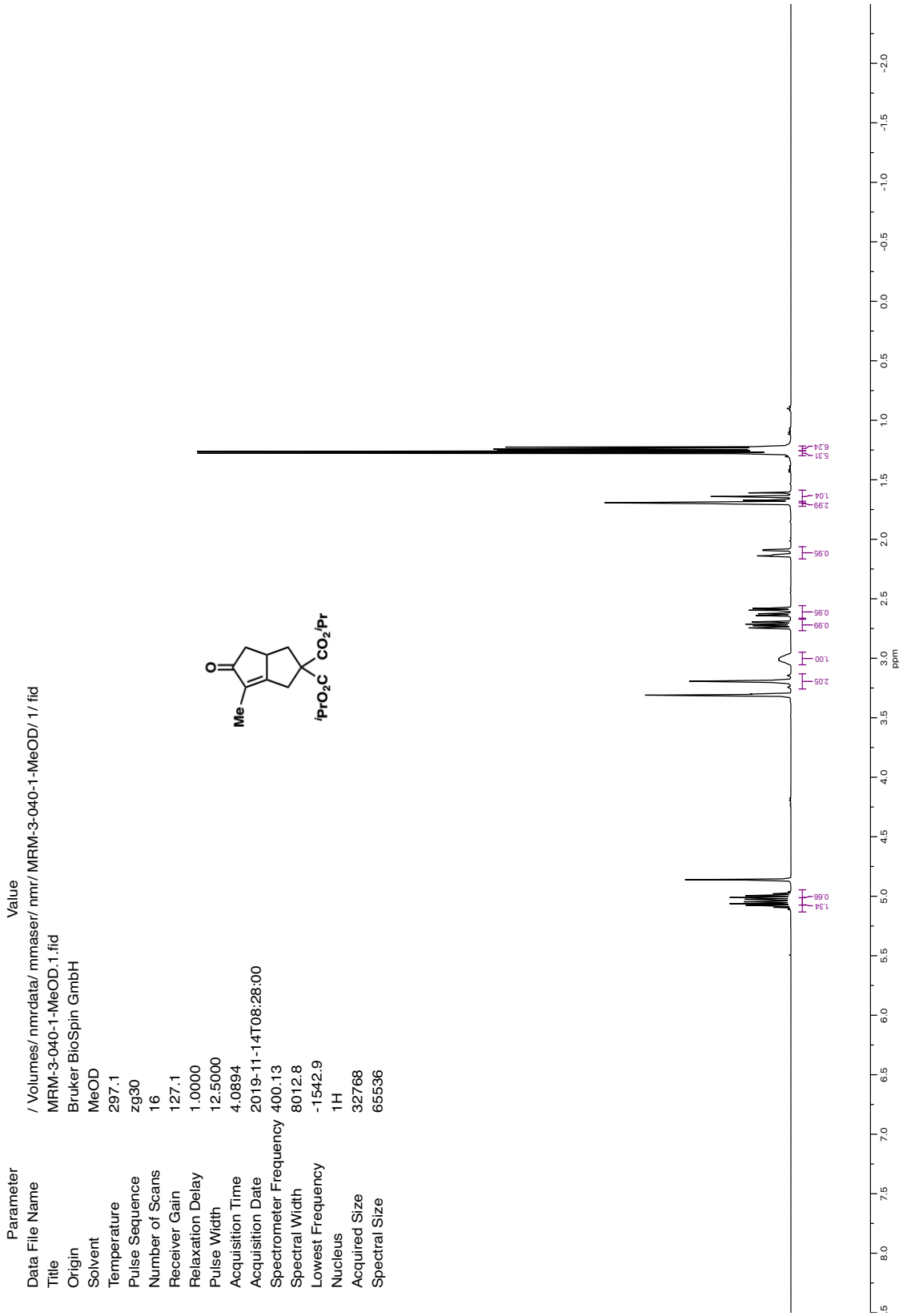

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-151-SM/ 2/ fid
Title SED-2-151-SM.2.fid
Origin Bruker BioSpin GmbH
Solvent Dioxane
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 43.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-01T11:20:28
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1870.1
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
O
Me
O
Me
Me
Br
206.21
188.84
134.61
132.28
131.83
131.52
123.11
79.52
75.35
67.09 Dioxane
66.88 Dioxane
66.44 Dioxane
66.23 Dioxane
66.81 Dioxane
66.66 Dioxane
51.12
49.15
29.82
25.93
25.22
ppm

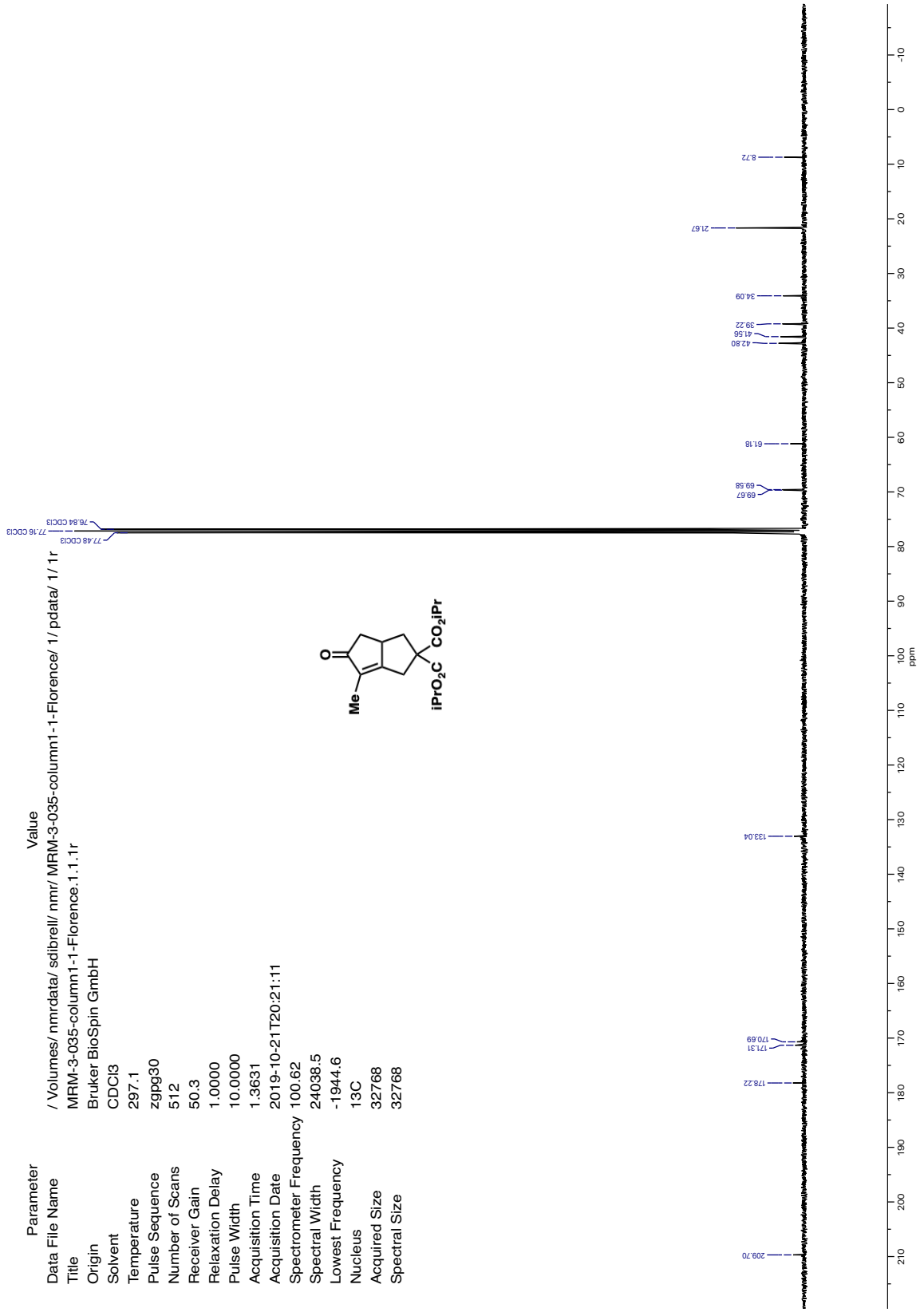

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-253B/ 1/ fid
Title MRM-1-253B-2.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2018-05-31T08:33:27
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
O
O
Ph
Ph
N
F
0.99
5.15
0.96
3.13
2.10
1.03
1.06
1.13
1.06
1.05
3.00
ppm

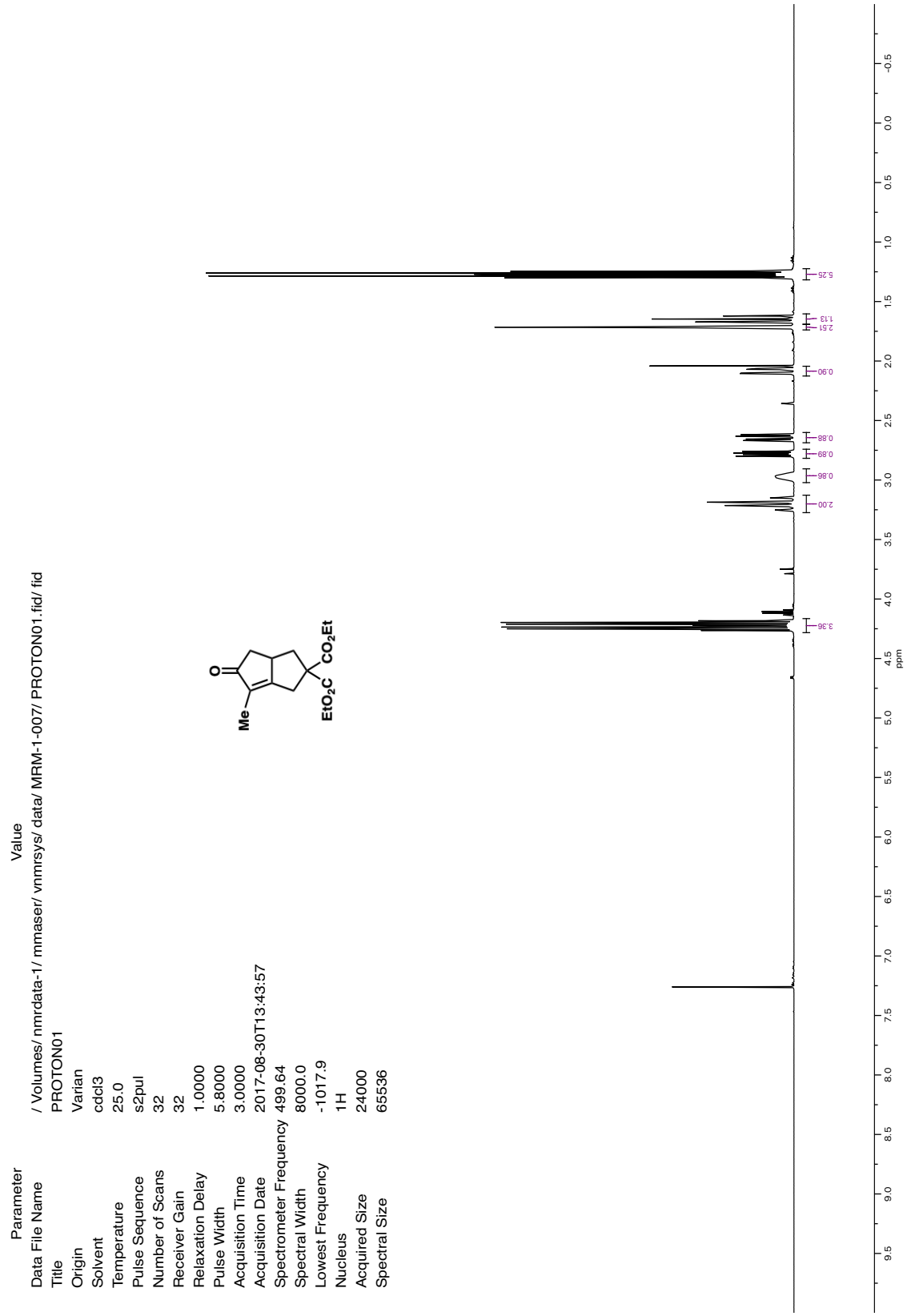

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-253B/ 2/ fid
Title MRM-1-253B-2.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 128
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-05-31T08:39:09
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1946.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
O
O
Ph
Ph
N
F
206.84
186.03
164.27
161.87
147.82
147.67
143.10
142.03
141.57
141.49
135.07
128.66
127.68
124.09
124.04
108.79
108.42
87.95
78.79
50.28
49.55
27.34
ppm

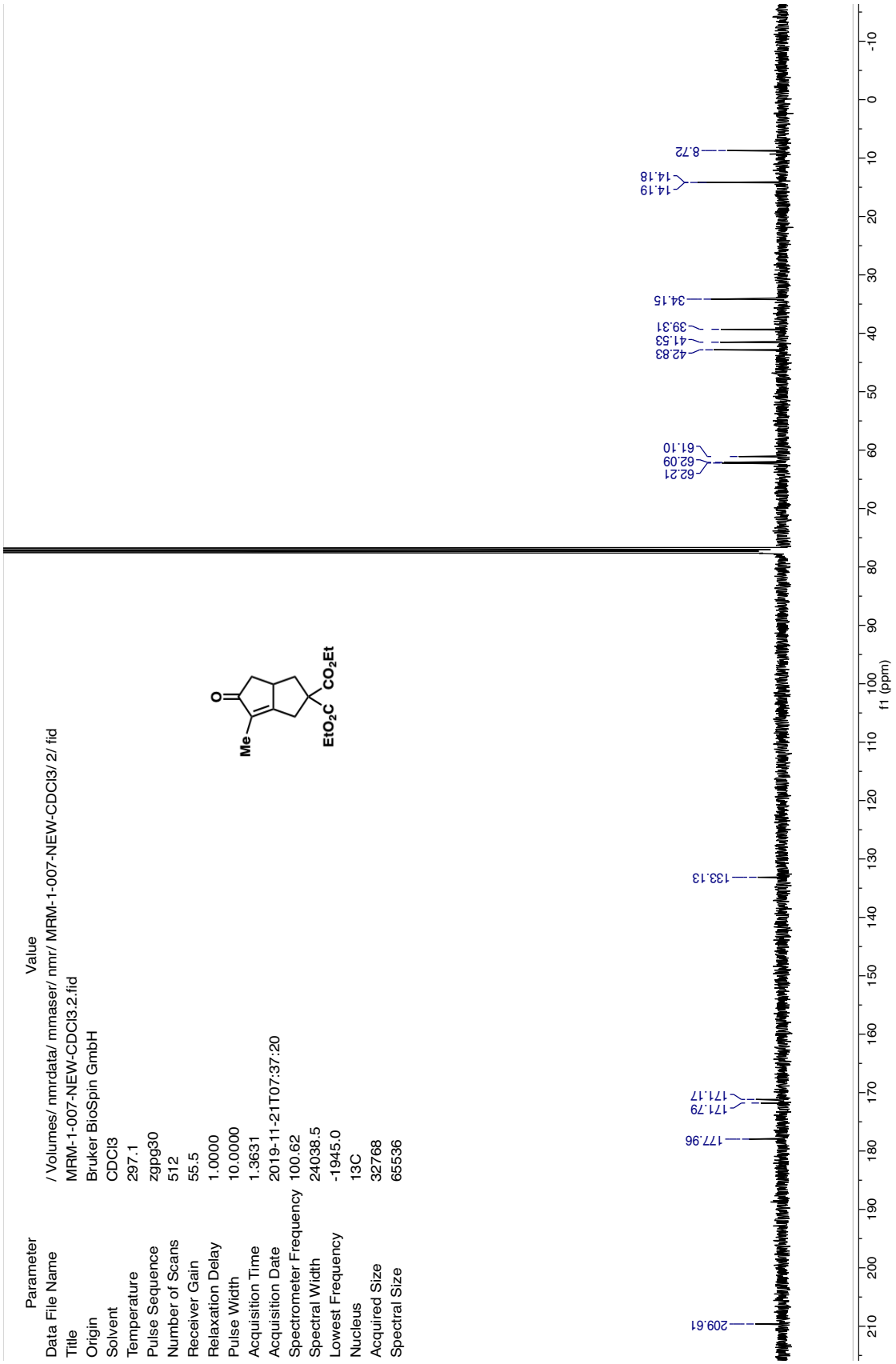

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-253A/ 1/ fid
Title
MRM-1-253A.1.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
zg30
Number of Scans
16
Receiver Gain
197.4
Relaxation Delay
1.0000
Pulse Width
11.7000
Acquisition Time
4.0894
Acquisition Date
2018-05-31T08:09:44
Spectrometer Frequency
400.13
Spectral Width
8012.8
Lowest Frequency
-1545.3
Nucleus
1H
Acquired Size
32768
Spectral Size
65536
Me
O
O
Ph
Ph
F3C
F3C
2.90
1.02
1.02
1.00
0.96
1.90
0.97
2.03
1.94
5.11
1.00
ppm

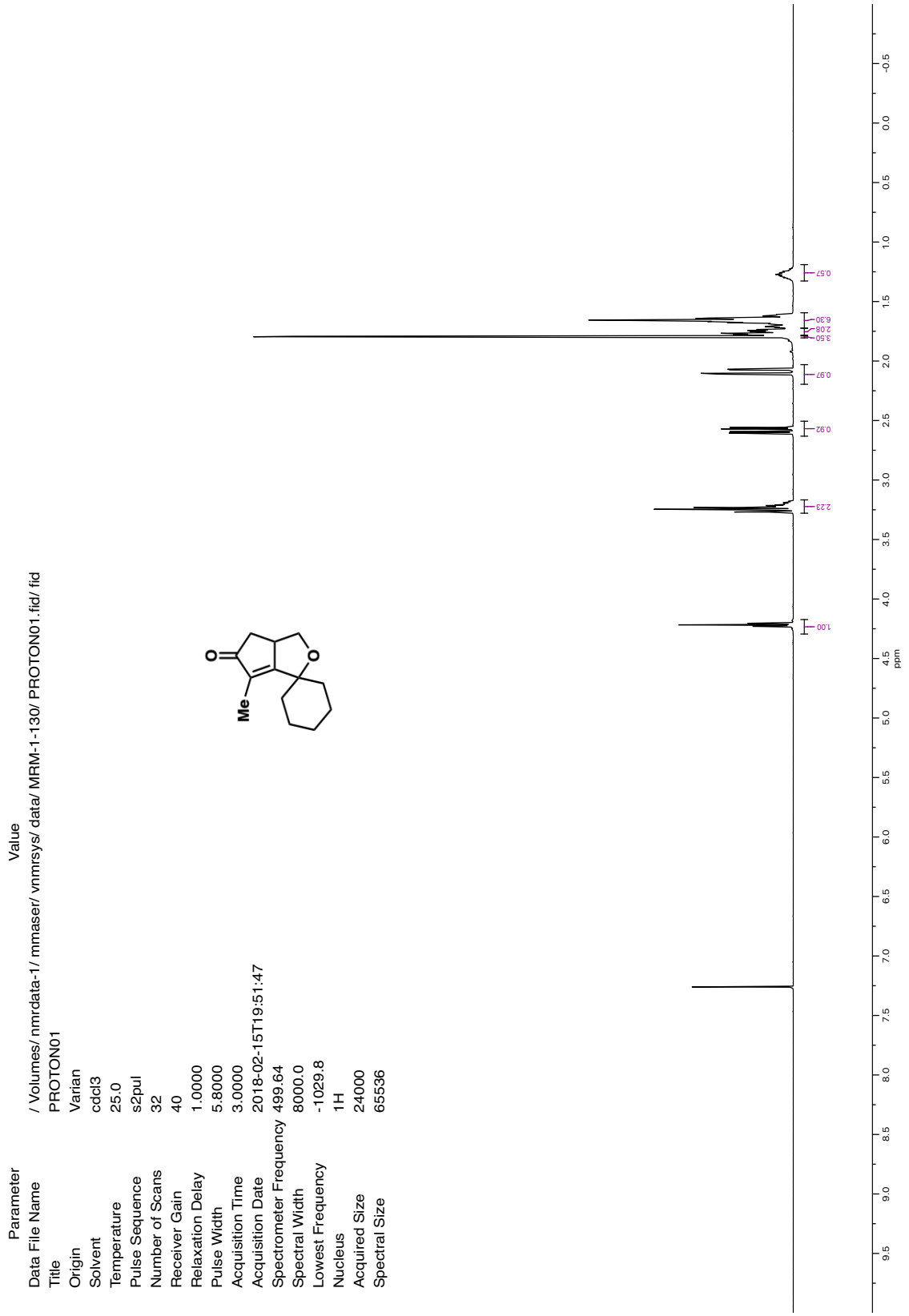

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-253A/ 2/ fid
Title MRM-1-253A.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 128
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-05-31T08:16:49
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
O
O
Ph
Ph
F3C
F3C
206.16
186.80
142.83
141.75
136.76
132.24
131.17
128.66
128.08
127.68
127.52
124.44
121.72
121.90
87.88
78.76
50.27
49.67
27.29
ppm

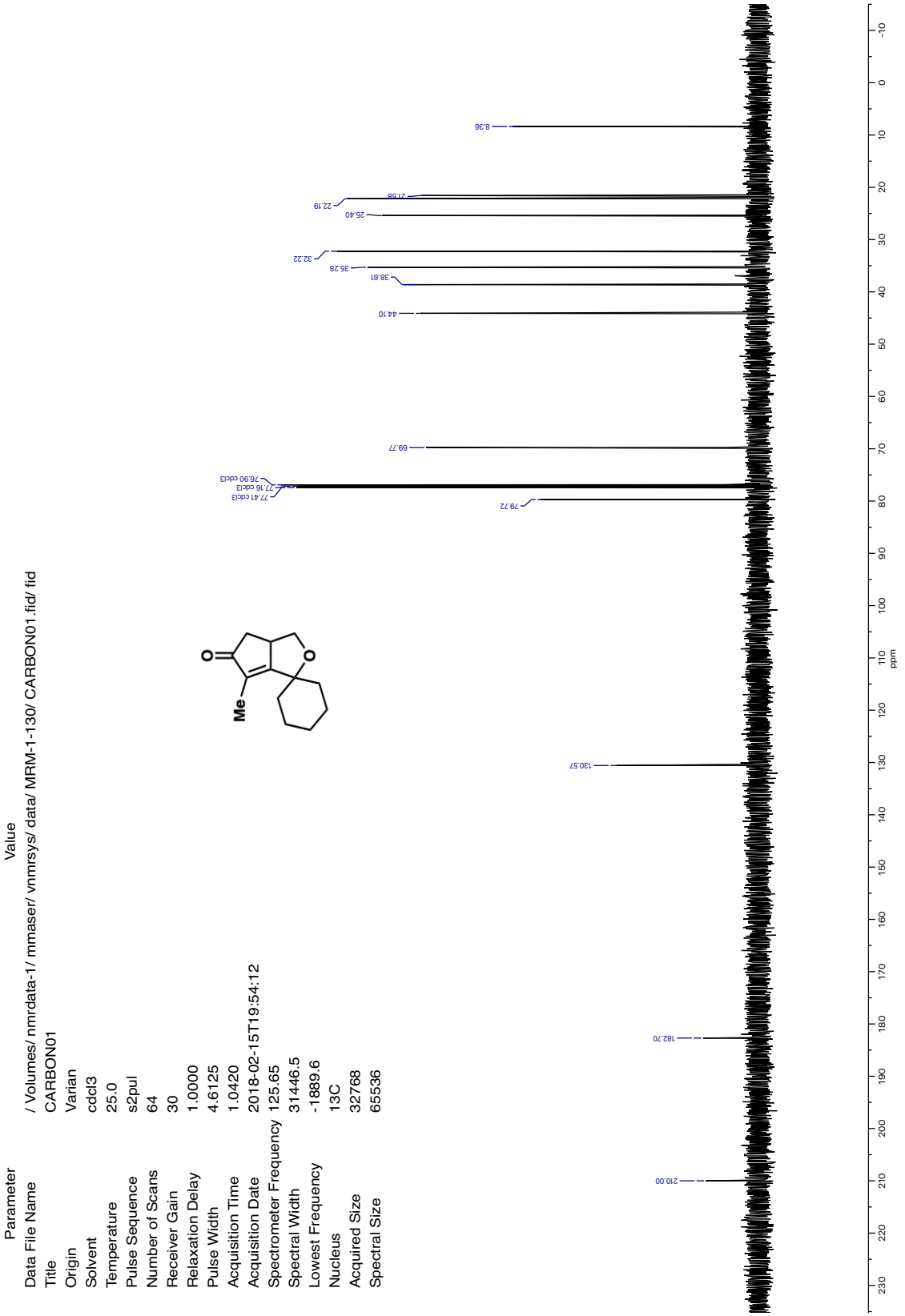

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-040-1-MeOD/ 1/ fid
Title MRM-3-040-1-MeOD.1.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 127.1
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-14T08:28:00
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1542.9
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
Me
iPrO2C
CO2iPr
1.34
0.66
2.05
1.00
0.99
0.95
0.95
2.99
1.04
6.24
5.31
ppm

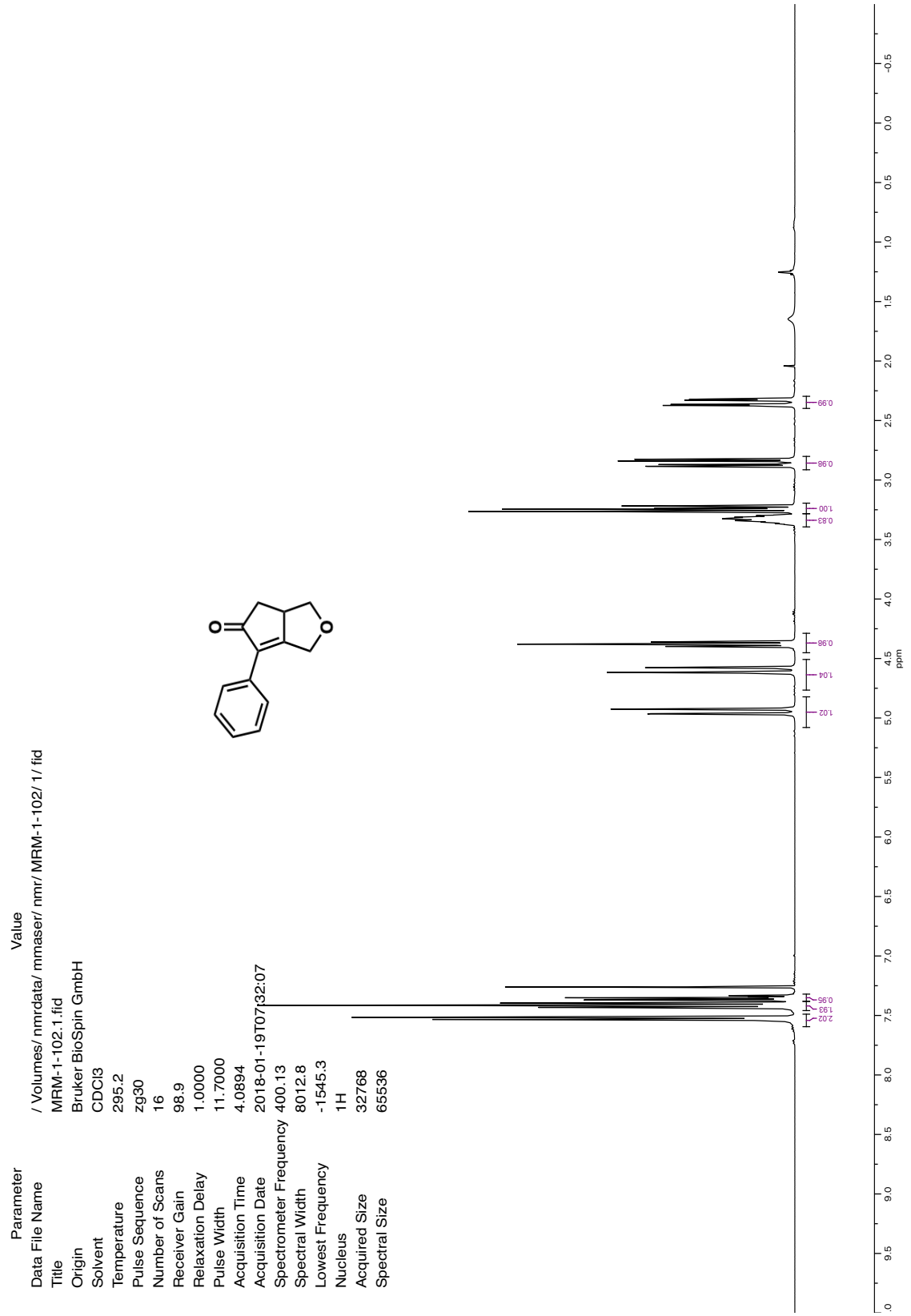

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ MRM-3-035-column1-1-Florence/ 1/ pdata/ 1/ 1r
Title MRM-3-035-column1-1-Florence.1.1.1r
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-10-21T20:21:11
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.6
Nucleus 13C
Acquired Size 32768
Spectral Size 32768
Me
iPrO2C
CO2iPr
O
77.16 CDCl3
76.84 CDCl3
77.48 CDCl3
209.70
178.22
171.31
170.69
133.04
69.67
69.58
61.18
42.80
41.56
39.22
34.09
21.67
8.72
ppm

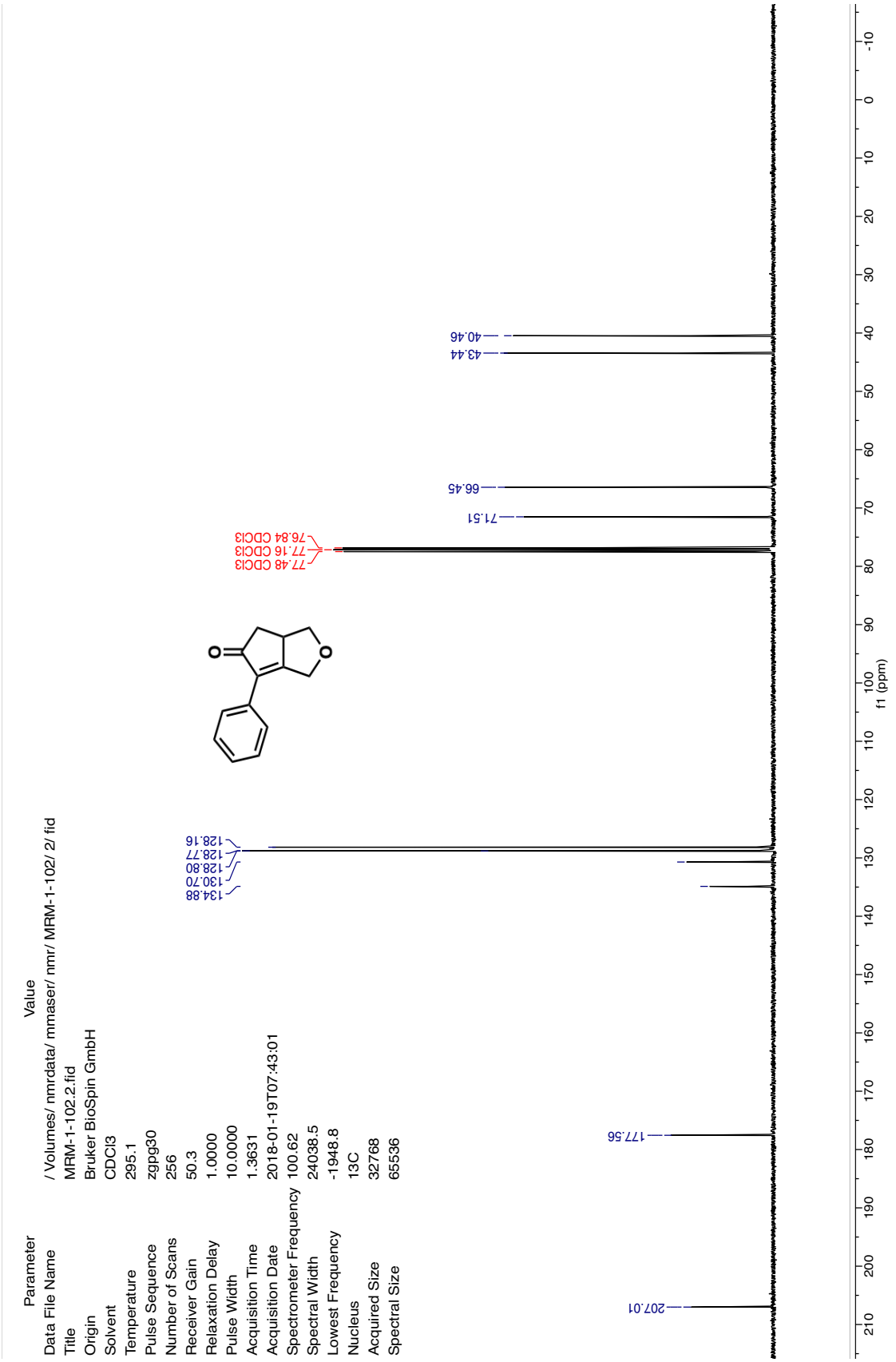

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-007/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 32
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2017-08-30T13:43:57
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1017.9
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
O
Me
EtO2C CO2Et
5.25
1.13
2.51
0.90
0.88
0.89
0.86
2.00
3.36
ppm

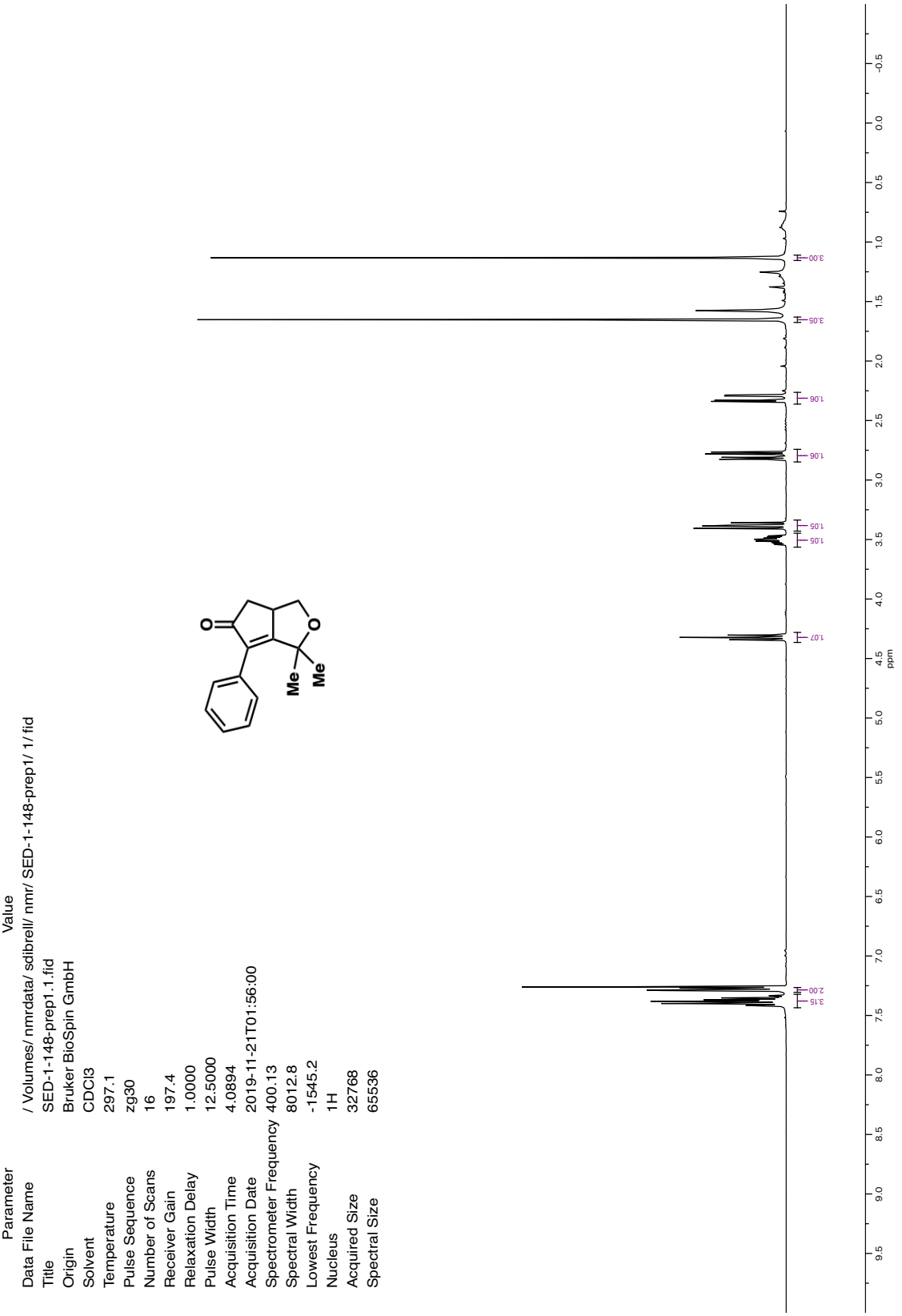

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-007-NEW-CDCl3/ 2/ fid
Title MRM-1-007-NEW-CDCl3.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-21T07:37:20
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
O
Me
EtO2C
CO2Et
209.61
177.96
171.79
171.17
133.13
62.21
62.09
61.10
42.83
41.53
39.31
34.15
14.19
14.18
8.72
f1 (ppm)

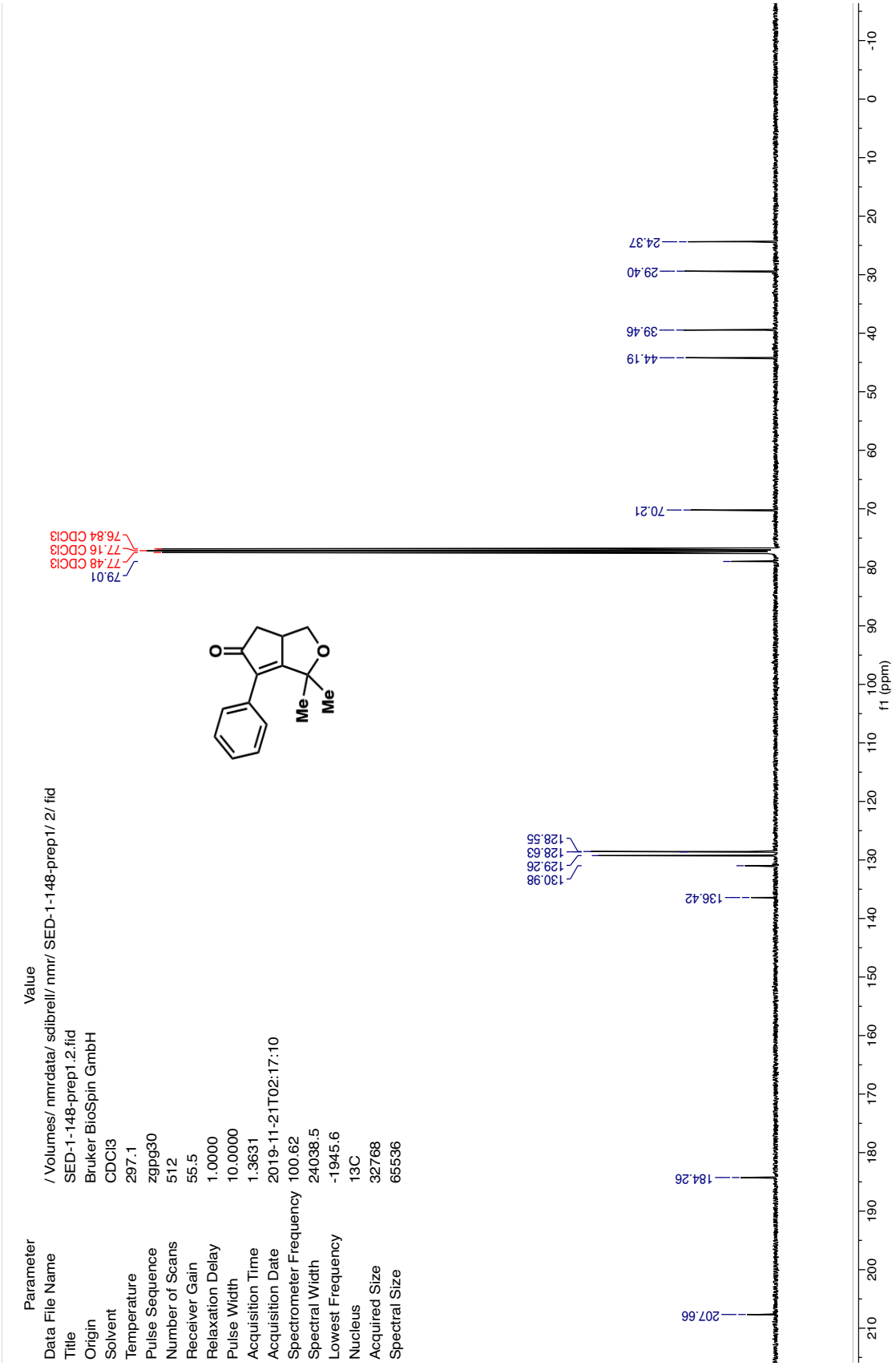

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-130/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 40
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2018-02-15T19:51:47
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1029.8
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
Me
O
O
1.00
2.23
0.92
0.97
3.50
2.08
6.30
0.57
ppm

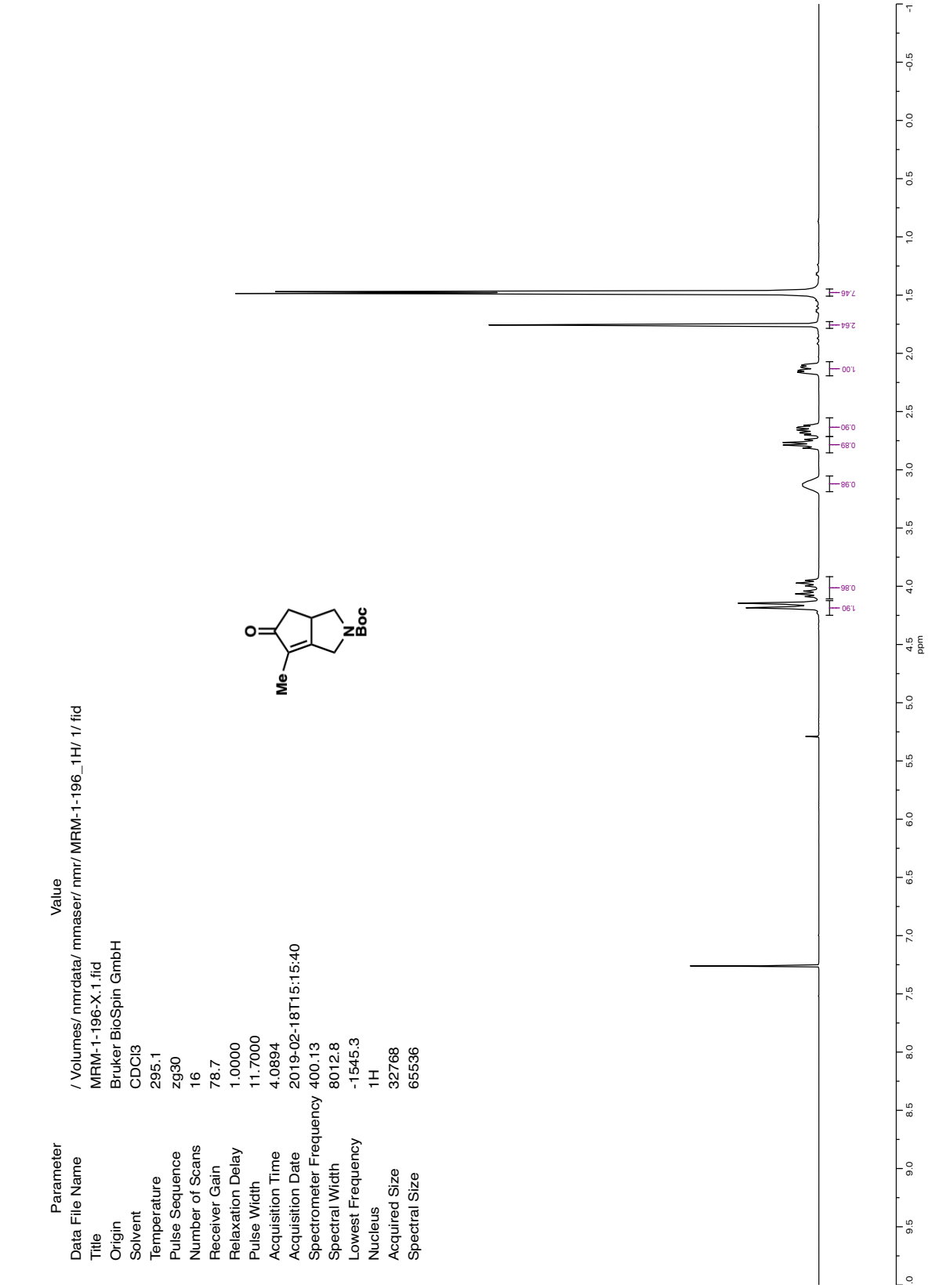

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-130/ CARBON01.fid/ fid
Title CARBON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 64
Receiver Gain 30
Relaxation Delay 1.0000
Pulse Width 4.6125
Acquisition Time 1.0420
Acquisition Date 2018-02-15T19:54:12
Spectrometer Frequency 125.65
Spectral Width 31446.5
Lowest Frequency -1889.6
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
210.00
182.70
130.57
79.72
77.41 cdcl3
77.16 cdcl3
76.90 cdcl3
69.77
44.10
38.61
35.28
32.22
25.40
22.19
21.58
8.36
ppm

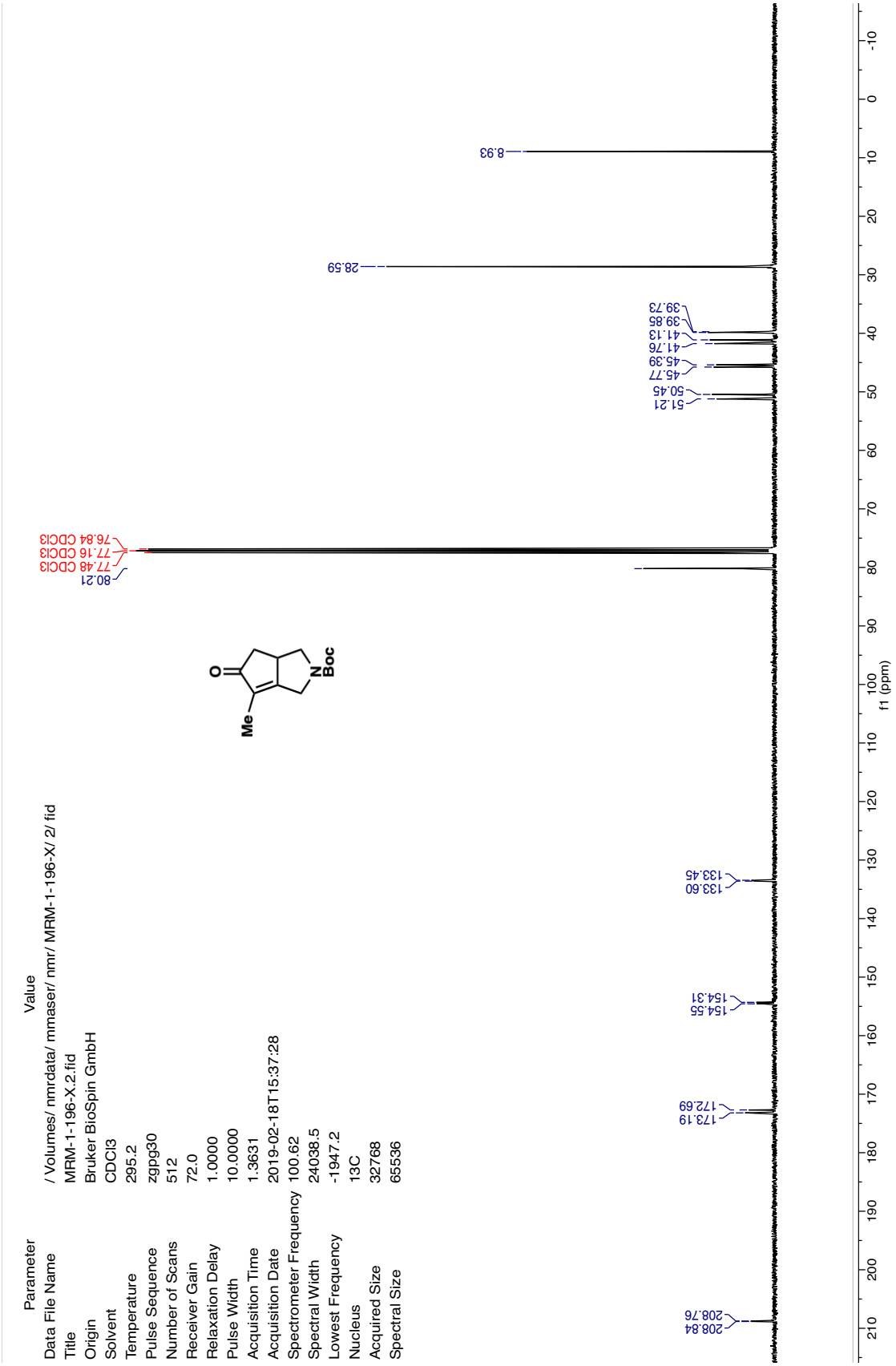

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-102/ 1/ fid
Title MRM-1-102.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 98.9
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2018-01-19T07:32:07
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
0.99
0.98
1.00
0.83
0.98
1.04
1.02
0.96
1.93
2.02
ppm

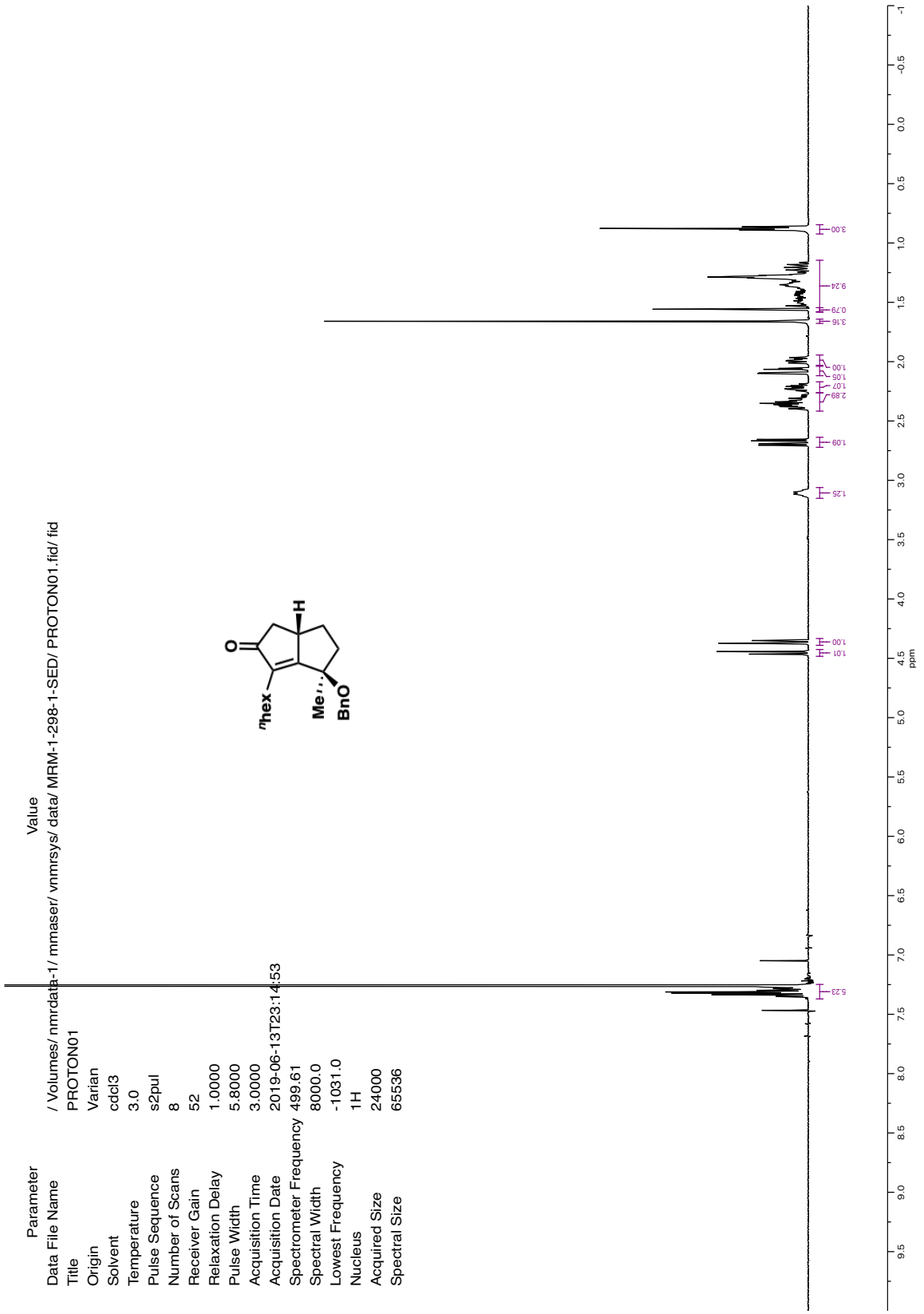

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-102/ 2/ fid
Title MRM-1-102.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.1
Pulse Sequence zgpg30
Number of Scans 256
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-01-19T07:43:01
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1948.8
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
207.01
177.56
134.88
130.70
128.80
128.77
128.16
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
71.51
66.45
43.44
40.46
f1 (ppm)

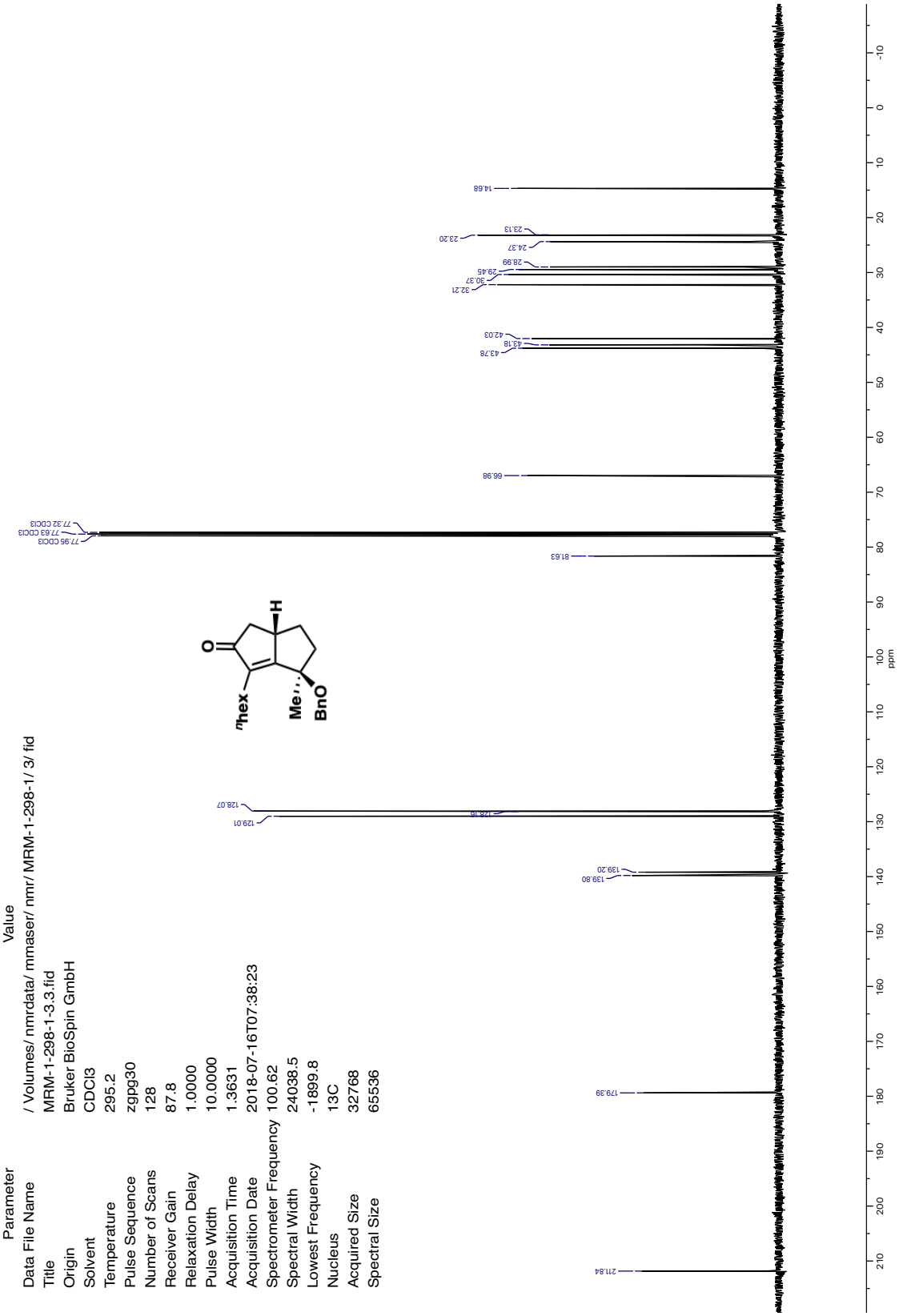

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-148-prep1/ 1/ fid
Title SED-1-148-prep1.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-21T01:56:00
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
O
Me
Me
3.00
3.05
1.06
1.06
1.05
1.05
1.07
2.00
3.15
ppm

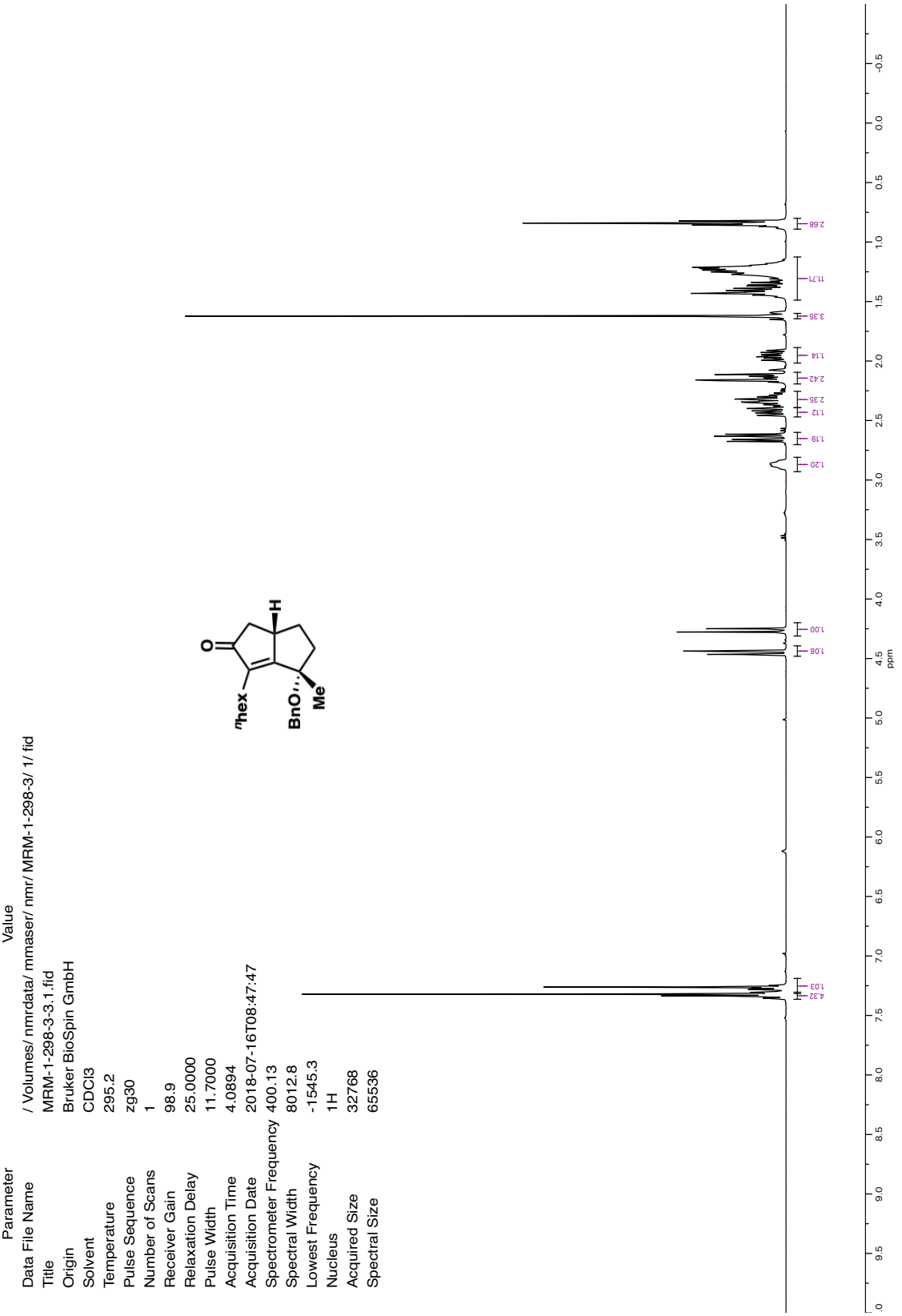

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-148-prep1/ 2/ fid
Title
SED-1-148-prep1.2.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
zgpg30
Number of Scans
512
Receiver Gain
55.5
Relaxation Delay
1.0000
Pulse Width
10.0000
Acquisition Time
1.3631
Acquisition Date
2019-11-21T02:17:10
Spectrometer Frequency
100.62
Spectral Width
24038.5
Lowest Frequency
-1945.6
Nucleus
13C
Acquired Size
32768
Spectral Size
65536
O
O
Me
Me
79.01
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
70.21
44.19
39.46
29.40
24.37
130.98
129.26
128.63
128.55
136.42
184.26
207.66
f1 (ppm)

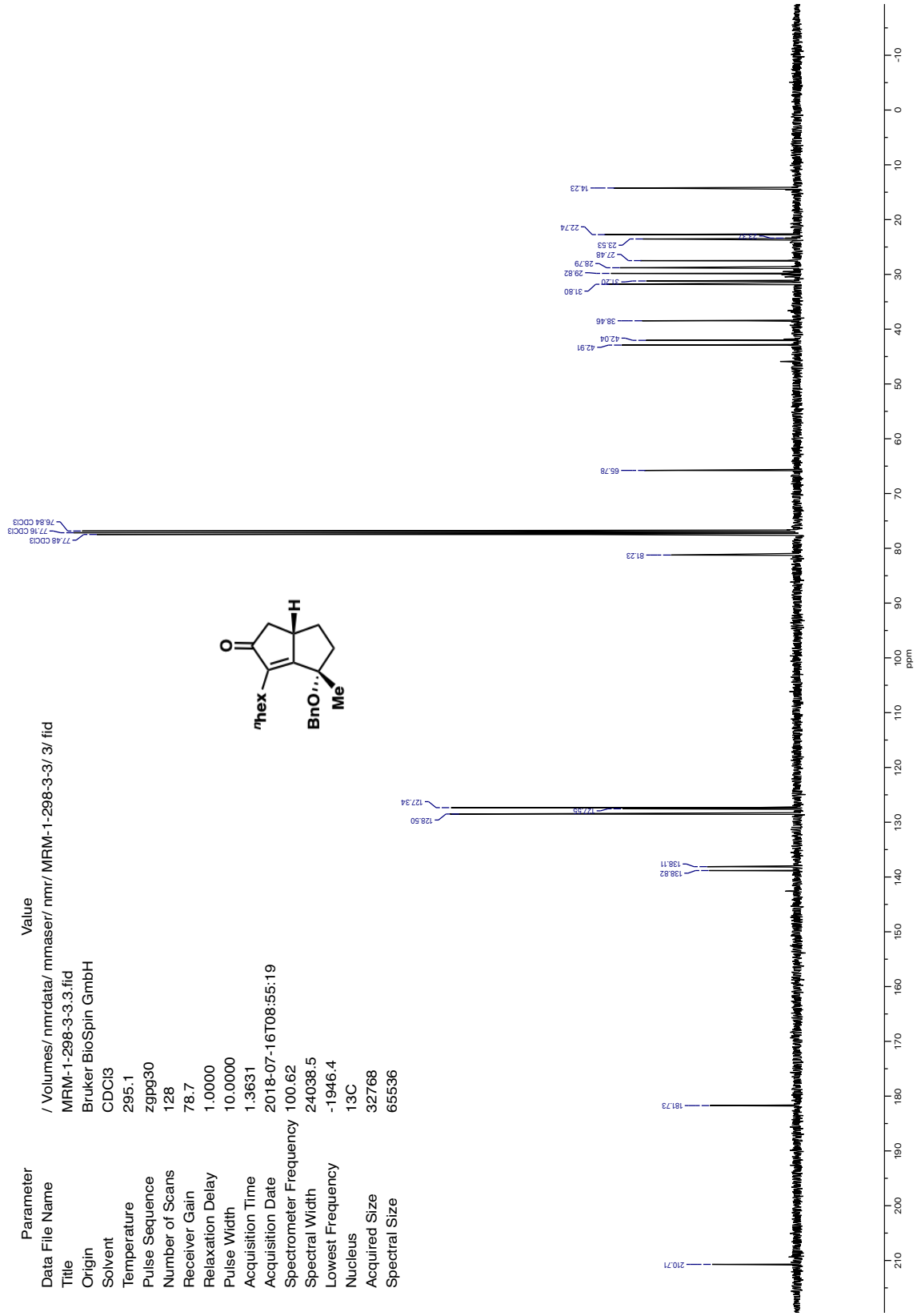

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-196_1H/ 1/ fid
Title MRM-1-196-X.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-02-18T15:15:40
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
N
Boc
Me
7.46
2.64
1.00
0.90
0.89
0.98
0.86
1.90
ppm

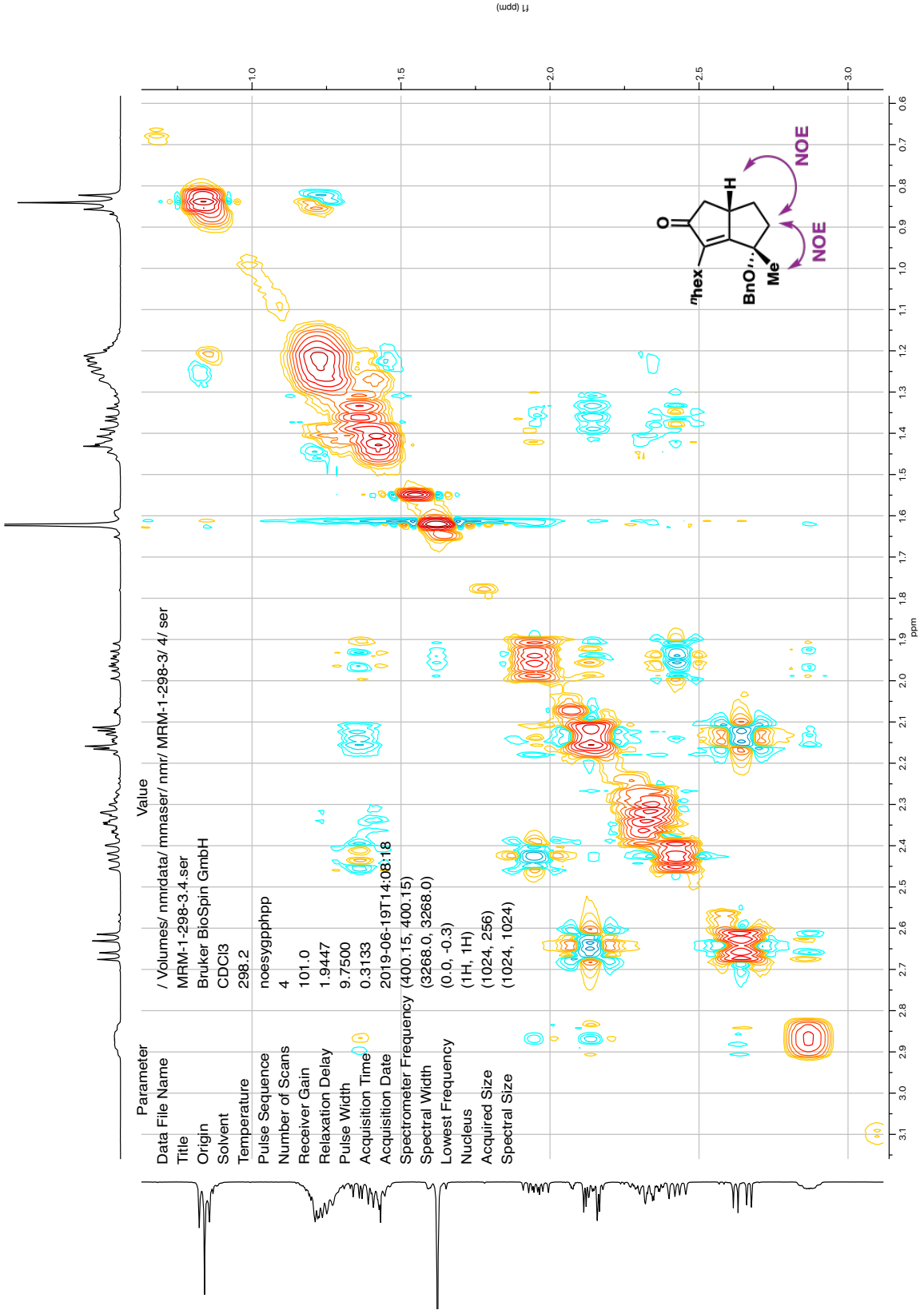

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-196-X/ 2/ fid
Title
MRM-1-196-X.2.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
295.2
Pulse Sequence
zgpg30
Number of Scans
512
Receiver Gain
72.0
Relaxation Delay
1.0000
Pulse Width
10.0000
Acquisition Time
1.3631
Acquisition Date
2019-02-18T15:37:28
Spectrometer Frequency
100.62
Spectral Width
24038.5
Lowest Frequency
-1947.2
Nucleus
13C
Acquired Size
32768
Spectral Size
65536
Me
Boc
N
O
80.21
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
51.21
50.45
45.77
45.39
41.76
41.13
39.85
39.73
28.59
8.93
208.84
208.76
173.19
172.69
154.55
154.31
133.60
133.45
f1 (ppm)

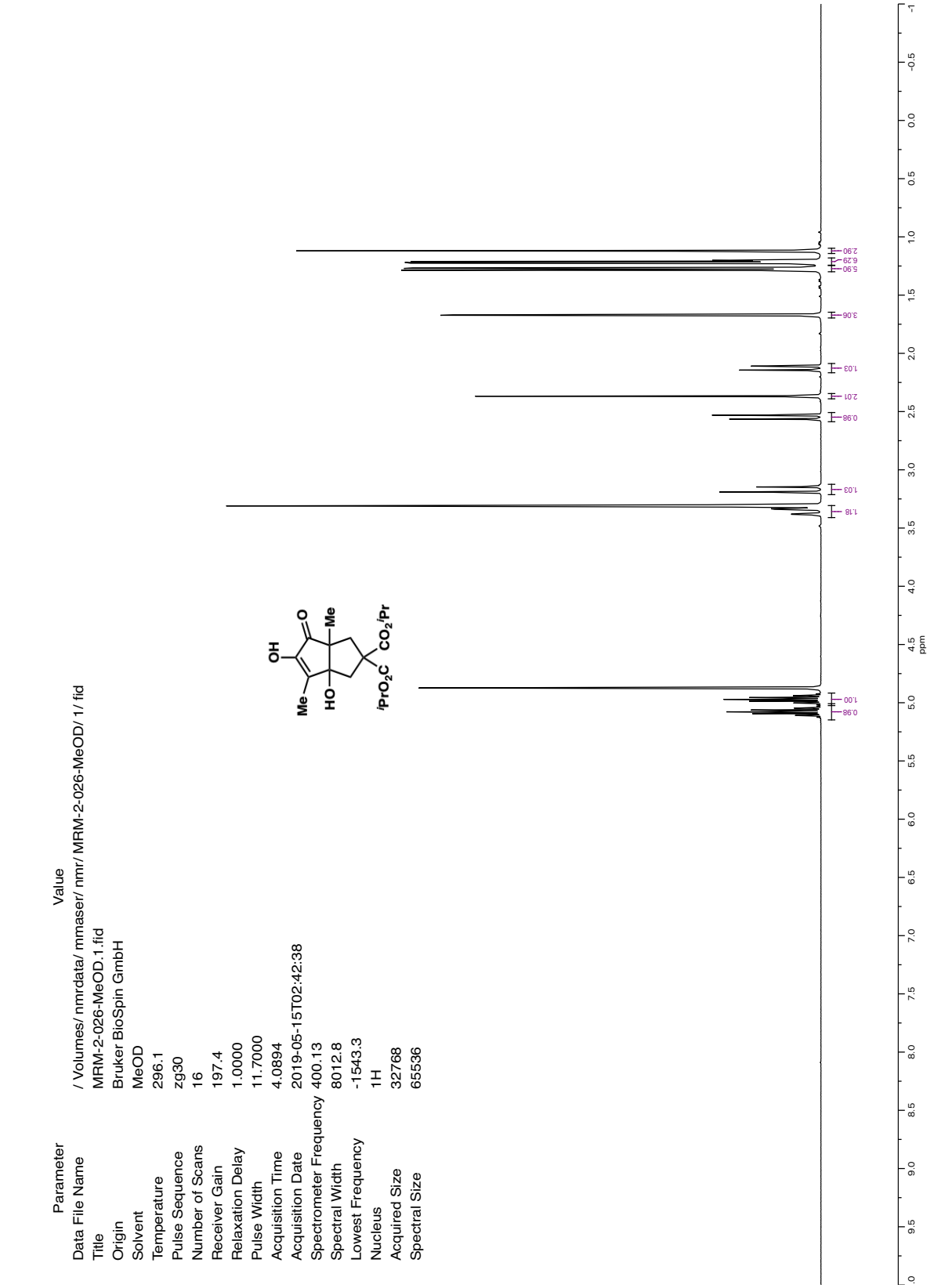

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-298-1-SED/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 3.0
Pulse Sequence s2pul
Number of Scans 8
Receiver Gain 52
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2019-06-13T23:14:53
Spectrometer Frequency 499.61
Spectral Width 8000.0
Lowest Frequency -1031.0
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
O
H
nhex
Me
BnO
5.23
1.00
1.01
1.25
1.09
2.89
1.07
1.05
1.00
3.16
0.79
9.24
3.00
ppm

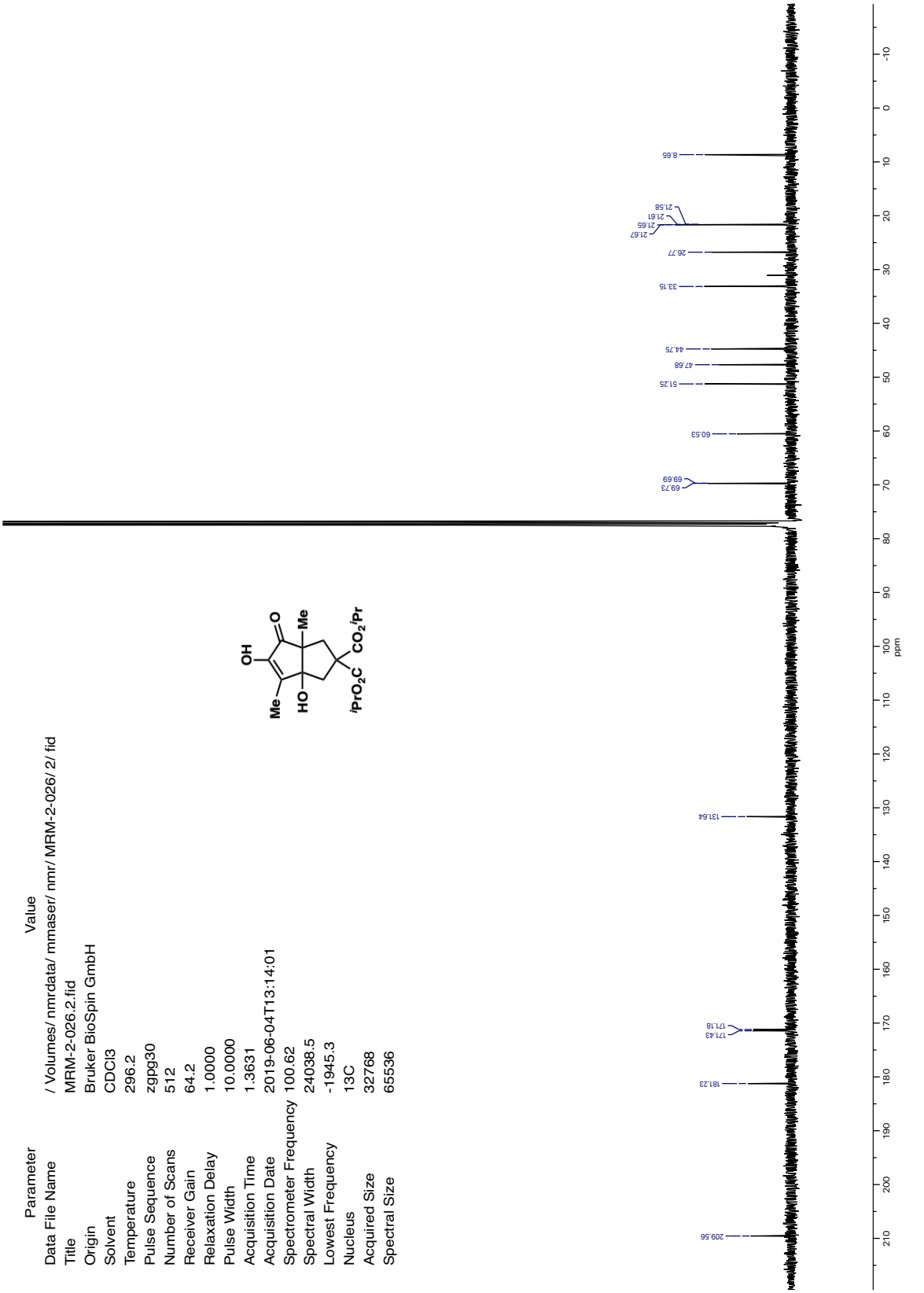

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-298-1/ 3/ fid
Title MRM-1-298-1-3.3.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.2
Pulse Sequence zgpg30
Number of Scans 128
Receiver Gain 87.8
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-07-16T07:38:23
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1899.8
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
211.84
179.39
139.80
139.20
129.01
128.07
77.32 CDCl3
77.63 CDCl3
77.96 CDCl3
81.63
66.98
43.78
43.18
42.03
32.21
30.37
29.45
28.99
24.37
23.20
23.13
14.68
O
H
nhex
Me
BnO
ppm

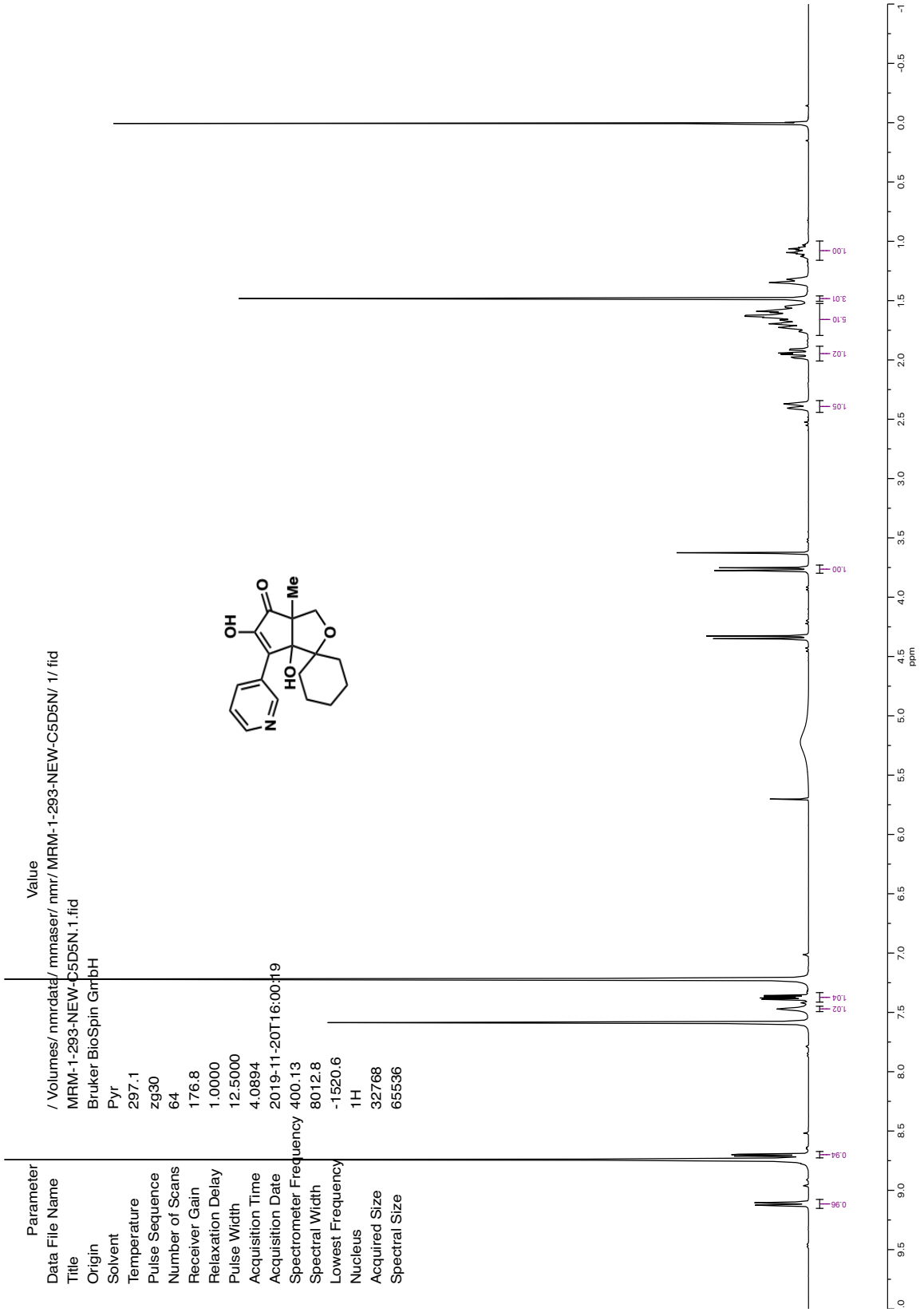

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-298-3/ 1/ fid
Title MRM-1-298-3-3.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.2
Pulse Sequence zg30
Number of Scans 1
Receiver Gain 98.9
Relaxation Delay 25.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2018-07-16T08:47:47
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
H
nhex
BnO
Me
4.32
1.03
1.08
1.00
1.20
1.19
1.12
2.35
2.42
1.14
3.35
11.71
2.68
ppm

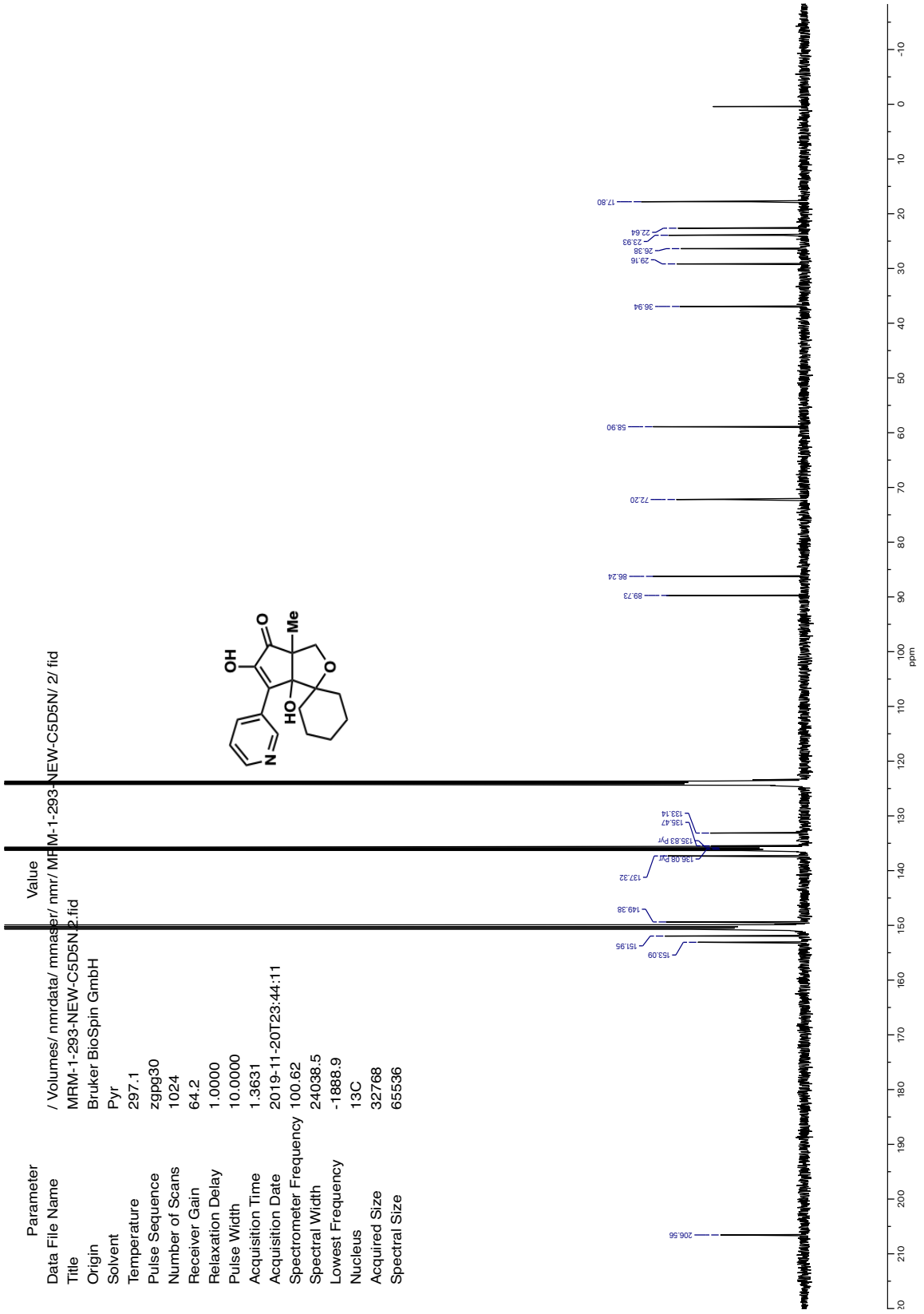

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-298-3-3/ 3/ fid
Title MRM-1-298-3-3.3.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.1
Pulse Sequence zgpg30
Number of Scans 128
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-07-16T08:55:19
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1946.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
O
H
nhex
BnO
Me
14.23
22.74
23.53
27.48
28.79
29.82
31.20
31.80
38.46
42.04
42.91
65.78
81.23
76.84 CDCl3
77.16 CDCl3
77.48 CDCl3
127.34
128.50
138.11
138.82
181.73
210.71
ppm

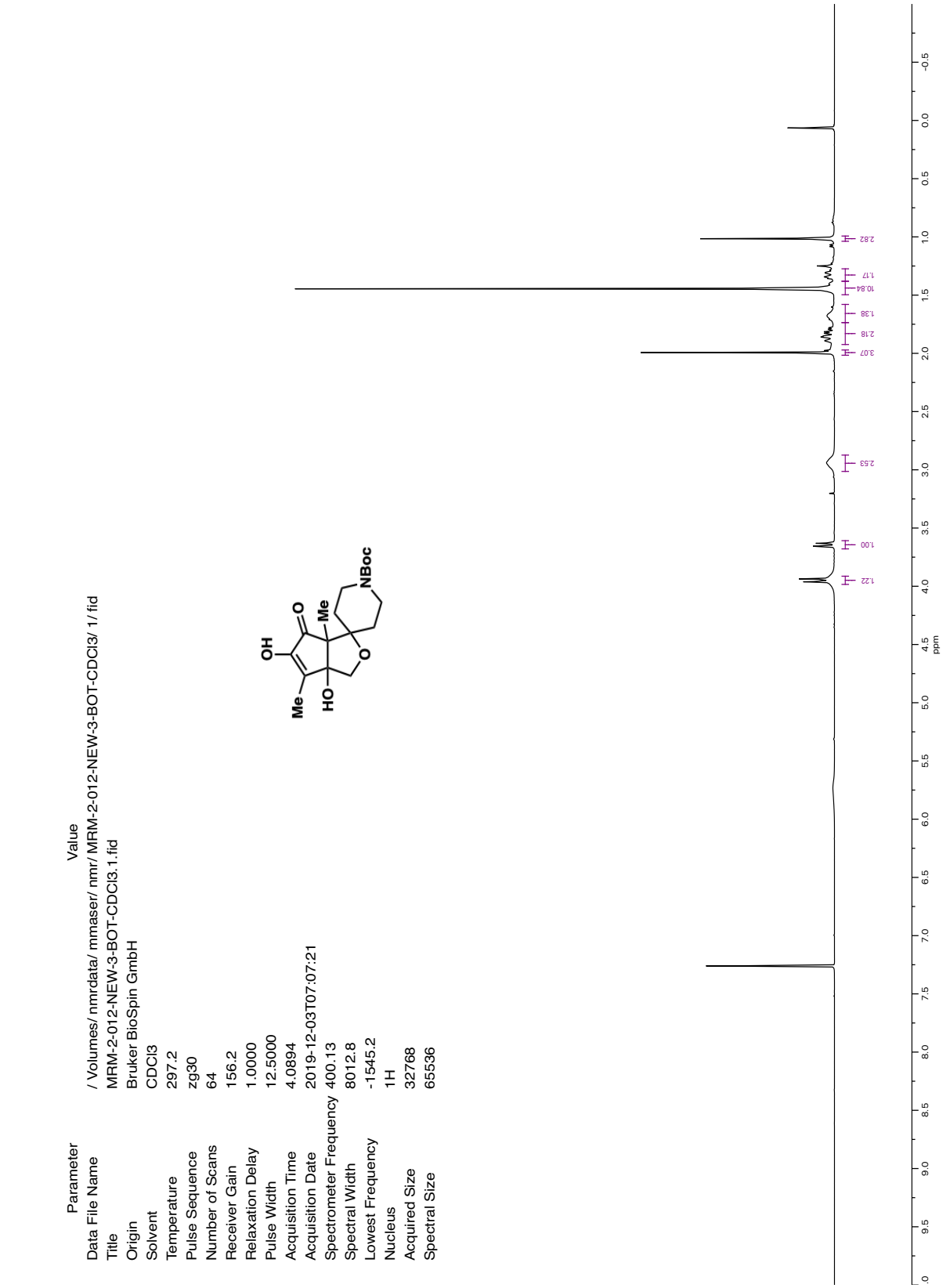

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-298-3/ 4/ ser
Title MRM-1-298-3.4.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 298.2
Pulse Sequence noesygpphpp
Number of Scans 4
Receiver Gain 101.0
Relaxation Delay 1.9447
Pulse Width 9.7500
Acquisition Time 0.3133
Acquisition Date 2019-06-19T14:08:18
Spectrometer Frequency (400.15, 400.15)
Spectral Width (3268.0, 3268.0)
Lowest Frequency (0.0, -0.3)
Nucleus (1H, 1H)
Acquired Size (1024, 256)
Spectral Size (1024, 1024)
f1 (ppm)
ppm
NOE
NOE
H
O
nhex
BnO
Me

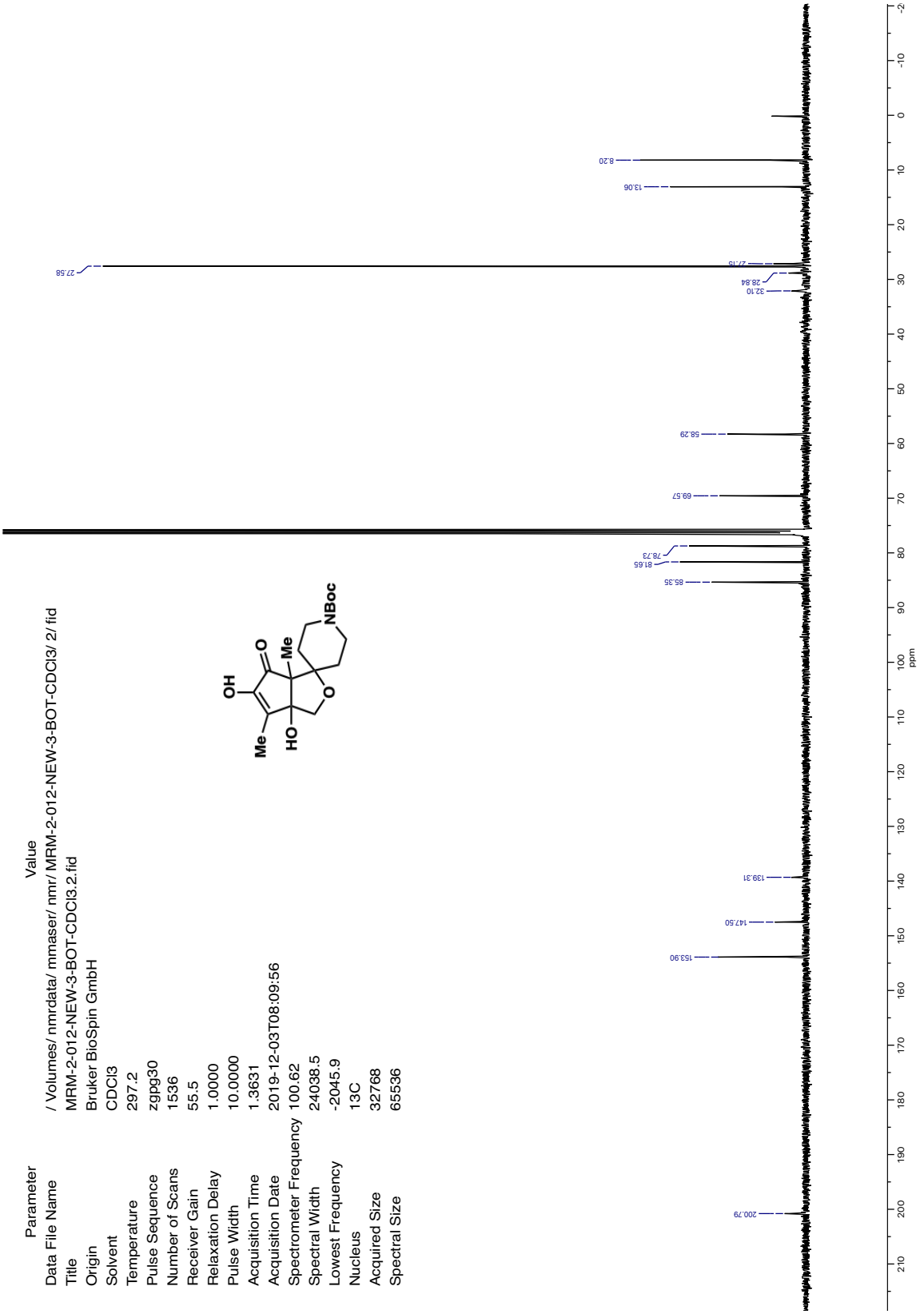

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-026-MeOD/ 1/ fid
Title MRM-2-026-MeOD.1.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 296.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-05-15T02:42:38
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1543.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
Me
Me
HO
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
0.98
1.00
1.18
1.03
0.98
2.01
1.03
3.06
2.90
6.29
5.90
ppm

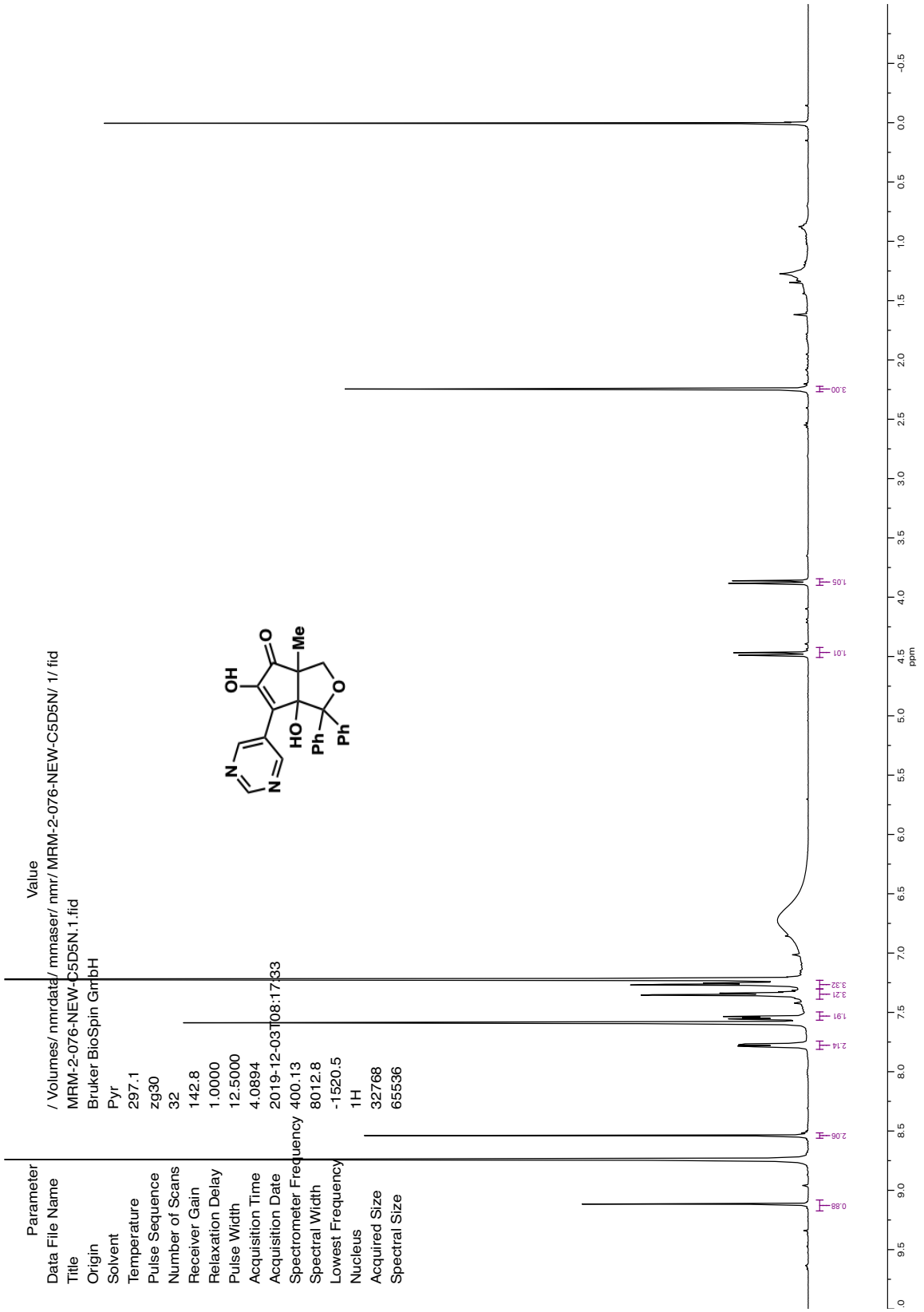

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-026/ 2/ fid
Title MRM-2-026.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-06-04T13:14:01
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
Me
Me
HO
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
209.56
181.23
171.18
171.43
131.64
69.69
69.73
60.53
51.25
47.68
44.75
33.15
26.77
21.67
21.65
21.61
21.58
8.65
ppm

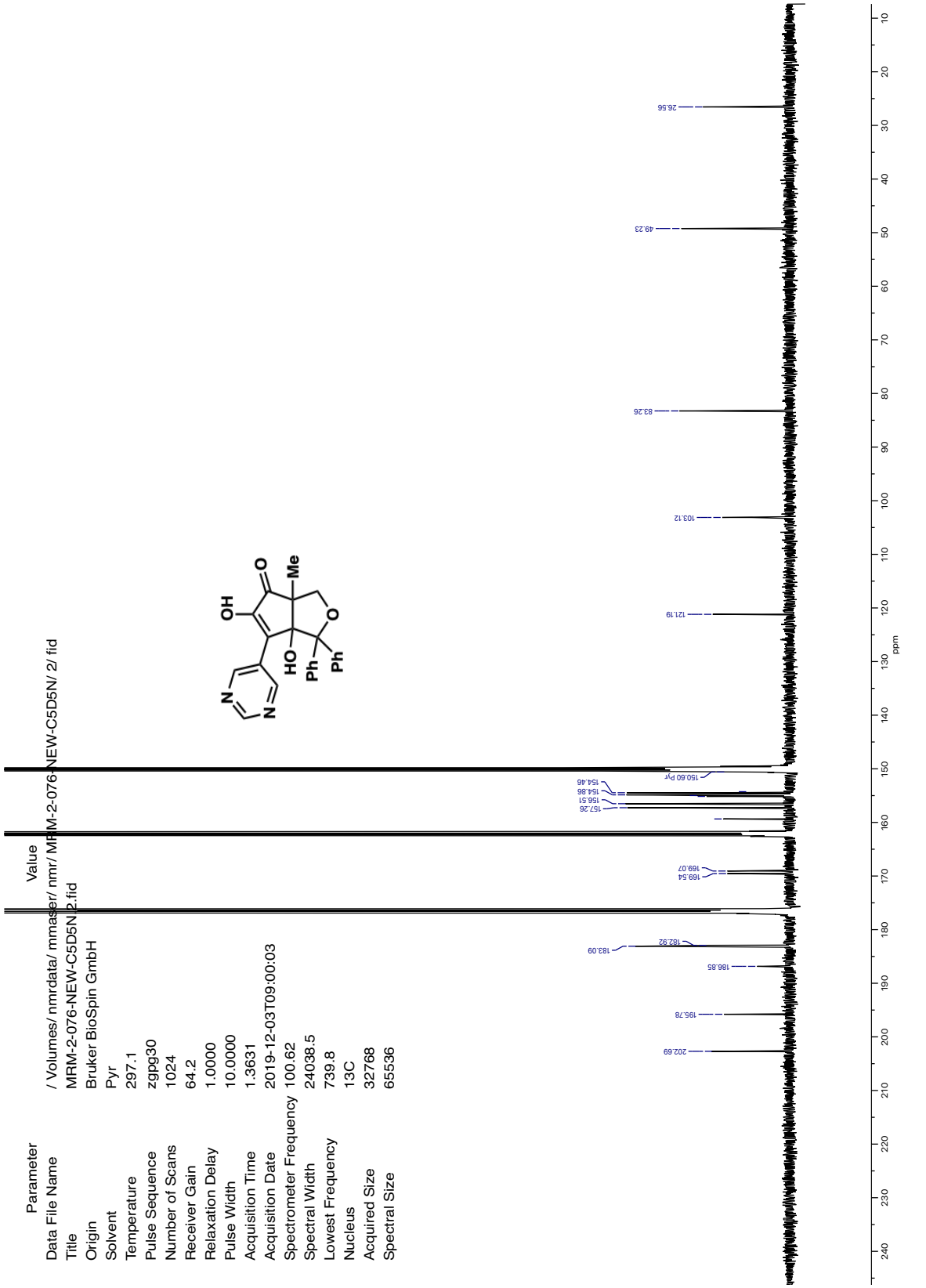

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-293-NEW-C5D5N/ 1/ fid
Title
MRM-1-293-NEW-C5D5N.1.fid
Origin
Bruker BioSpin GmbH
Solvent
Pyr
Temperature
297.1
Pulse Sequence
zg30
Number of Scans
64
Receiver Gain
176.8
Relaxation Delay
1.0000
Pulse Width
12.5000
Acquisition Time
4.0894
Acquisition Date
2019-11-20T16:00:19
Spectrometer Frequency
400.13
Spectral Width
8012.8
Lowest Frequency
-1520.6
Nucleus
1H
Acquired Size
32768
Spectral Size
65536
OH
O
Me
HO
O
N
0.96
0.94
1.02
1.04
1.00
1.05
1.02
5.10
3.01
1.00
ppm

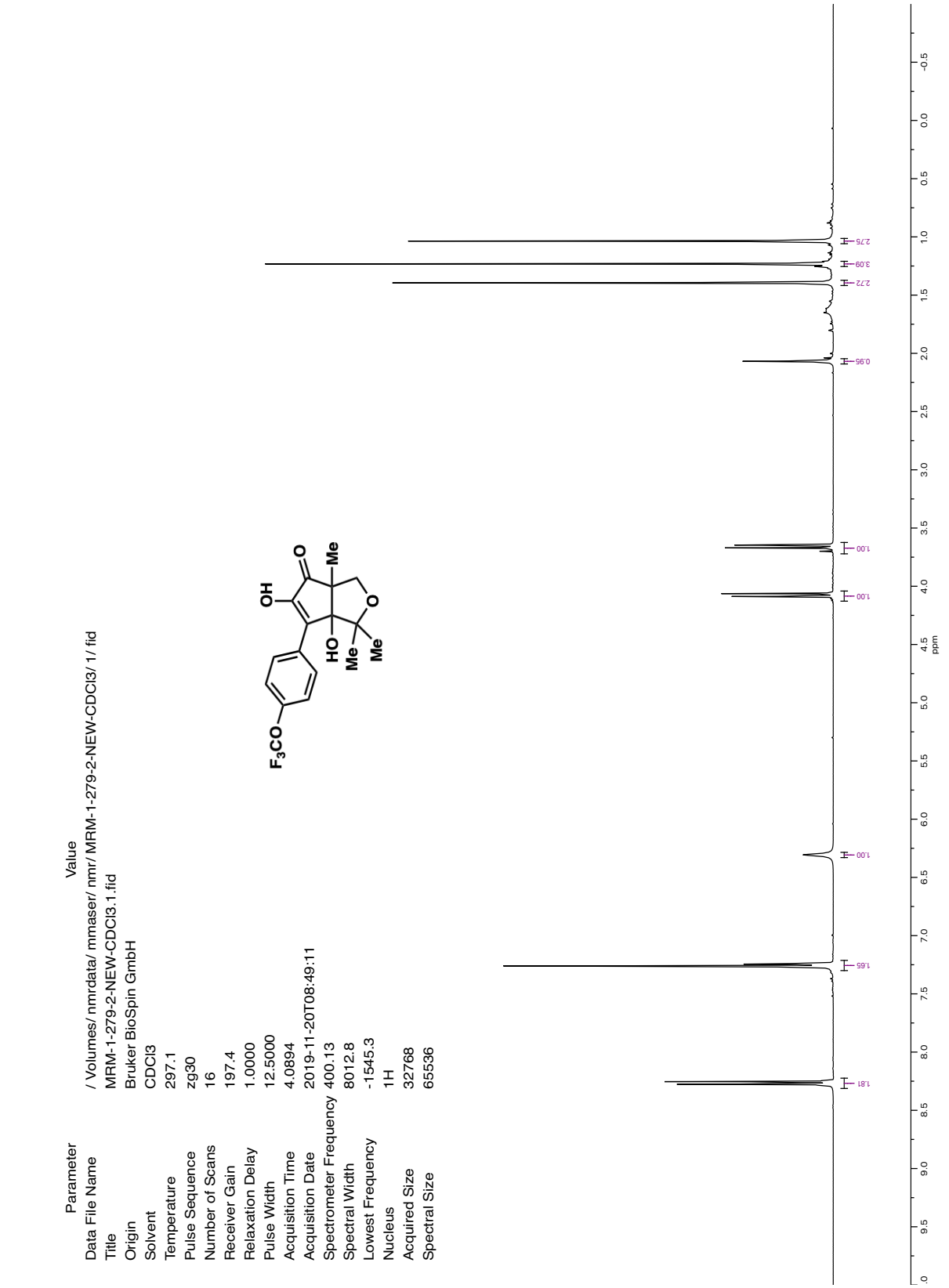

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaster/ nmr/ MRM-1-293-NEW-C5D5N/ 2/ fid
Title
MRM-1-293-NEW-C5D5N 2.fid
Origin
Bruker BioSpin GmbH
Solvent
Pyr
Temperature
297.1
Pulse Sequence
zgpg30
Number of Scans
1024
Receiver Gain
64.2
Relaxation Delay
1.0000
Pulse Width
10.0000
Acquisition Time
1.3631
Acquisition Date
2019-11-20T23:44:11
Spectrometer Frequency
100.62
Spectral Width
24038.5
Lowest Frequency
-1888.9
Nucleus
13C
Acquired Size
32768
Spectral Size
65536
OH
O
Me
HO
O
N
206.56
153.09
151.95
149.38
137.32
136.08 Pyr
135.83 Pyr
135.47
133.14
89.73
86.24
72.20
58.90
36.94
29.16
26.38
23.93
22.64
17.80
ppm

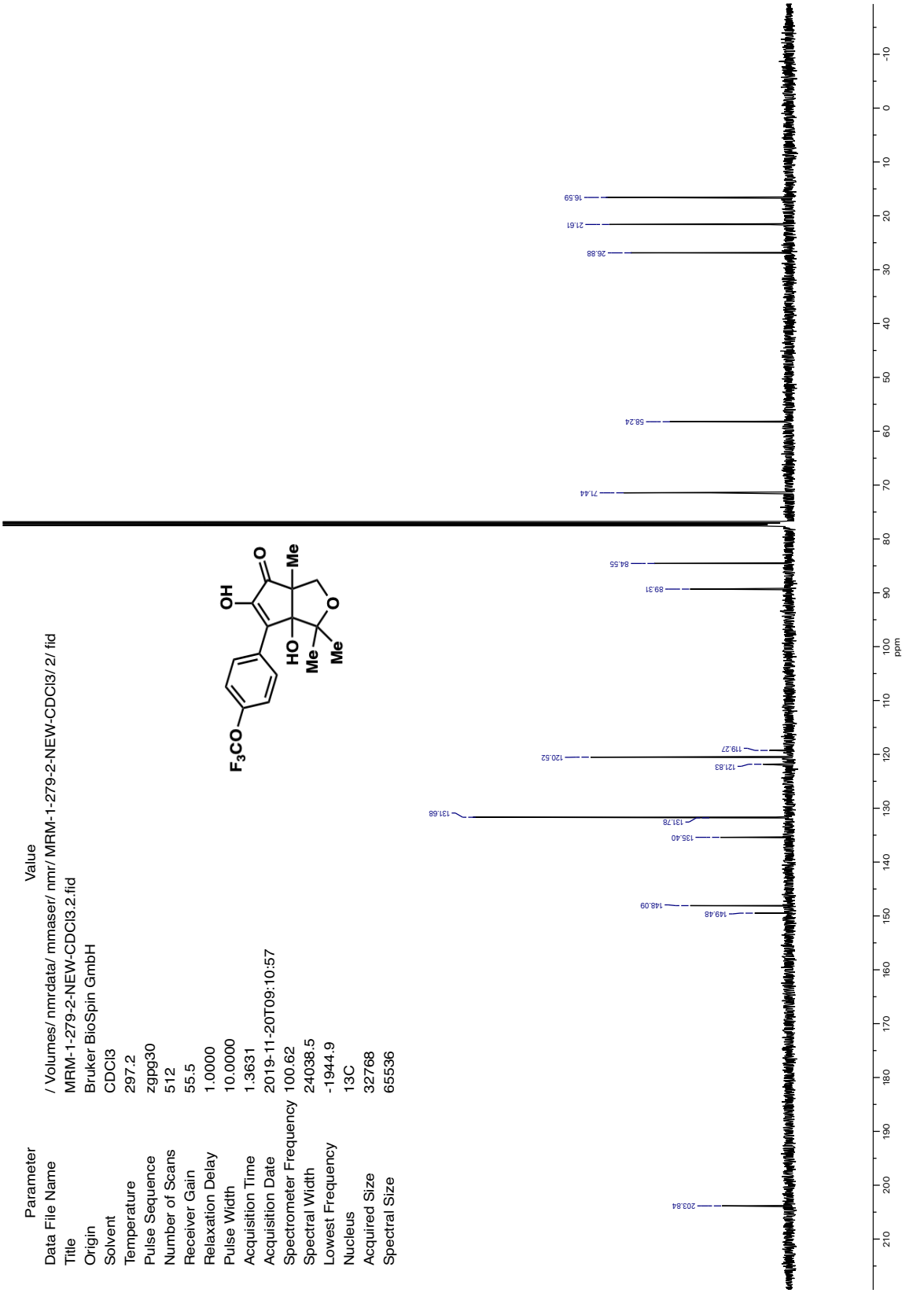

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-012-NEW-3-BOT-CDCl3/ 1/ fid
Title MRM-2-012-NEW-3-BOT-CDCl3.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 64
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-12-03T07:07:21
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
Me
NBoc
Me
HO
O
1.22
1.00
2.53
3.07
2.18
1.38
10.84
1.17
2.82
ppm

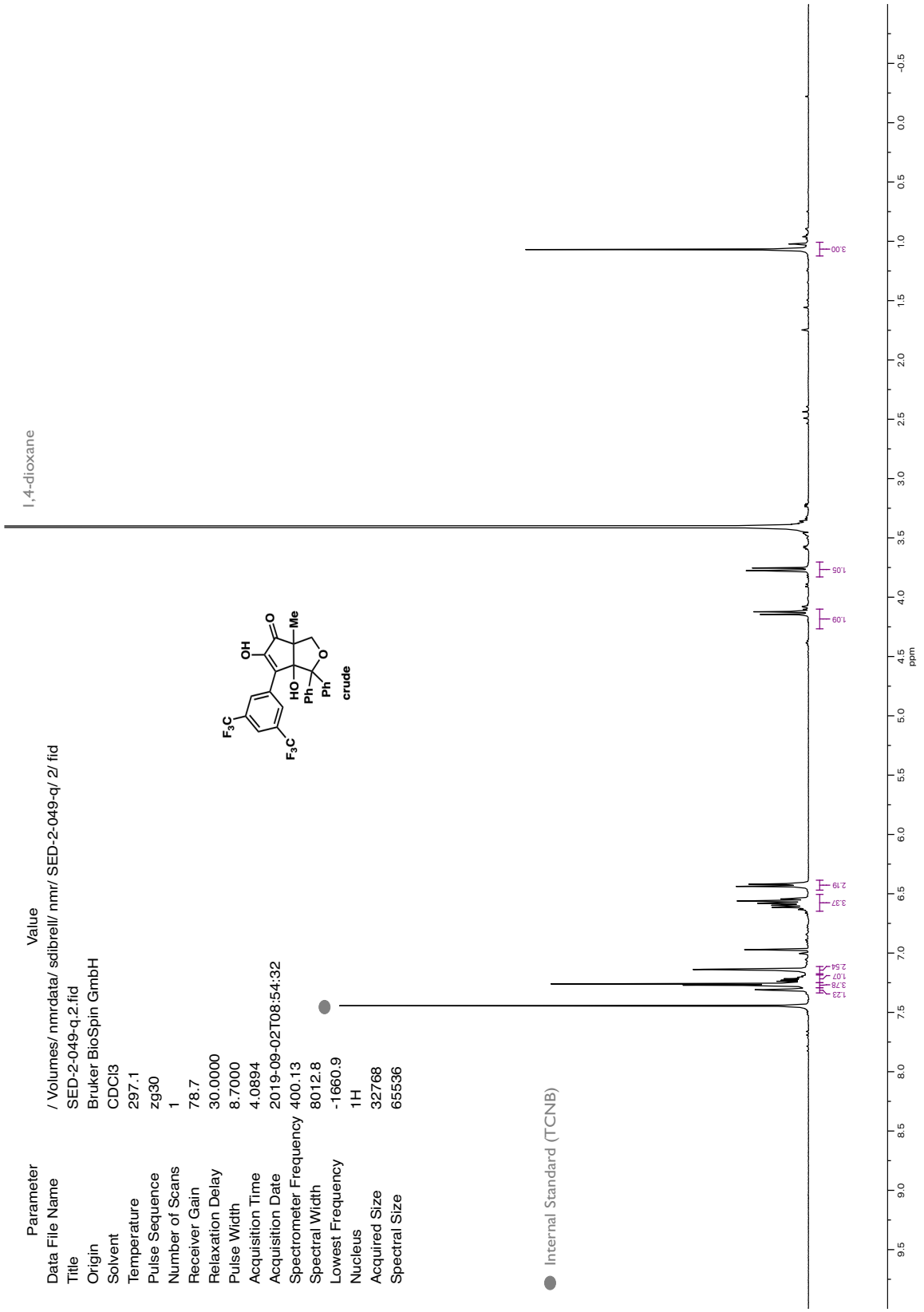

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-012-NEW-3-BOT-CDCl3/ 2/ fid
Title MRM-2-012-NEW-3-BOT-CDCl3.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1536
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-12-03T08:09:56
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -2045.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
Me
NBoc
Me
HO
O
27.58
8.20
13.06
27.19
28.84
32.10
58.29
69.57
78.73
81.65
85.35
139.31
147.50
153.90
200.79
ppm

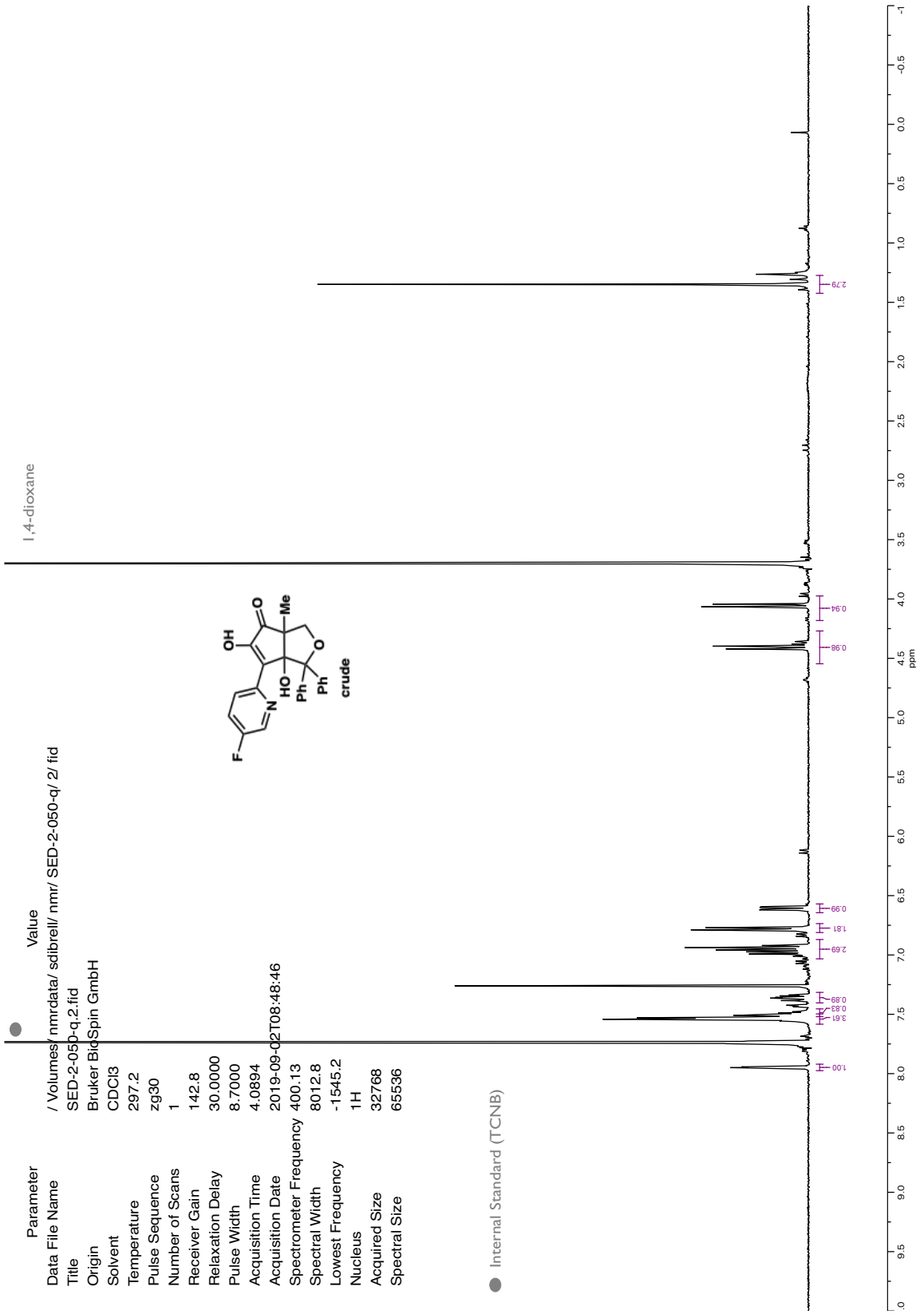

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-076-NEW-C5D5N/ 1/ fid
Title
MRM-2-076-NEW-C5D5N.1.fid
Origin
Bruker BioSpin GmbH
Solvent
Pyr
Temperature
297.1
Pulse Sequence
zg30
Number of Scans
32
Receiver Gain
142.8
Relaxation Delay
1.0000
Pulse Width
12.5000
Acquisition Time
4.0894
Acquisition Date
2019-12-03T08:17:33
Spectrometer Frequency
400.13
Spectral Width
8012.8
Lowest Frequency
-1520.5
Nucleus
1H
Acquired Size
32768
Spectral Size
65536
OH
O
Me
HO
Ph
Ph
O
N
N
0.88
2.06
2.14
1.91
3.21
3.32
1.01
1.05
3.00
ppm

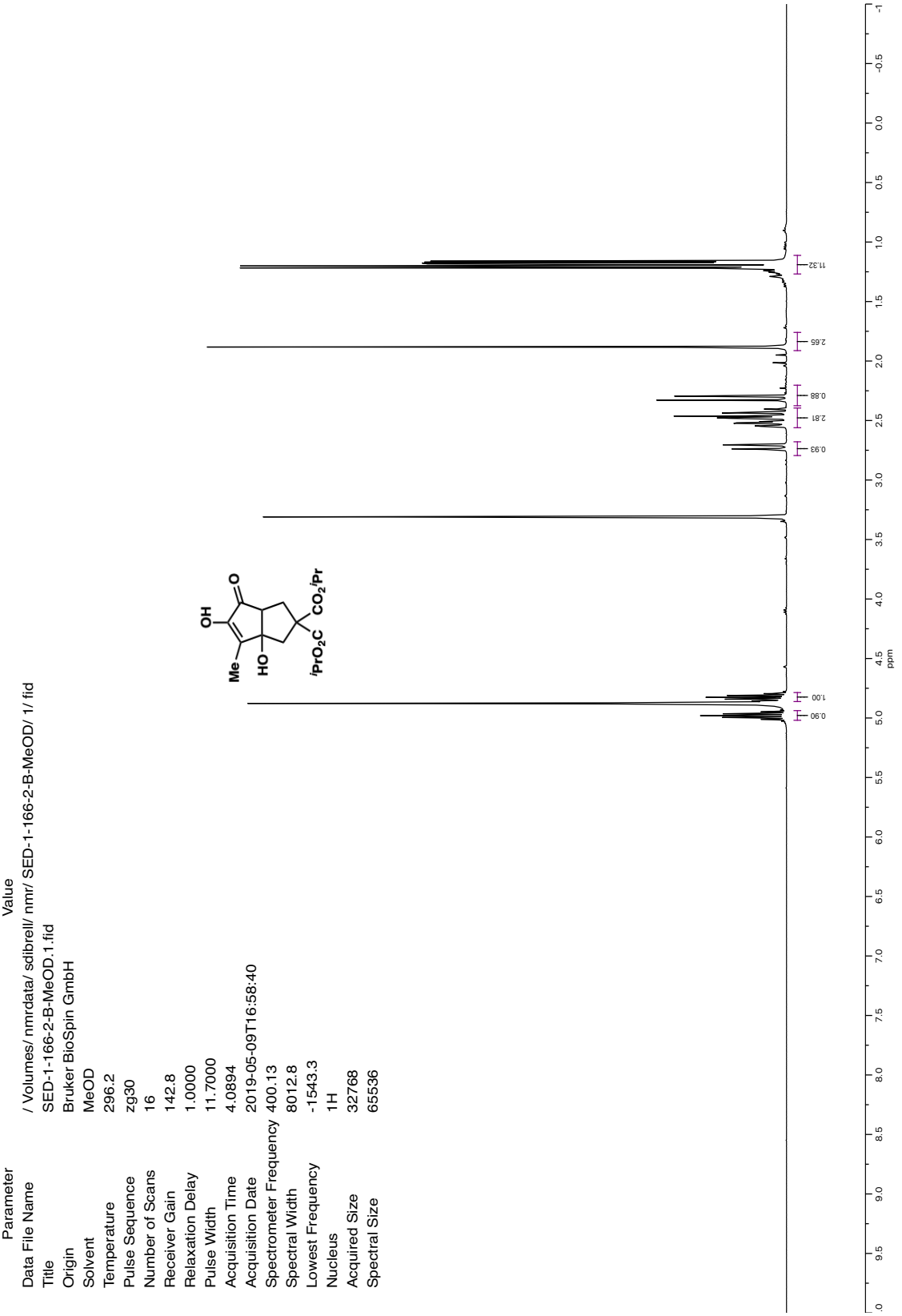

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaster/ nmr/ MRM-2-076-NEW-C5D5N/ 2/ fid
Title
MRM-2-076-NEW-C5D5N 2.fid
Origin
Bruker BioSpin GmbH
Solvent
Pyr
Temperature
297.1
Pulse Sequence
zgpg30
Number of Scans
1024
Receiver Gain
64.2
Relaxation Delay
1.0000
Pulse Width
10.0000
Acquisition Time
1.3631
Acquisition Date
2019-12-03T09:00:03
Spectrometer Frequency
100.62
Spectral Width
24038.5
Lowest Frequency
739.8
Nucleus
13C
Acquired Size
32768
Spectral Size
65536
OH
O
Me
HO
Ph
Ph
O
N
N
26.56
49.23
83.26
103.12
121.19
150.60 Pyr
154.46
154.86
156.51
157.26
169.54
169.07
182.92
183.09
186.85
195.78
202.69
ppm

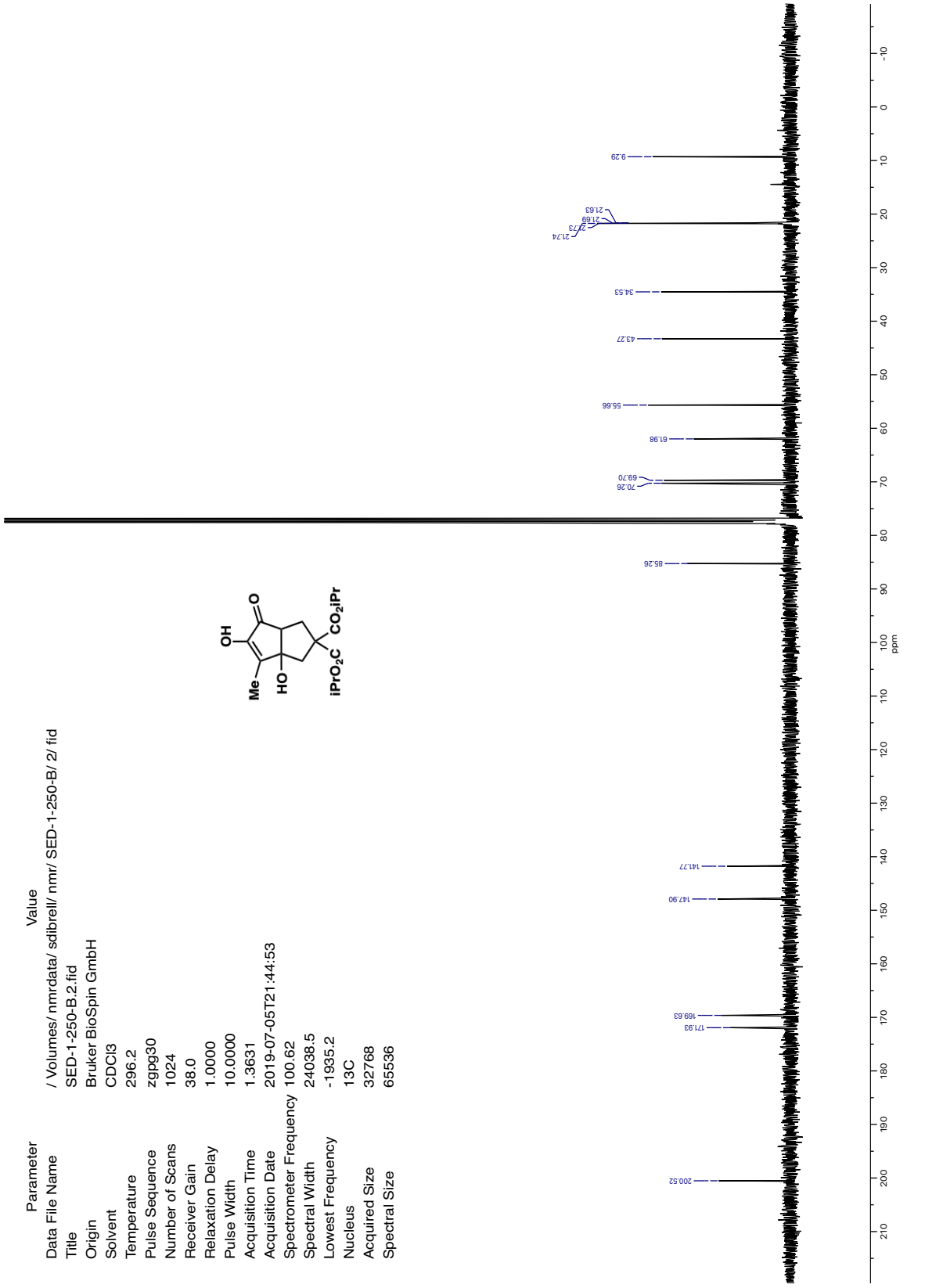

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-279-2-NEW-CDCl3/ 1/ fid
Title
MRM-1-279-2-NEW-CDCl3.1.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
zg30
Number of Scans
16
Receiver Gain
197.4
Relaxation Delay
1.0000
Pulse Width
12.5000
Acquisition Time
4.0894
Acquisition Date
2019-11-20T08:49:11
Spectrometer Frequency
400.13
Spectral Width
8012.8
Lowest Frequency
-1545.3
Nucleus
1H
Acquired Size
32768
Spectral Size
65536
OH
O
Me
HO
Me
Me
O
F3CO
1.81
1.65
1.00
1.00
1.00
0.96
2.72
3.09
2.75
ppm

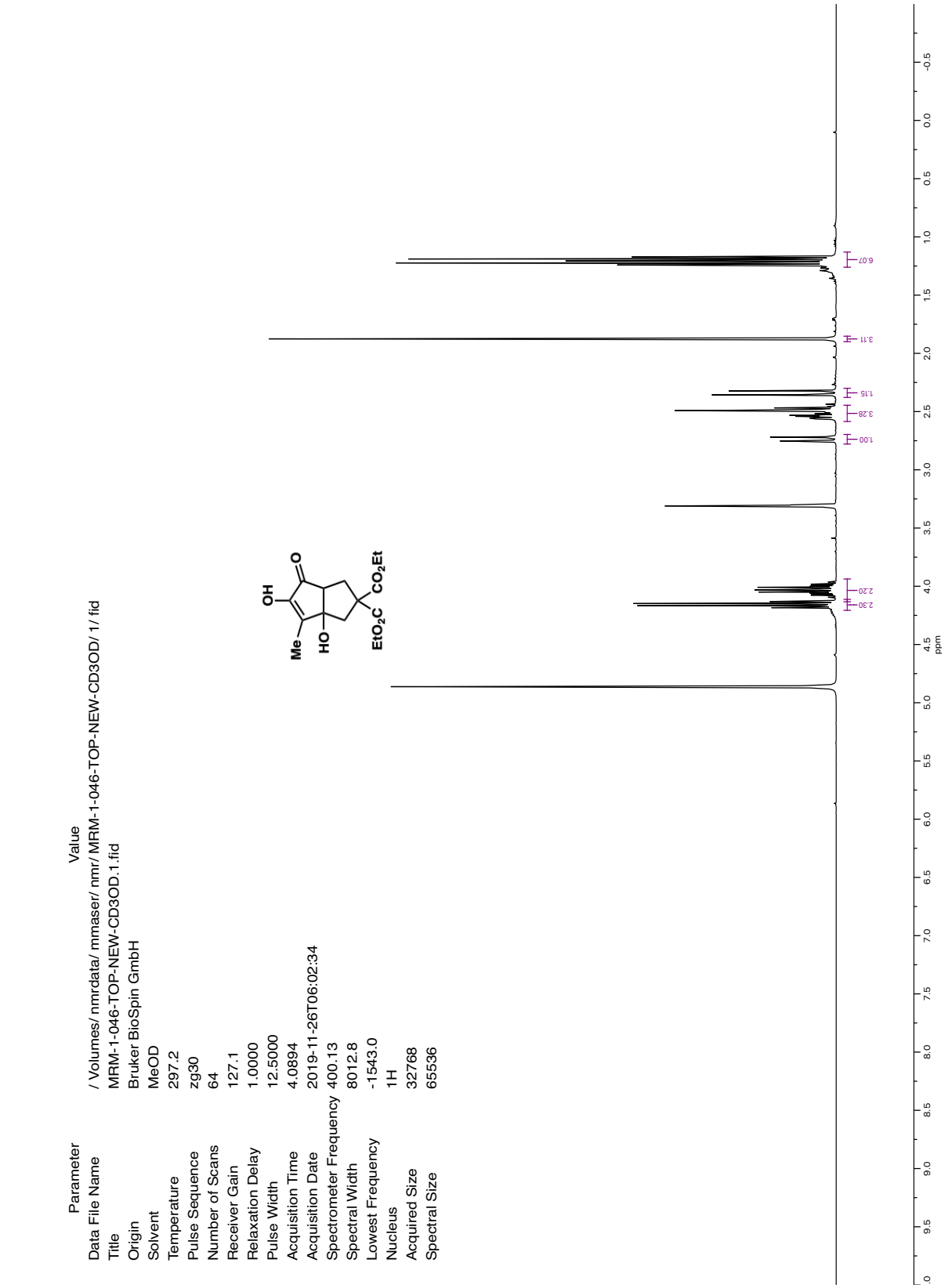

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-279-2-NEW-CDCl3/ 2/ fid
Title MRM-1-279-2-NEW-CDCl3.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-20T09:10:57
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
Me
HO
Me
Me
O
F3CO
16.59
21.61
26.88
58.24
71.44
84.55
89.31
119.27
120.52
121.83
131.68
131.78
135.40
148.09
149.48
203.84
ppm

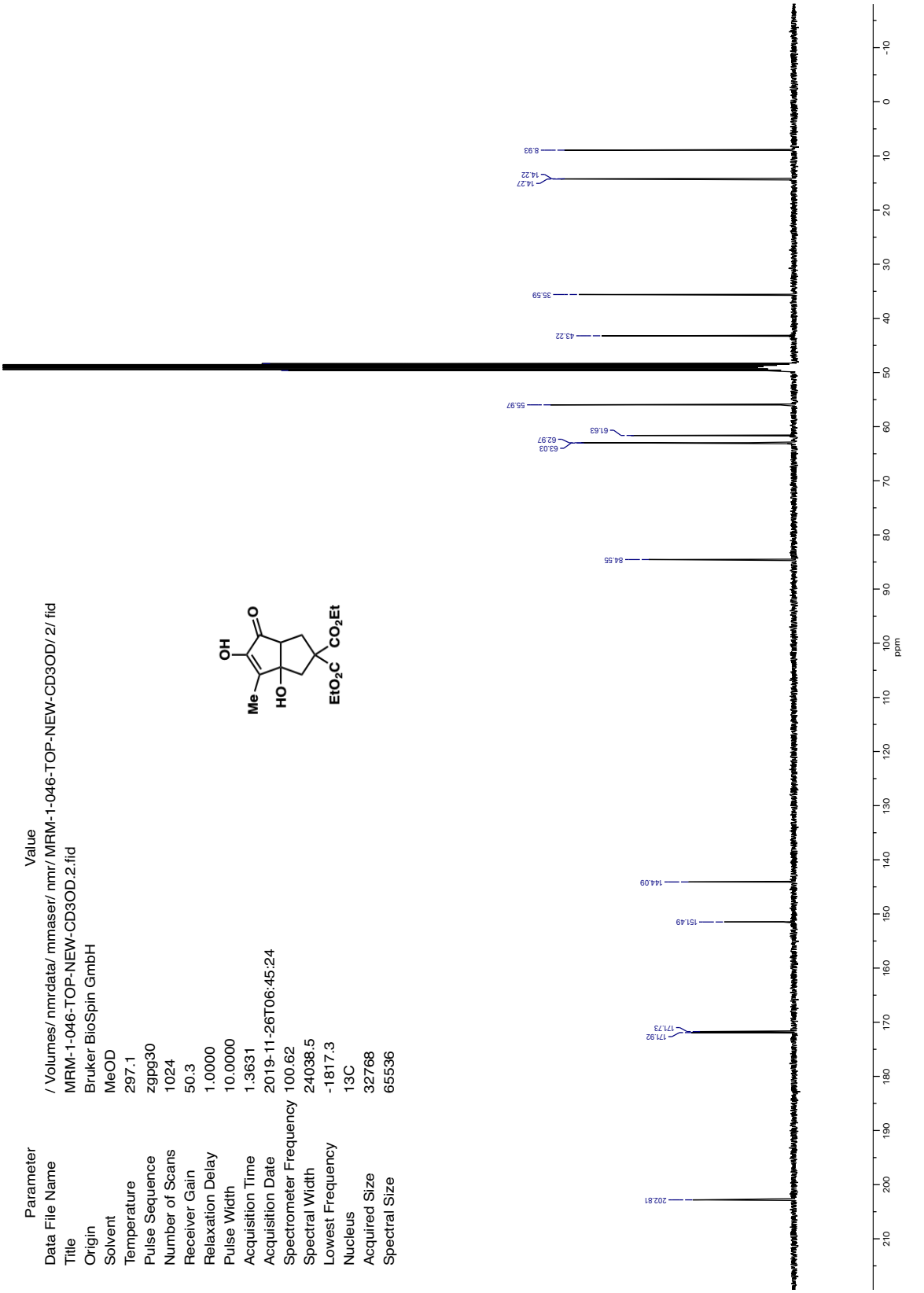

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-049-q/ 2/ fid
Title SED-2-049-q.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 1
Receiver Gain 78.7
Relaxation Delay 30.0000
Pulse Width 8.7000
Acquisition Time 4.0894
Acquisition Date 2019-09-02T08:54:32
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1660.9
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Internal Standard (TCNB)
1,4-dioxane
crude
3.00
1.05
1.09
2.19
3.37
1.23
3.78
1.07
2.54
ppm

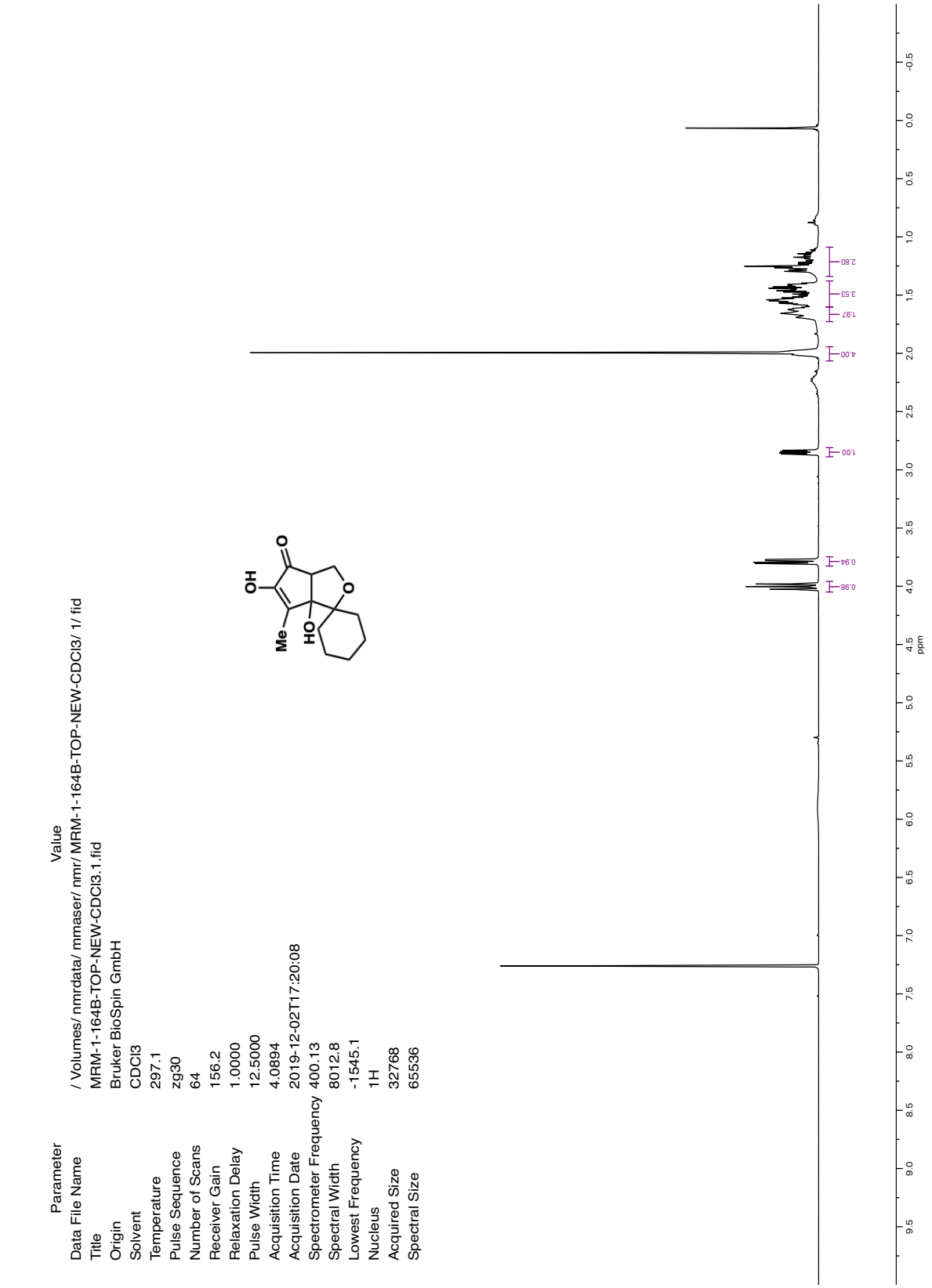

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-050-q/ 2/ fid
Title SED-2-050-q.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 1
Receiver Gain 142.8
Relaxation Delay 30.0000
Pulse Width 8.7000
Acquisition Time 4.0894
Acquisition Date 2019-09-02T08:48:46
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Internal Standard (TCNB)
1,4-dioxane
crude
1.00
3.61
0.83
0.89
2.69
1.81
0.99
0.98
0.94
2.79
ppm

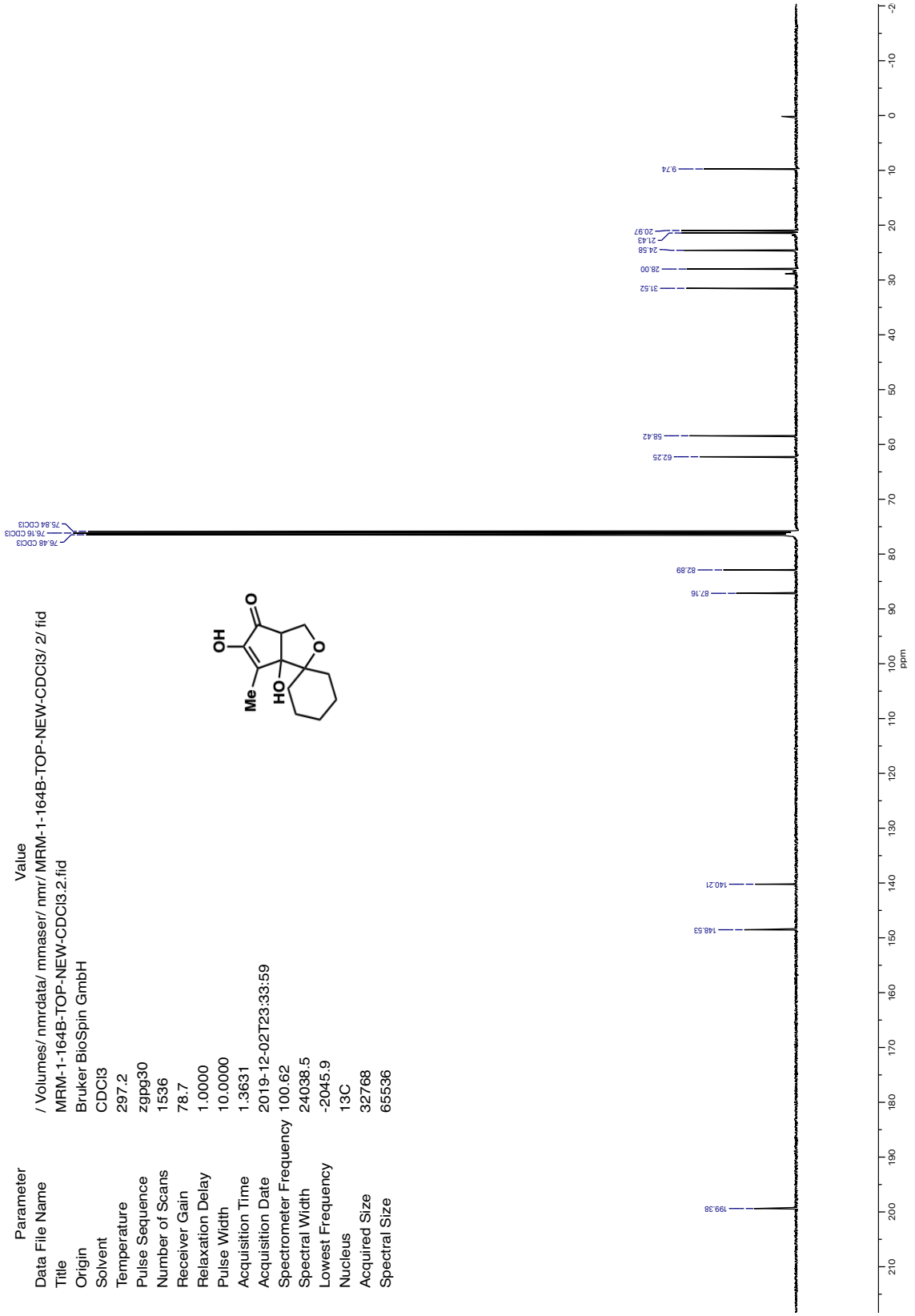

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-166-2-B-MeOD/ 1/ fid
Title SED-1-166-2-B-MeOD.1.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 296.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 142.8
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-05-09T16:58:40
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1543.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
Me
HO
iPrO2C
CO2iPr
11.32
2.65
0.88
2.81
0.93
1.00
0.90
ppm

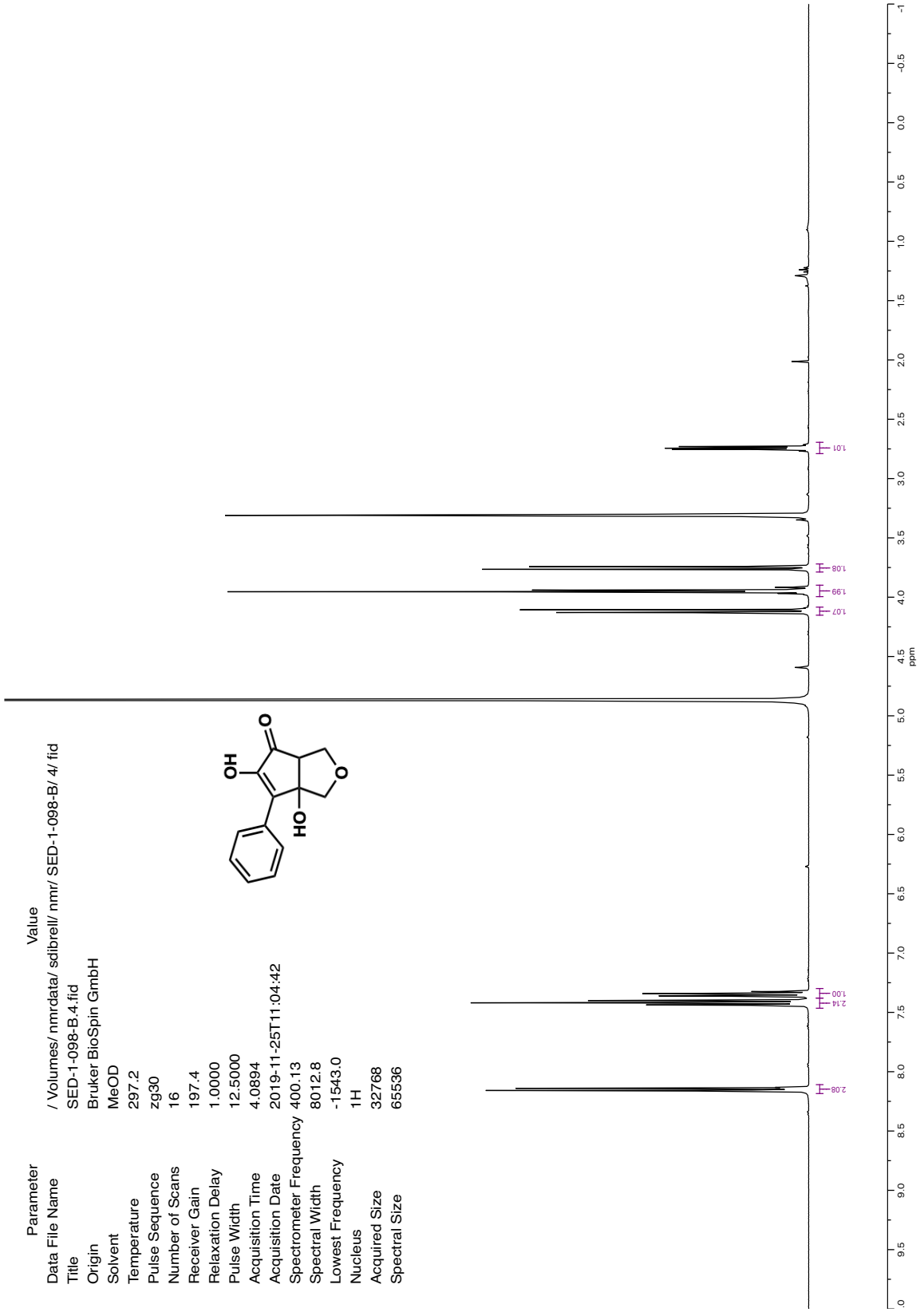

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-250-B/ 2/ fid
Title SED-1-250-B.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 38.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-07-05T21:44:53
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1935.2
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
Me
HO
iPrO2C
CO2iPr
200.52
171.93
169.63
147.90
141.77
85.26
70.26
69.70
61.98
55.66
43.27
34.53
21.74
21.73
21.68
21.63
9.29
ppm

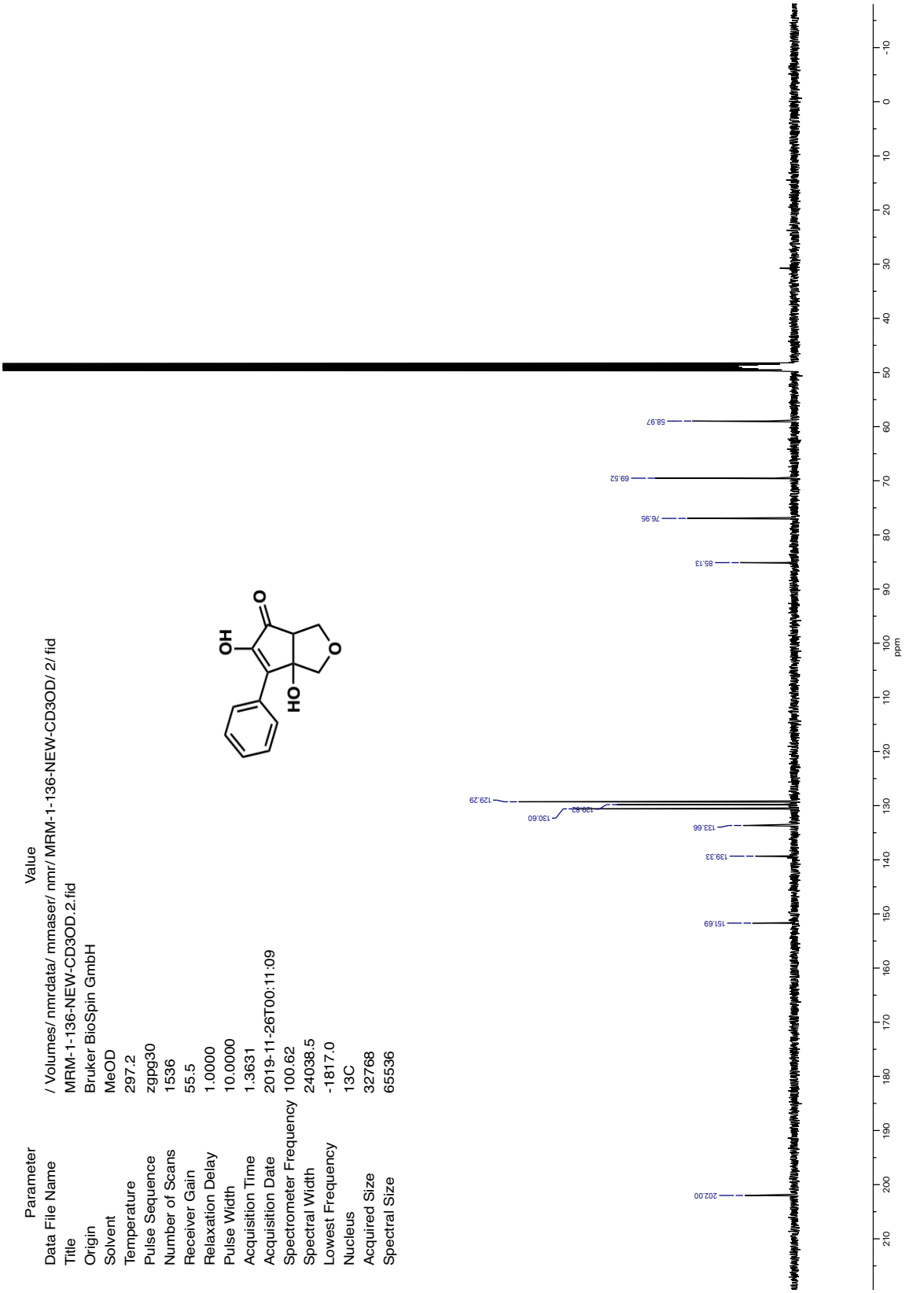

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-046-TOP-NEW-CD3OD/ 1/ fid
Title MRM-1-046-TOP-NEW-CD3OD.1.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 297.2
Pulse Sequence zg30
Number of Scans 64
Receiver Gain 127.1
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-26T06:02:34
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1543.0
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
Me
HO
EtO2C
CO2Et
6.07
3.11
1.15
3.28
1.00
2.20
2.30
ppm

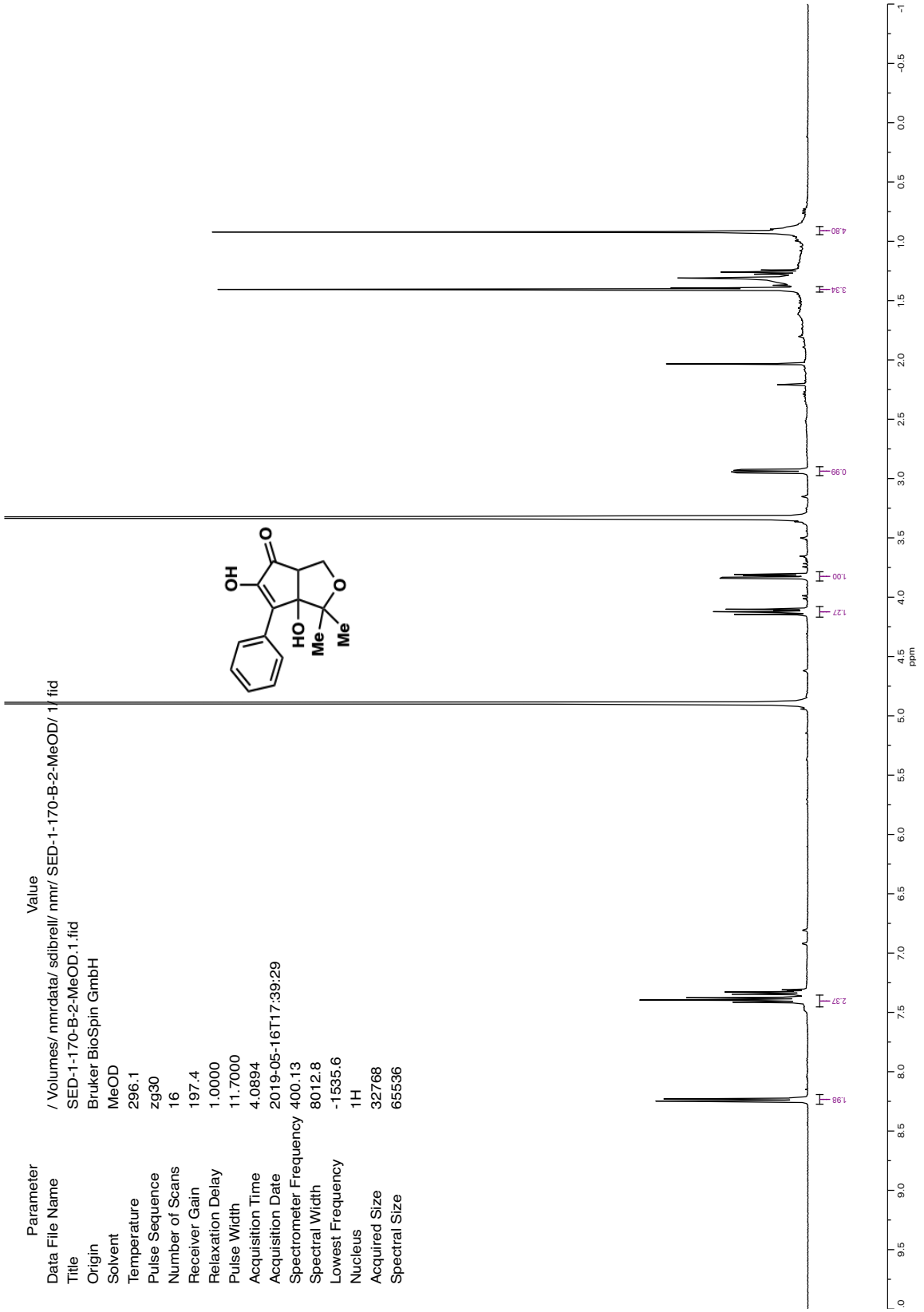

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-046-TOP-NEW-CD3OD/ 2/ fid
Title MRM-1-046-TOP-NEW-CD3OD.2.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-26T06:45:24
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1817.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
Me
HO
EtO2C
CO2Et
202.81
171.92
171.73
151.49
144.09
84.55
63.03
62.97
61.63
55.97
43.22
35.59
14.27
14.22
8.93
ppm

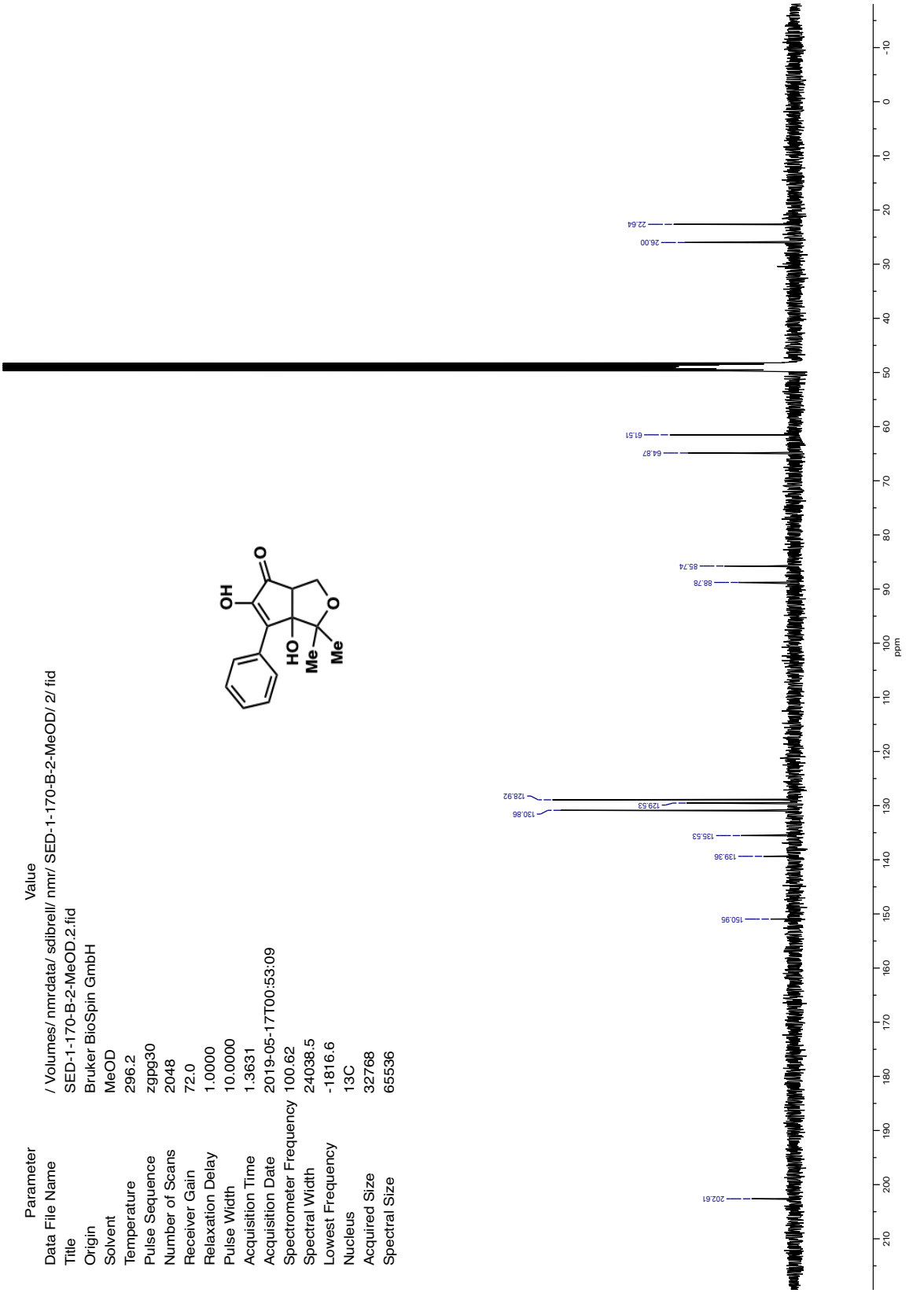

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-164B-TOP-NEW-CDCl3/ 1/ fid
Title
MRM-1-164B-TOP-NEW-CDCl3.1.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
zg30
Number of Scans
64
Receiver Gain
156.2
Relaxation Delay
1.0000
Pulse Width
12.5000
Acquisition Time
4.0894
Acquisition Date
2019-12-02T17:20:08
Spectrometer Frequency
400.13
Spectral Width
8012.8
Lowest Frequency
-1545.1
Nucleus
1H
Acquired Size
32768
Spectral Size
65536
OH
O
Me
HO
O
0.98
0.94
1.00
4.00
1.97
3.53
2.80
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
ppm

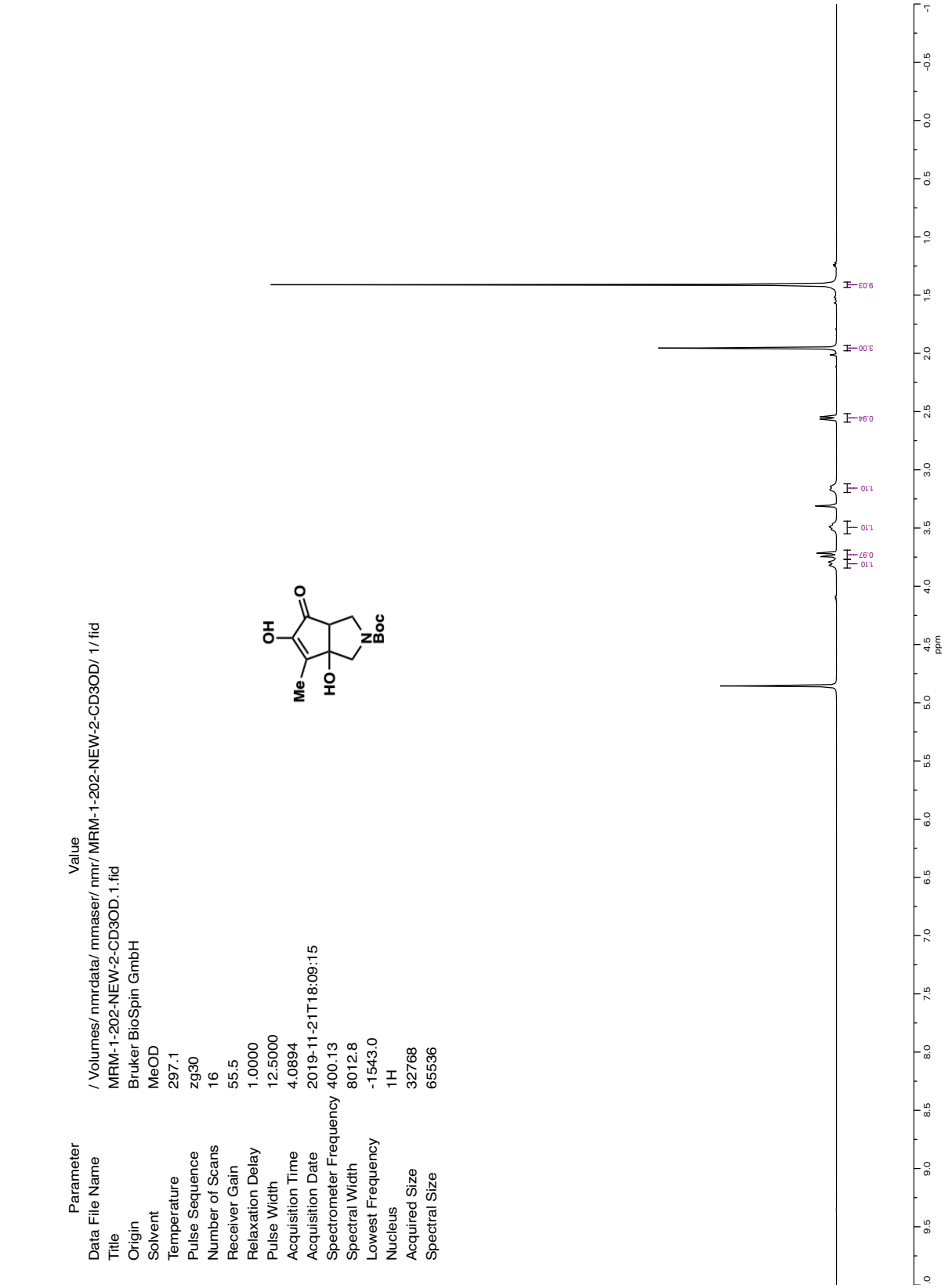

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-164B-TOP-NEW-CDCl3/ 2/ fid
Title
MRM-1-164B-TOP-NEW-CDCl3.2.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.2
Pulse Sequence
zgpg30
Number of Scans
1536
Receiver Gain
78.7
Relaxation Delay
1.0000
Pulse Width
10.0000
Acquisition Time
1.3631
Acquisition Date
2019-12-02T23:33:59
Spectrometer Frequency
100.62
Spectral Width
24038.5
Lowest Frequency
-2045.9
Nucleus
13C
Acquired Size
32768
Spectral Size
65536
OH
O
Me
HO
O
76.84 CDCl3
76.16 CDCl3
76.48 CDCl3
199.38
148.53
140.21
87.16
82.89
62.25
58.42
31.52
28.00
24.58
21.43
20.97
9.74
ppm

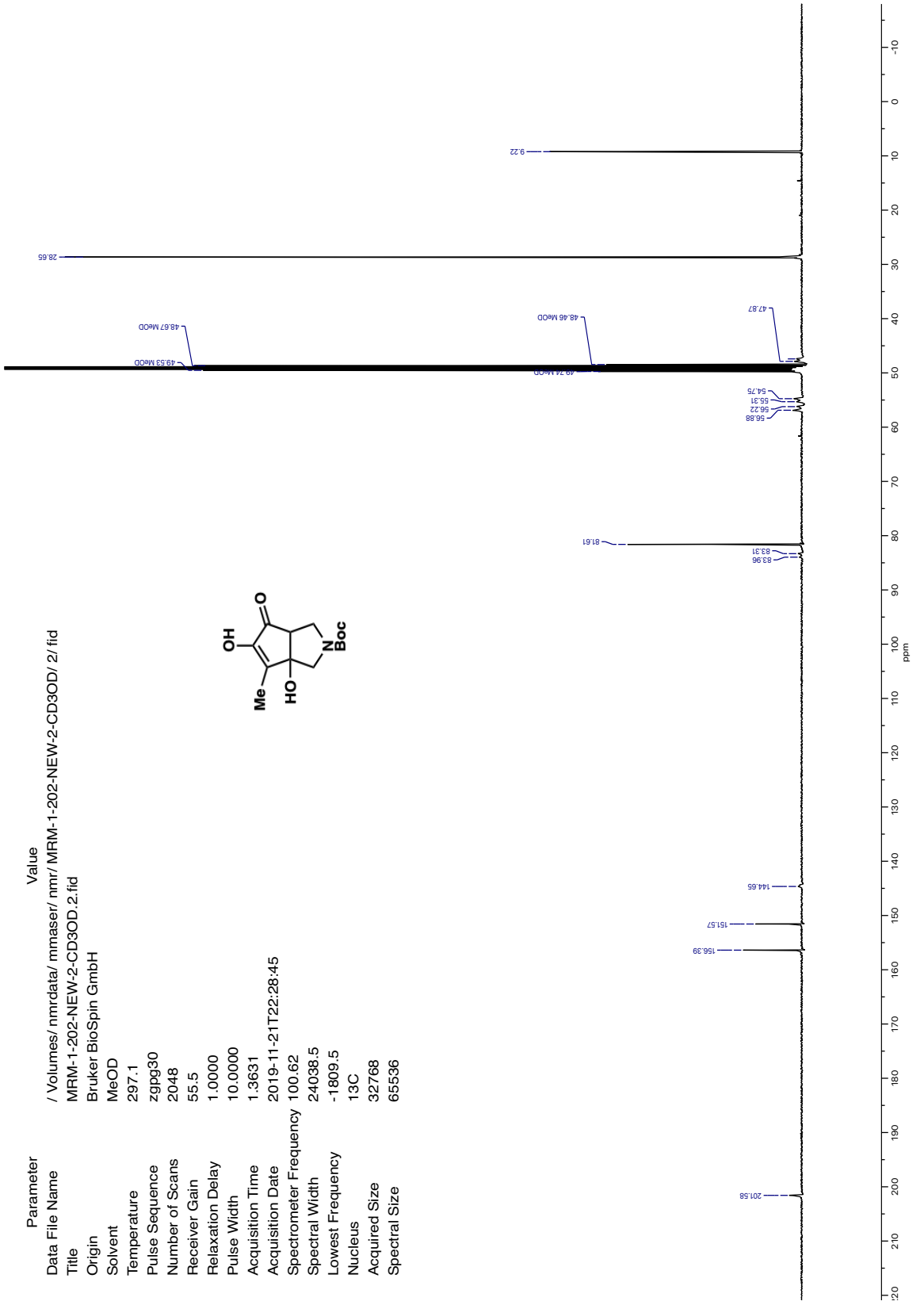

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-098-B/ 4/ fid
Title SED-1-098-B.4.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-25T11:04:42
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1543.0
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
HO
O
2.08
2.14
1.00
1.01
1.08
1.99
1.07
ppm

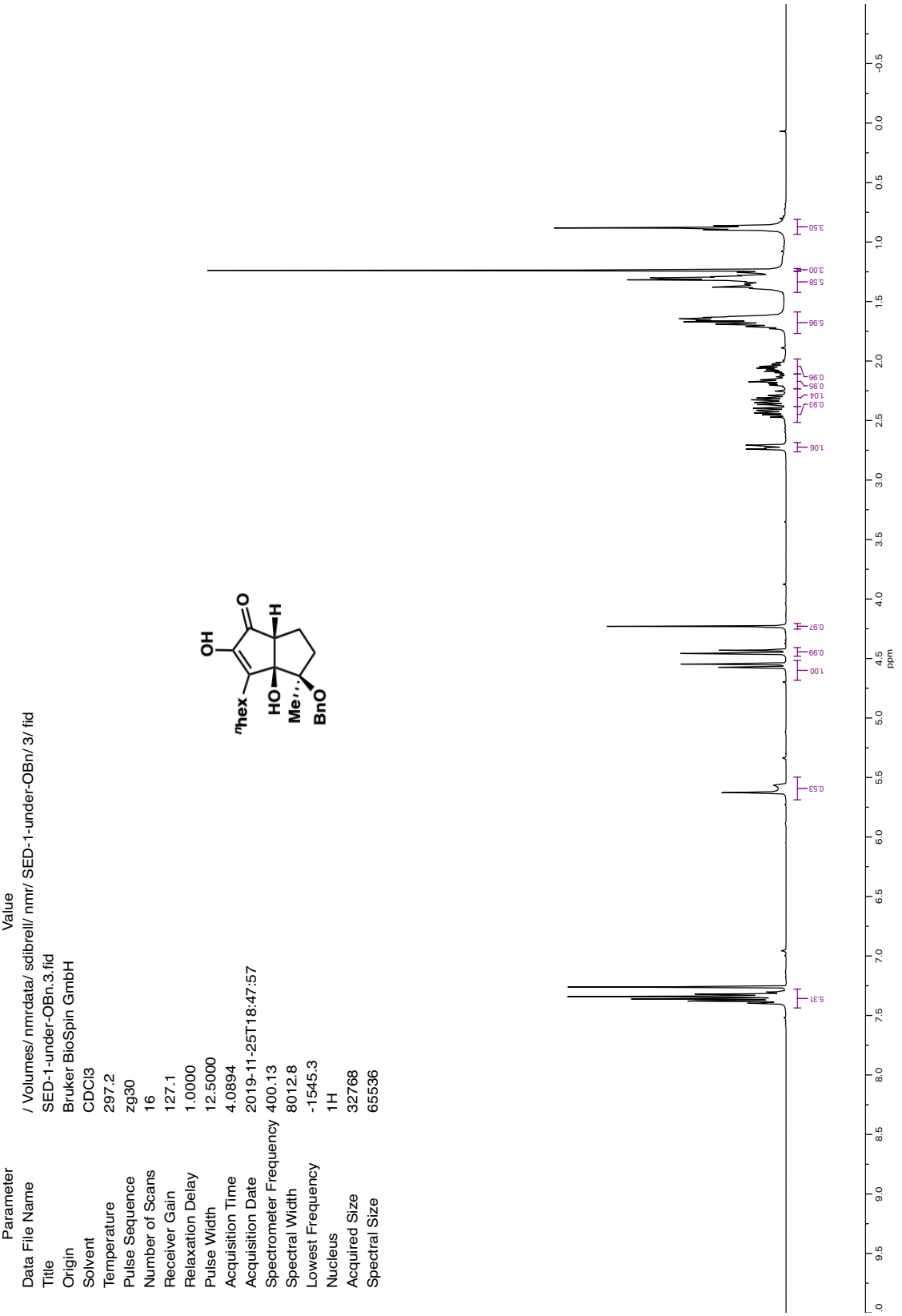

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-136-NEW-CD3OD/ 2/ fid
Title MRM-1-136-NEW-CD3OD.2.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1536
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-26T00:11:09
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1817.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
HO
O
58.97
69.52
76.95
85.13
129.29
130.60
130.82
133.66
139.33
151.69
202.00
ppm

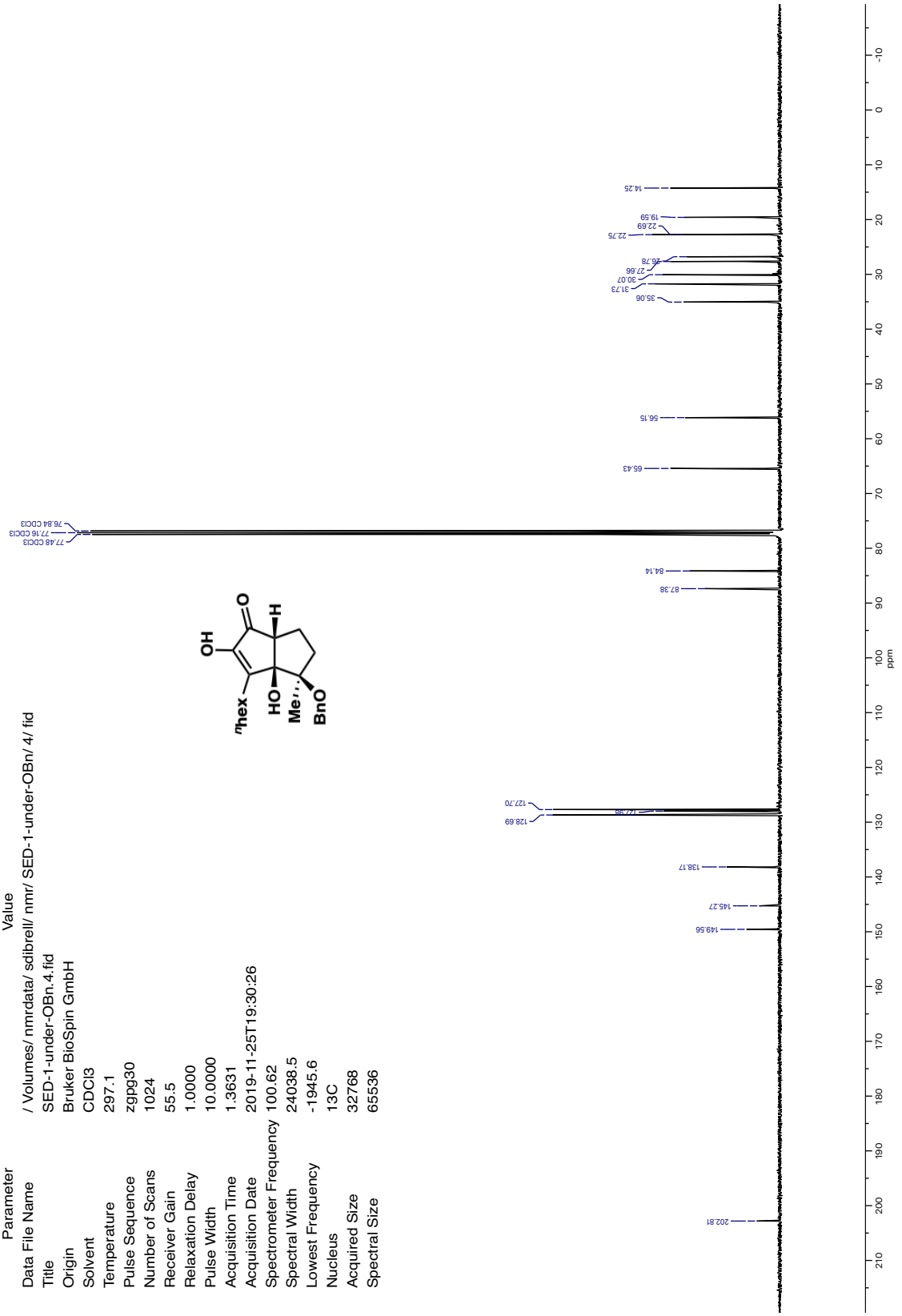

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-170-B-2-MeOD/ 1/ fid
Title SED-1-170-B-2-MeOD.1.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 296.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-05-16T17:39:29
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1535.6
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
HO
Me
Me
O
1.98
2.37
1.27
1.00
0.99
3.34
4.80
ppm

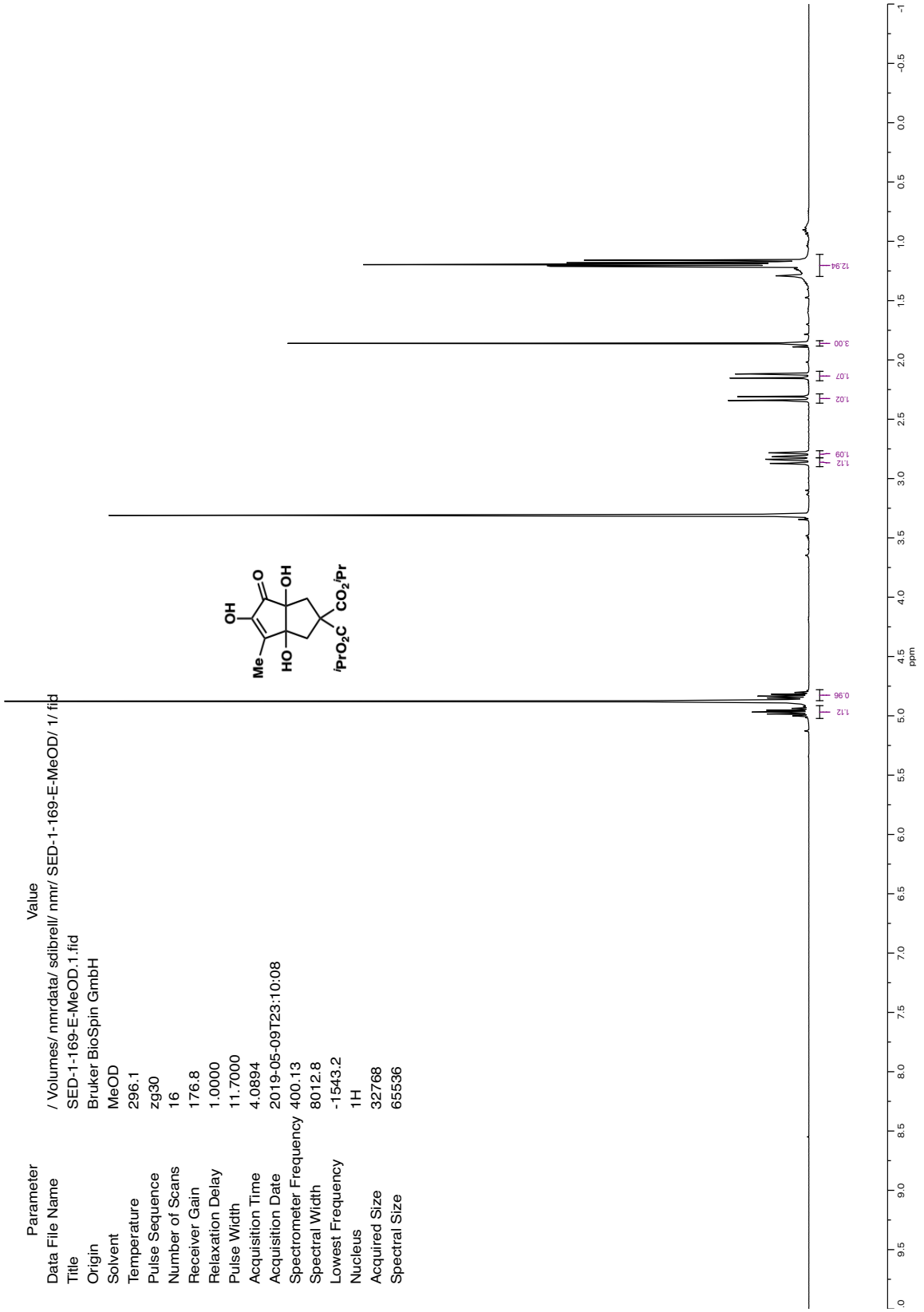

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-170-B-2-MeOD/ 2/ fid
Title SED-1-170-B-2-MeOD.2.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 296.2
Pulse Sequence zgpg30
Number of Scans 2048
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-05-17T00:53:09
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1816.6
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
HO
Me
Me
O
22.64
26.00
61.51
64.87
85.74
88.78
128.92
129.53
130.86
135.53
139.36
150.95
202.61
ppm

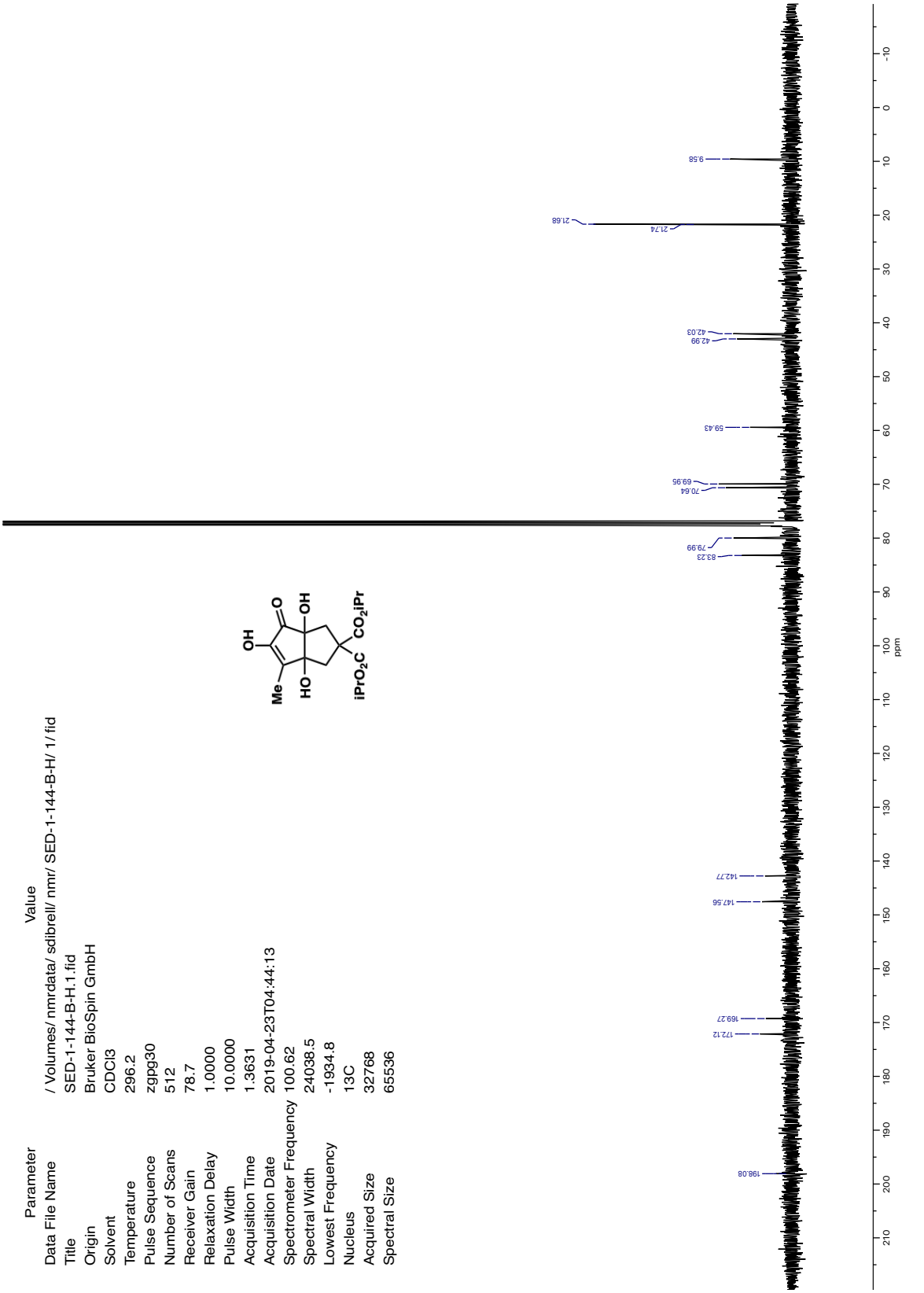

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-202-NEW-2-CD3OD/ 1/ fid
Title MRM-1-202-NEW-2-CD3OD.1.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-21T18:09:15
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1543.0
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
Me
HO
N
Boc
9.03
3.00
0.94
1.10
1.10
0.97
1.10
ppm

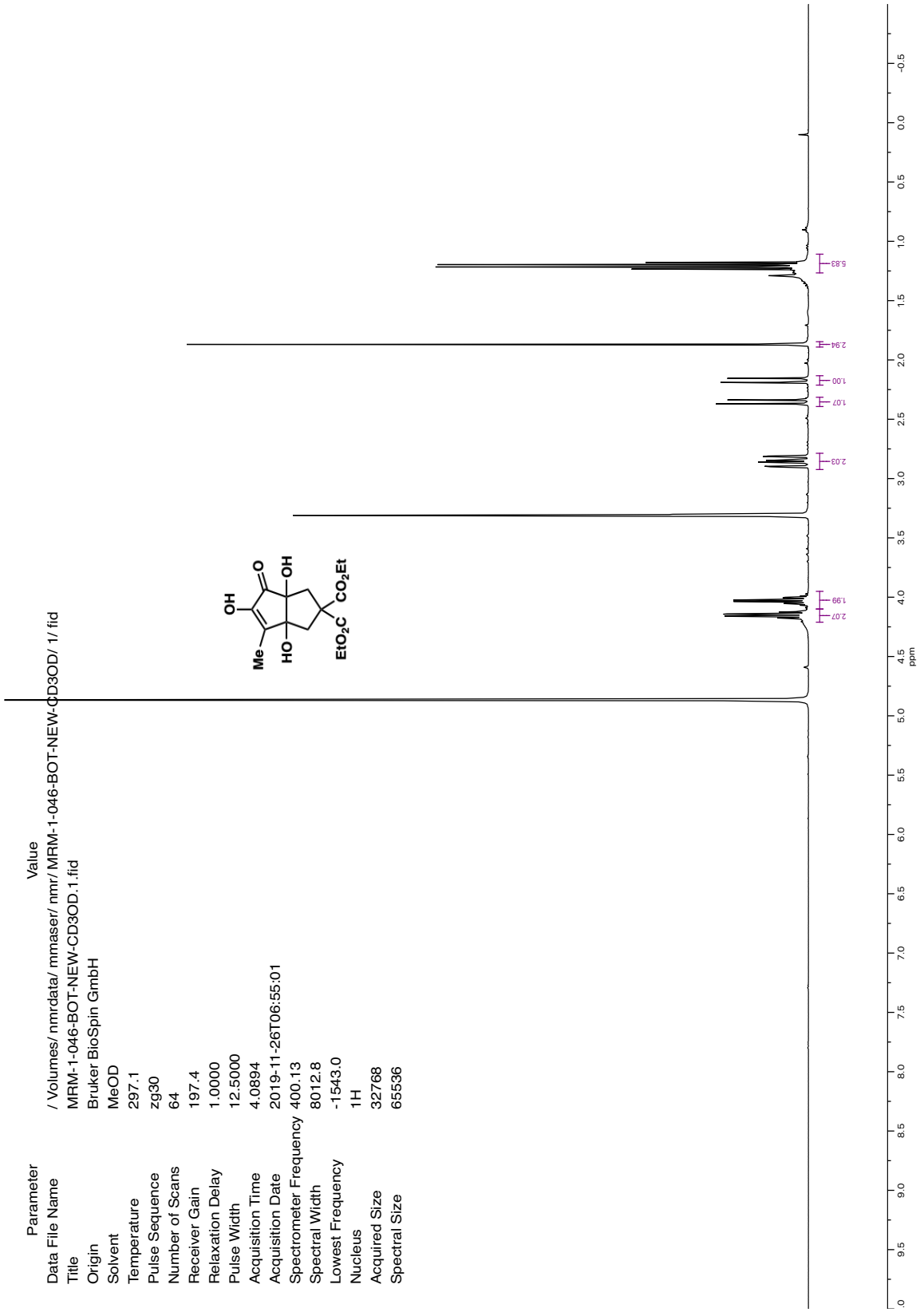

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-202-NEW-2-CD3OD/ 2/ fid
Title MRM-1-202-NEW-2-CD3OD.2.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 2048
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-21T22:28:45
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1809.5
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
Me
HO
N
Boc
9.22
28.65
47.87
48.67 MeOD
49.53 MeOD
48.46 MeOD
49.24 MeOD
54.75
55.31
56.22
56.88
81.61
83.31
83.96
144.65
151.57
156.39
201.58
ppm

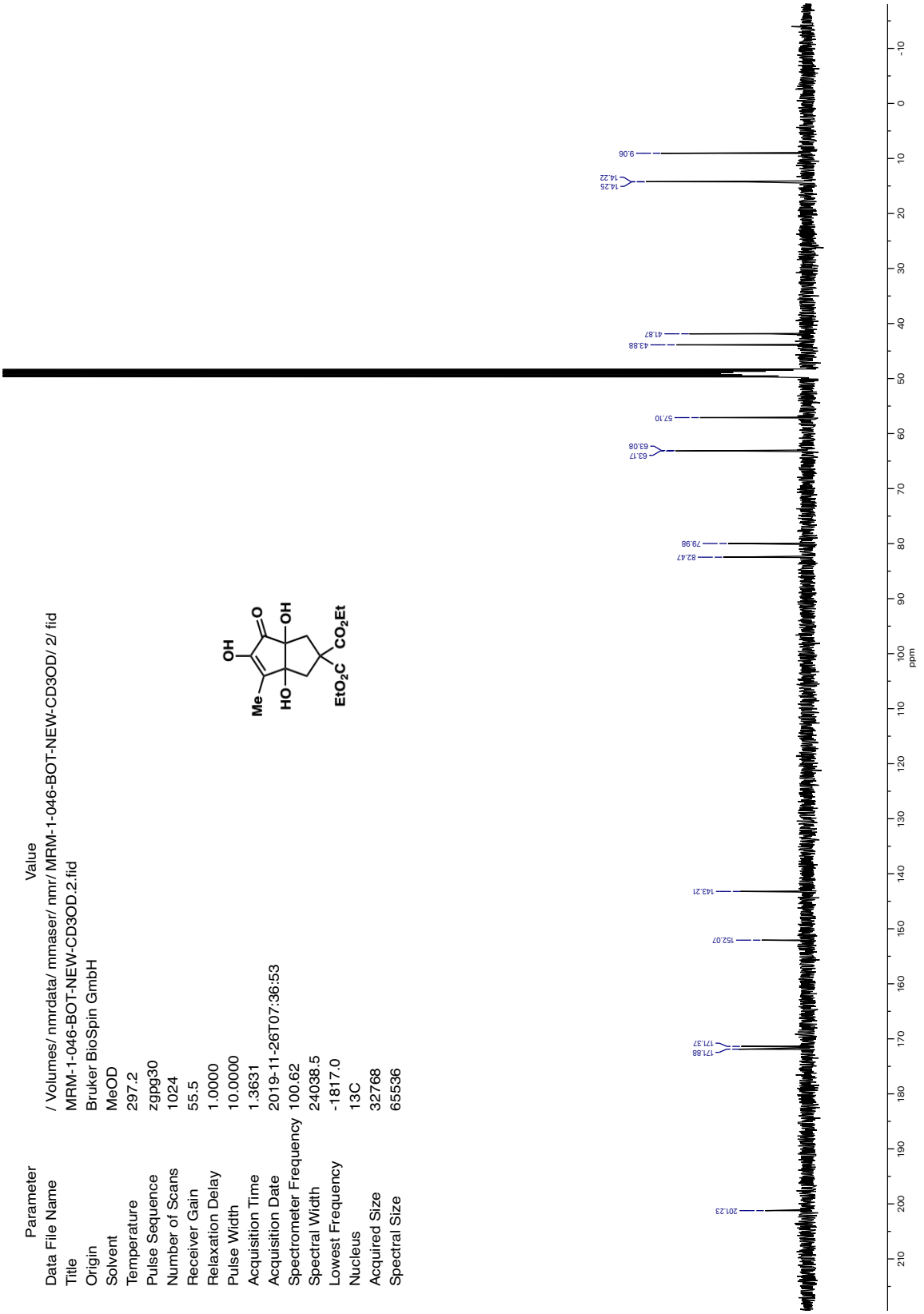

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-under-OBn/ 3/ fid
Title SED-1-under-OBn.3.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 127.1
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-25T18:47:57
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
H
HO
Me
BnO
nhex
5.31
0.53
0.97
0.99
1.00
1.06
0.93
1.04
0.95
0.96
5.96
5.58
3.00
3.50
ppm

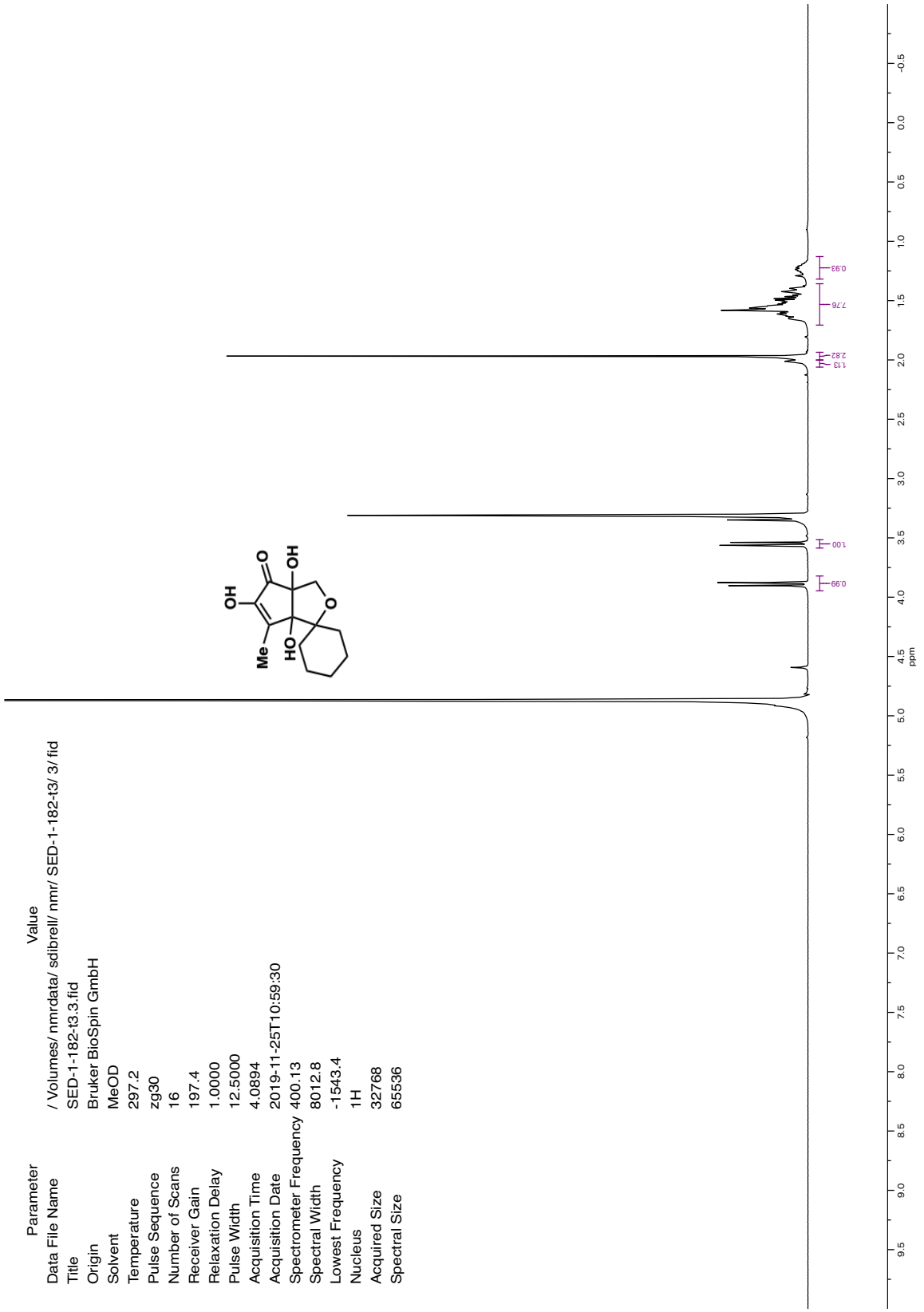

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-under-OBn/ 4/ fid
Title SED-1-under-OBn.4.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-25T19:30:26
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.6
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
H
HO
Me
BnO
nhex
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
202.81
149.56
145.27
138.17
128.69
127.98
127.70
87.38
84.14
65.43
56.15
35.06
31.73
30.07
27.66
26.78
22.75
22.69
19.59
14.25
ppm

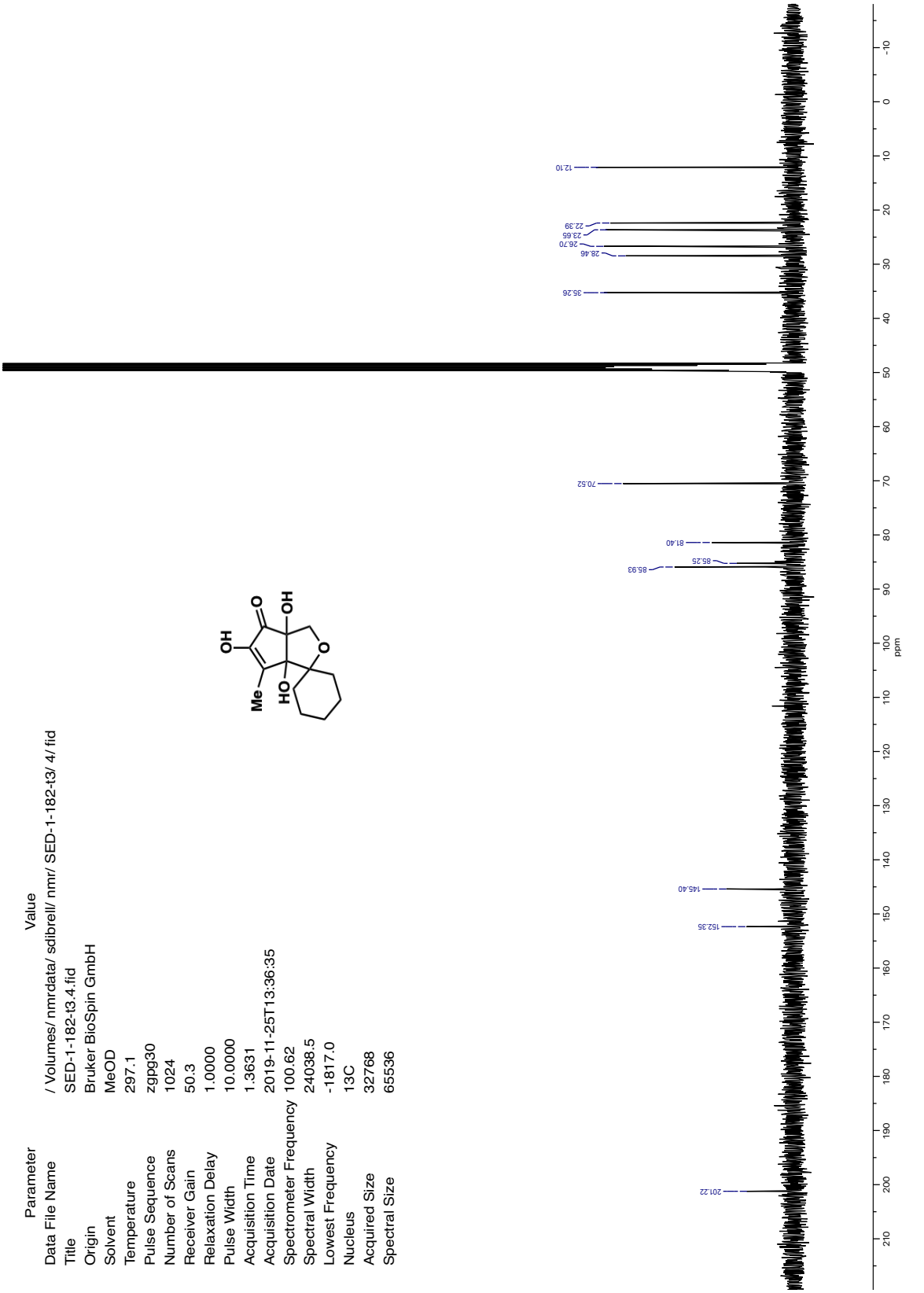

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-169-E-MeOD/ 1/ fid
Title SED-1-169-E-MeOD.1.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 296.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 176.8
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-05-09T23:10:08
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1543.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
OH
Me
HO
iPrO2C
CO2iPr
0.96
1.12
1.09
1.12
1.02
1.07
3.00
12.94
ppm

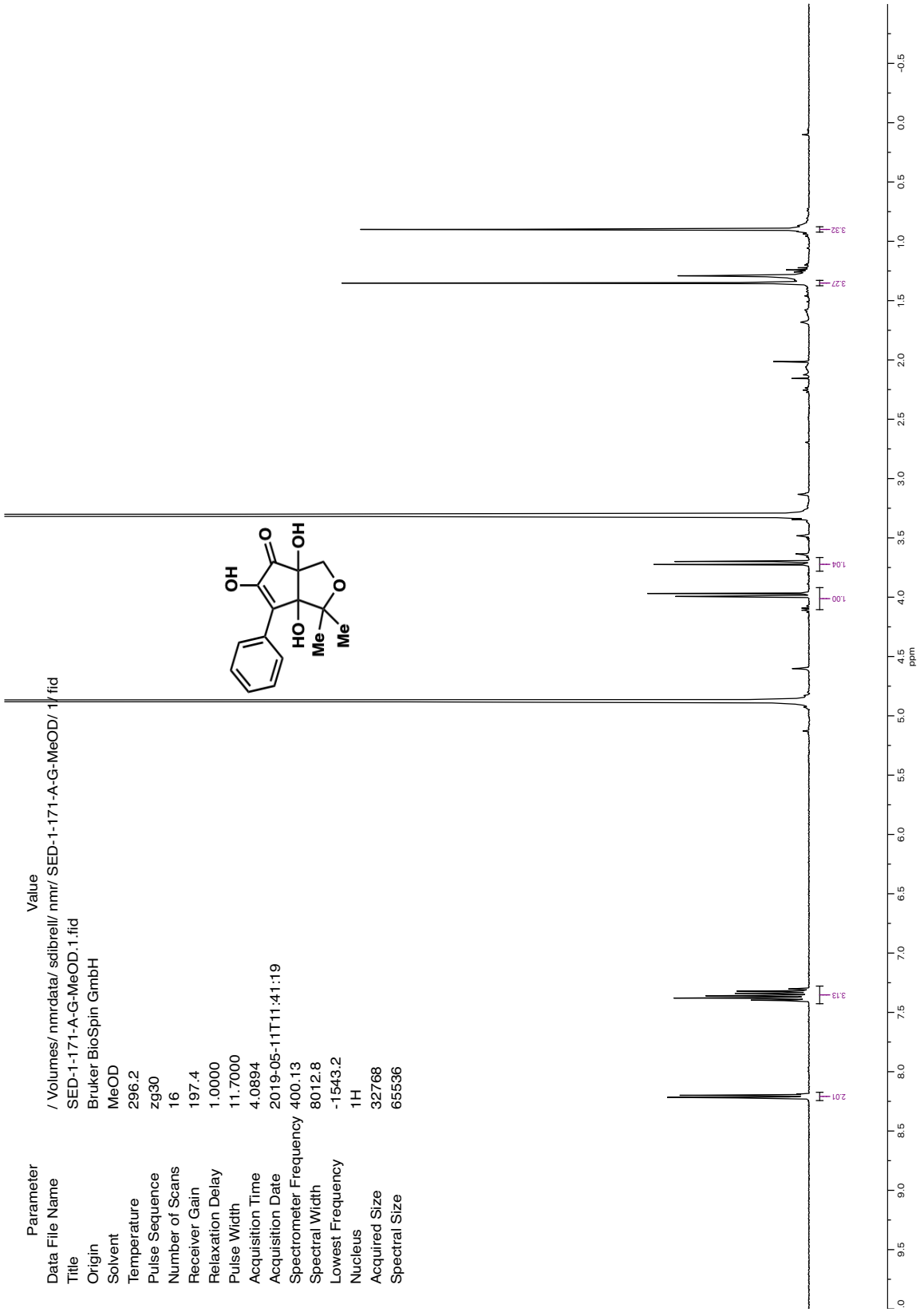

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-144-B-H/ 1/ fid
Title SED-1-144-B-H.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-04-23T04:44:13
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1934.8
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
OH
Me
HO
iPrO2C
CO2iPr
198.08
172.12
169.27
147.56
142.77
83.23
79.99
70.64
69.36
59.43
42.99
42.03
21.74
21.68
9.58
ppm

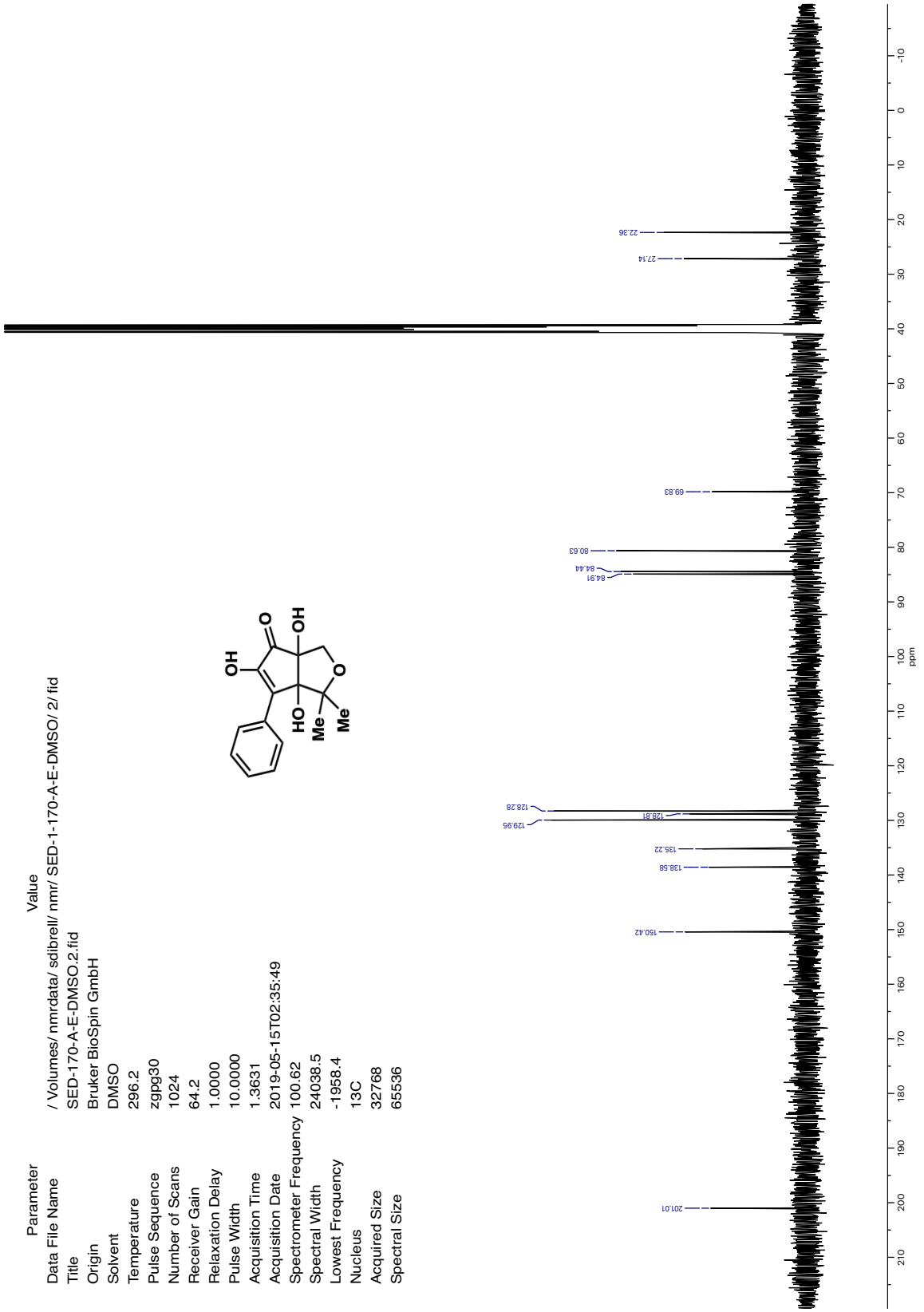

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-046-BOT-NEW-CD3OD/ 1/ fid
Title MRM-1-046-BOT-NEW-CD3OD.1.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 297.1
Pulse Sequence zg30
Number of Scans 64
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-26T06:55:01
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1543.0
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
OH
Me
HO
EtO2C
CO2Et
5.83
2.94
1.00
1.07
2.03
1.99
2.07
ppm

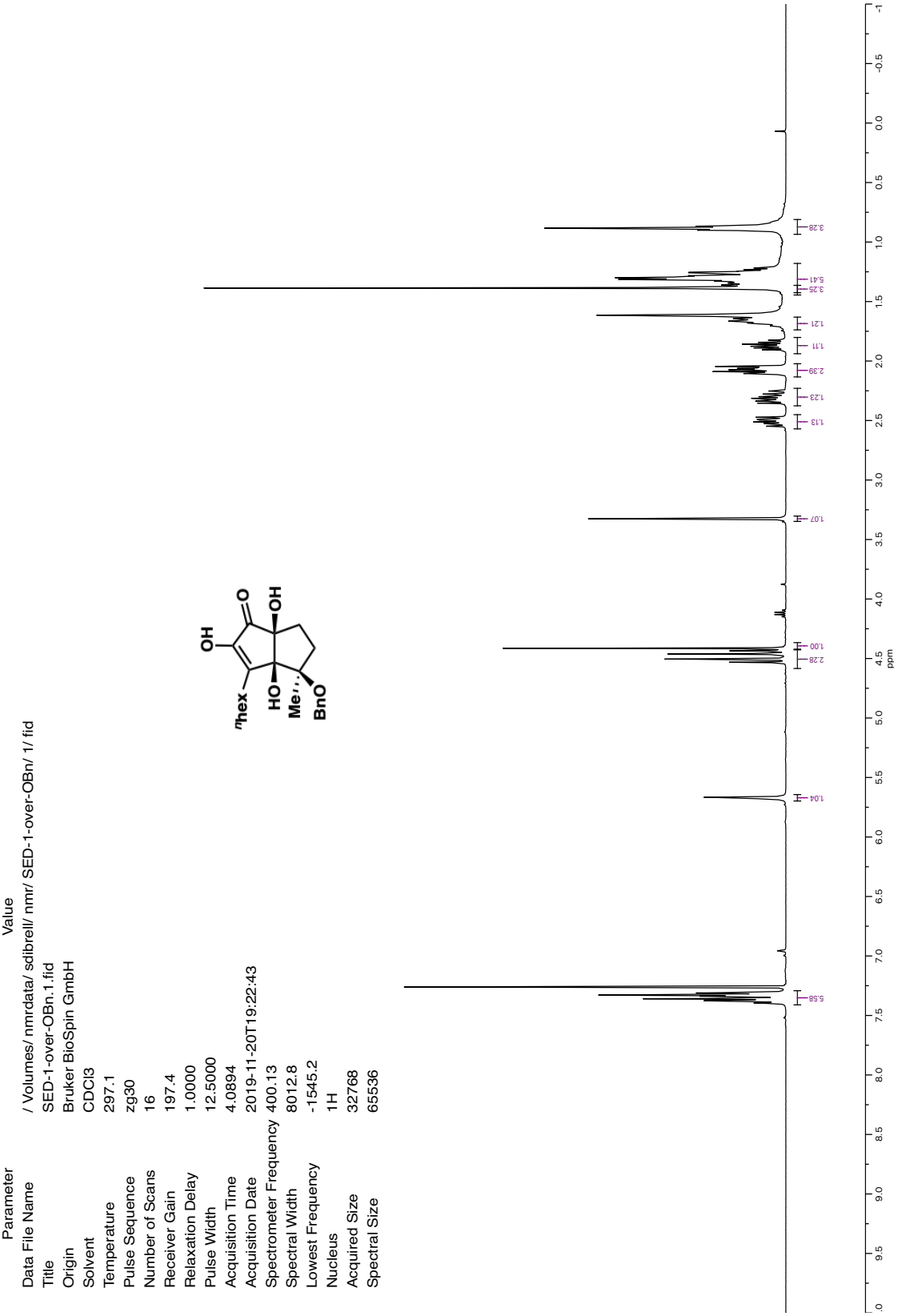

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-046-BOT-NEW-CD3OD/ 2/ fid
Title MRM-1-046-BOT-NEW-CD3OD.2.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-26T07:36:53
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1817.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
OH
Me
HO
EtO2C
CO2Et
9.06
14.22
14.25
41.87
43.88
57.10
63.08
63.17
79.98
82.47
143.21
152.07
171.37
171.88
201.23
ppm

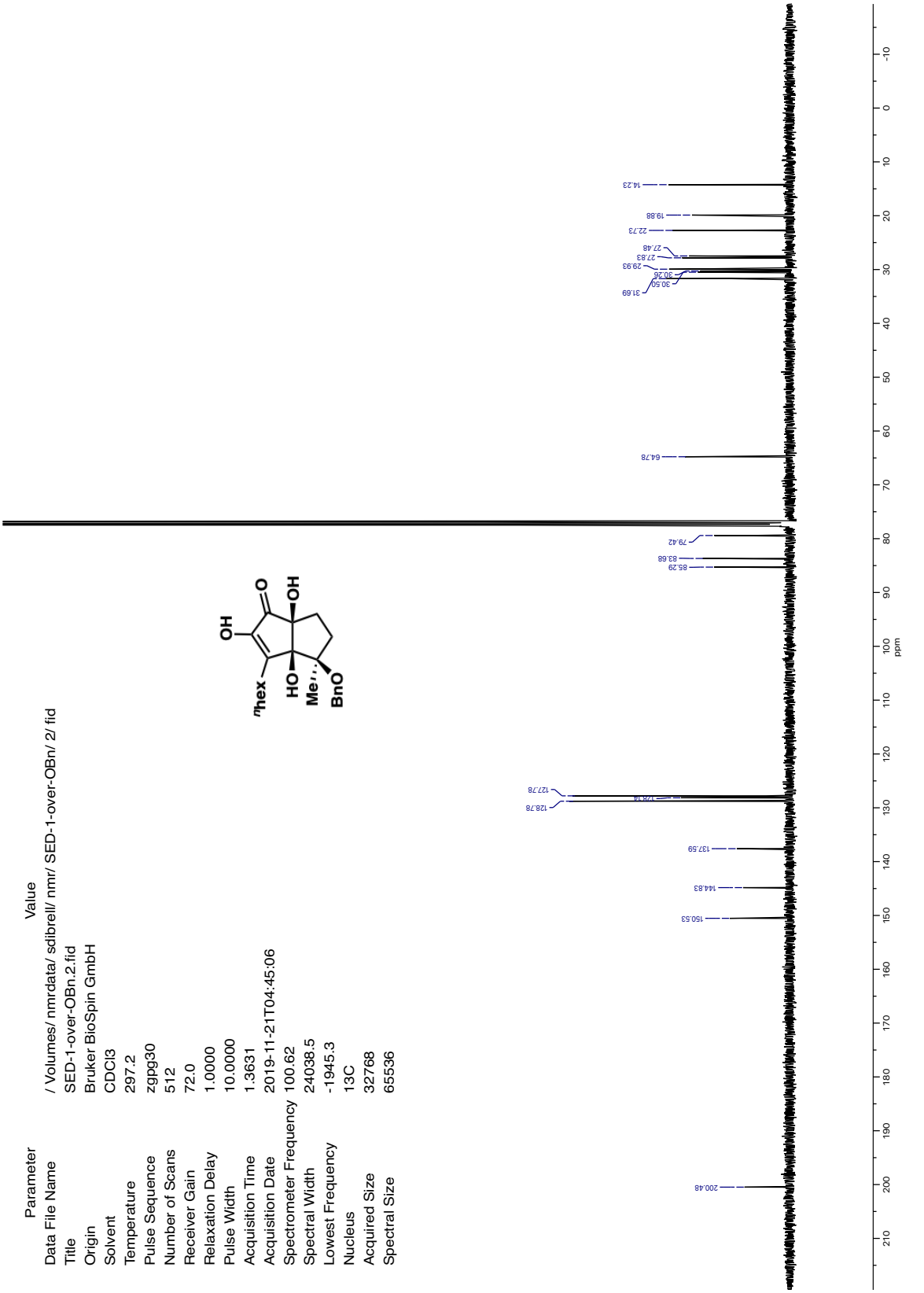

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-182-t3/ 3/ fid
Title SED-1-182-t3.3.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-25T10:59:30
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1543.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
OH
Me
HO
O
0.99
1.00
1.13
2.82
7.76
0.93
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
ppm

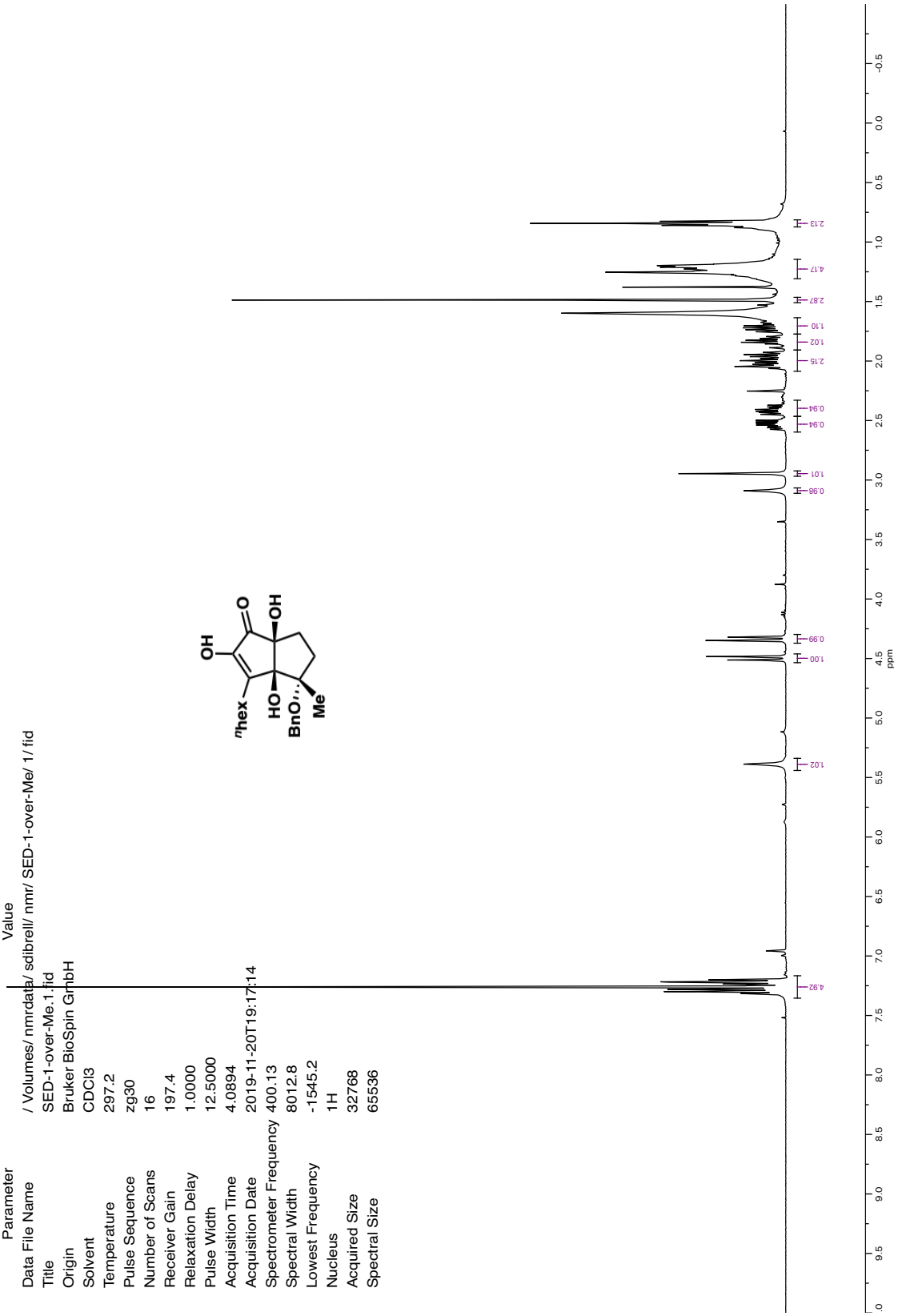

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-182-t3/ 4/ fid
Title SED-1-182-t3.4.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-25T13:36:35
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1817.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
OH
Me
HO
O
12.10
22.39
23.65
26.70
28.46
35.26
70.52
81.40
85.25
85.93
145.40
152.35
201.22
ppm

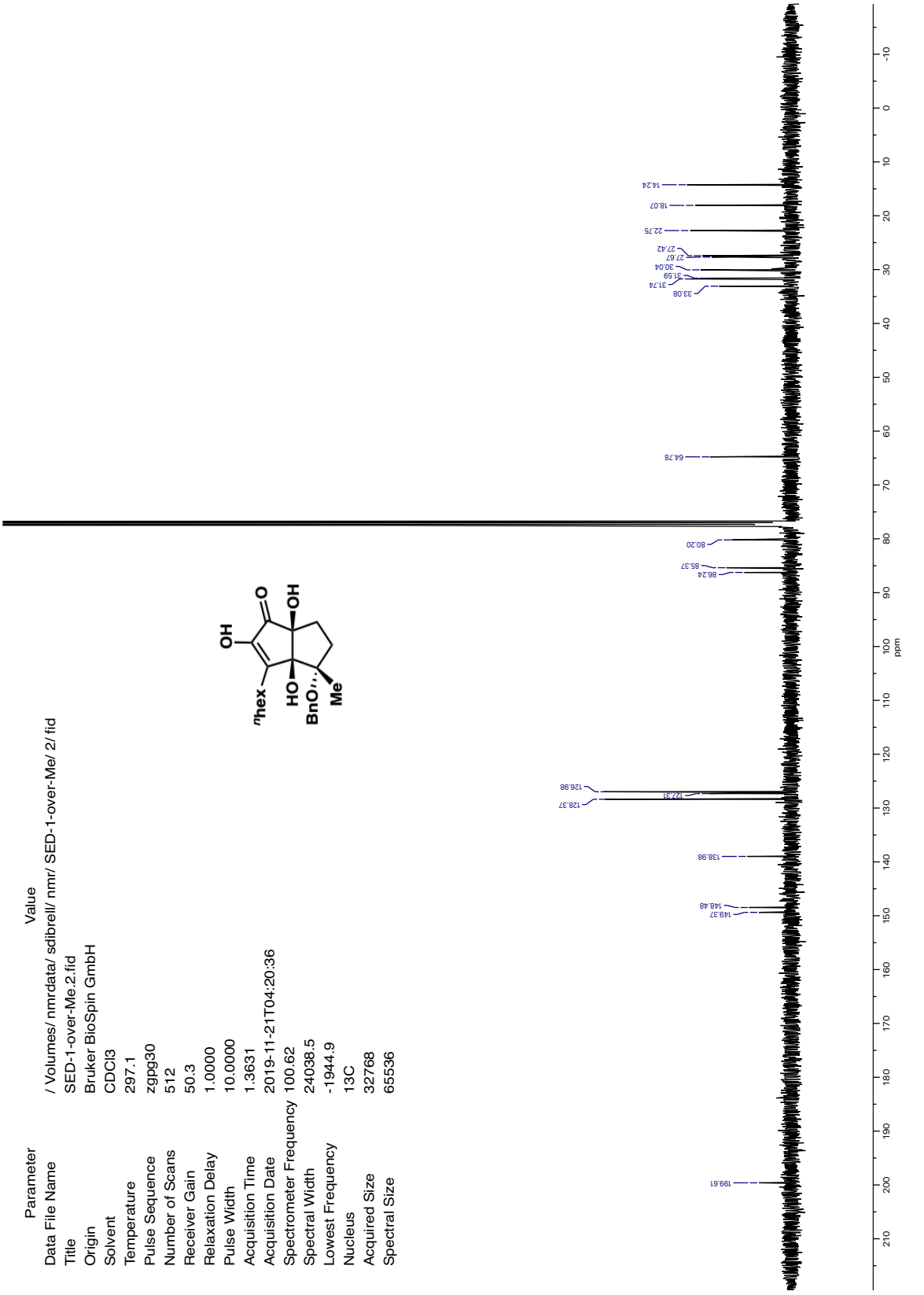

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-171-A-G-MeOD/ 1/ fid
Title SED-1-171-A-G-MeOD.1.fid
Origin Bruker BioSpin GmbH
Solvent MeOD
Temperature 296.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-05-11T11:41:19
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1543.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
OH
HO
Me
Me
O
2.01
3.13
1.00
1.04
3.27
3.32
ppm

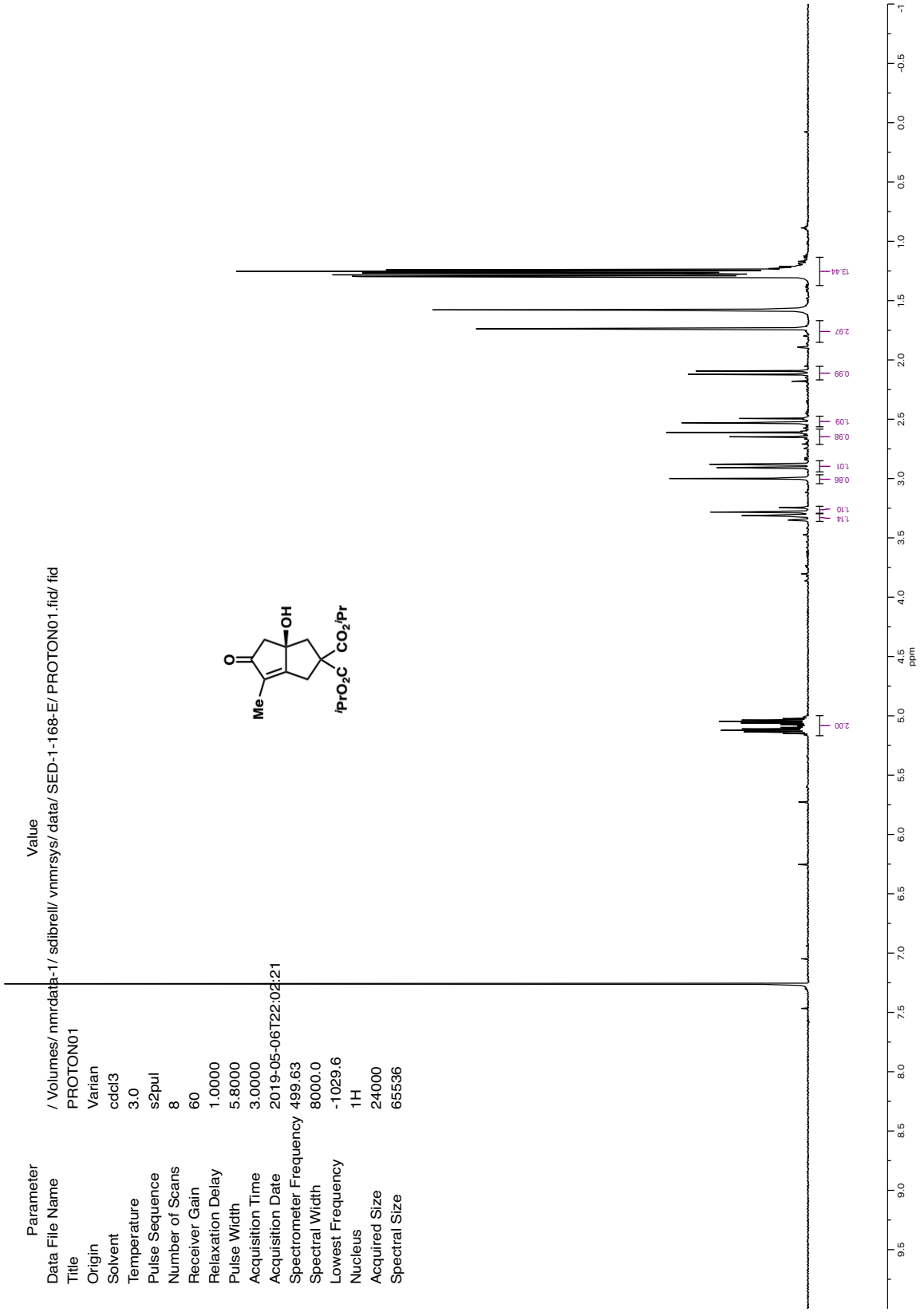

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-170-A-E-DMSO/ 2/ fid
Title SED-170-A-E-DMSO.2.fid
Origin Bruker BioSpin GmbH
Solvent DMSO
Temperature 296.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-05-15T02:35:49
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1958.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
OH
HO
Me
Me
O
22.36
27.14
69.83
80.63
84.44
84.91
128.28
128.81
129.95
135.22
138.58
150.42
201.01
ppm

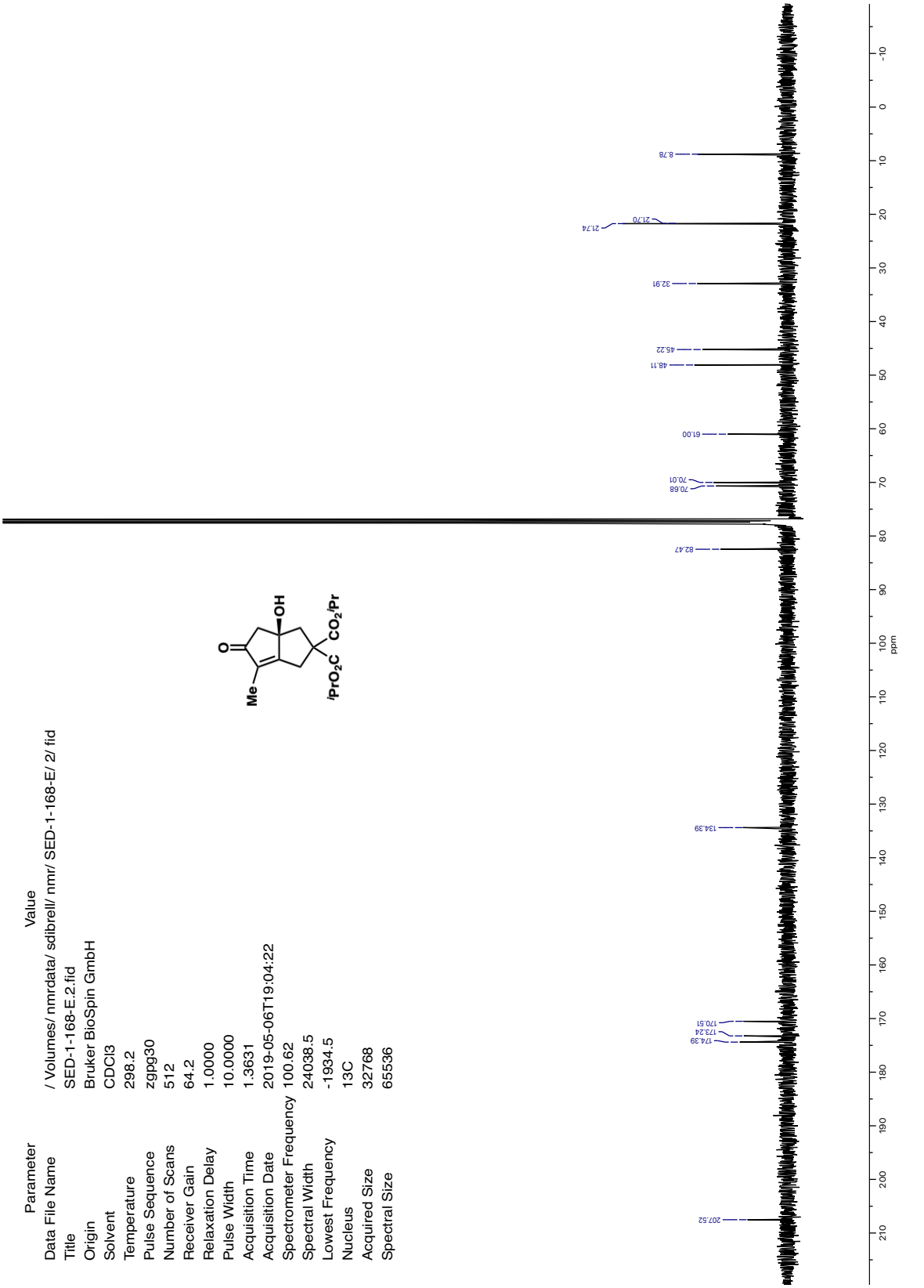

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-over-OBn/ 1/ fid
Title SED-1-over-OBn.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-20T19:22:43
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
OH
HO
Me
BnO
nhex
5.58
1.04
2.28
1.00
1.07
1.13
1.23
2.39
1.11
1.21
3.25
5.41
3.28
ppm

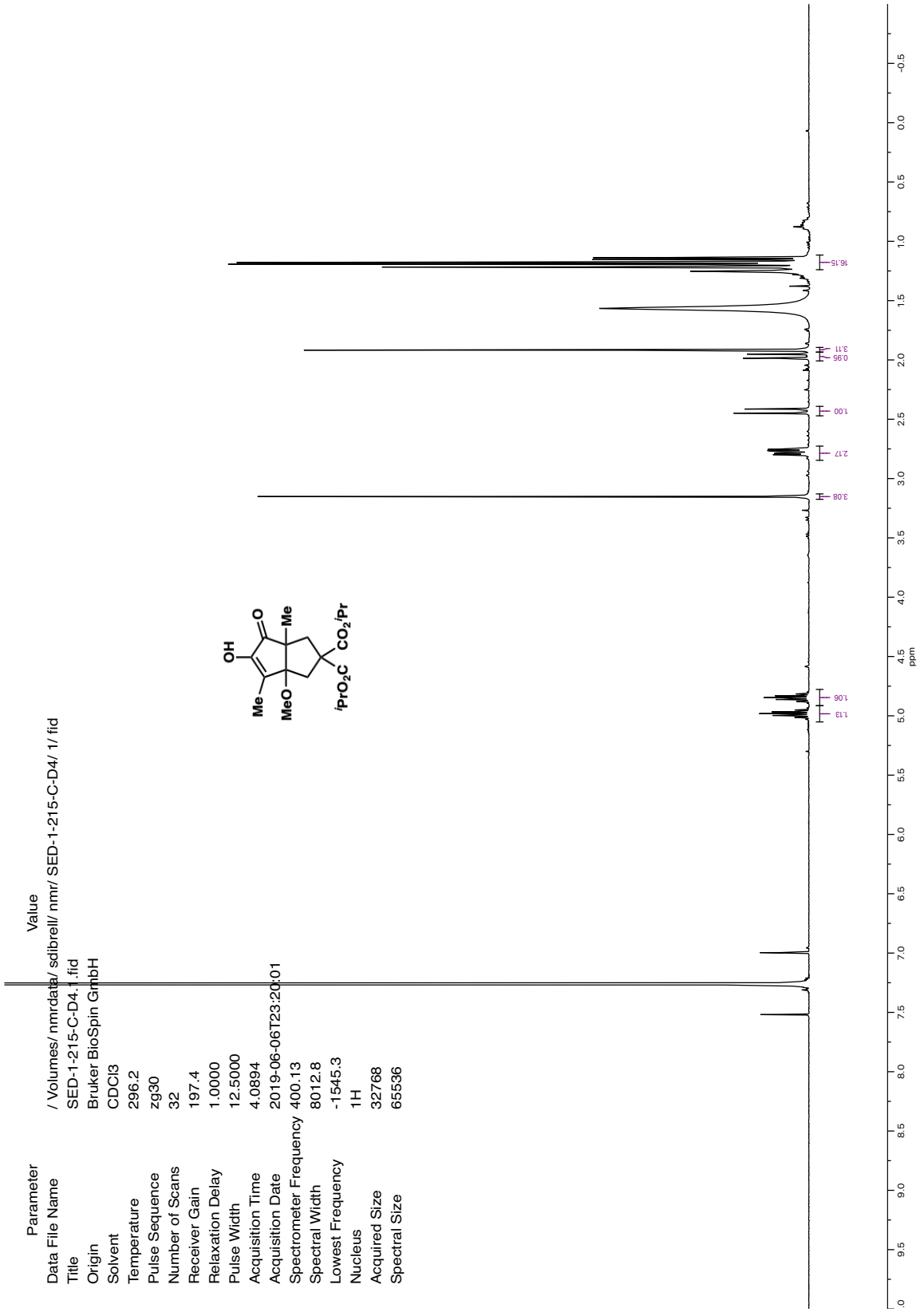

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-over-OBn/ 2/ fid
Title SED-1-over-OBn.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-21T04:45:06
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
OH
HO
Me
BnO
nhex
14.23
19.88
22.73
27.48
27.83
29.93
30.26
30.59
31.69
64.78
79.42
83.68
85.29
127.78
128.10
128.78
137.59
144.83
150.53
200.48
ppm

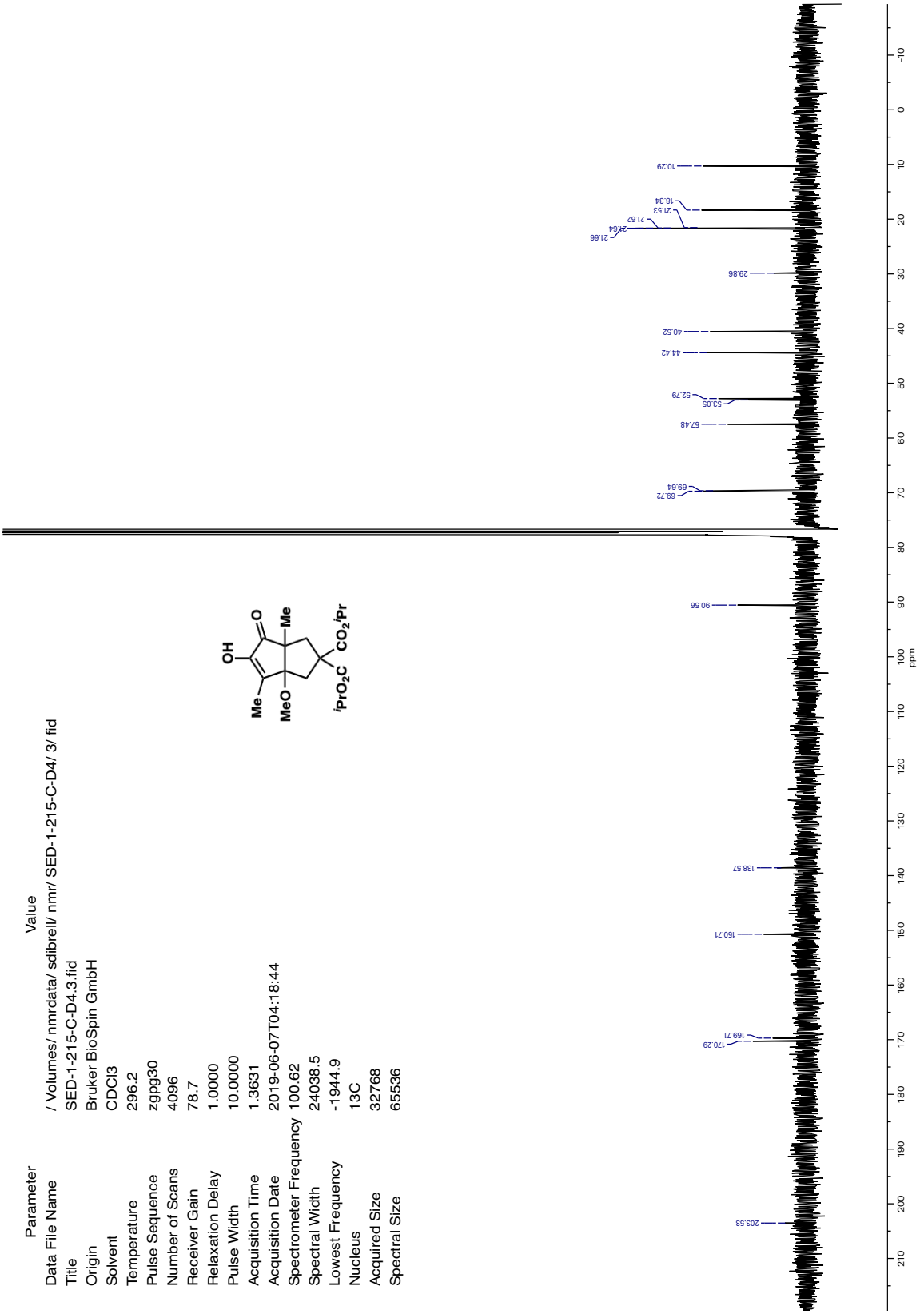

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-over-Me/ 1/ fid
Title SED-1-over-Me.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-20T19:17:14
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
OH
HO
BnO
Me
nhex
4.92
1.02
0.99
1.00
0.98
1.01
0.94
0.94
2.15
1.02
1.10
2.87
4.17
2.13
ppm

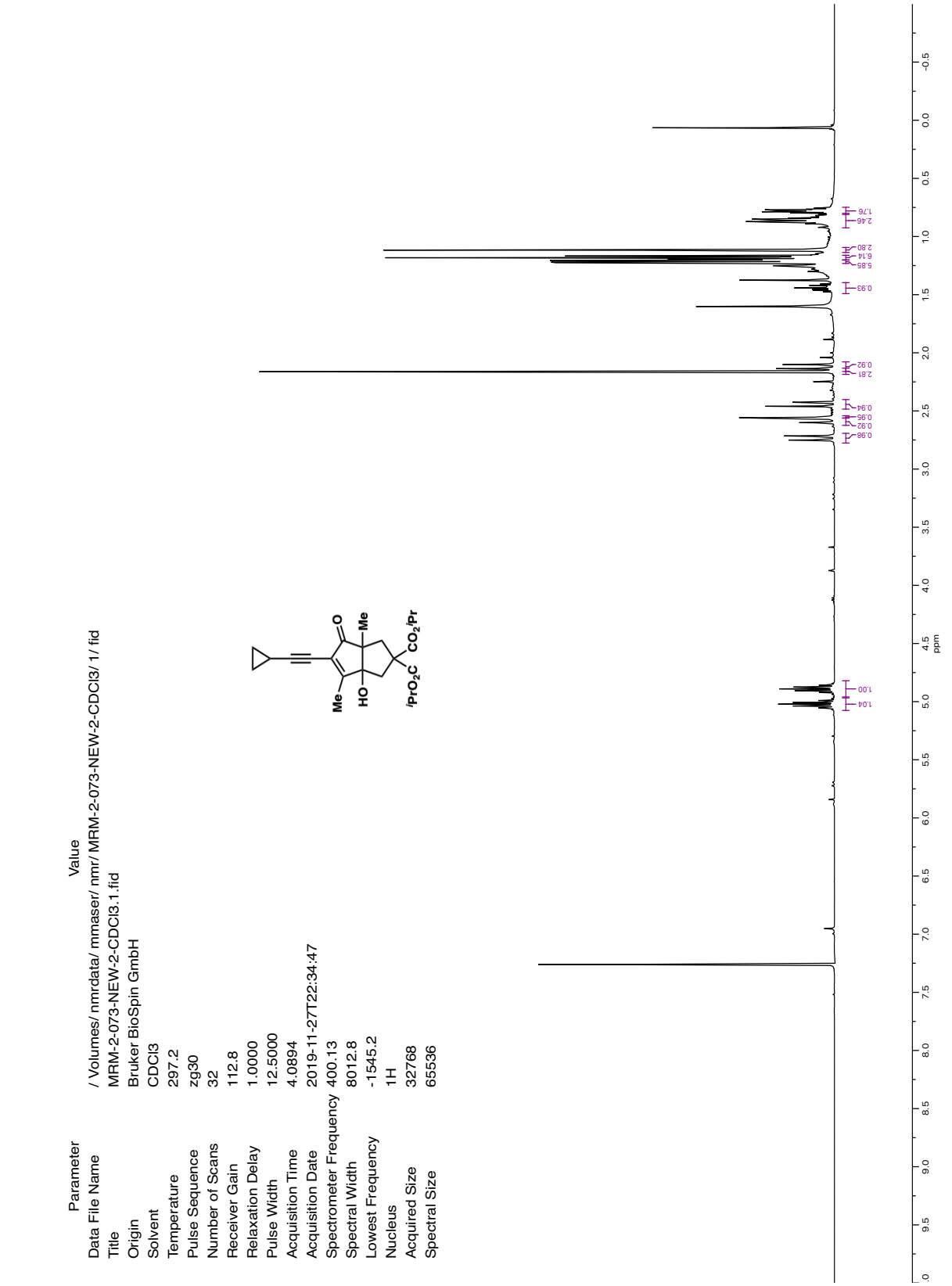

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-over-Me/ 2/ fid
Title SED-1-over-Me.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-21T04:20:36
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
OH
HO
BnO
Me
nhex
199.61
149.37
148.48
138.98
128.37
126.98
86.24
85.37
80.20
64.78
33.08
31.74
31.59
30.04
27.67
27.42
22.75
18.07
14.24
ppm

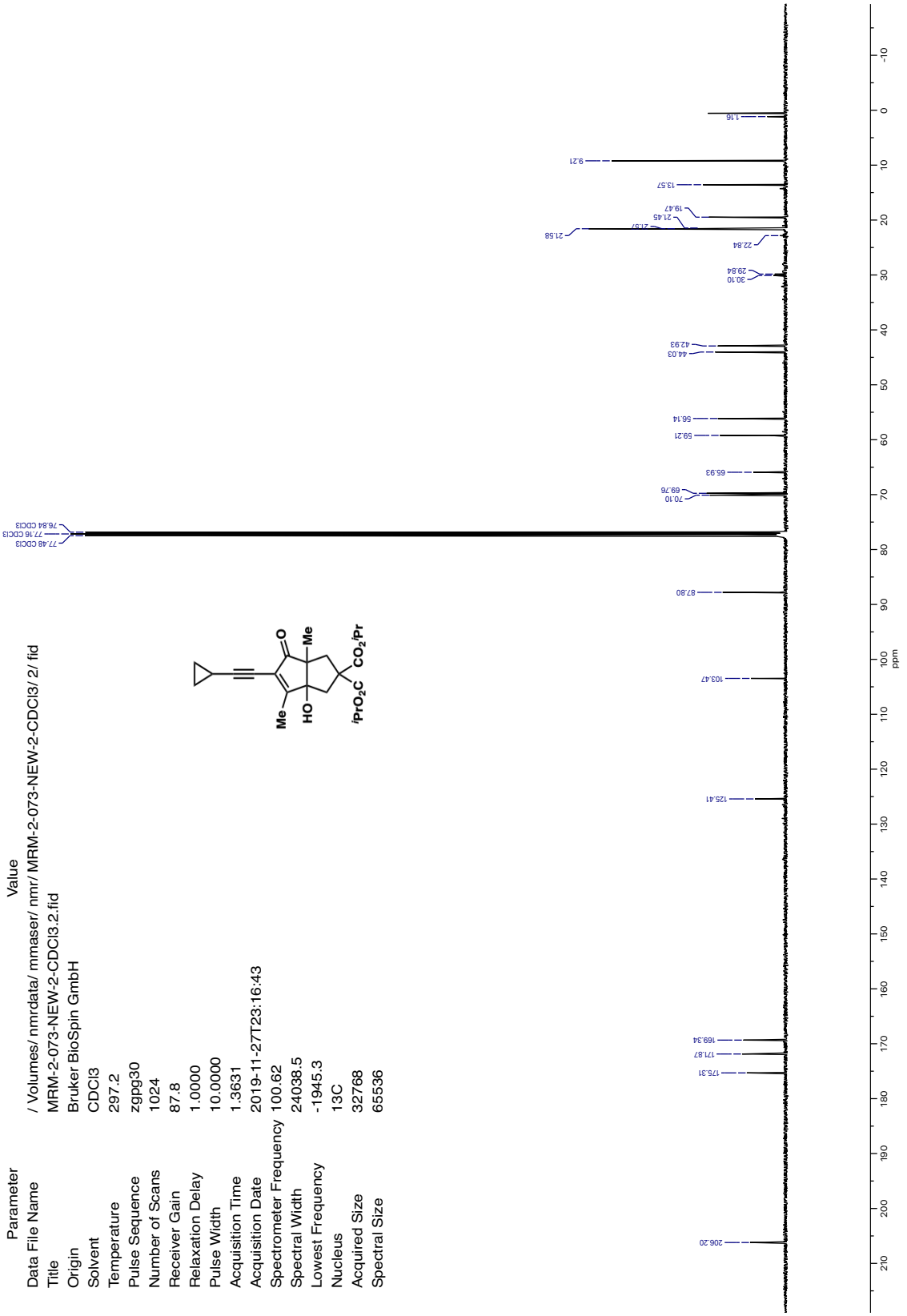

Parameter Value
Data File Name / Volumes/ nmrdata-1/ sdibrell/ vnmrsys/ data/ SED-1-168-E/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 3.0
Pulse Sequence s2pul
Number of Scans 8
Receiver Gain 60
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2019-05-06T22:04:21
Spectrometer Frequency 499.63
Spectral Width 8000.0
Lowest Frequency -1029.6
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
O
OH
Me
iPrO2C
CO2iPr
13.44
2.97
0.99
1.09
0.98
1.01
0.86
1.14
1.10
2.00
ppm

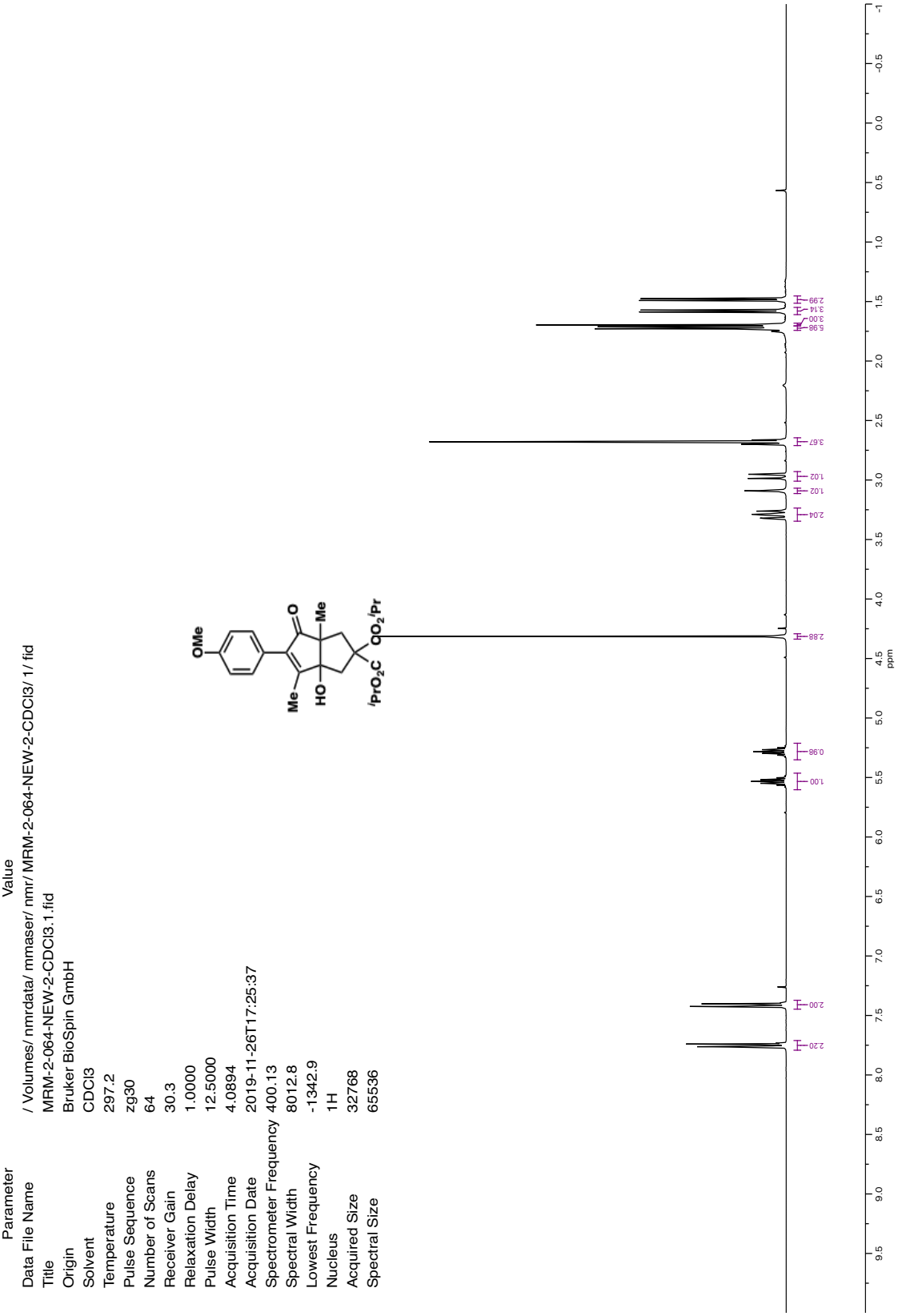

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-168-E/ 2/ fid
Title SED-1-168-E.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 298.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-05-06T19:04:22
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1934.5
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
O
OH
iPrO2C
CO2iPr
207.52
174.39
173.24
170.51
134.39
82.47
70.68
70.01
61.00
48.11
45.22
32.91
21.74
21.70
8.78
ppm

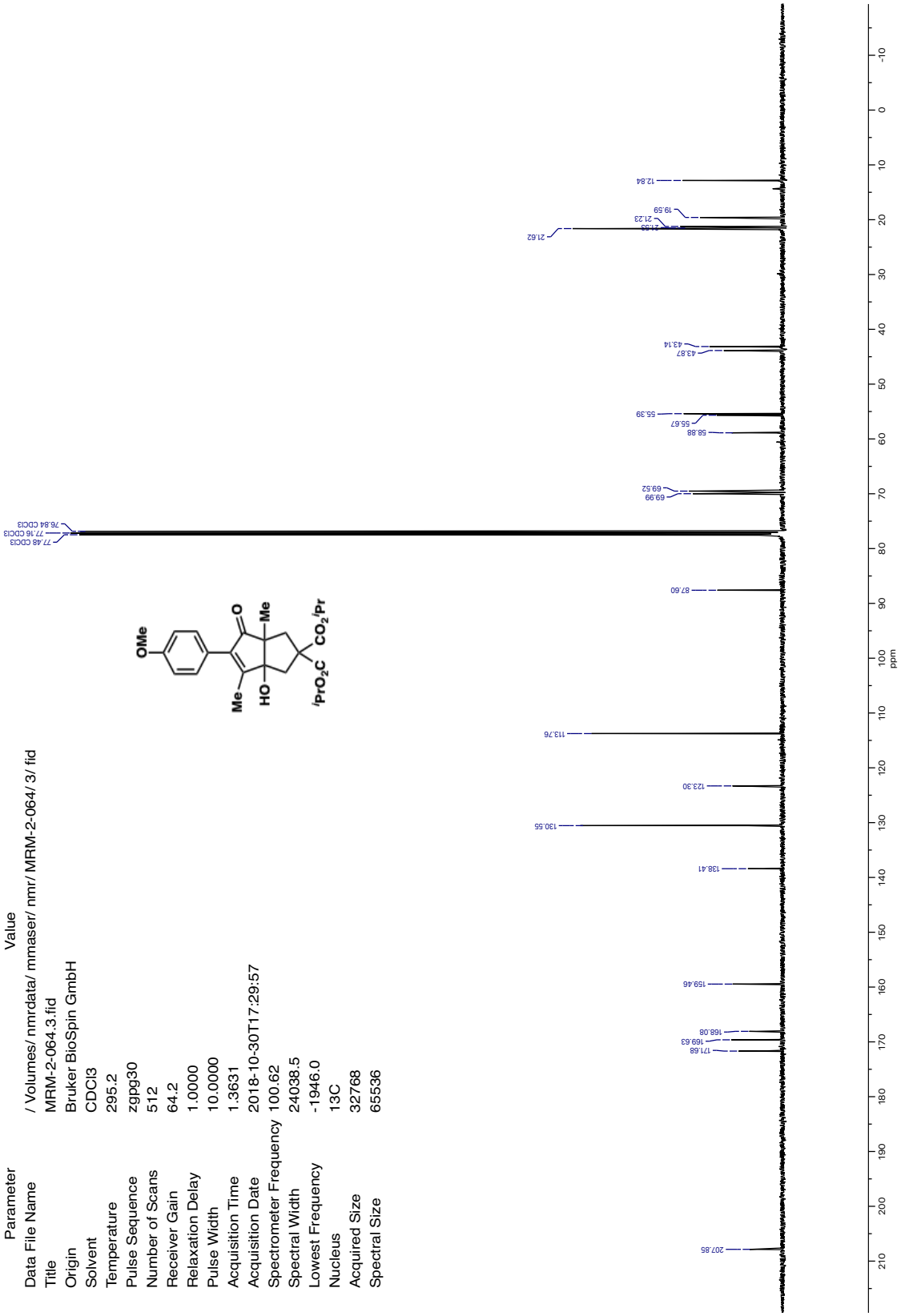

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-215-C-D4/ 1/ fid
Title SED-1-215-C-D4 1 .fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.2
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-06-06T23:20:01
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
Me
Me
MeO
iPrO2C
CO2iPr
16.15
3.11
0.95
1.00
2.17
3.08
1.06
1.13
ppm

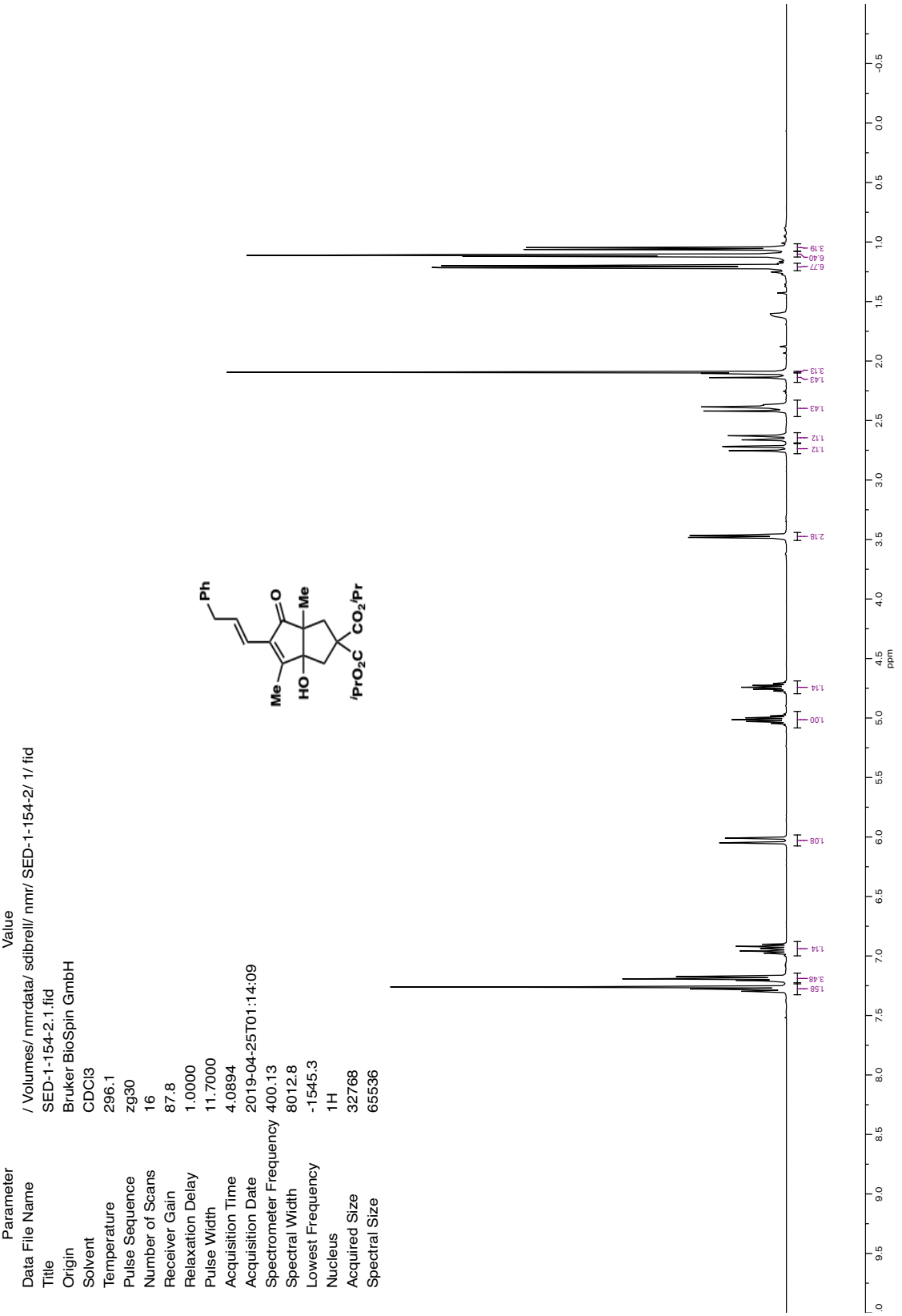

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-215-C-D4/ 3/ fid
Title SED-1-215-C-D4.3.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.2
Pulse Sequence zgpg30
Number of Scans 4096
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-06-07T04:18:44
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
O
Me
Me
MeO
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
10.29
18.34
21.53
21.62
21.64
21.66
29.86
40.52
44.42
52.79
53.05
57.48
69.64
69.72
90.56
138.57
150.71
169.71
170.29
203.53
ppm

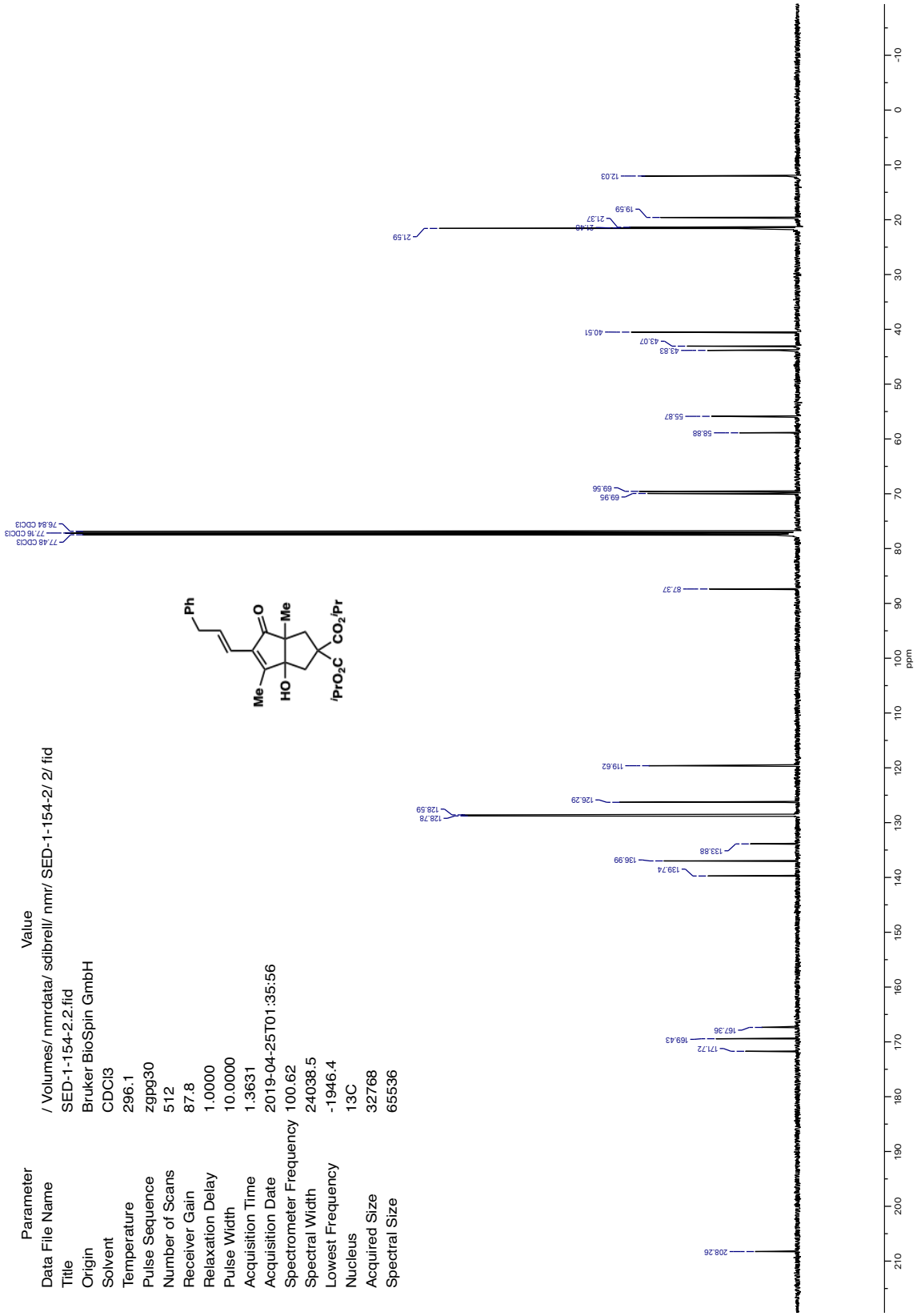

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-073-NEW-2-CDCl3/ 1/ fid
Title MRM-2-073-NEW-2-CDCl3.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 112.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-27T22:34:47
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
Me
HO
Me
iPrO2C
CO2iPr
1.00
1.04
0.98
0.92
0.95
0.94
2.81
0.92
0.93
5.85
6.14
2.80
2.46
1.76
ppm


Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-073-NEW-2-CDCl3/ 2/ fid
Title MRM-2-073-NEW-2-CDCl3.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 87.8
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-27T23:16:43
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
O
Me
HO
Me
iPrO2C
CO2iPr
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
206.20
175.31
171.87
169.34
125.41
103.47
87.80
70.10
69.76
65.93
59.21
56.14
44.03
42.93
30.10
29.84
22.84
21.58
21.57
21.45
19.47
13.57
9.21
1.16
ppm

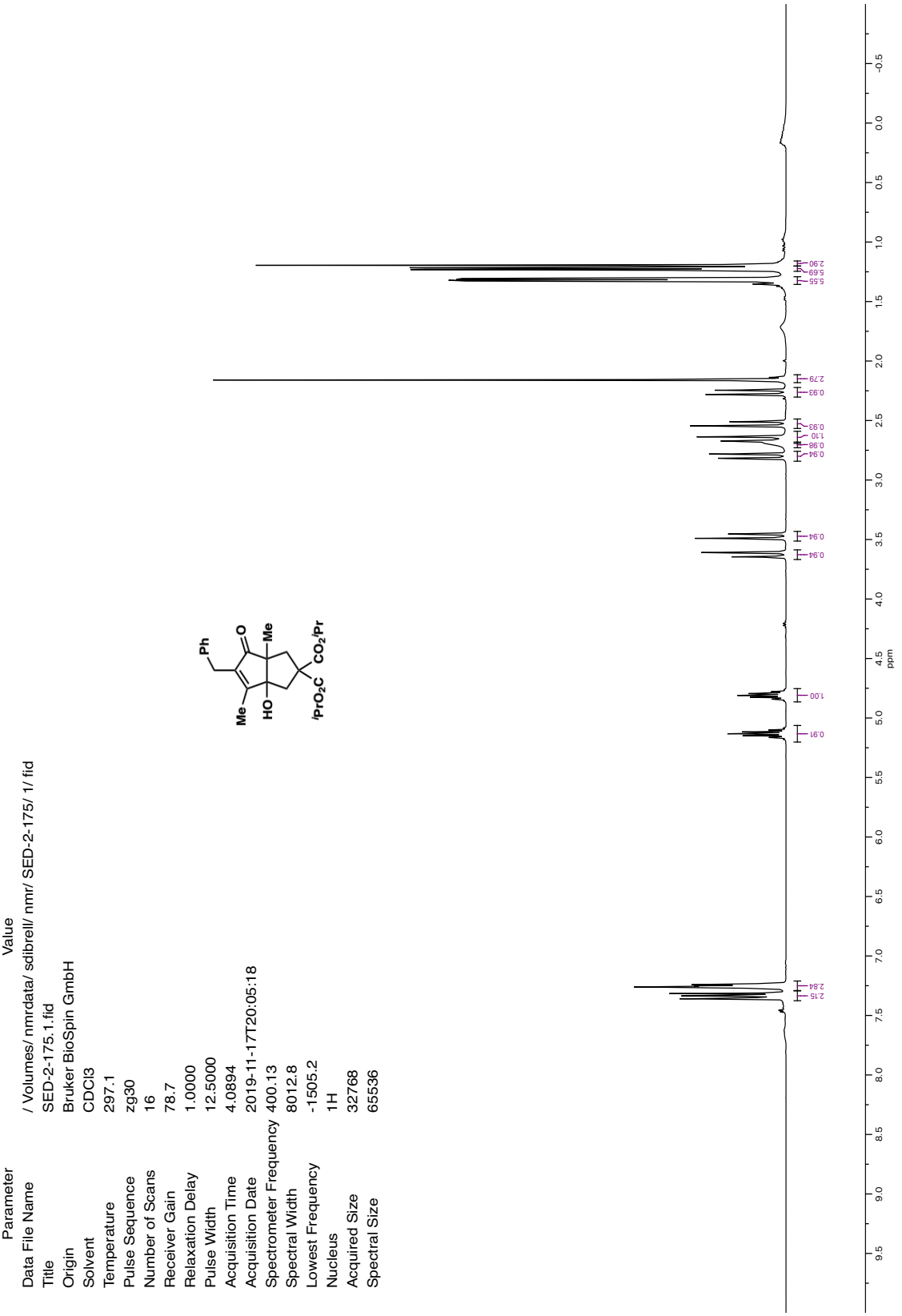

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-064-NEW-2-CDCl3/ 1/ fid
Title MRM-2-064-NEW-2-CDCl3.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 64
Receiver Gain 30.3
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-26T17:25:37
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1342.9
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OMe
O
Me
Me
HO
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
2.20
2.00
1.00
0.98
2.88
2.04
1.02
1.02
3.67
2.99
3.14
3.00
5.98
ppm

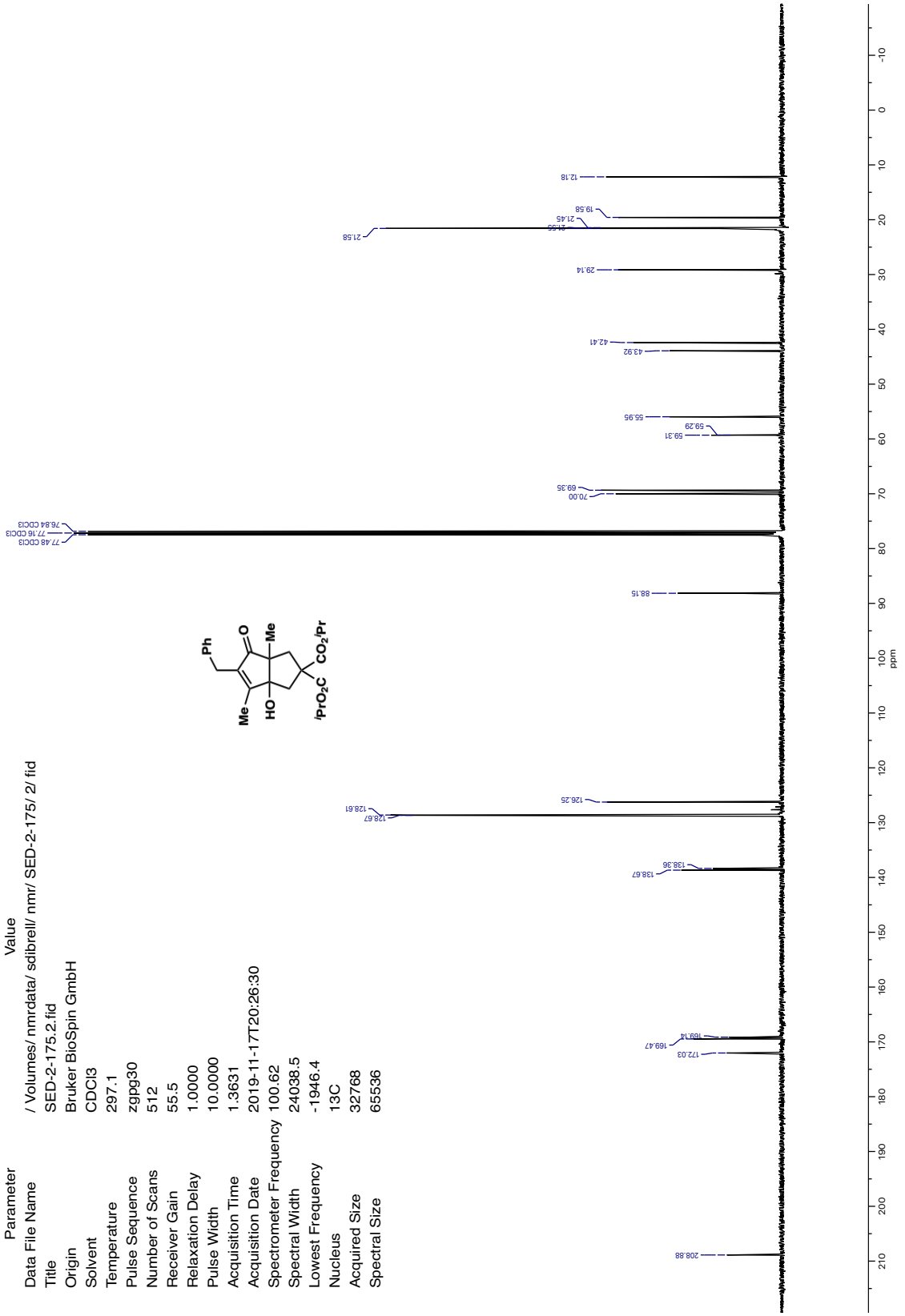

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-064/ 3/ fid
Title MRM-2-064.3.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-10-30T17:29:57
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1946.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OMe
O
Me
Me
HO
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
207.85
171.68
169.63
168.08
159.46
138.41
130.55
123.30
113.76
87.60
69.99
69.52
58.88
55.67
55.39
43.87
43.14
21.62
21.35
21.23
19.59
12.84
ppm

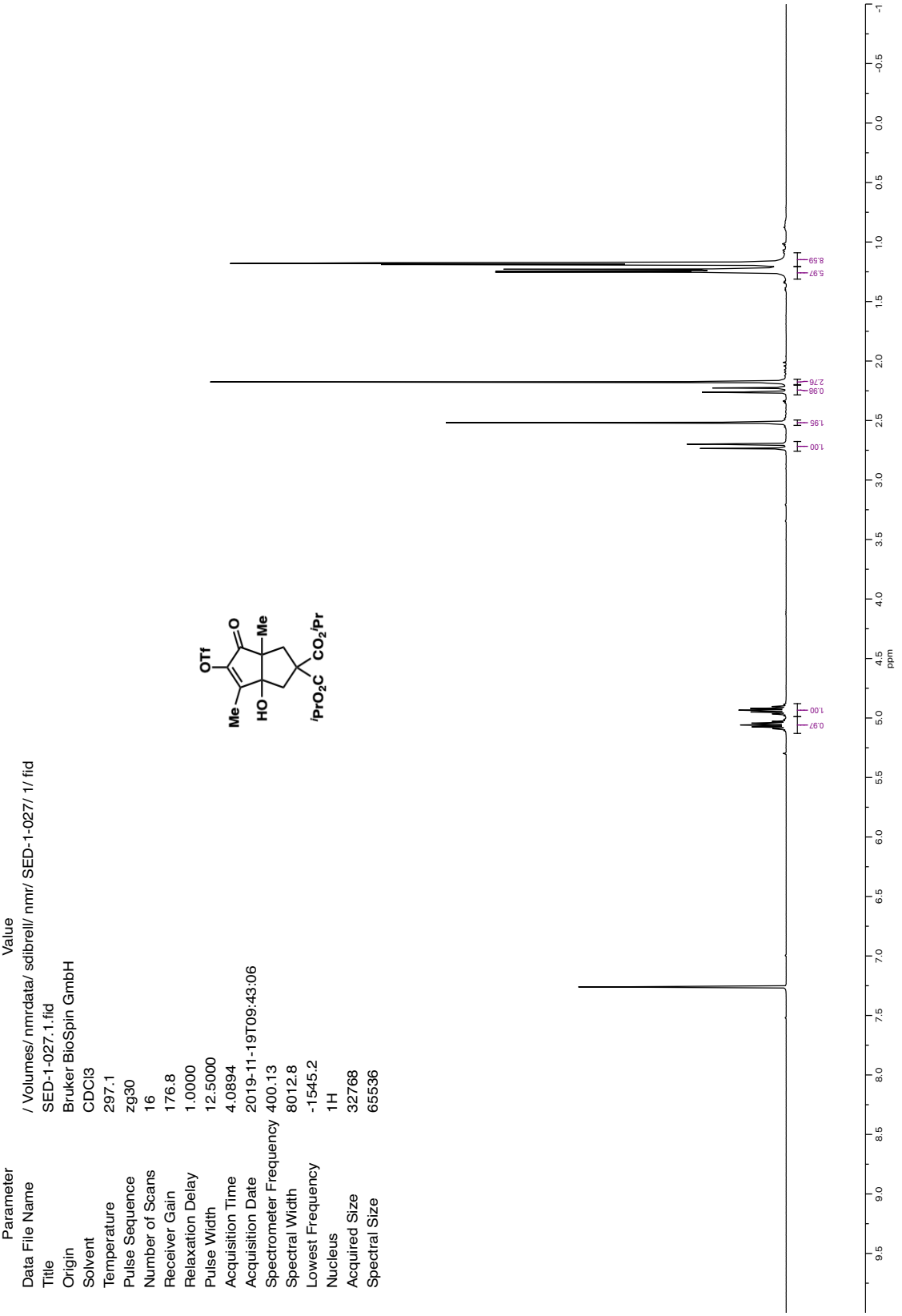

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-154-2/ 1/ fid
Title SED-1-154-2.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 87.8
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-04-25T01:14:09
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Ph
O
Me
Me
HO
iPrO2C
CO2iPr
1.58
3.48
1.14
1.08
1.14
1.00
2.18
1.12
1.43
1.43
3.13
6.77
6.40
3.19
ppm

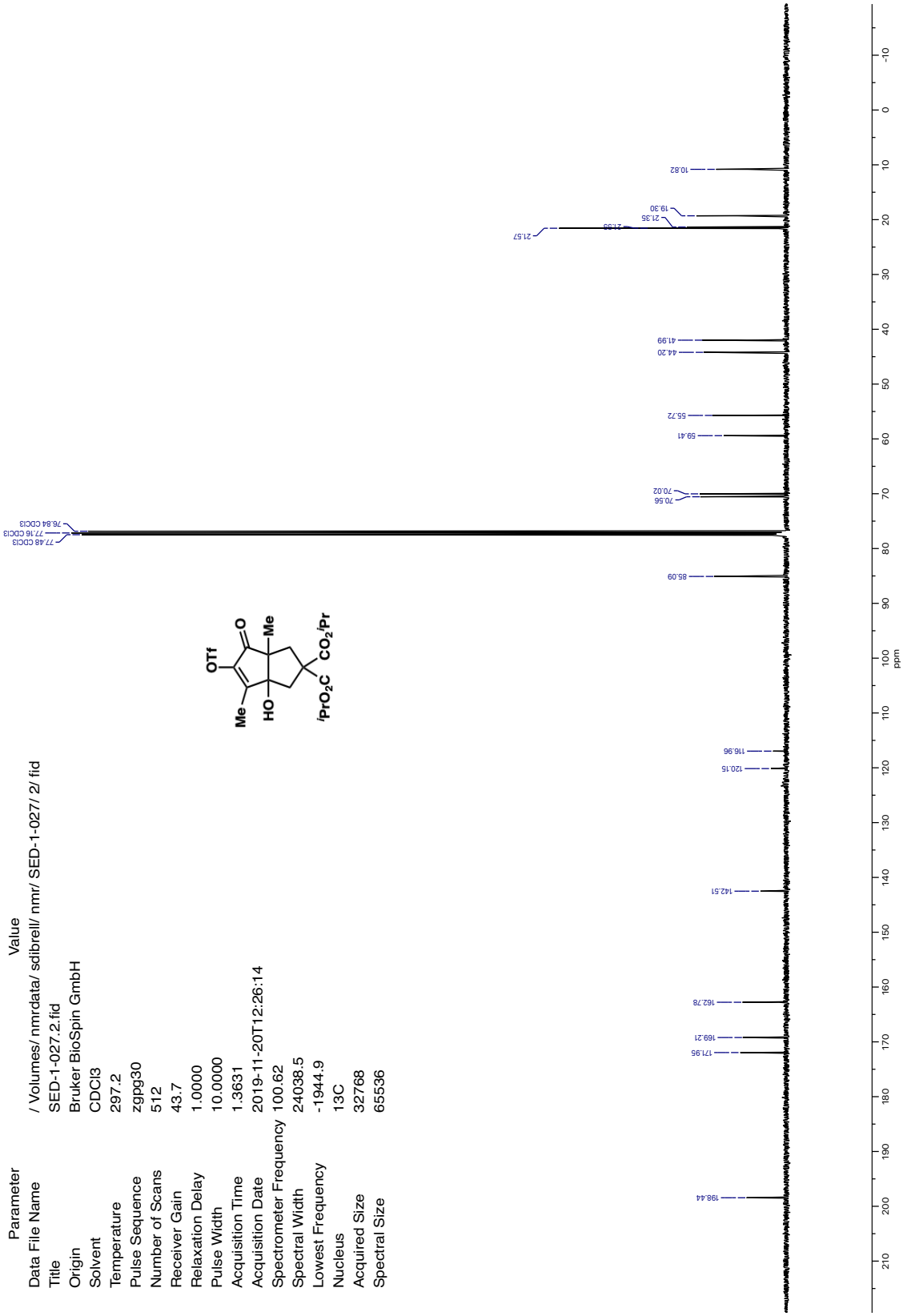


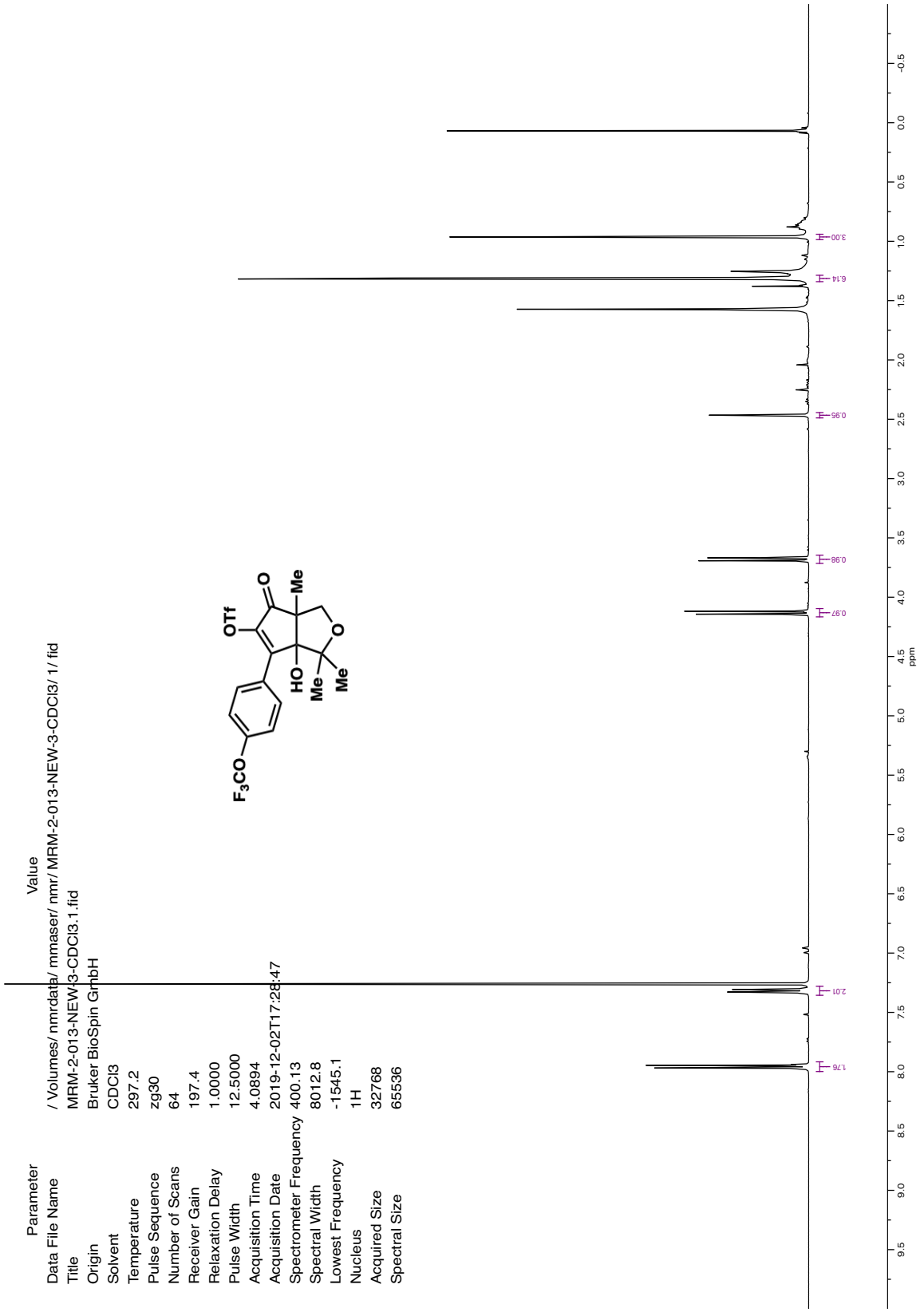
Parameter	Value
Data File Name	/ Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-154-2/ 2/ fid
Title	SED-1-154-2.2.fid
Origin	Bruker BioSpin GmbH
Solvent	CDCl3
Temperature	296.1
Pulse Sequence	zgpg30
Number of Scans	512
Receiver Gain	87.8
Relaxation Delay	1.0000
Pulse Width	10.0000
Acquisition Time	1.3631
Acquisition Date	2019-04-25T01:35:56
Spectrometer Frequency	100.62
Spectral Width	24038.5
Lowest Frequency	-1946.4
Nucleus	13C
Acquired Size	32768
Spectral Size	65536

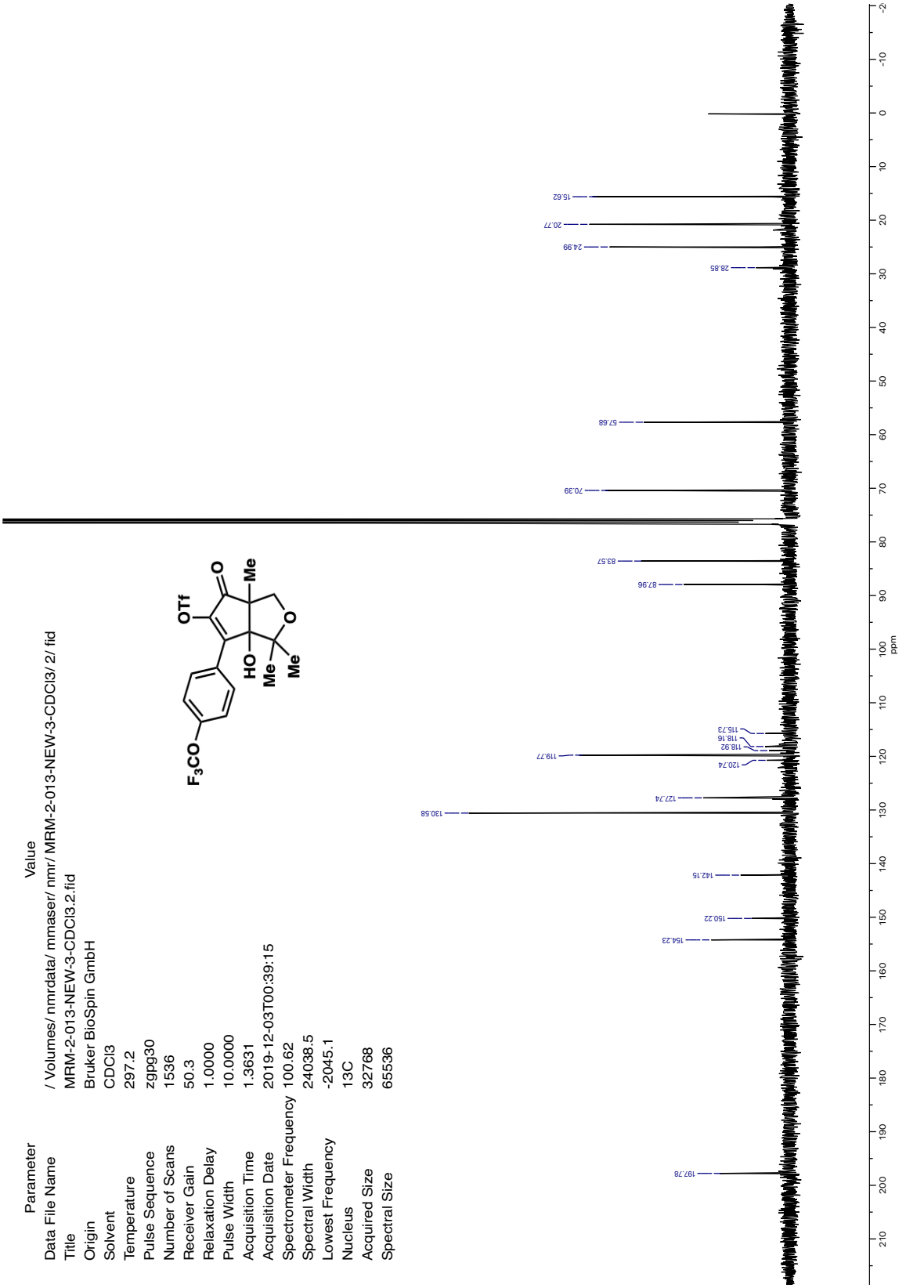

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-175/ 1/ fid
Title SED-2-175.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-17T20:05:18
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1505.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Ph
O
Me
Me
HO
iPrO2C
CO2iPr
2.15
2.84
0.91
1.00
0.94
0.94
0.94
0.98
1.10
0.93
0.93
2.79
2.90
5.69
5.55
ppm

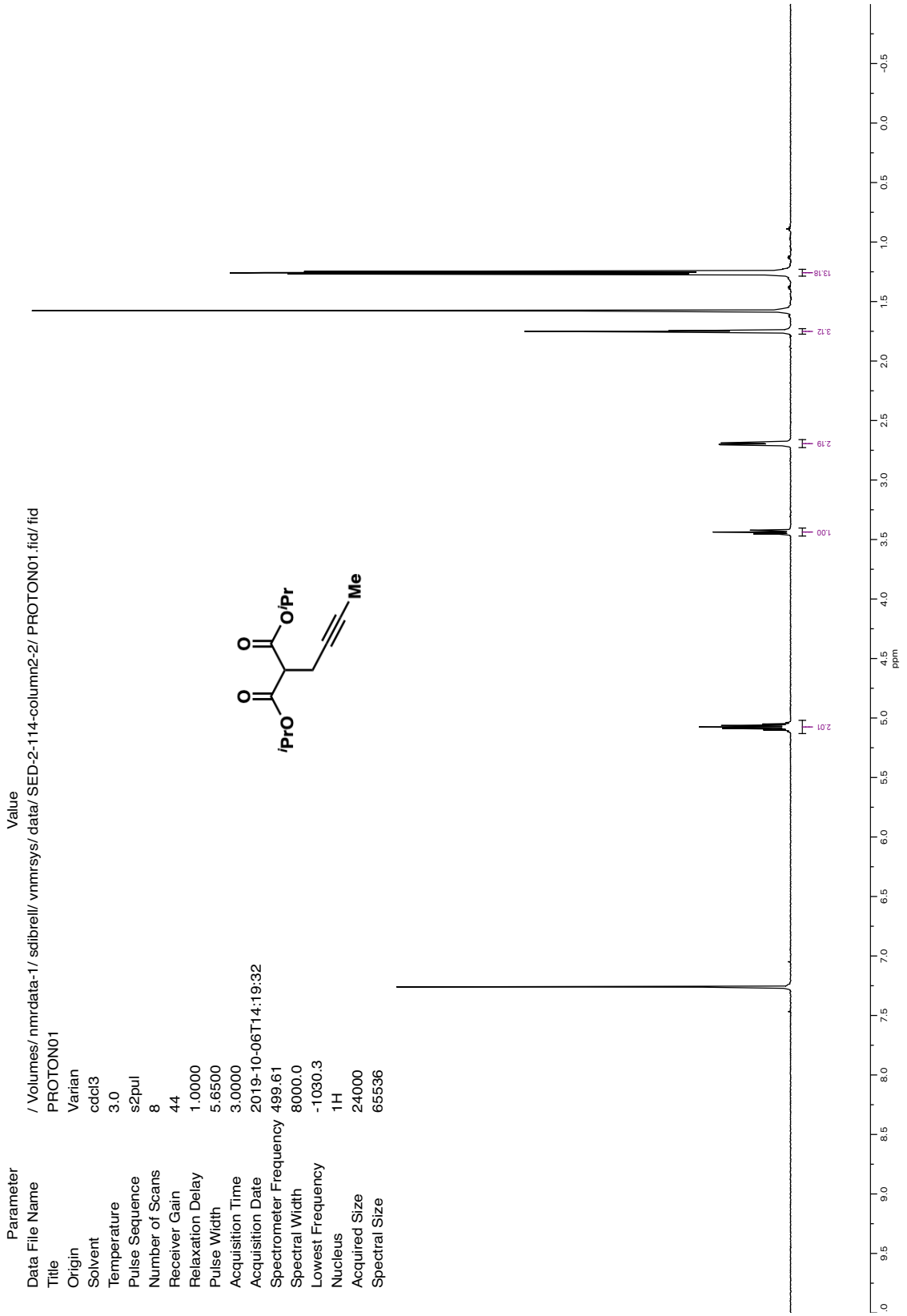

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-175/ 2/ fid
Title SED-2-175.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-17T20:26:30
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1946.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Ph
O
Me
Me
HO
iPrO2C
CO2iPr
208.88
172.03
169.74
169.47
138.67
138.36
128.67
128.61
126.25
88.15
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
70.00
69.35
59.31
59.29
55.95
43.92
42.41
29.14
21.58
21.55
21.45
19.58
12.18
ppm

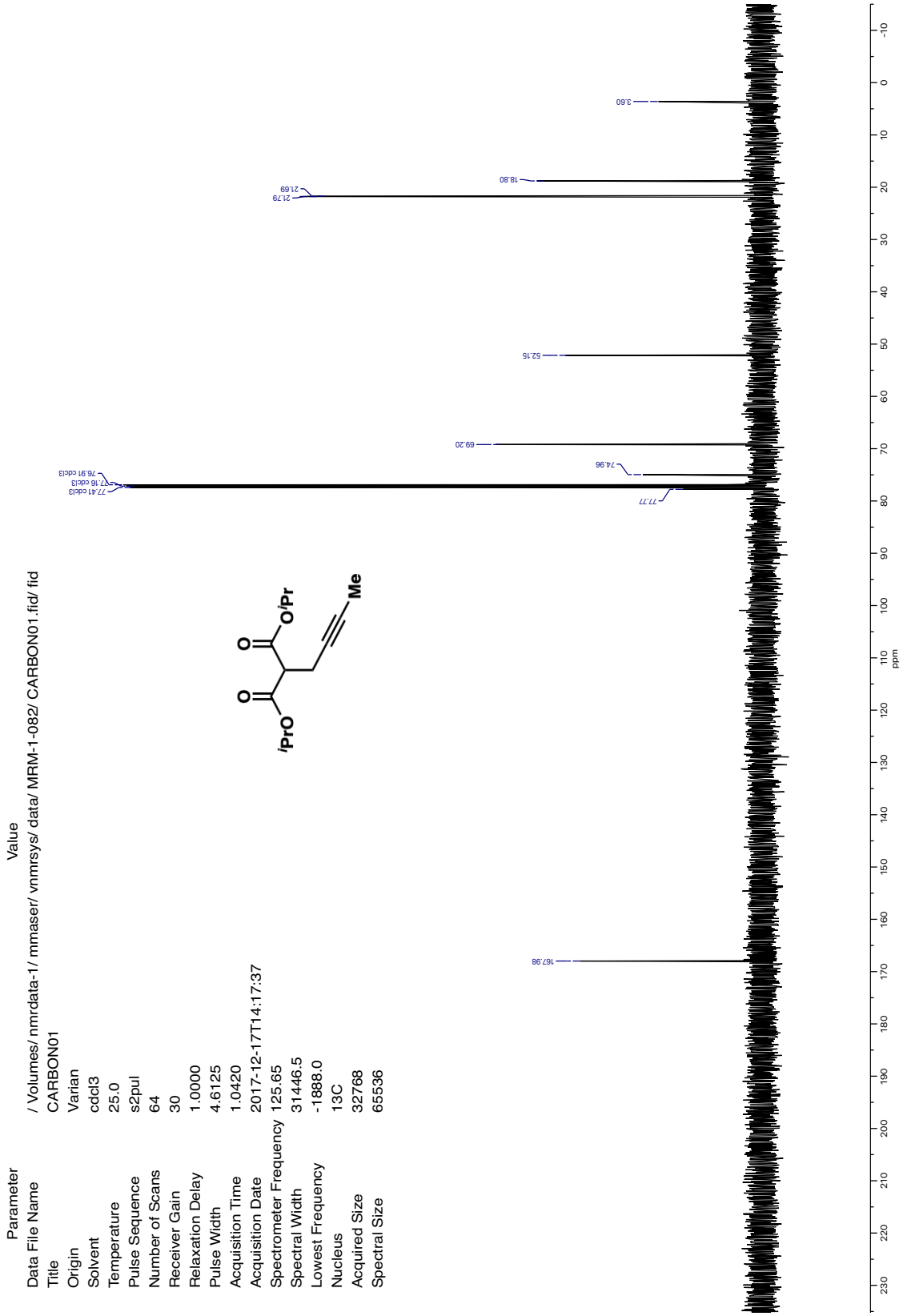

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-027/ 1/ fid
Title SED-1-027.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 176.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-19T09:43:06
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OTf
O
Me
Me
HO
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
8.59
5.97
2.78
0.98
1.95
1.00
1.00
0.97
ppm

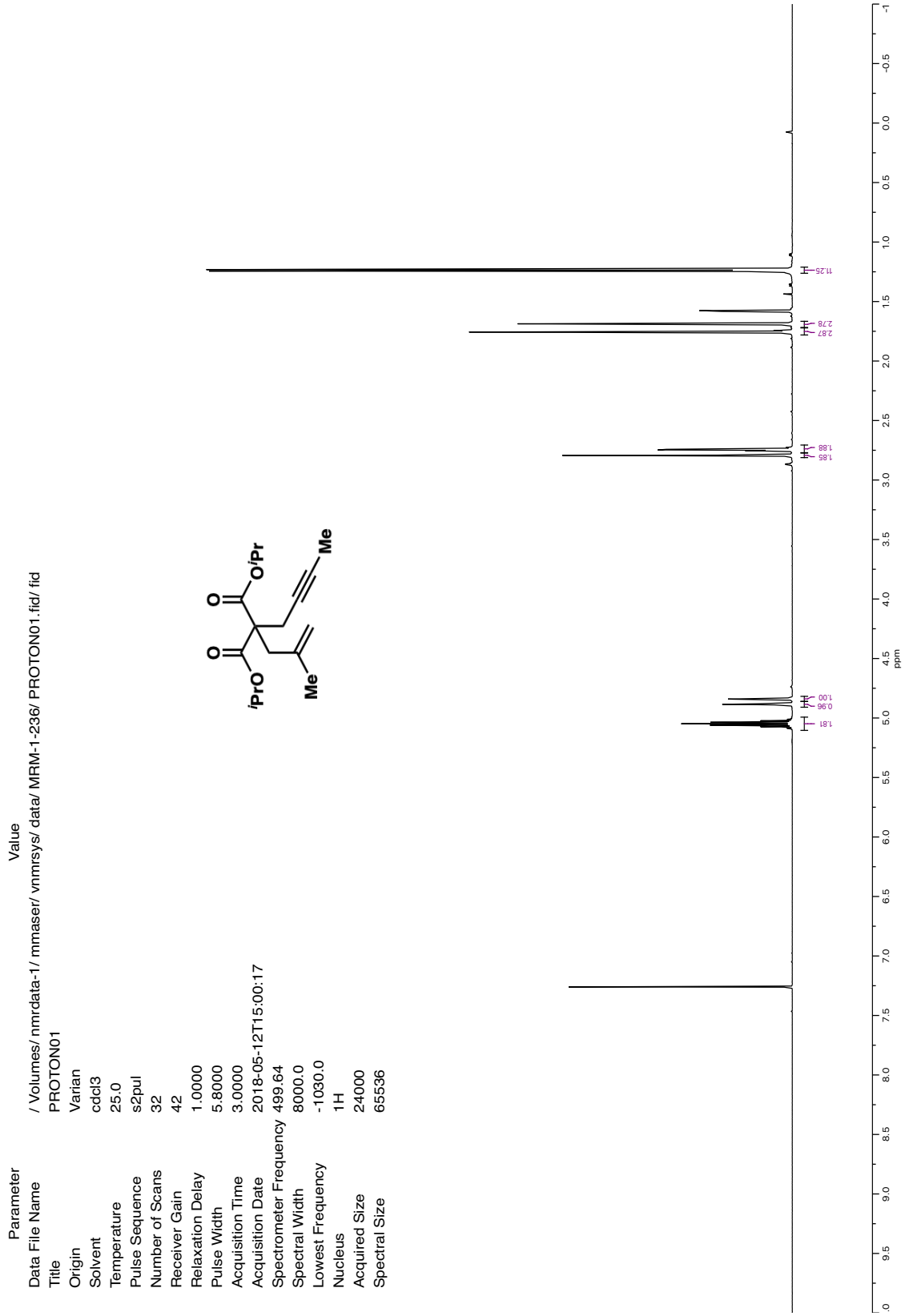

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-027/ 2/ fid
Title SED-1-027.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 43.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-20T12:26:14
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OTf
O
Me
Me
HO
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
198.44
171.95
169.21
162.78
142.51
120.15
116.96
85.09
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
70.56
70.02
59.41
55.72
44.20
41.99
21.57
21.39
21.35
19.30
10.82
ppm

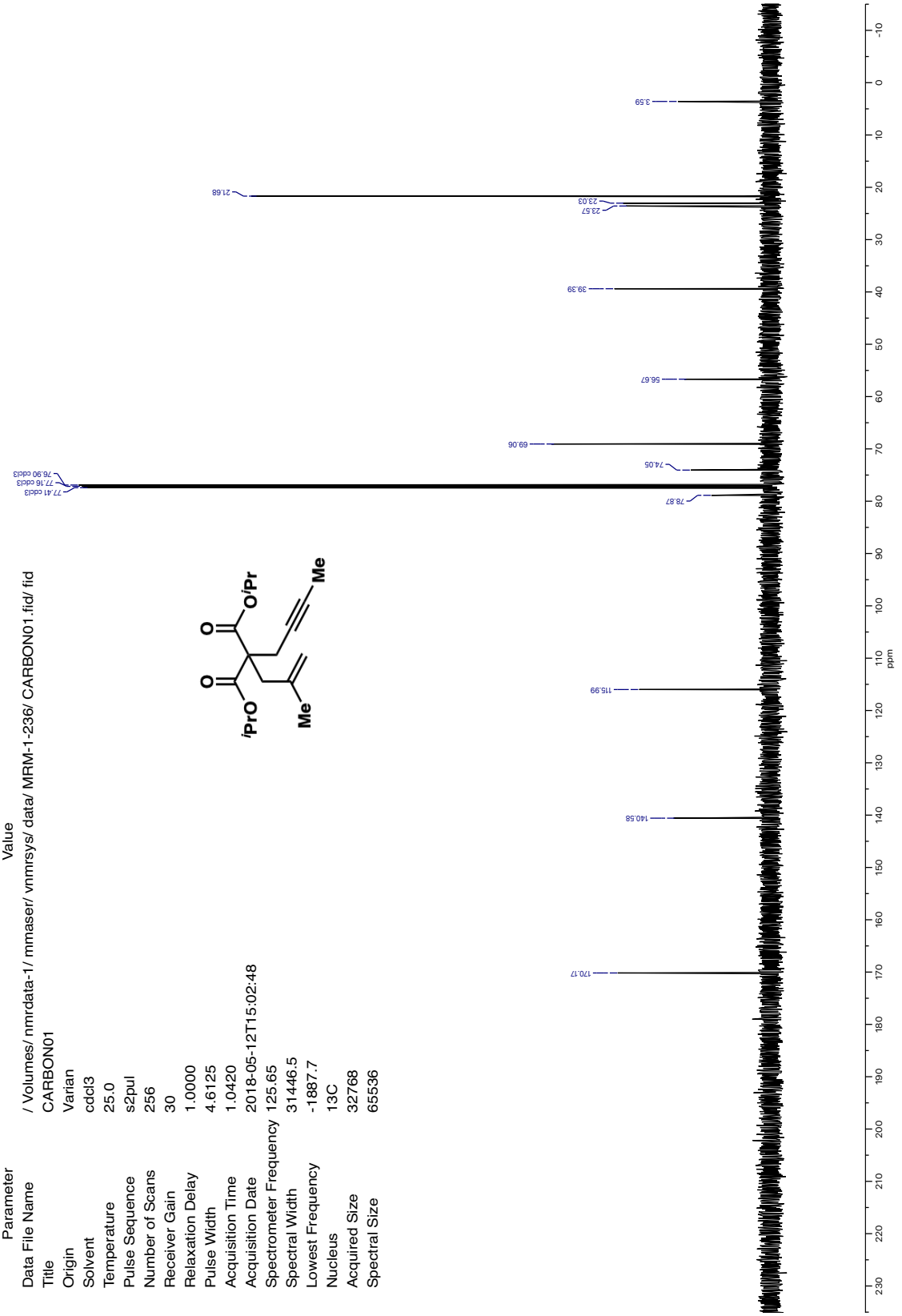

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-013-NEW-3-CDCl3/ 1/ fid
Title
MRM-2-013-NEW-3-CDCl3.1.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.2
Pulse Sequence
zg30
Number of Scans
64
Receiver Gain
197.4
Relaxation Delay
1.0000
Pulse Width
12.5000
Acquisition Time
4.0894
Acquisition Date
2019-12-02T17:28:47
Spectrometer Frequency
400.13
Spectral Width
8012.8
Lowest Frequency
-1545.1
Nucleus
1H
Acquired Size
32768
Spectral Size
65536
OTf
O
Me
HO
Me
Me
O
F3CO
1.76
2.01
0.97
0.98
0.96
6.14
3.00
ppm

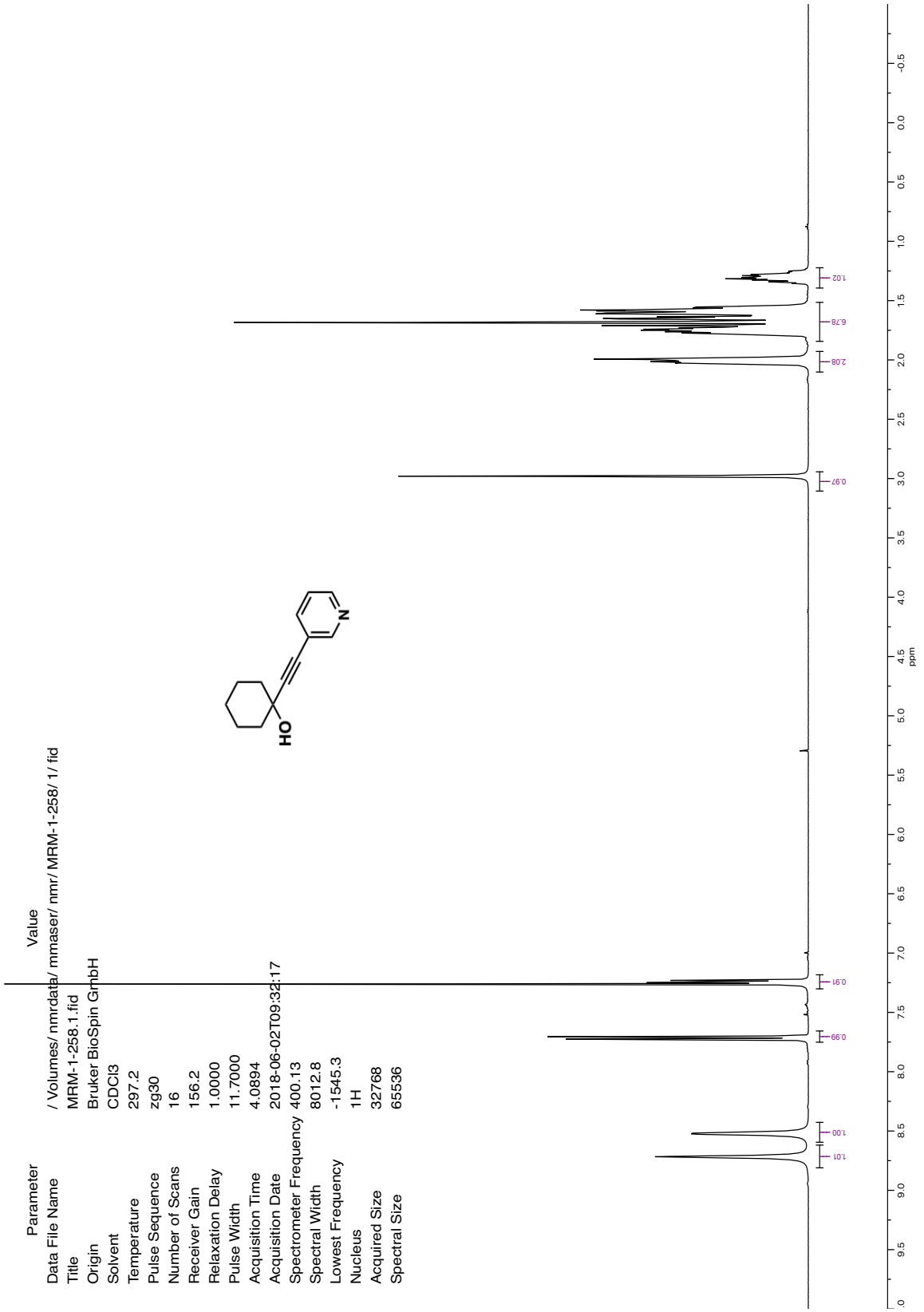

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-013-NEW-3-CDCl3/ 2/ fid
Title MRM-2-013-NEW-3-CDCl3.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1536
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-12-03T00:39:15
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -2045.1
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OTf
O
Me
HO
Me
Me
O
F3CO
15.62
20.77
24.99
28.85
57.68
70.39
83.57
87.96
115.73
118.16
118.92
119.77
120.74
127.74
130.58
142.15
150.22
154.23
197.78
ppm

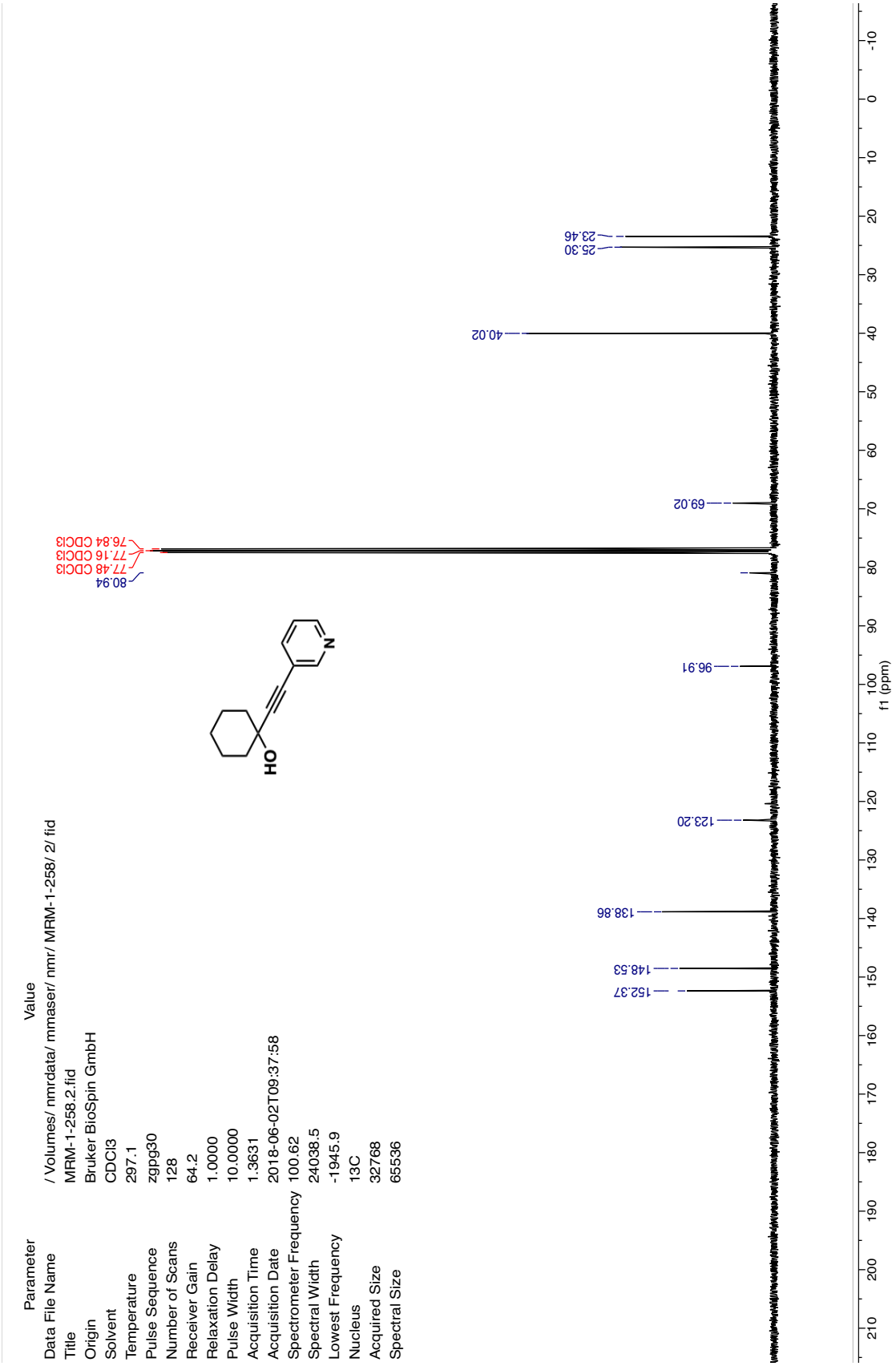

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ sdibrell/ vnmrsys/ data/ SED-2-114-column2-2/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 3.0
Pulse Sequence s2pul
Number of Scans 8
Receiver Gain 44
Relaxation Delay 1.0000
Pulse Width 5.6500
Acquisition Time 3.0000
Acquisition Date 2019-10-06T14:19:32
Spectrometer Frequency 499.61
Spectral Width 8000.0
Lowest Frequency -1030.3
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
O
O
O
O
iPrO
iPr
Me
13.18
3.12
2.19
1.00
2.01
ppm

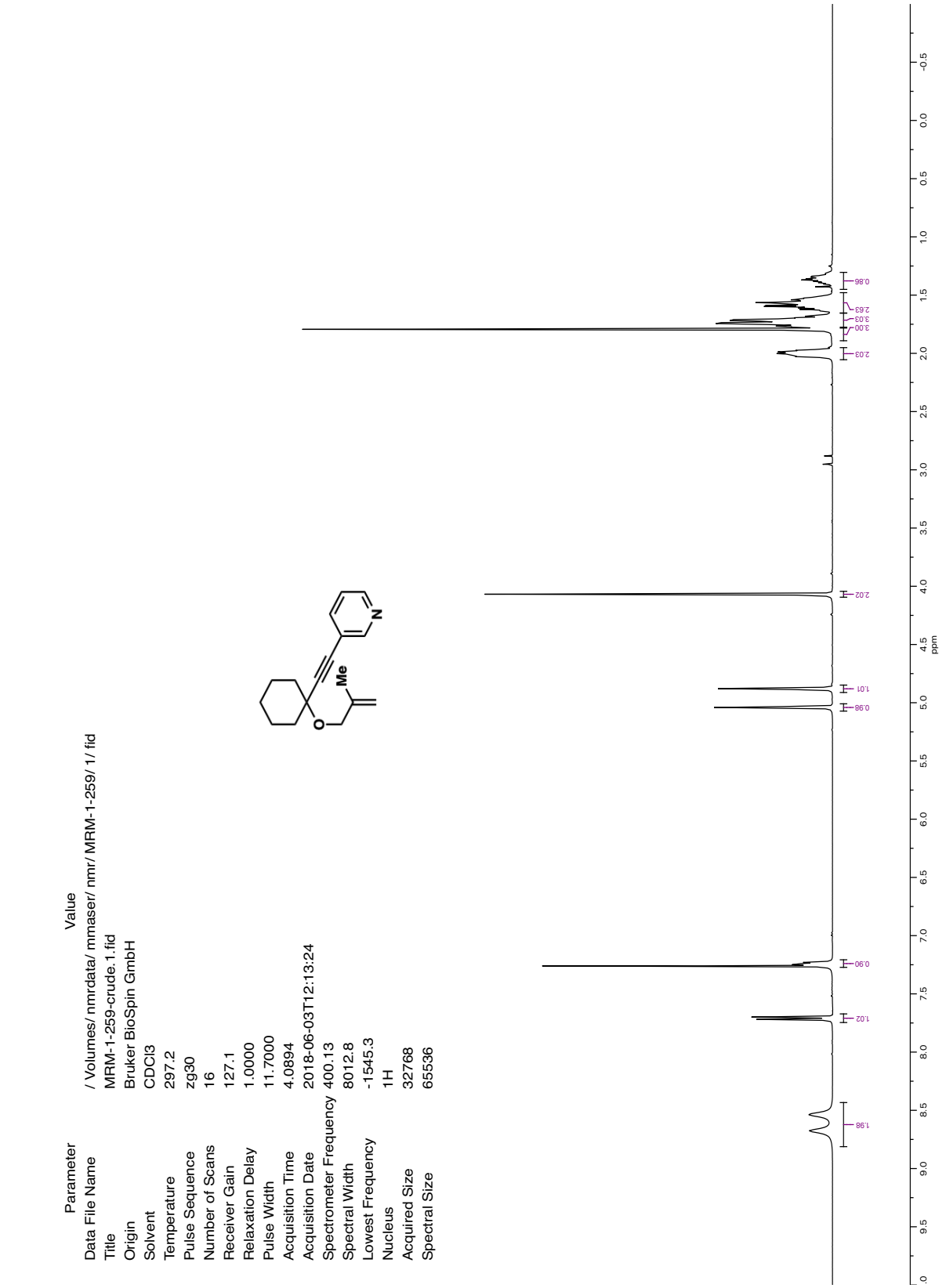

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-082/ CARBON01.fid/ fid
Title CARBON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 64
Receiver Gain 30
Relaxation Delay 1.0000
Pulse Width 4.6125
Acquisition Time 1.0420
Acquisition Date 2017-12-17T14:17:37
Spectrometer Frequency 125.65
Spectral Width 31446.5
Lowest Frequency -1888.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
O
O
O
O
Me
77.41 cdcl3
77.16 cdcl3
76.91 cdcl3
167.98
77.77
74.96
69.20
52.15
21.79
21.69
18.80
3.60
ppm

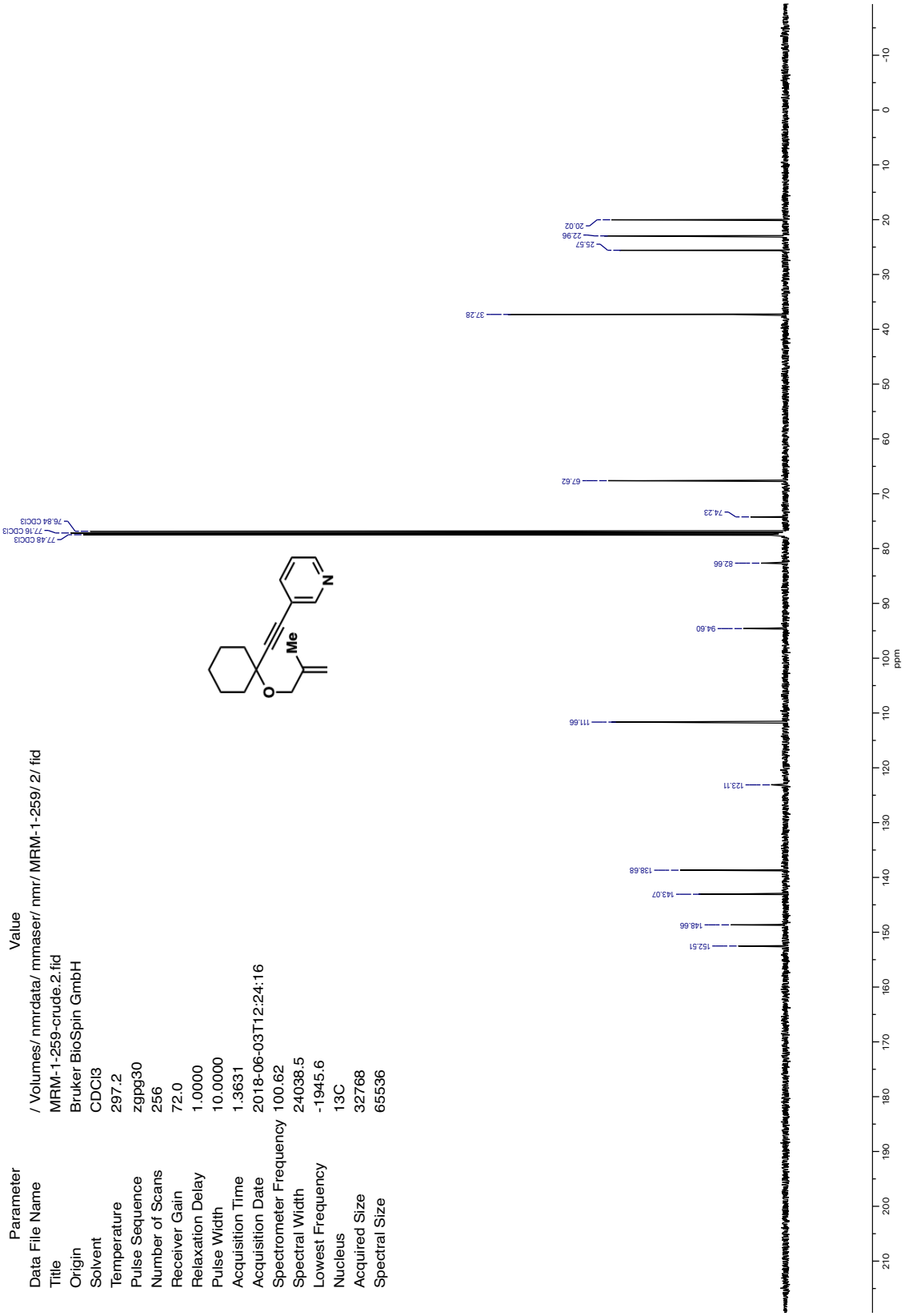

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-236/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 42
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2018-05-12T15:00:17
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1030.0
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
iPrO
O
O
OiPr
Me
Me
11.25
2.87
2.78
1.85
1.88
1.81
0.96
1.00
ppm

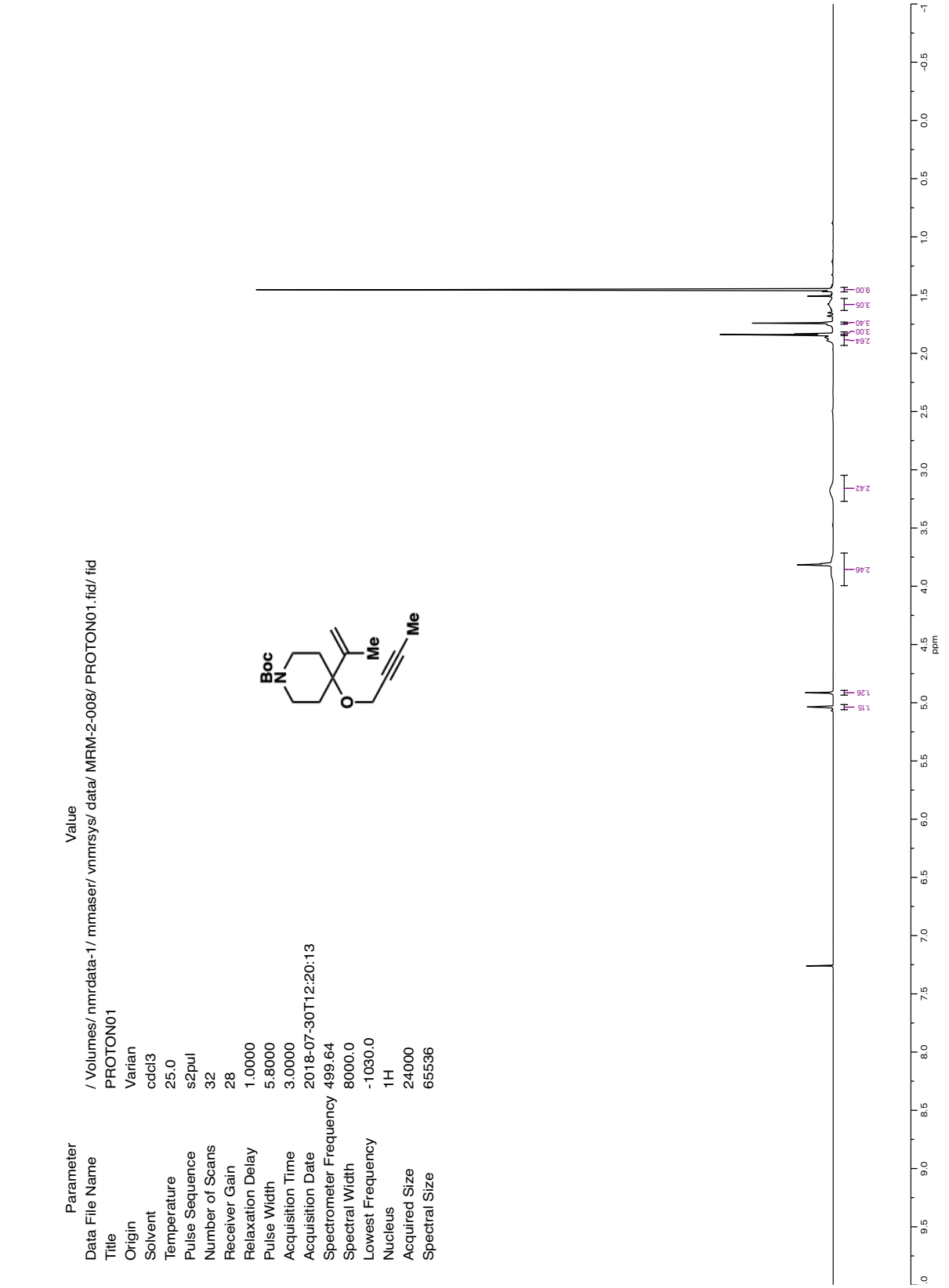

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-236/ CARBON01.fid/ fid
Title CARBON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 256
Receiver Gain 30
Relaxation Delay 1.0000
Pulse Width 4.6125
Acquisition Time 1.0420
Acquisition Date 2018-05-12T15:02:48
Spectrometer Frequency 125.65
Spectral Width 31446.5
Lowest Frequency -1887.7
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
77.41 cdcl3
77.16 cdcl3
76.90 cdcl3
170.17
140.58
115.99
78.87
74.05
69.06
56.67
39.39
23.57
23.03
21.68
3.59
O
O
O
O
Me
Me
ppm

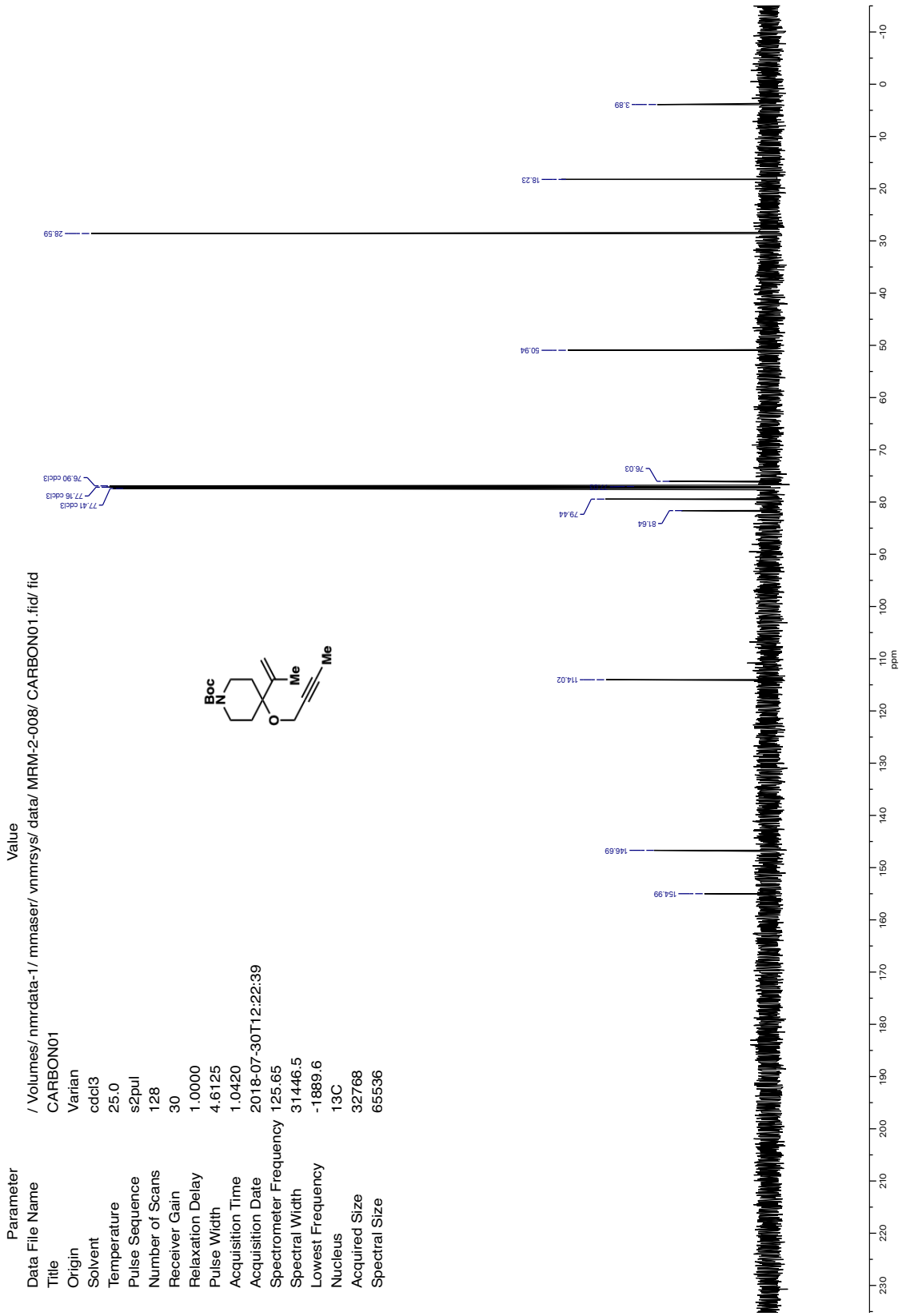

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-258/ 1/ fid
Title MRM-1-258.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2018-06-02T09:32:17
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
HO
N
1.01
1.00
0.99
0.91
0.97
2.08
6.78
1.02
ppm

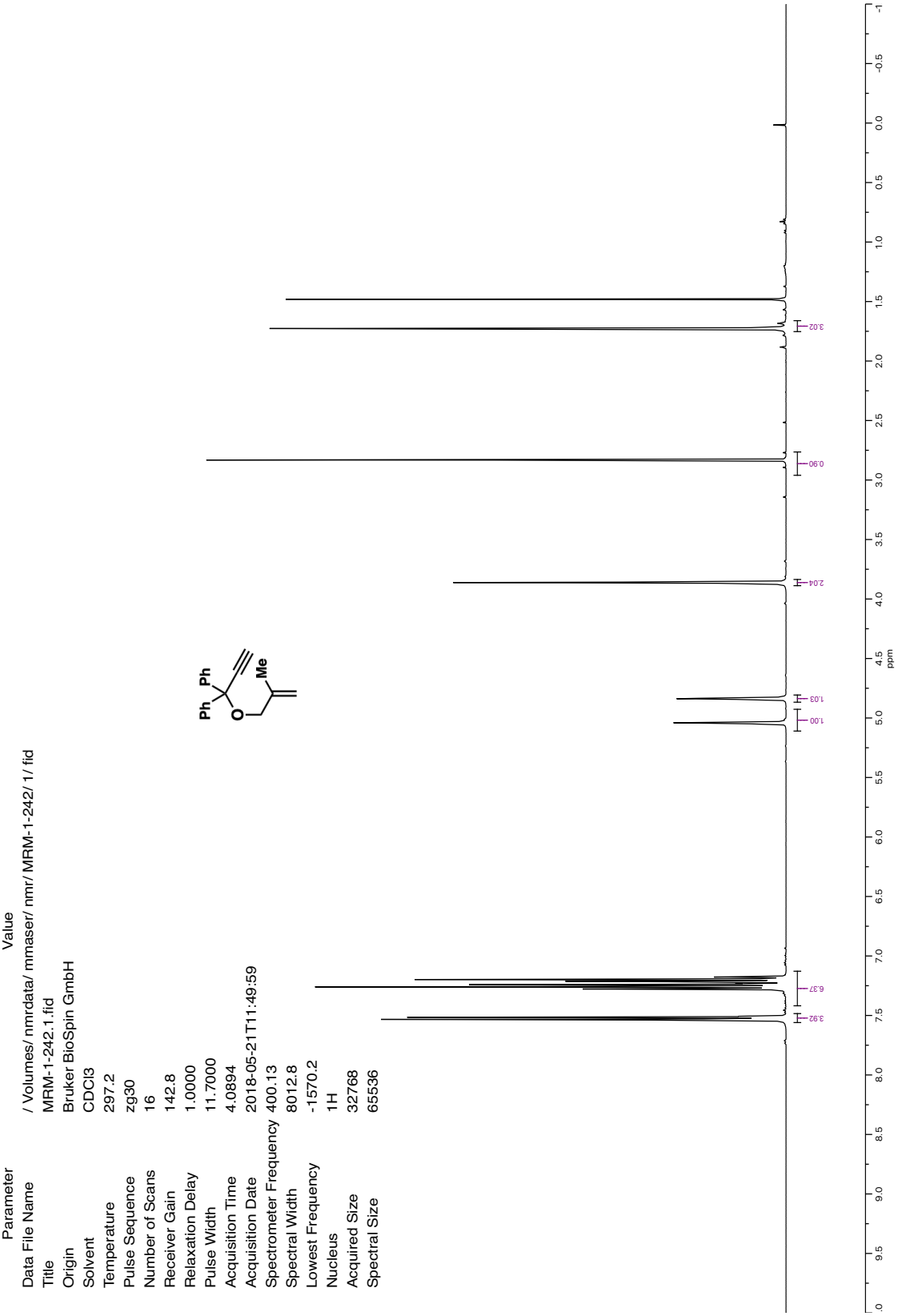

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-258/ 2/ fid
Title MRM-1-258.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 128
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-06-02T09:37:58
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
HO
N
80.94
76.84 CDCl3
77.16 CDCl3
77.48 CDCl3
23.46
25.30
40.02
69.02
96.91
123.20
138.86
148.53
152.37
f1 (ppm)

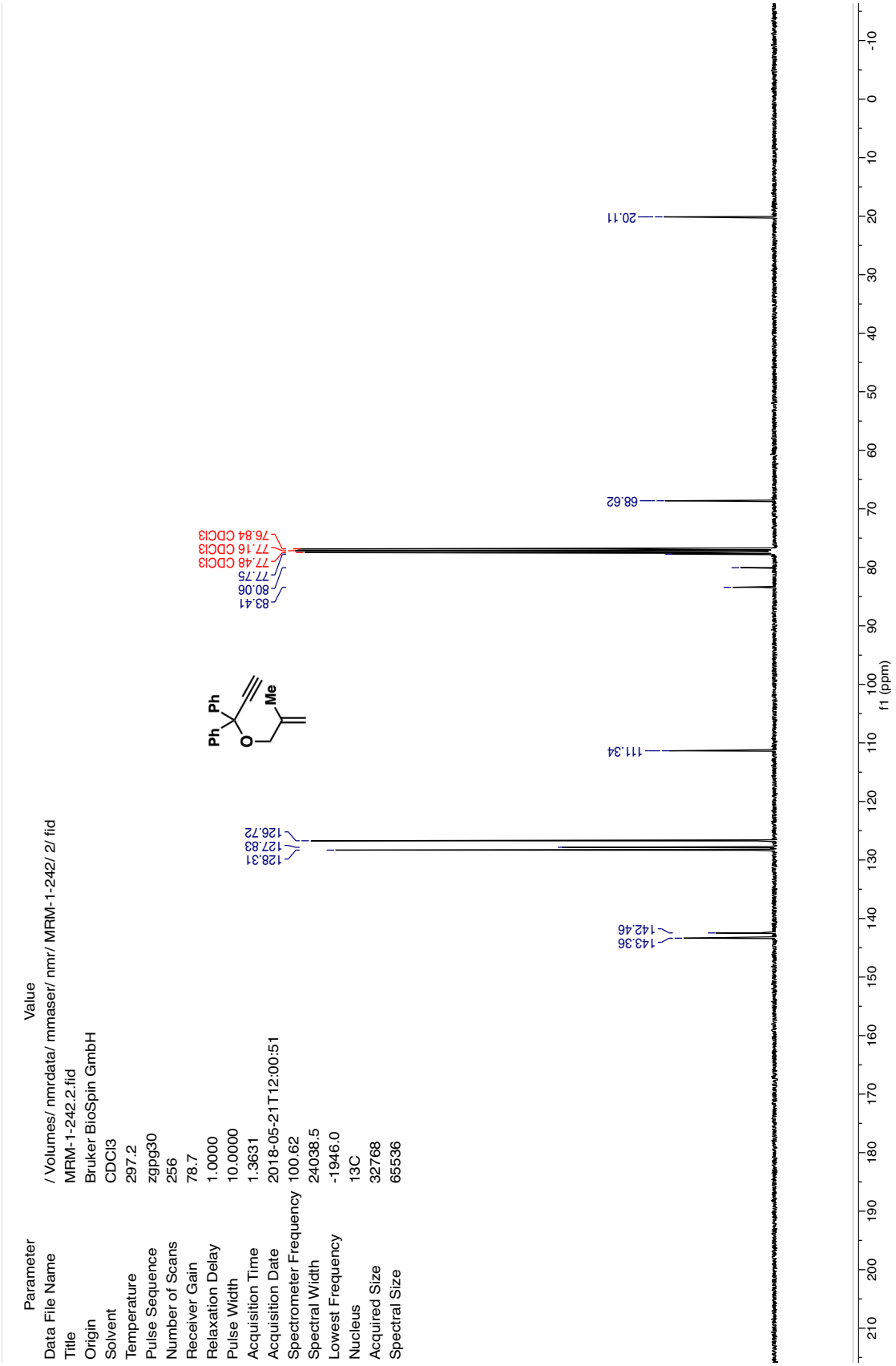

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-259/ 1/ fid
Title MRM-1-259-crude.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 127.1
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2018-06-03T12:13:24
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
O
N
1.98
1.02
0.90
2.02
1.01
0.98
2.03
3.00
3.03
2.63
0.86
ppm

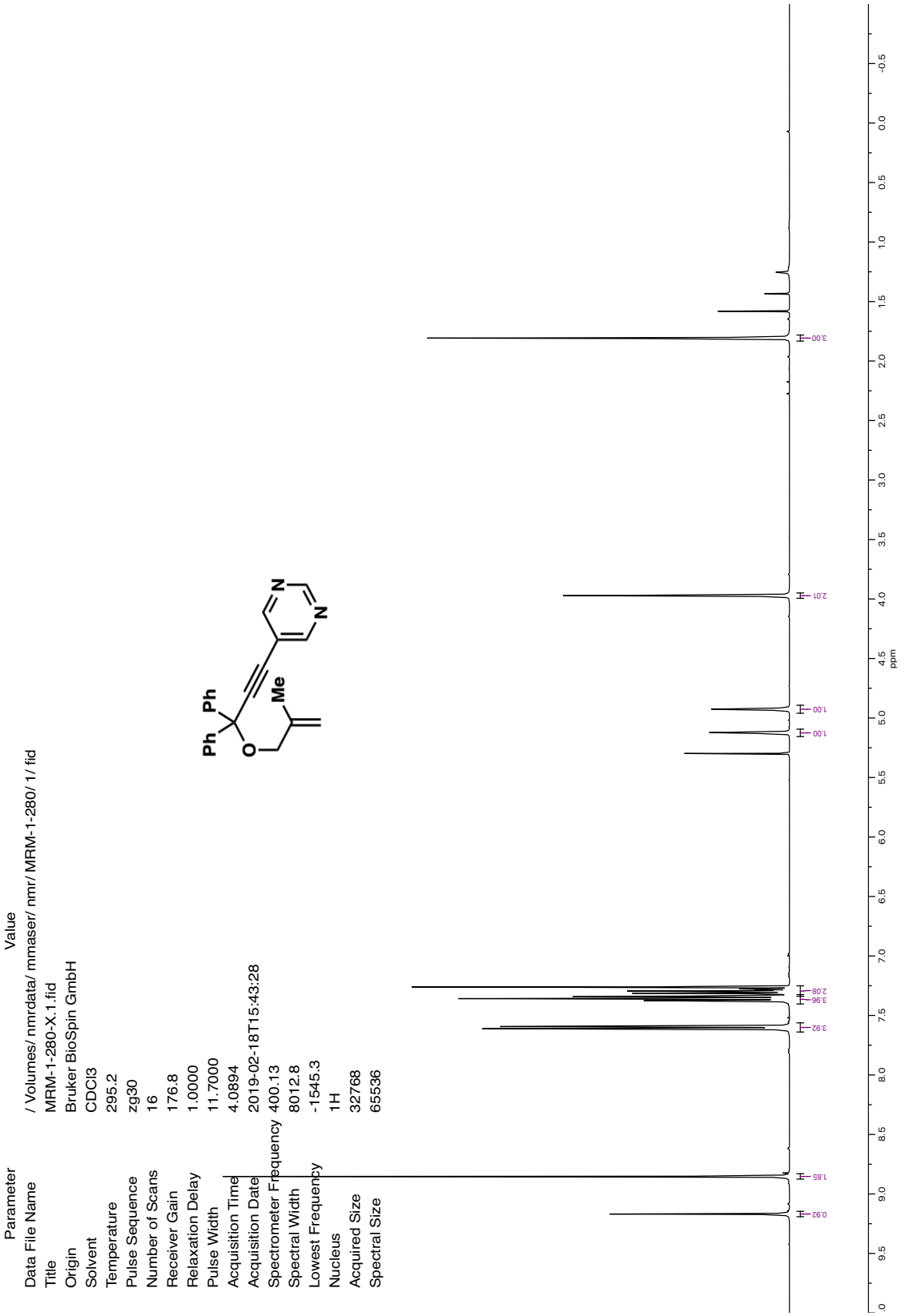

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-259/ 2/ fid
Title MRM-1-259-crude.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 256
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-06-03T12:24:16
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.6
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
N
O
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
152.51
148.66
143.07
138.68
123.11
111.66
94.60
82.66
74.23
67.62
37.28
25.57
22.96
20.02
ppm

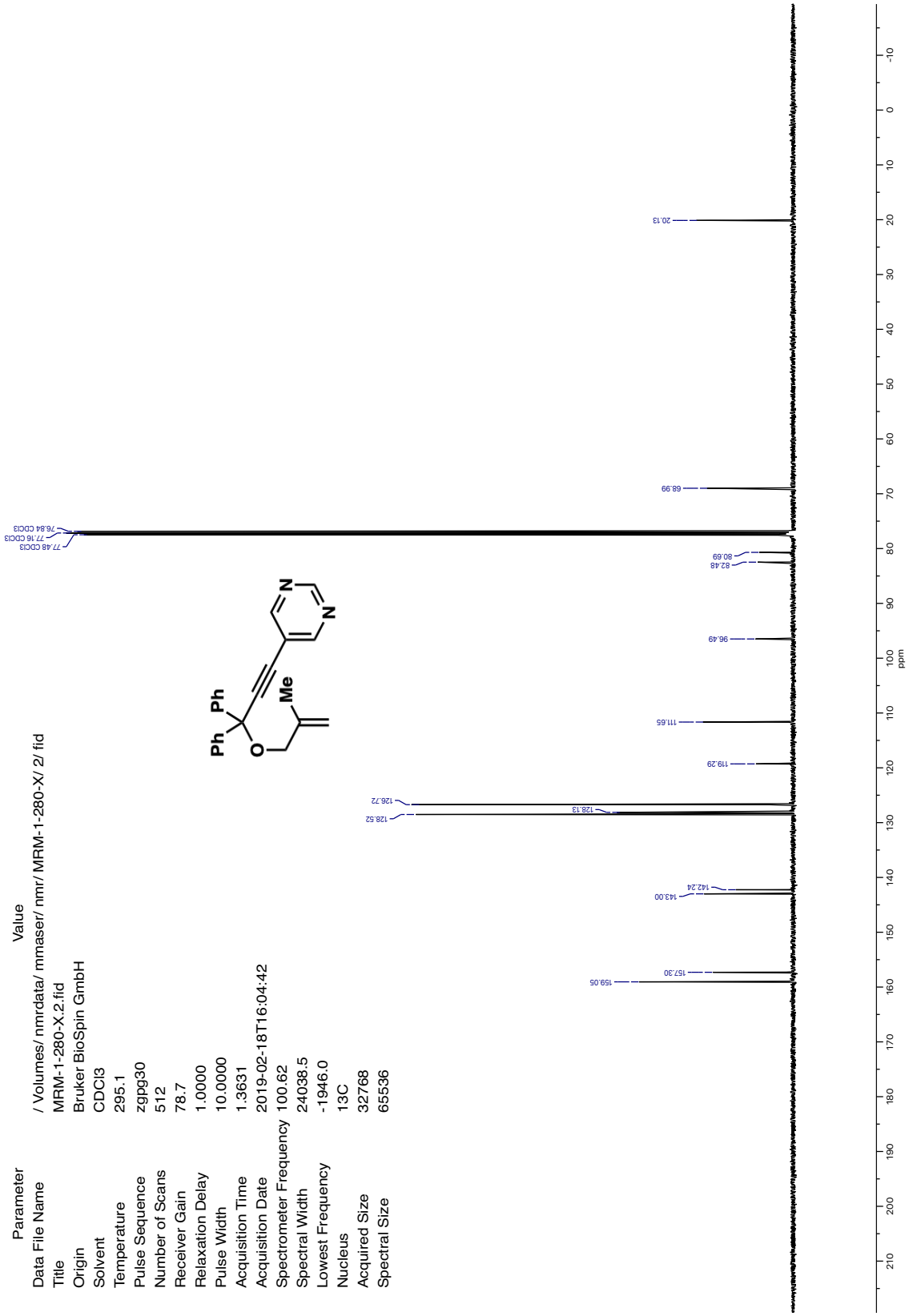

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-2-008/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 28
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2018-07-30T12:20:13
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1030.0
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
Boc
N
O
Me
Me
9.00
3.05
3.40
3.00
2.64
2.42
2.46
1.26
1.15
ppm

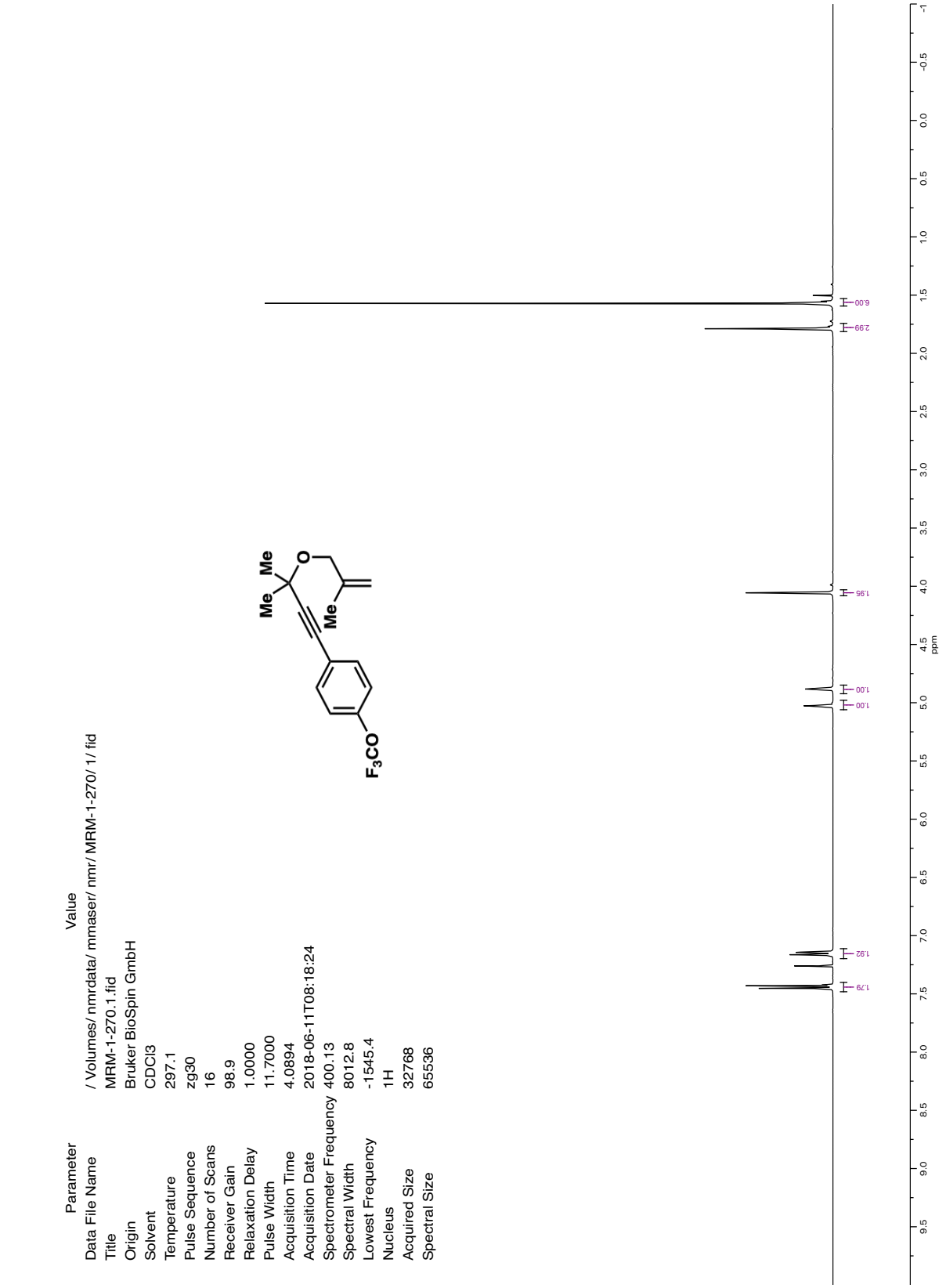

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-2-008/ CARBON01.fid/ fid
Title CARBON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 128
Receiver Gain 30
Relaxation Delay 1.0000
Pulse Width 4.6125
Acquisition Time 1.0420
Acquisition Date 2018-07-30T12:22:39
Spectrometer Frequency 125.65
Spectral Width 31446.5
Lowest Frequency -1889.6
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Boc
N
O
Me
Me
3.89
18.23
28.59
50.94
76.03
76.90 cdcl3
77.16 cdcl3
77.41 cdcl3
79.44
81.64
114.02
146.69
154.99
ppm

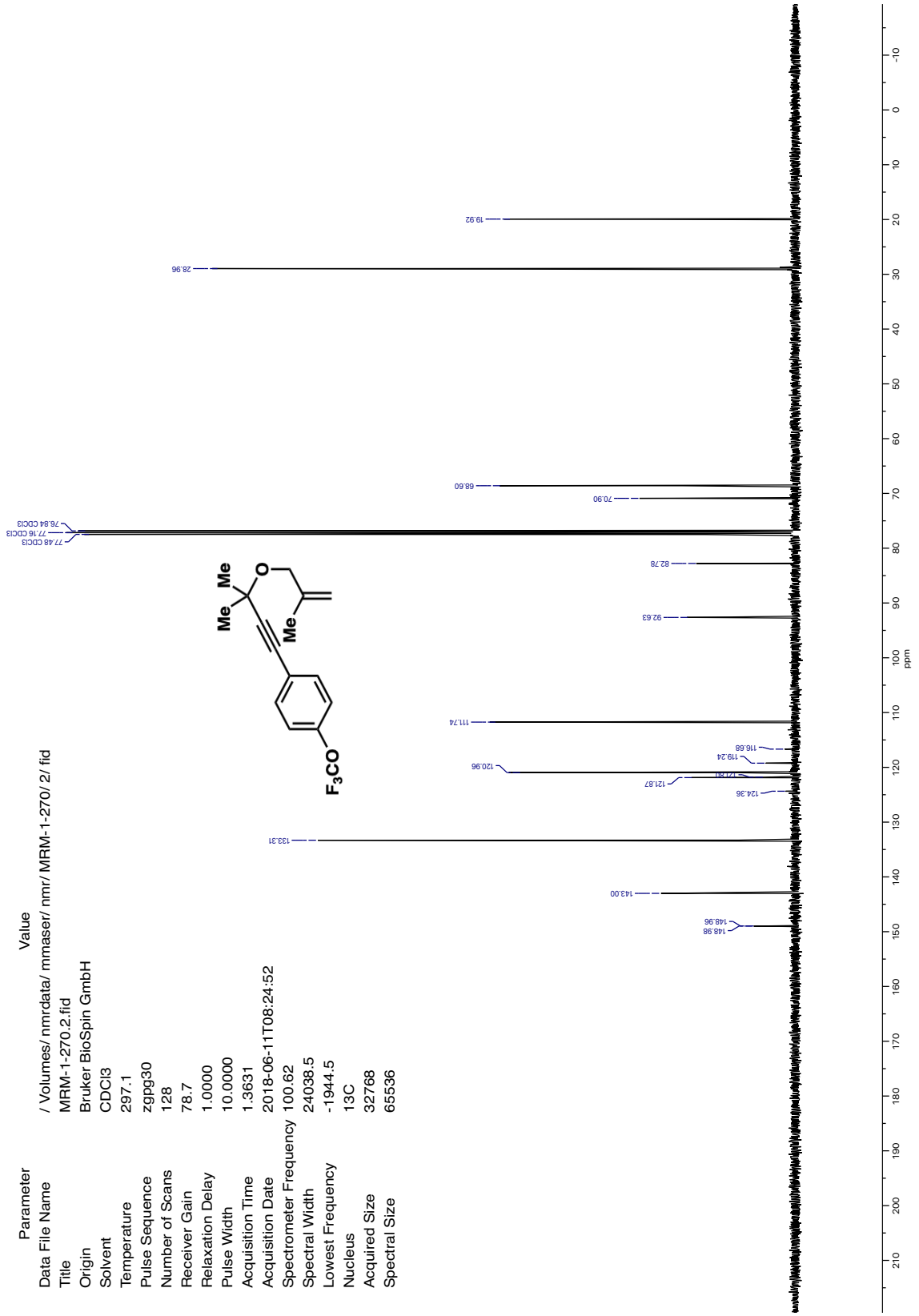

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-242/ 1/ fid
Title MRM-1-242.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 142.8
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2018-05-21T11:49:59
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1570.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Ph
Ph
O
Me
3.92
6.37
1.00
1.03
2.04
0.90
3.02
ppm

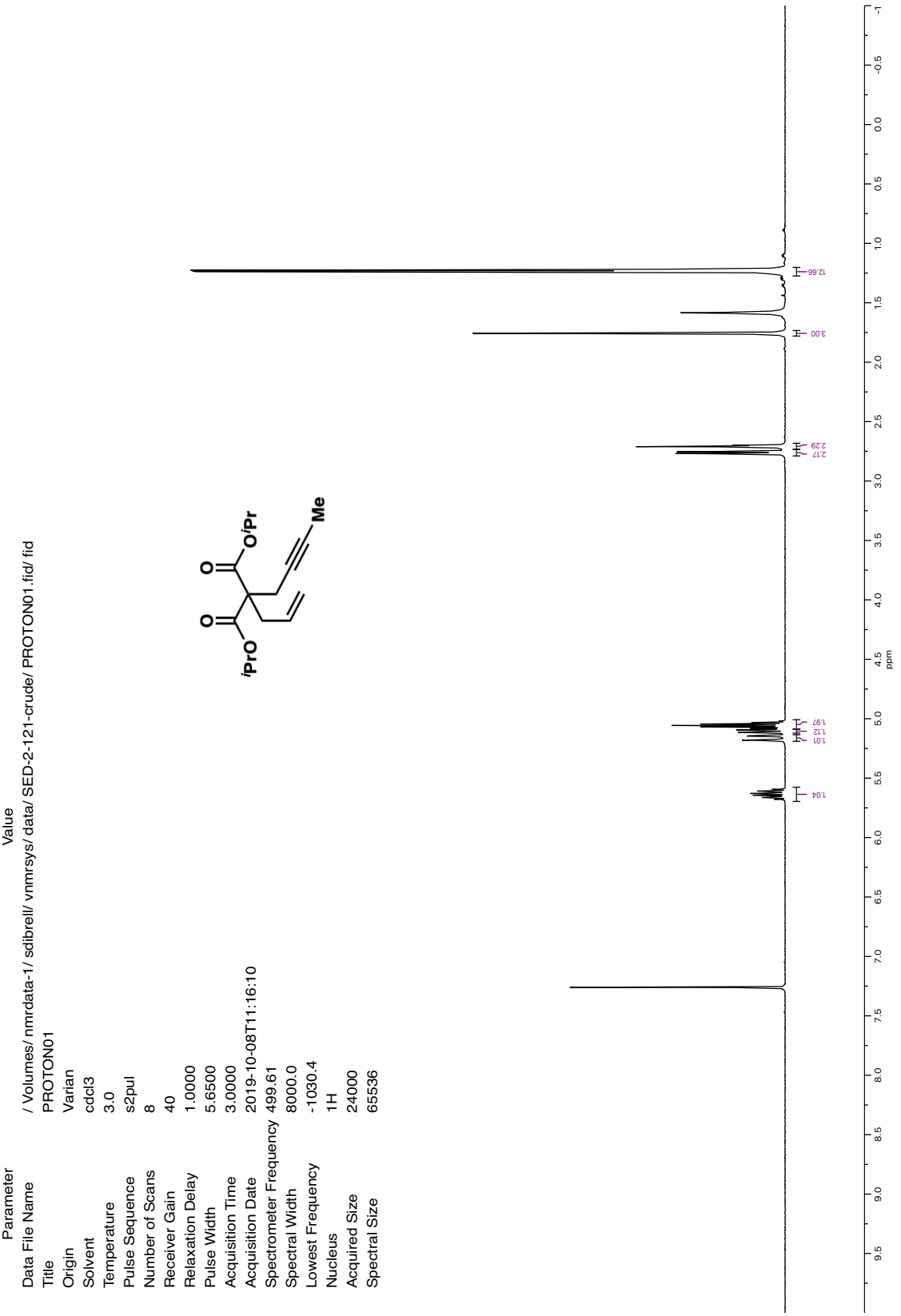

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-242/ 2/ fid
Title MRM-1-242.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 256
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-05-21T12:00:51
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1946.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Ph
Ph
Me
O
20.11
68.62
111.34
126.72
127.83
128.31
142.46
143.36
76.84 CDCl3
77.16 CDCl3
77.48 CDCl3
77.75
80.06
83.41
f1 (ppm)

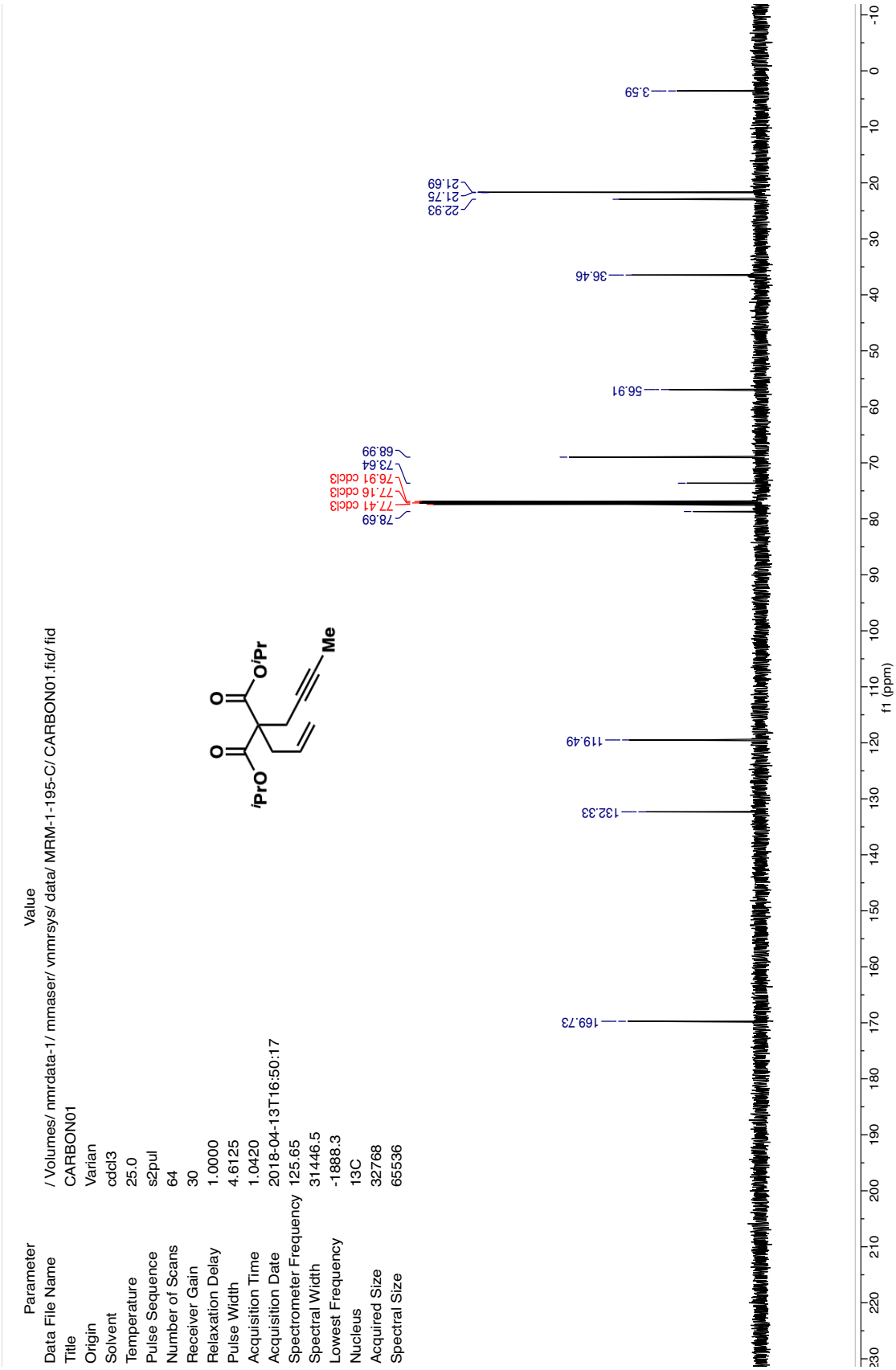

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-280/ 1/ fid
Title
MRM-1-280-X.1.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
295.2
Pulse Sequence
zg30
Number of Scans
16
Receiver Gain
176.8
Relaxation Delay
1.0000
Pulse Width
11.7000
Acquisition Time
4.0894
Acquisition Date
2019-02-18T15:43:28
Spectrometer Frequency
400.13
Spectral Width
8012.8
Lowest Frequency
-1545.3
Nucleus
1H
Acquired Size
32768
Spectral Size
65536
Ph
Ph
Me
O
N
N
0.92
1.85
3.92
2.08
3.96
1.00
1.00
2.01
3.00
ppm

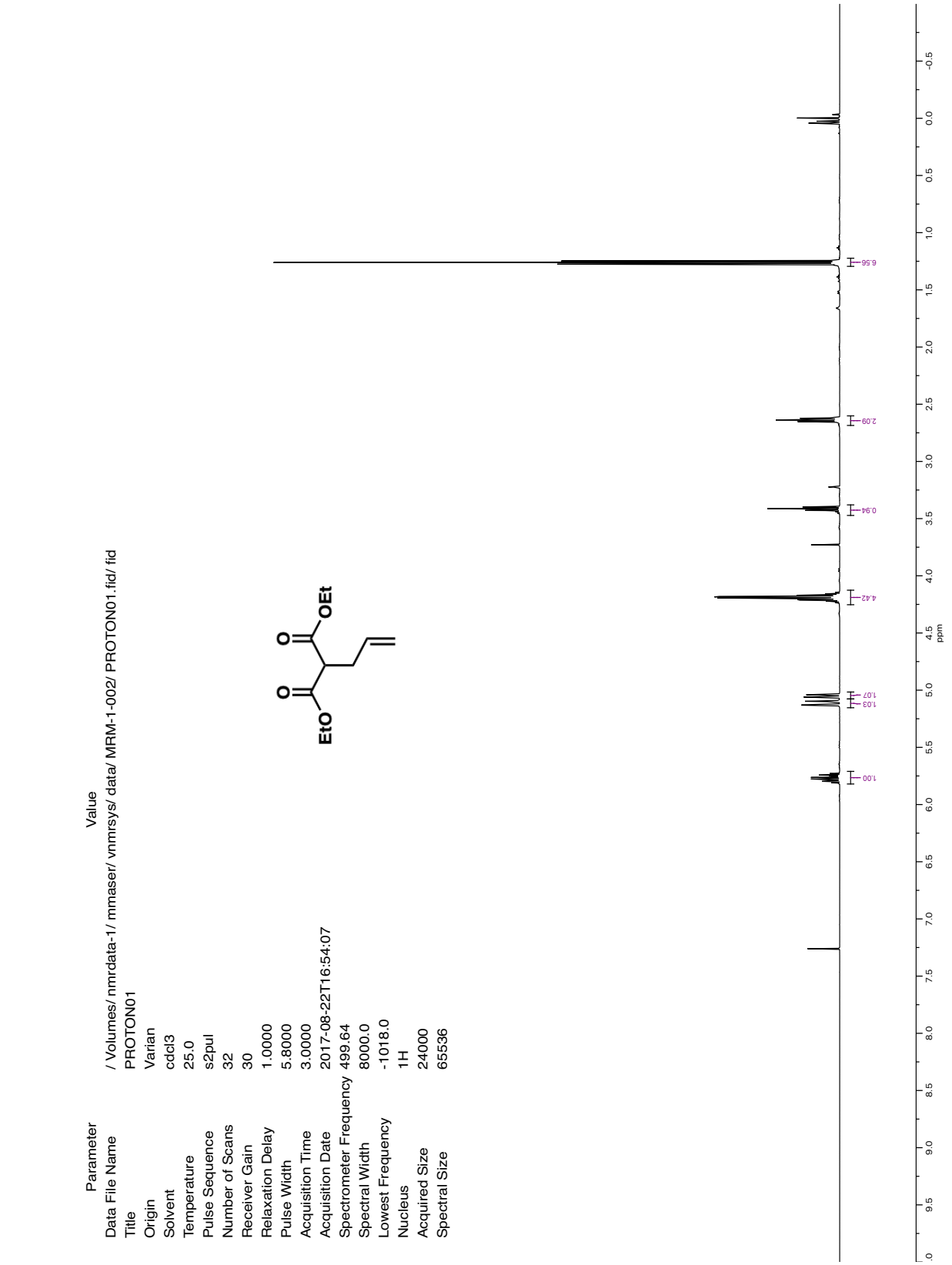

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-280-X/ 2/ fid
Title MRM-1-280-X.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-02-18T16:04:42
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1946.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Ph
Ph
Me
O
N
N
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
20.13
68.99
80.69
82.48
96.49
111.65
119.29
126.72
128.52
128.13
142.24
143.00
157.30
159.05
ppm

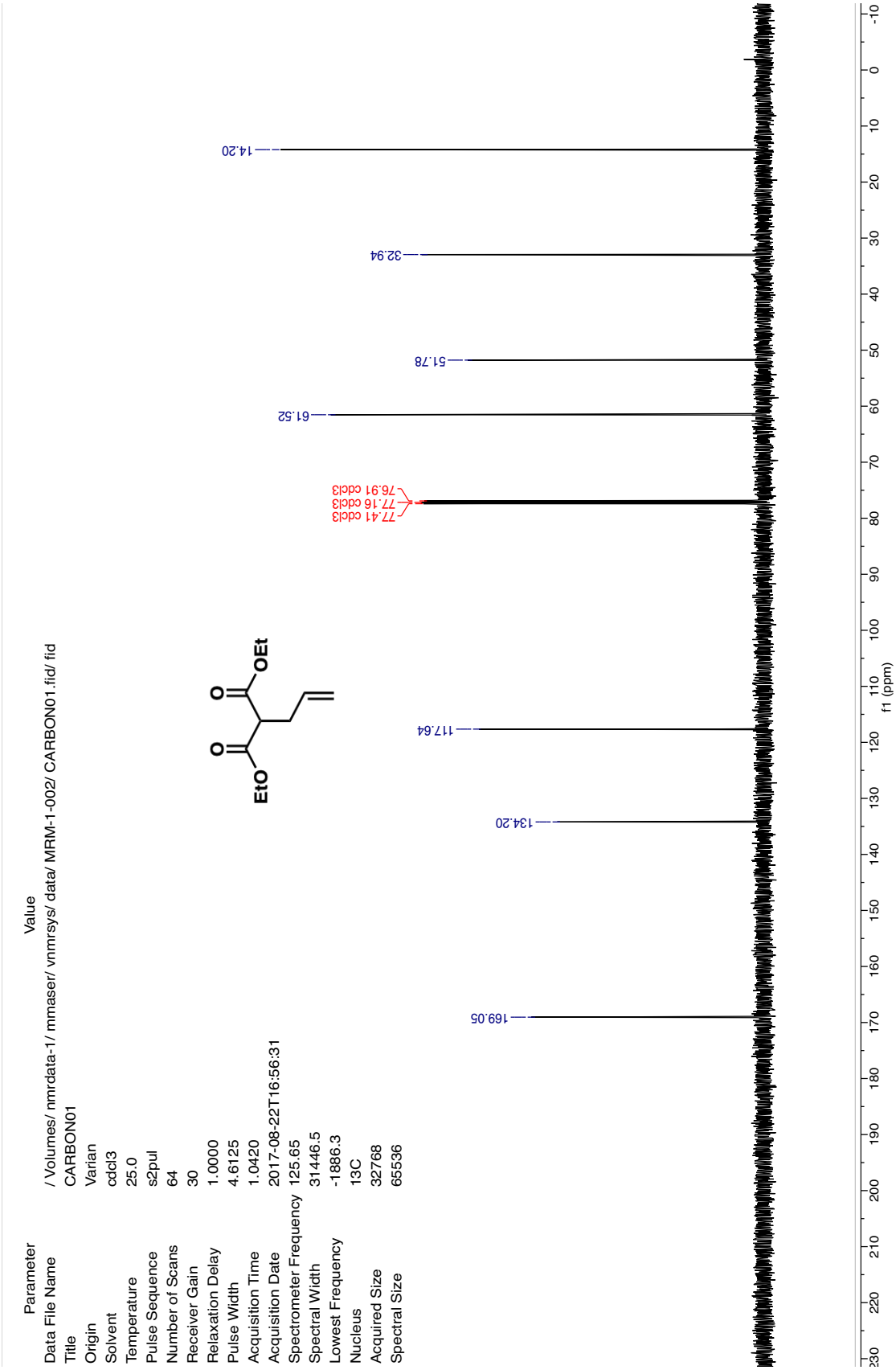

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-270/ 1/ fid
Title MRM-1-270.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 98.9
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2018-06-11T08:18:24
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
Me
O
Me
F3CO
6.00
2.99
1.95
1.00
1.00
1.92
1.79
ppm

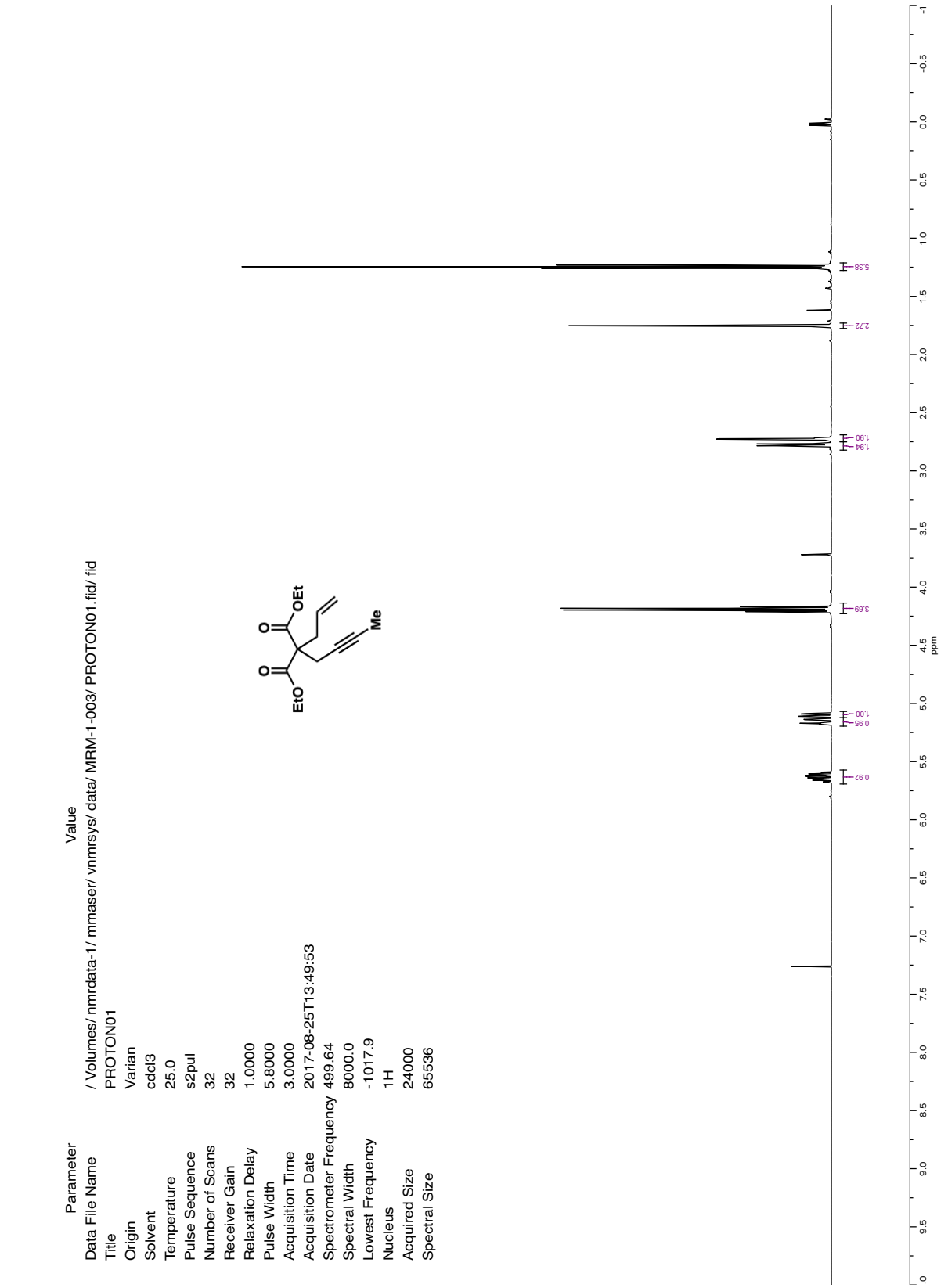

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-270/ 2/ fid
Title MRM-1-270.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 128
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-06-11T08:24:52
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.5
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
Me
O
Me
F3CO
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
19.92
28.96
68.60
70.90
82.78
92.63
111.74
116.68
119.24
120.96
121.87
124.36
133.31
143.00
148.98
148.96
ppm

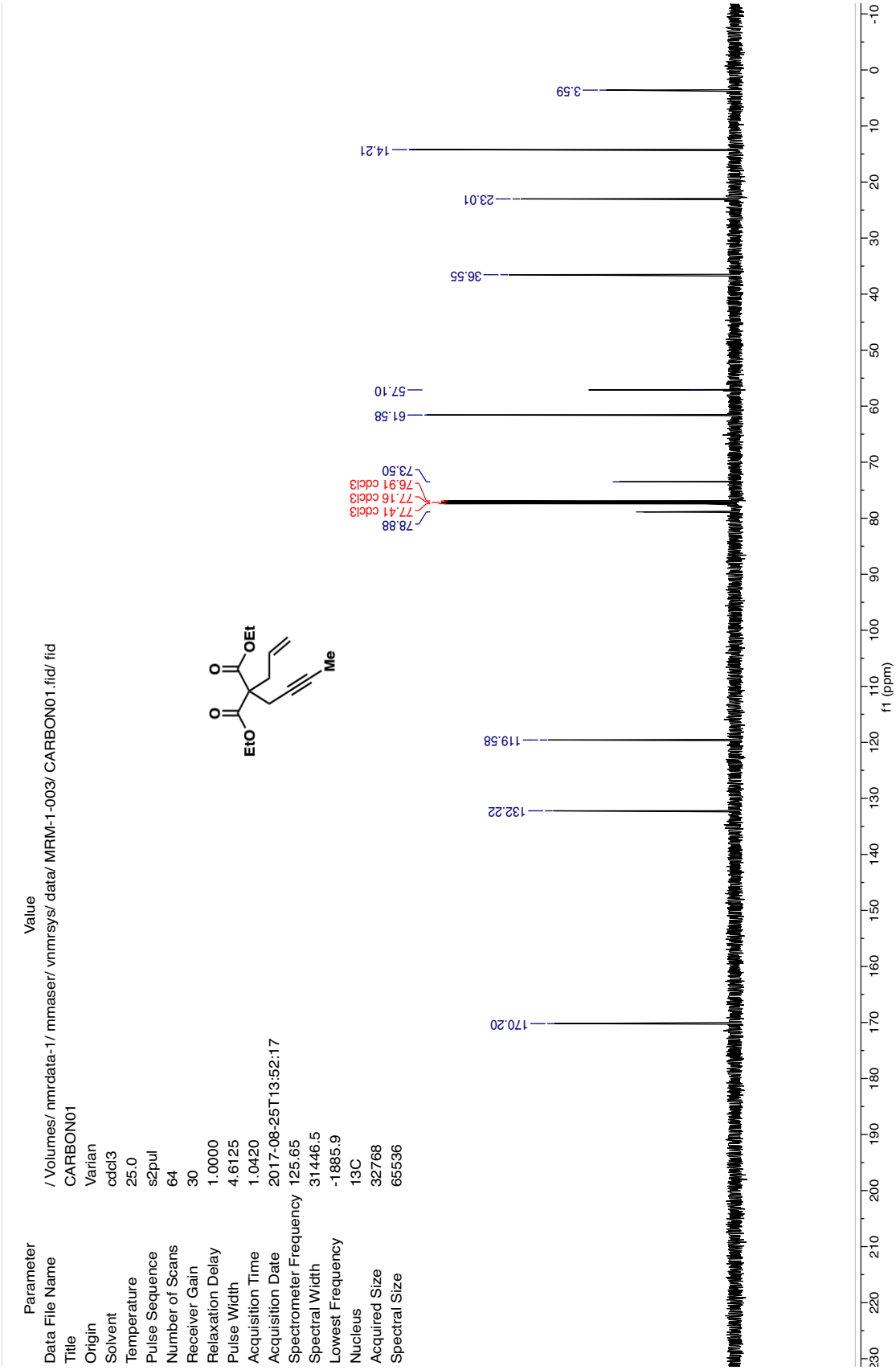

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ sdibrell/ vnmrsys/ data/ SED-2-121-crude/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 3.0
Pulse Sequence s2pul
Number of Scans 8
Receiver Gain 40
Relaxation Delay 1.0000
Pulse Width 5.6500
Acquisition Time 3.0000
Acquisition Date 2019-10-08T11:16:10
Spectrometer Frequency 499.61
Spectral Width 8000.0
Lowest Frequency -1030.4
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
O
O
OiPr
iPrO
Me
12.66
3.00
2.29
2.17
1.97
1.12
1.01
1.04
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1
ppm

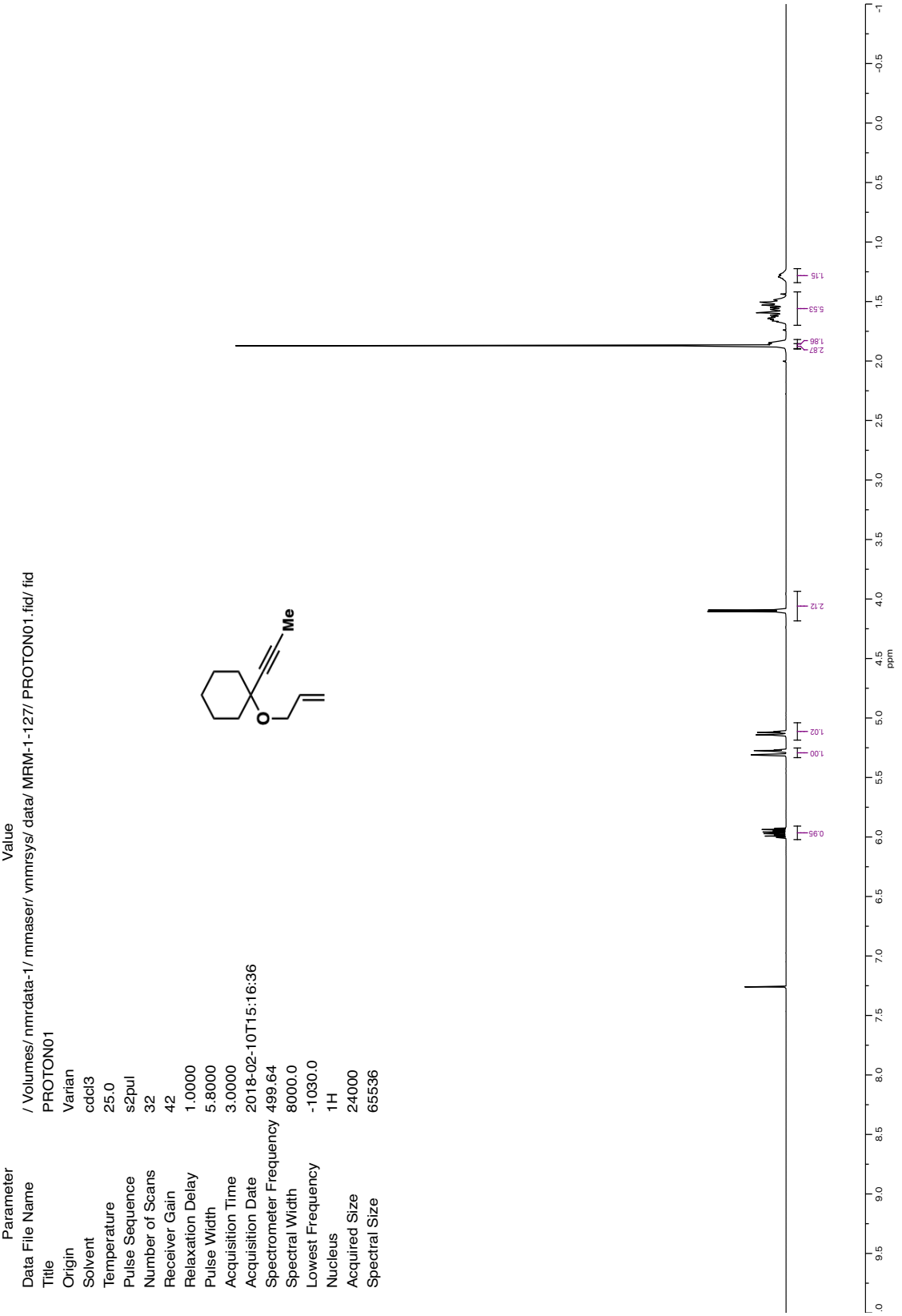

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-195-C/ CARBON01.fid/ fid
Title CARBON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 64
Receiver Gain 30
Relaxation Delay 1.0000
Pulse Width 4.6125
Acquisition Time 1.0420
Acquisition Date 2018-04-13T16:50:17
Spectrometer Frequency 125.65
Spectral Width 31446.5
Lowest Frequency -1888.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
O
O
OiPr
iPrO
Me
169.73
132.33
119.49
78.69
77.41 cdcl3
77.16 cdcl3
76.91 cdcl3
73.64
68.99
56.91
36.46
22.93
21.75
21.69
3.59
f1 (ppm)

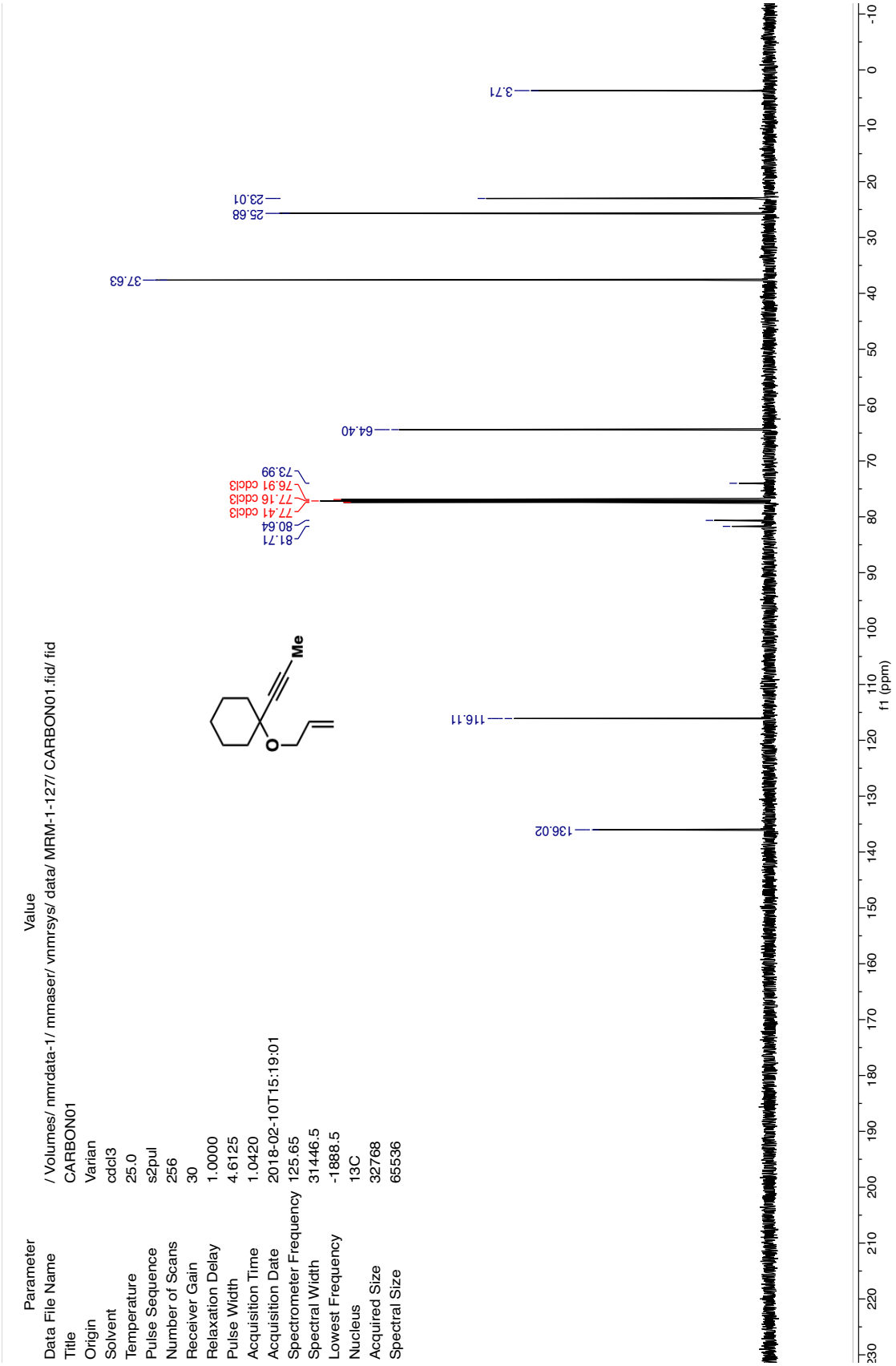

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-002/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 30
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2017-08-22T16:54:07
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1018.0
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
O
O
OEt
EtO
O
6.56
2.09
0.94
4.42
1.03
1.07
1.00
ppm


Parameter
Value
Data File Name
/ Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-002/ CARBON01.fid/ fid
Title
CARBON01
Origin
Varian
Solvent
cdcl3
Temperature
25.0
Pulse Sequence
s2pul
Number of Scans
64
Receiver Gain
30
Relaxation Delay
1.0000
Pulse Width
4.6125
Acquisition Time
1.0420
Acquisition Date
2017-08-22T16:56:31
Spectrometer Frequency
125.65
Spectral Width
31446.5
Lowest Frequency
-1886.3
Nucleus
13C
Acquired Size
32768
Spectral Size
65536
O
OEt
O
EtO
14.20
32.94
51.78
61.52
76.91 cdcl3
77.16 cdcl3
77.41 cdcl3
117.64
134.20
169.05
f1 (ppm)

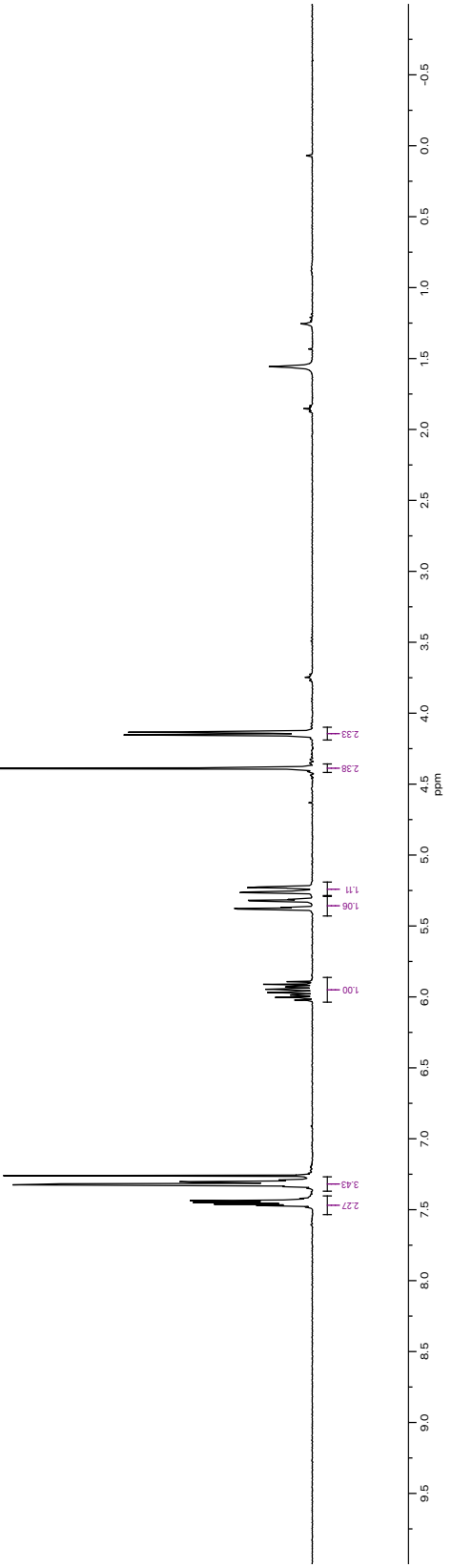

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-003/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 32
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2017-08-25T13:49:53
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1017.9
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
O
OEt
O
EtO
Me
5.38
2.72
1.94
1.90
3.69
0.96
1.00
0.92
ppm

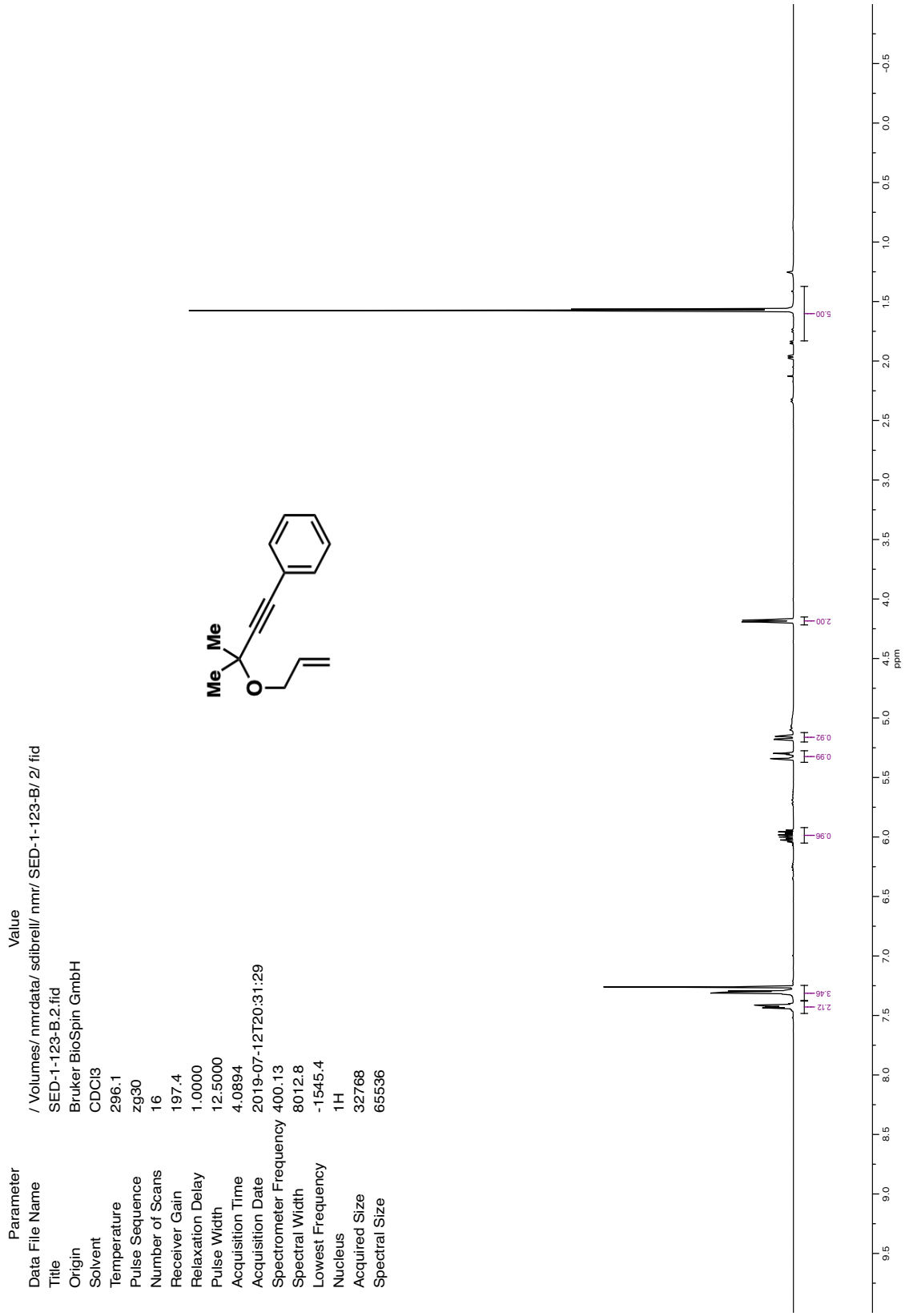

Parameter
Value
Data File Name
/ Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-003/ CARBON01.fid/ fid
Title
CARBON01
Origin
Varian
Solvent
cdcl3
Temperature
25.0
Pulse Sequence
s2pul
Number of Scans
64
Receiver Gain
30
Relaxation Delay
1.0000
Pulse Width
4.6125
Acquisition Time
1.0420
Acquisition Date
2017-08-25T13:52:17
Spectrometer Frequency
125.65
Spectral Width
31446.5
Lowest Frequency
-1885.9
Nucleus
13C
Acquired Size
32768
Spectral Size
65536
O
OEt
O
EtO
Me
3.59
14.21
23.01
36.55
57.10
61.58
73.50
76.91 cdcl3
77.16 cdcl3
77.41 cdcl3
78.88
119.58
132.22
170.20
f1 (ppm)


Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-127/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 42
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2018-02-10T15:16:36
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1030.0
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
Me
O
0.95
1.00
1.02
2.12
1.86
2.87
5.53
1.75
ppm

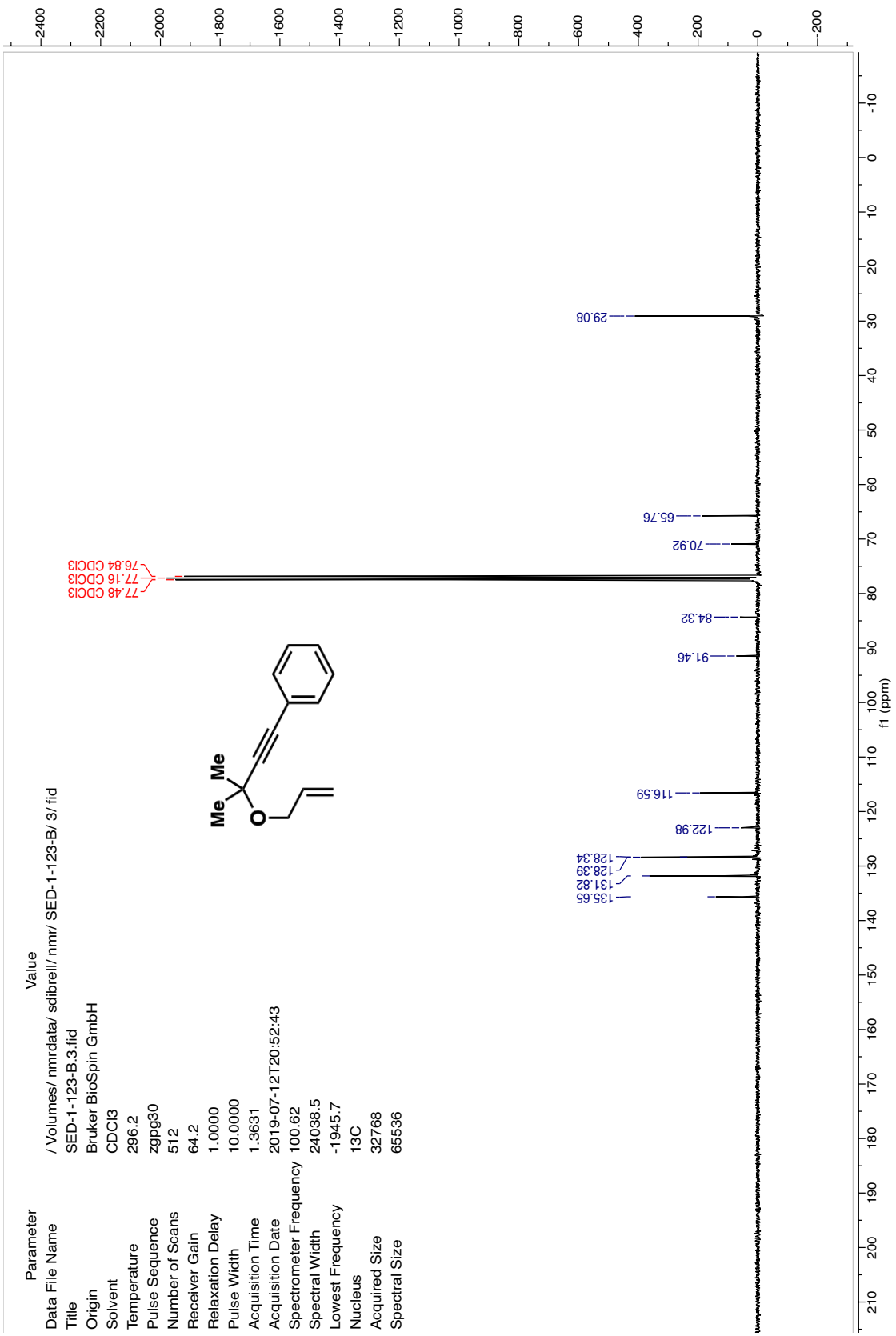

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-127/ CARBON01.fid/ fid
Title CARBON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 256
Receiver Gain 30
Relaxation Delay 1.0000
Pulse Width 4.6125
Acquisition Time 1.0420
Acquisition Date 2018-02-10T15:19:01
Spectrometer Frequency 125.65
Spectral Width 31446.5
Lowest Frequency -1888.5
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
3.71
23.01
25.68
37.63
64.40
73.99
76.91 cdcl3
77.16 cdcl3
77.41 cdcl3
80.64
81.71
116.11
136.02
f1 (ppm)

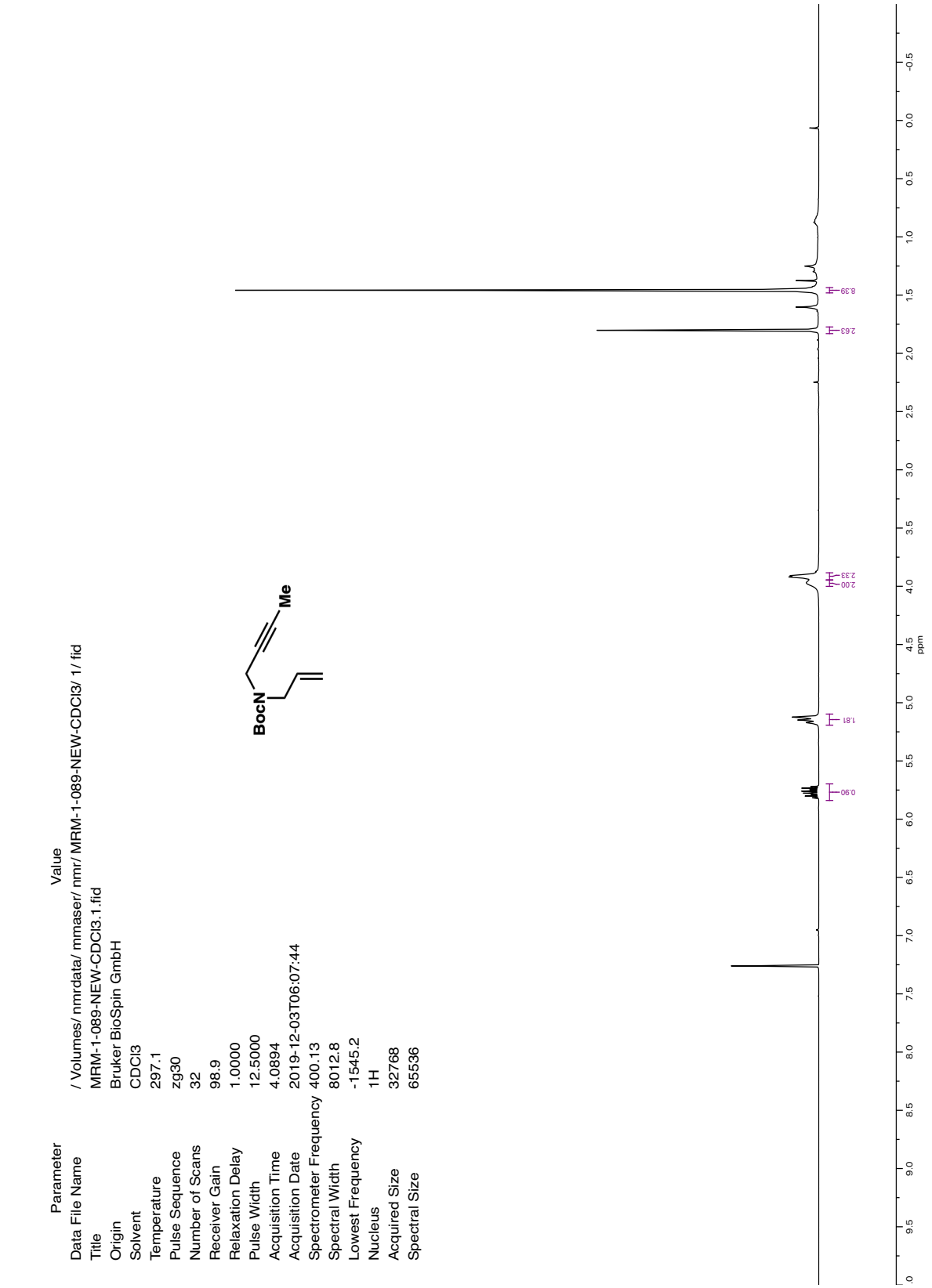
Parameter	Value
Data File Name	/ Volumes/ nmrdata-1/ sdibrell/ vnmrsys/ data/ SED-1-081/ PROTON03.fid/ fid
Title	PROTON03
Origin	Varian
Solvent	cdcl3
Temperature	25.0
Pulse Sequence	s2pul
Number of Scans	8
Receiver Gain	32
Relaxation Delay	1.0000
Pulse Width	5.1000
Acquisition Time	2.4999
Acquisition Date	2019-02-28T23:17:43
Spectrometer Frequency	300.09
Spectral Width	4796.2
Lowest Frequency	-597.2
Nucleus	1H
Acquired Size	11990
Spectral Size	65536

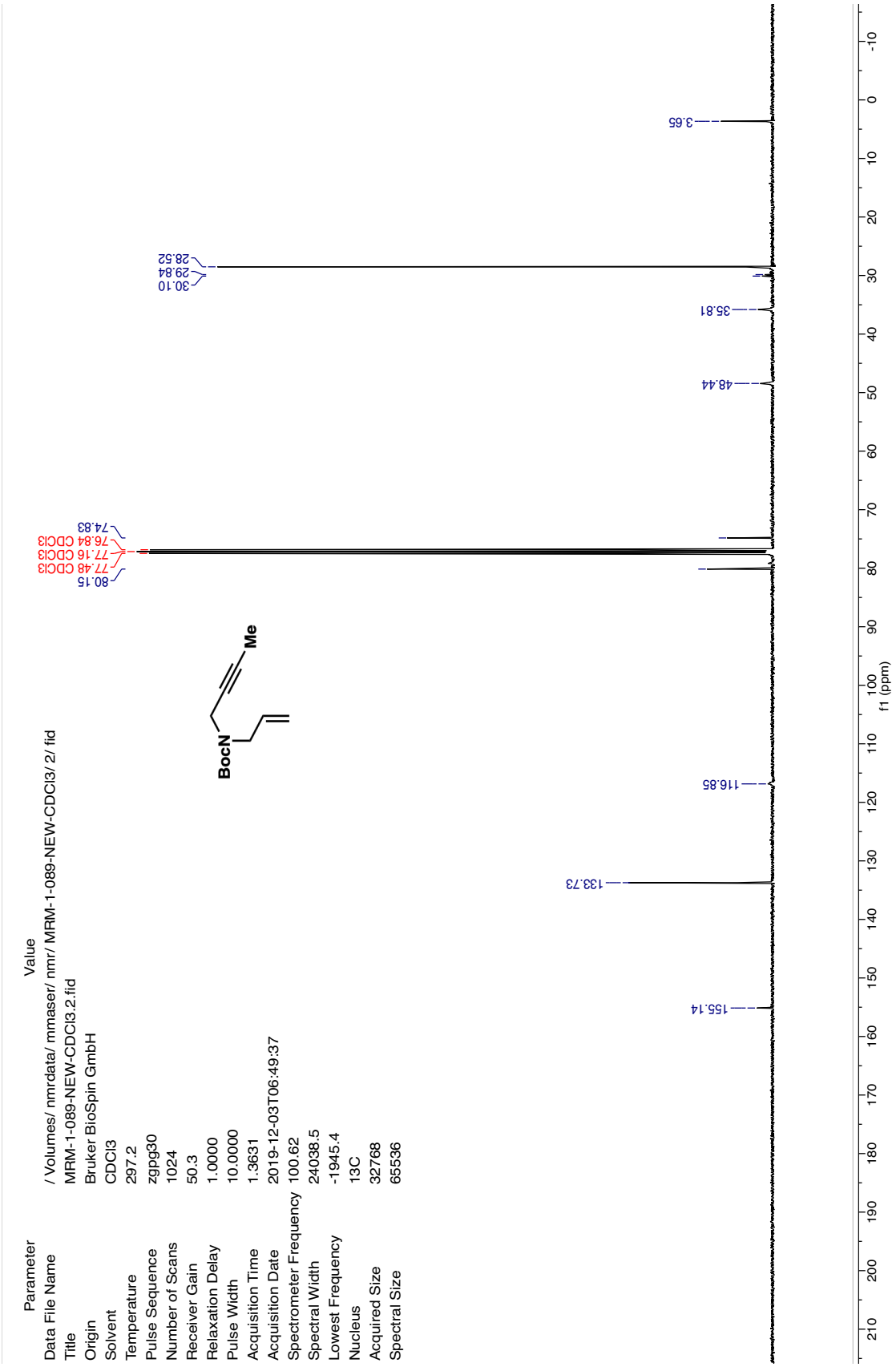


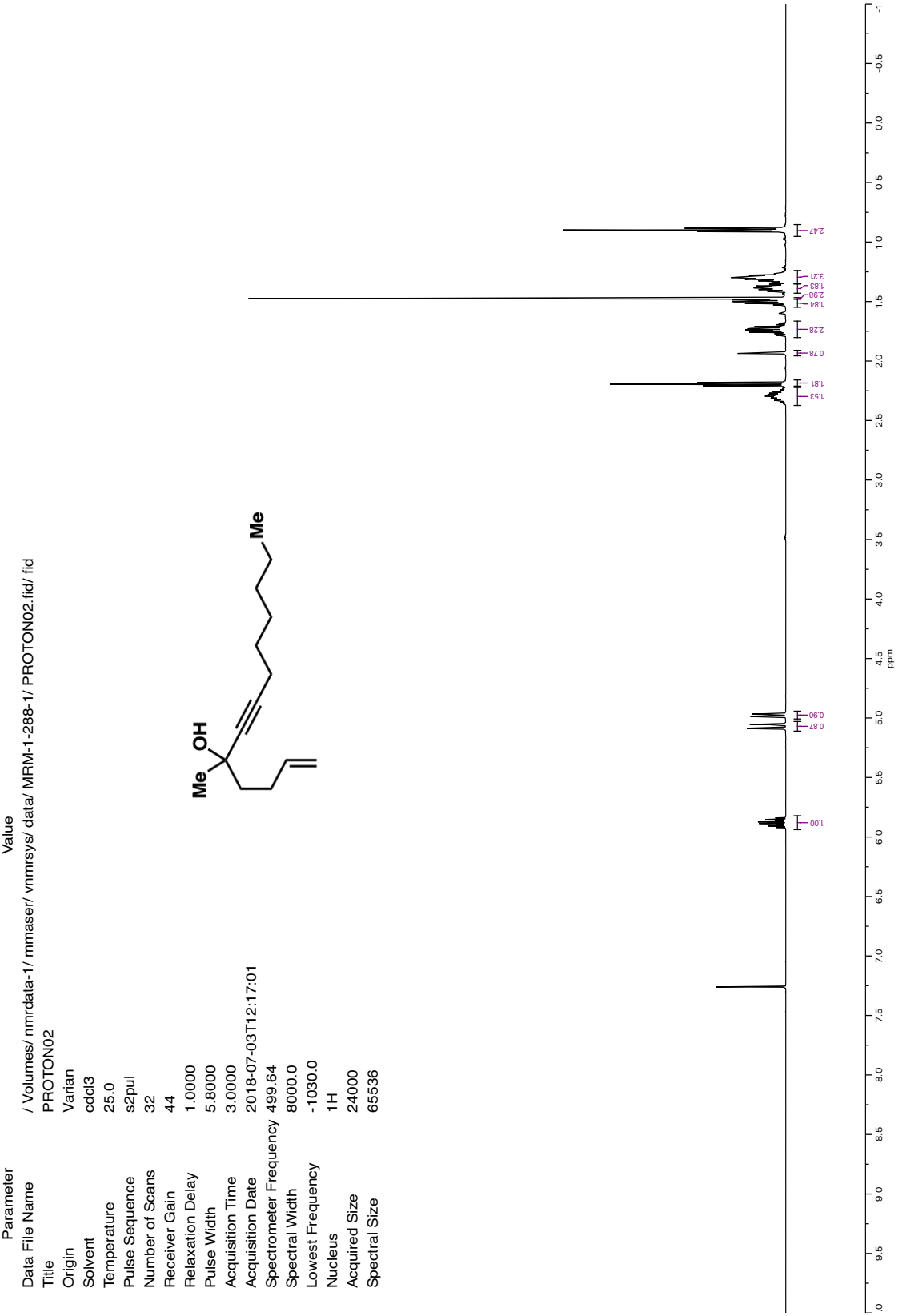


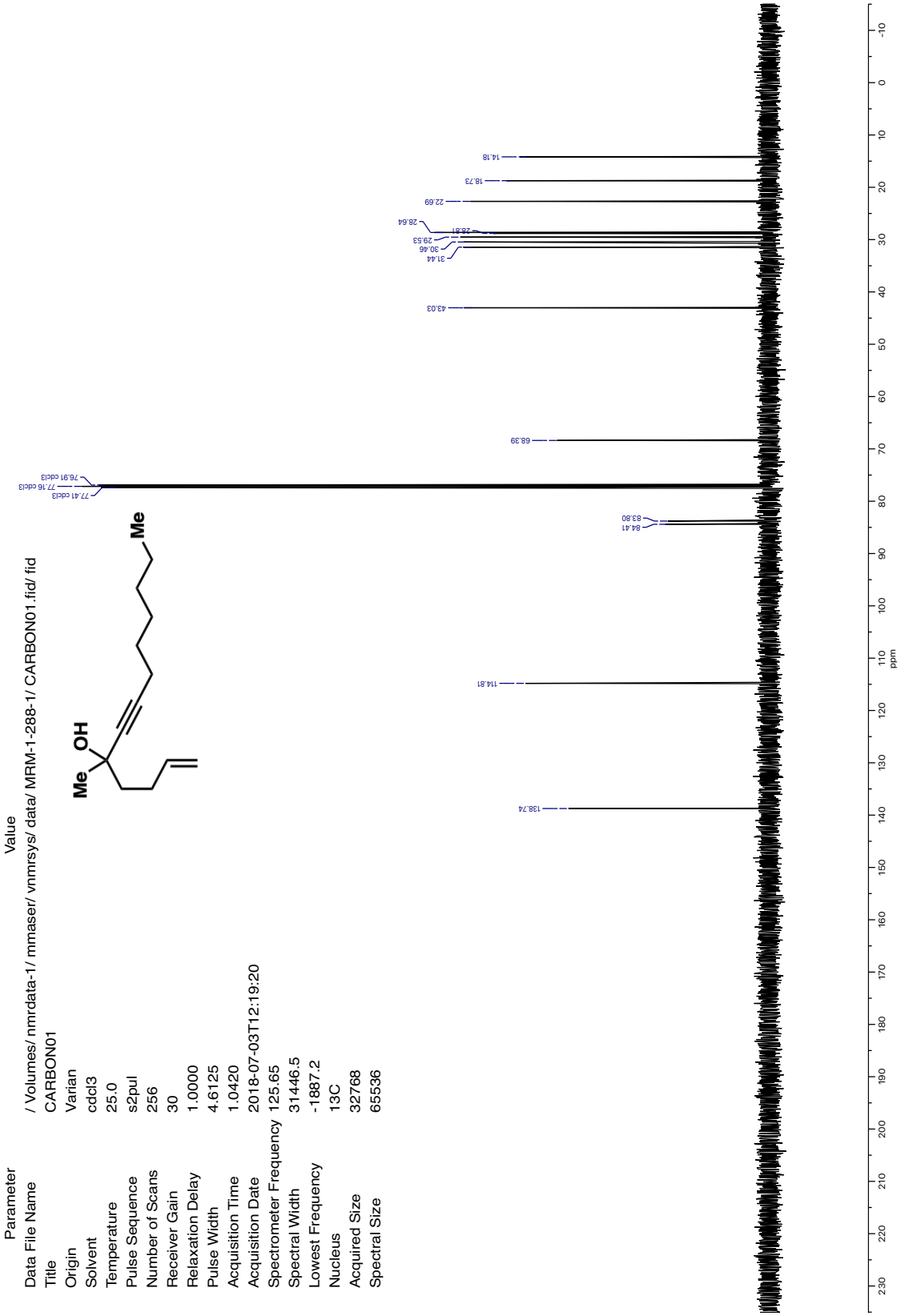
Parameter	Value
Data File Name	/ Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-123-B/ 2/ fid
Title	SED-1-123-B.2.fid
Origin	Bruker BioSpin GmbH
Solvent	CDCl3
Temperature	296.1
Pulse Sequence	zg30
Number of Scans	16
Receiver Gain	197.4
Relaxation Delay	1.0000
Pulse Width	12.5000
Acquisition Time	4.0894
Acquisition Date	2019-07-12T20:31:29
Spectrometer Frequency	400.13
Spectral Width	8012.8
Lowest Frequency	-1545.4
Nucleus	1H
Acquired Size	32768
Spectral Size	65536

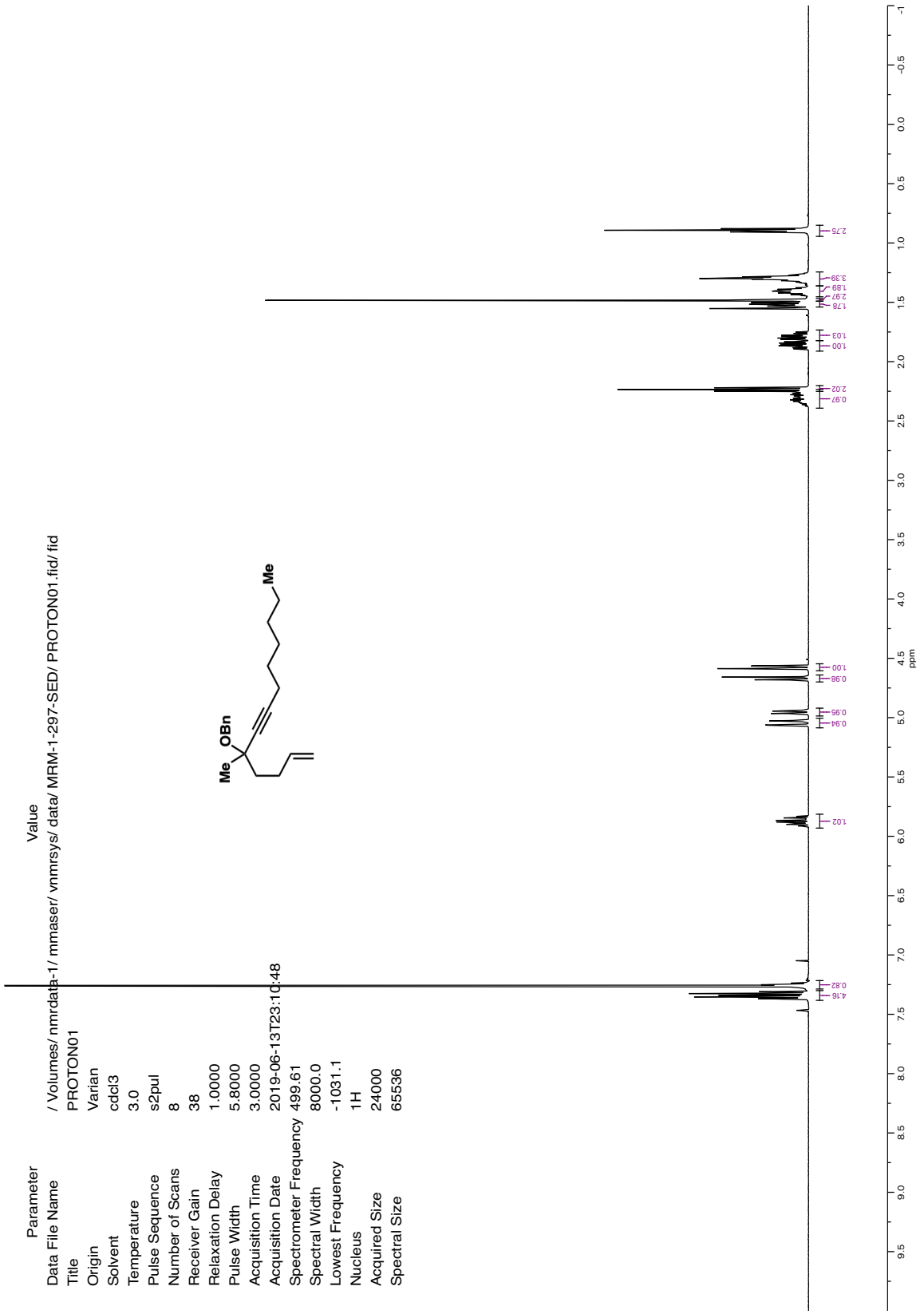

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-123-B/ 3/ fid
Title SED-1-123-B.3.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-07-12T20:52:43
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.7
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
Me
O
29.08
65.76
70.92
76.84 CDCl3
77.16 CDCl3
77.48 CDCl3
84.32
91.46
116.59
122.98
128.34
128.39
131.82
135.65
f1 (ppm)

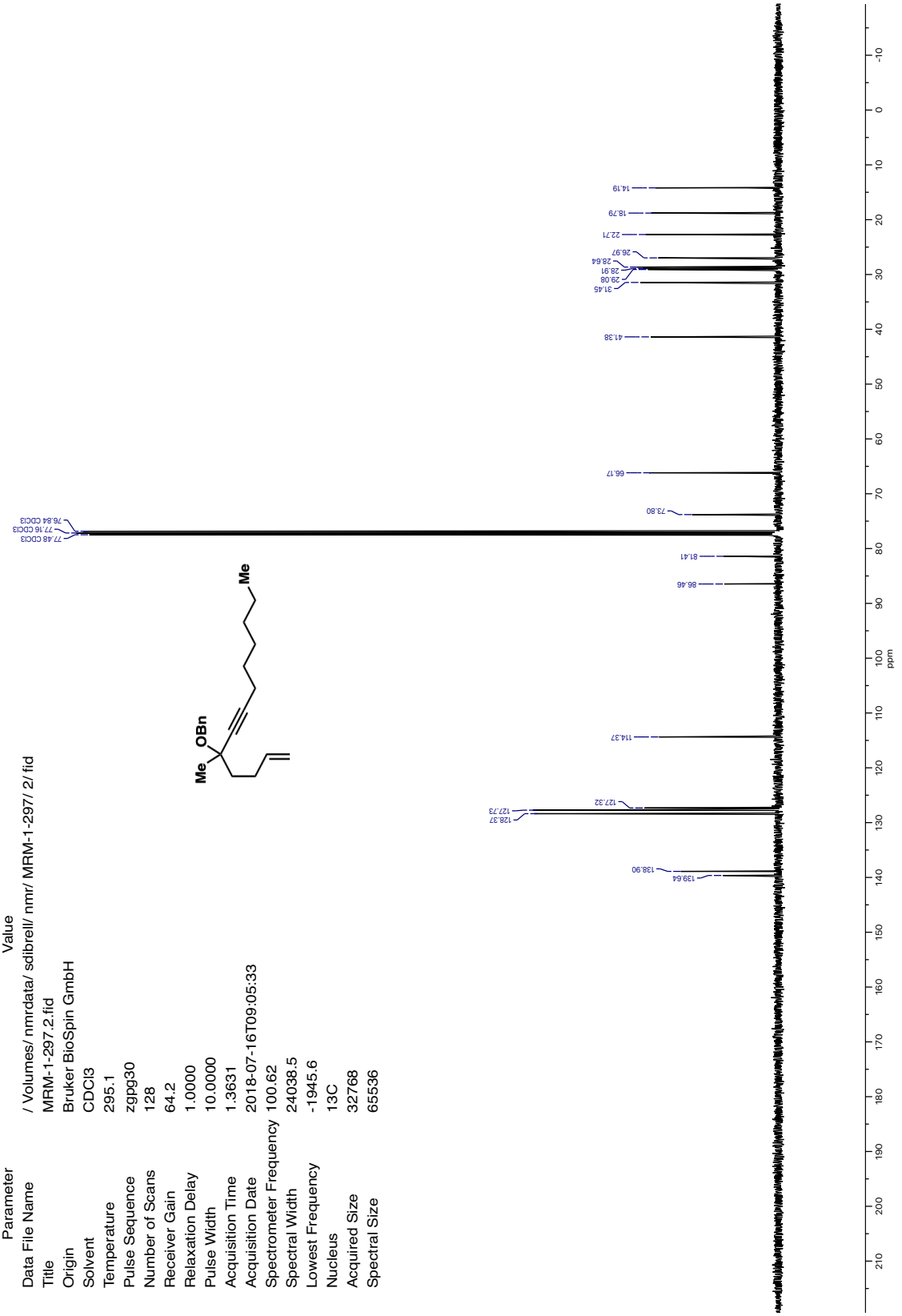

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-089-NEW-CDCl3/ 1/ fid
Title MRM-1-089-NEW-CDCl3.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 98.9
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-12-03T06:07:44
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
BocN
Me
0.90
1.81
2.00
2.33
2.63
8.39
ppm


Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-089-NEW-CDCl3/ 2/ fid
Title MRM-1-089-NEW-CDCl3.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-12-03T06:49:37
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
BocN
Me
80.15
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
74.83
30.10
29.84
28.52
3.65
35.81
48.44
116.85
133.73
155.14
f1 (ppm)

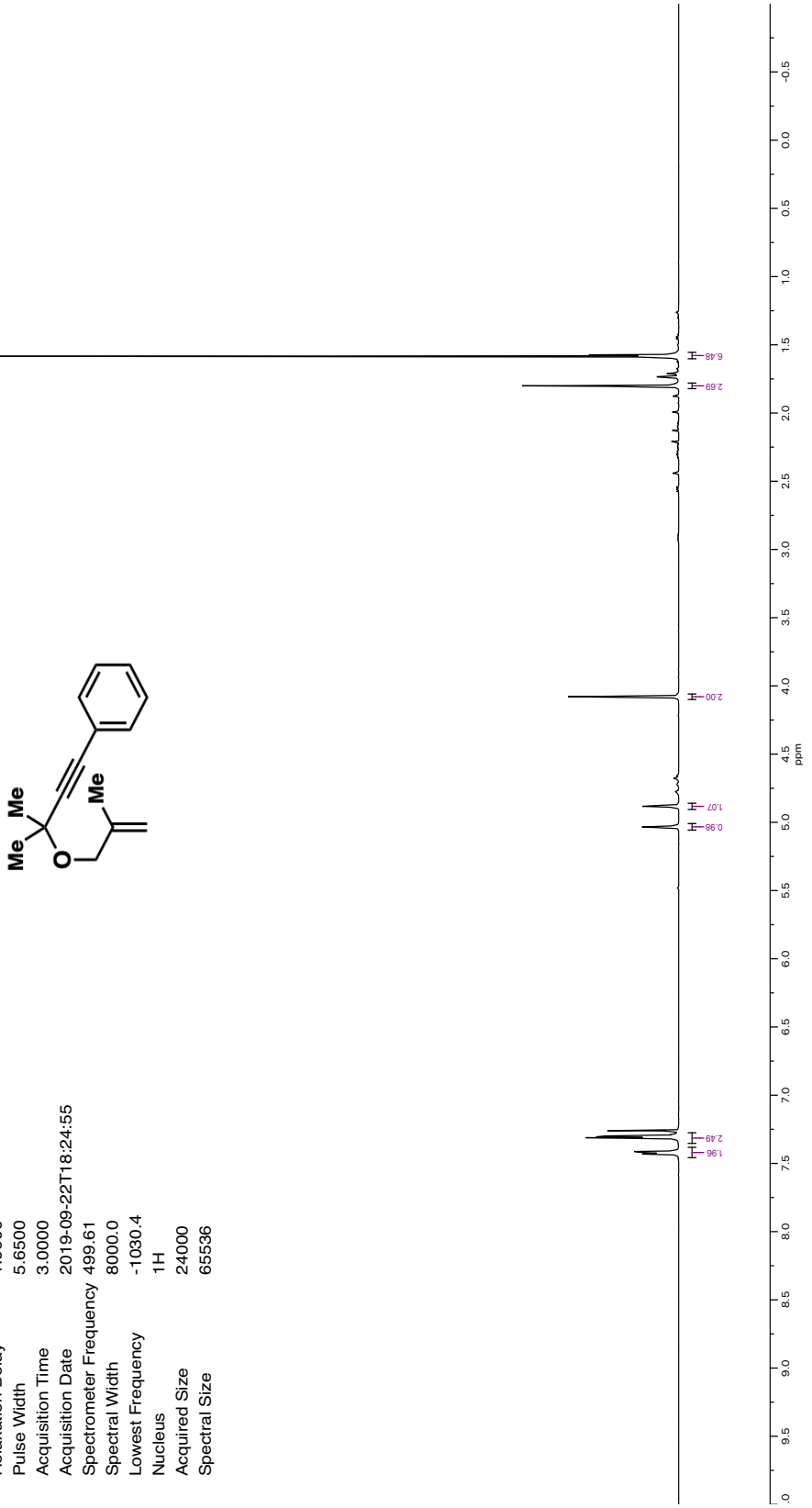

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-288-1/ PROTON02.fid/ fid
Title PROTON02
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 44
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2018-07-03T12:17:01
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1030.0
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
Me
OH
Me
1.00
0.87
0.90
1.53
1.81
0.78
2.28
1.84
2.98
1.83
3.21
2.47
ppm

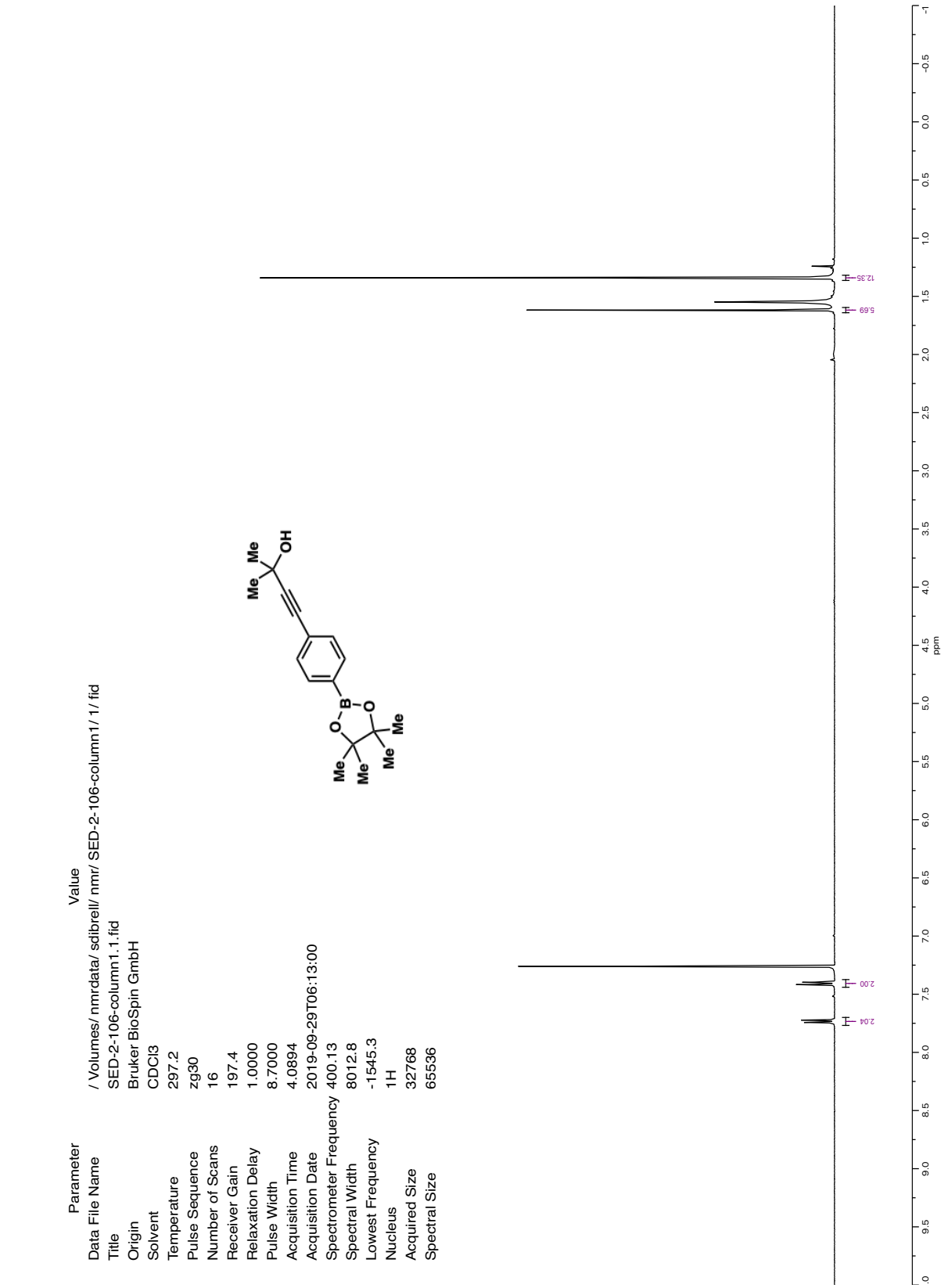

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-288-1/ CARBON01.fid/ fid
Title CARBON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 256
Receiver Gain 30
Relaxation Delay 1.0000
Pulse Width 4.6125
Acquisition Time 1.0420
Acquisition Date 2018-07-03T12:19:20
Spectrometer Frequency 125.65
Spectral Width 31446.5
Lowest Frequency -1887.2
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OH
Me
77.41 cdcl3
77.16 cdcl3
76.91 cdcl3
138.74
114.81
84.41
83.80
68.39
43.03
31.44
30.46
29.53
28.81
28.64
22.69
18.73
14.18
ppm

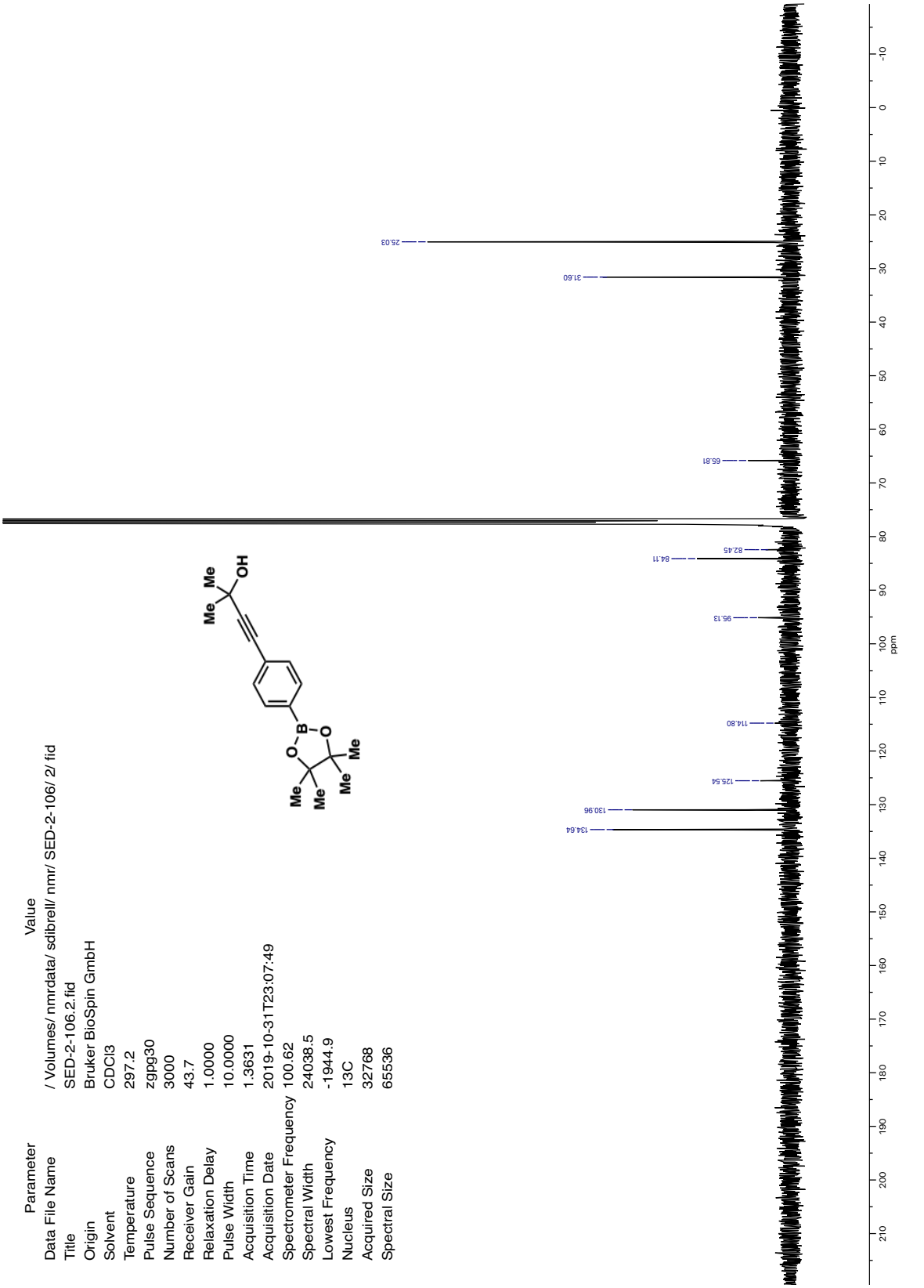

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ vnmrsys/ data/ MRM-1-297-SED/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 3.0
Pulse Sequence s2pul
Number of Scans 8
Receiver Gain 38
Relaxation Delay 1.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2019-06-13T23:10:48
Spectrometer Frequency 499.61
Spectral Width 8000.0
Lowest Frequency -1031.1
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
Me
OBn
Me
4.16
0.82
1.02
0.98
1.00
0.94
0.95
2.02
0.97
1.00
1.03
1.78
2.97
1.89
3.39
2.75
ppm

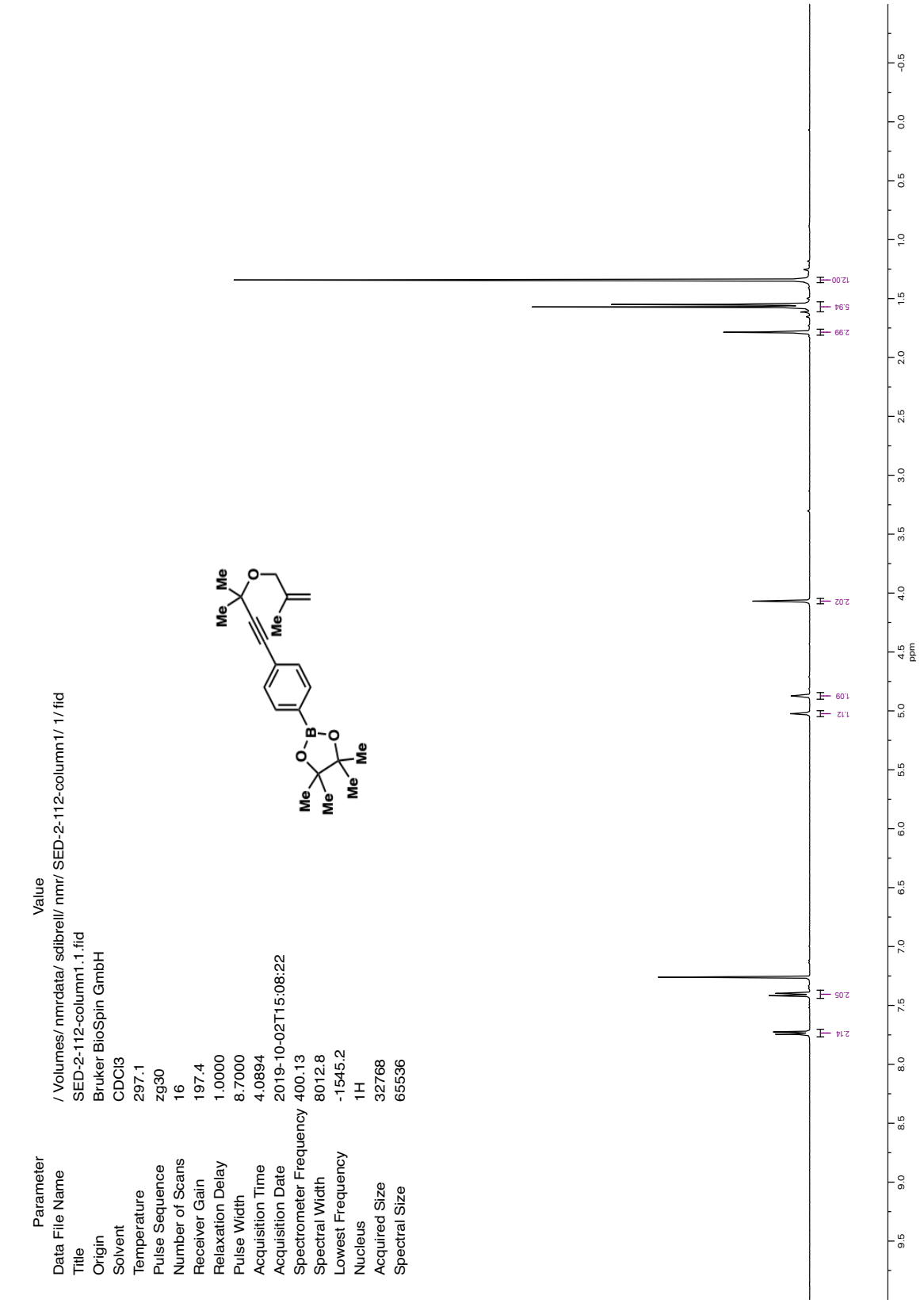

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ MRM-1-297/ 2/ fid
Title MRM-1-297.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.1
Pulse Sequence zgpg30
Number of Scans 128
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-07-16T09:05:33
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.6
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OBn
Me
76.84 CDCl3
77.16 CDCl3
77.48 CDCl3
14.19
18.79
22.71
26.97
28.64
28.91
29.08
31.45
41.38
66.17
73.80
81.41
86.46
114.37
127.32
127.73
128.37
138.64
138.90
ppm

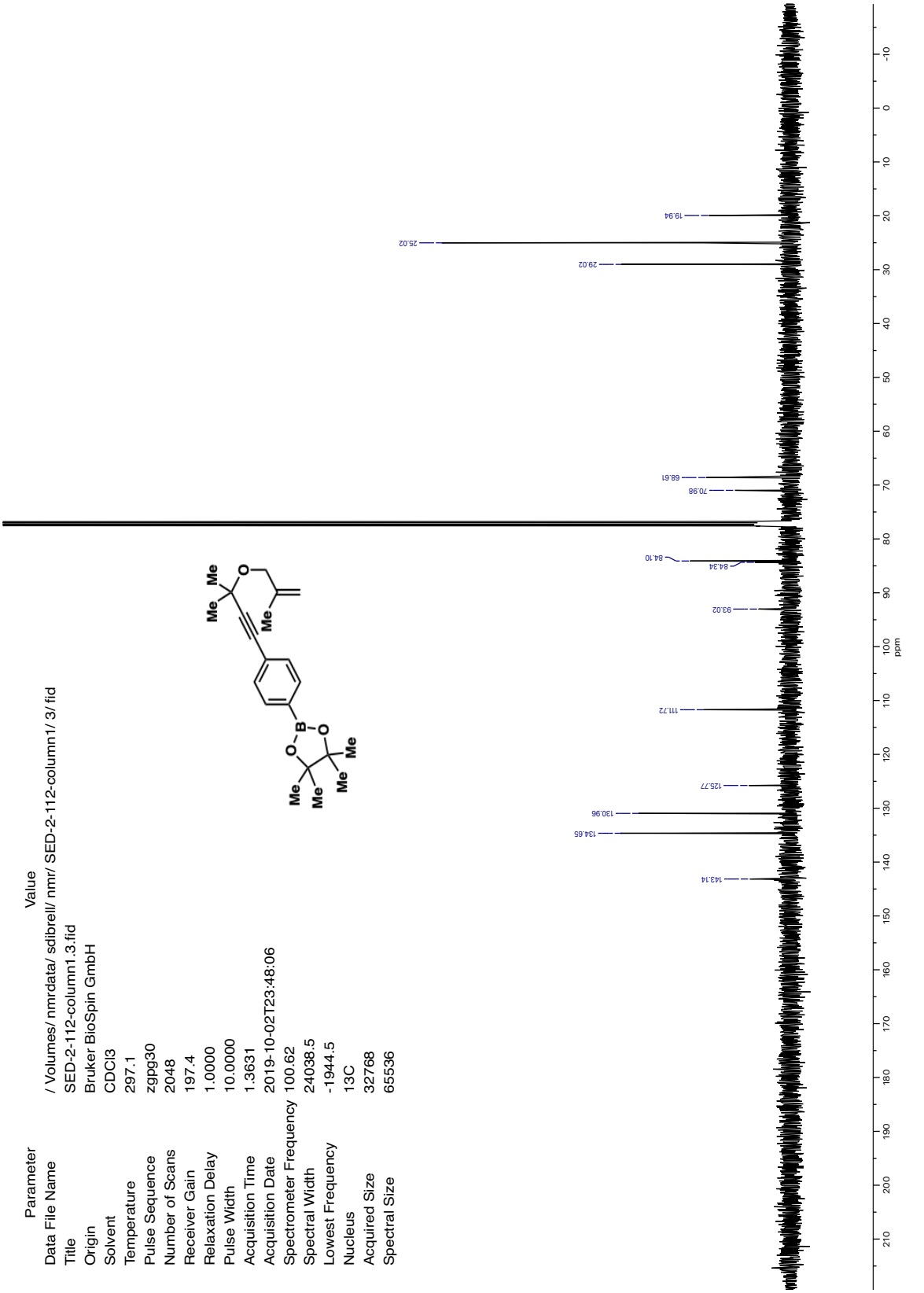
Parameter	Value
Data File Name	/ Volumes/ nmrdata-1/ sdibrell/ vnmrsys/ data/ SED-2-102-column2/ PROTON01.fid/ fid
Title	PROTON01
Origin	Varian
Solvent	cdcl3
Temperature	3.0
Pulse Sequence	s2pul
Number of Scans	8
Receiver Gain	36
Relaxation Delay	1.0000
Pulse Width	5.6500
Acquisition Time	3.0000
Acquisition Date	2019-09-22T18:24:55
Spectrometer Frequency	499.61
Spectral Width	8000.0
Lowest Frequency	-1030.4
Nucleus	1H
Acquired Size	24000
Spectral Size	65536

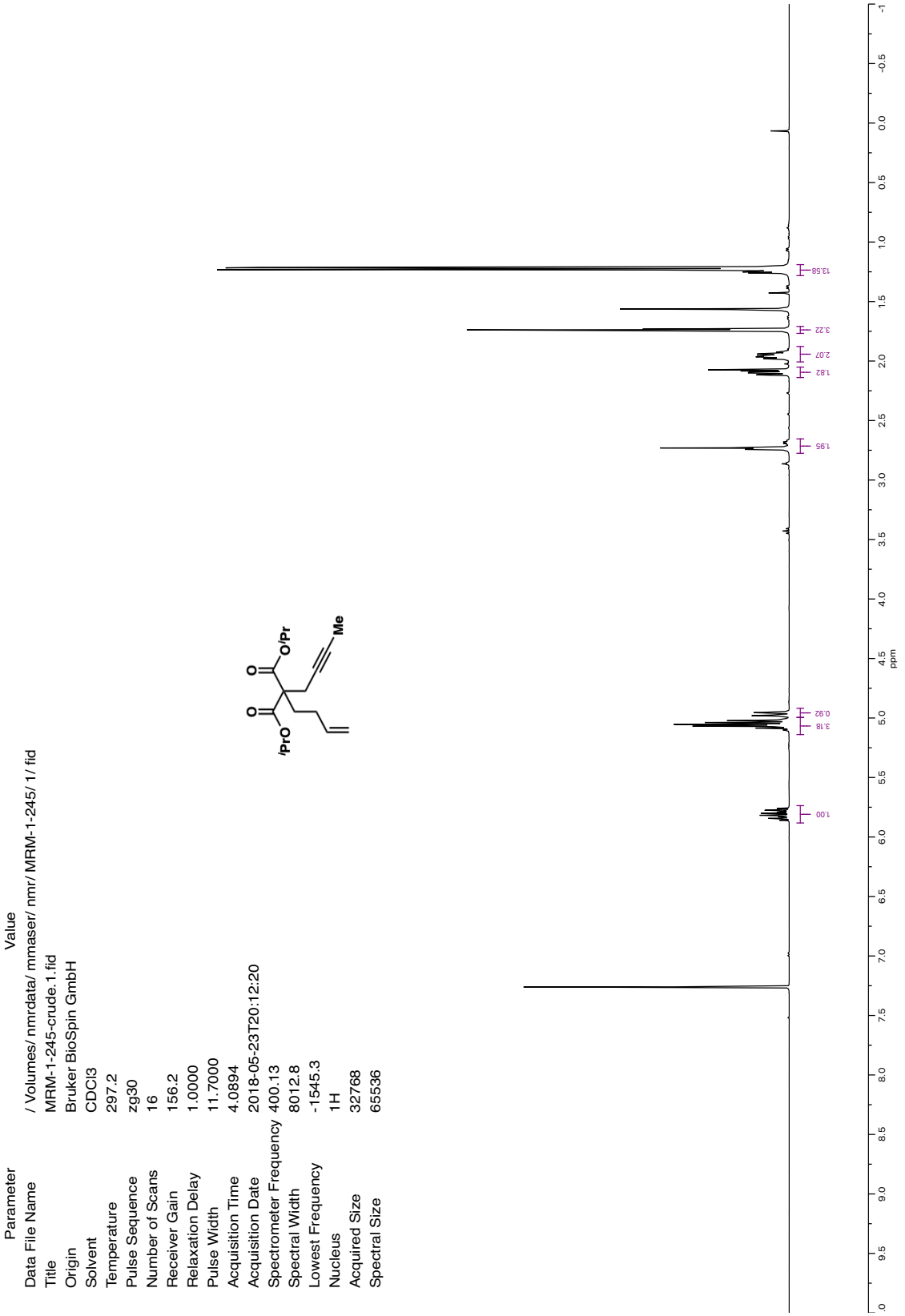


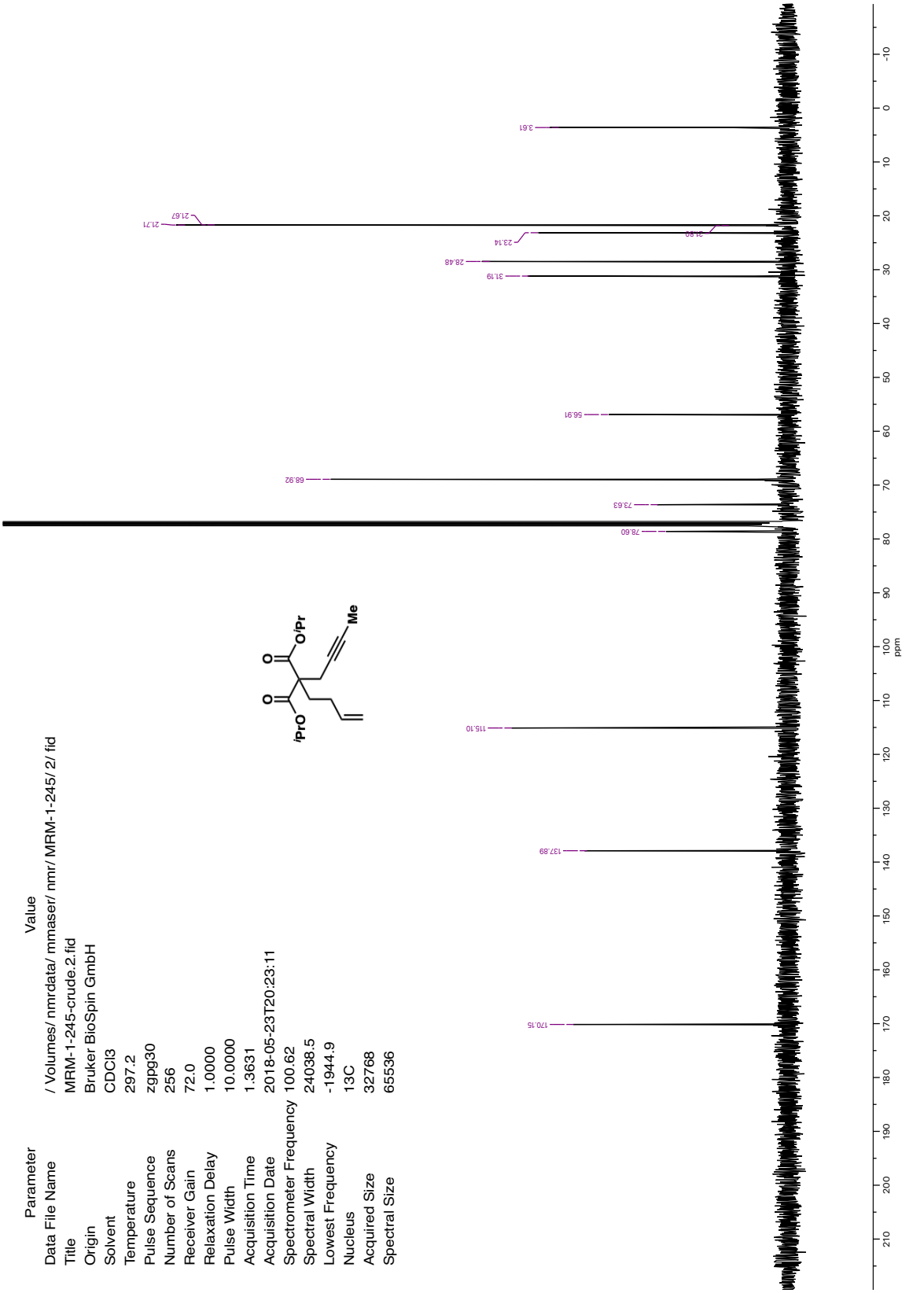

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-106-column1/ 1/ fid
Title SED-2-106-column1.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 8.7000
Acquisition Time 4.0894
Acquisition Date 2019-09-29T06:13:00
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
Me
OH
Me
Me
Me
Me
O
B
O
2.04
2.00
5.69
12.35
ppm

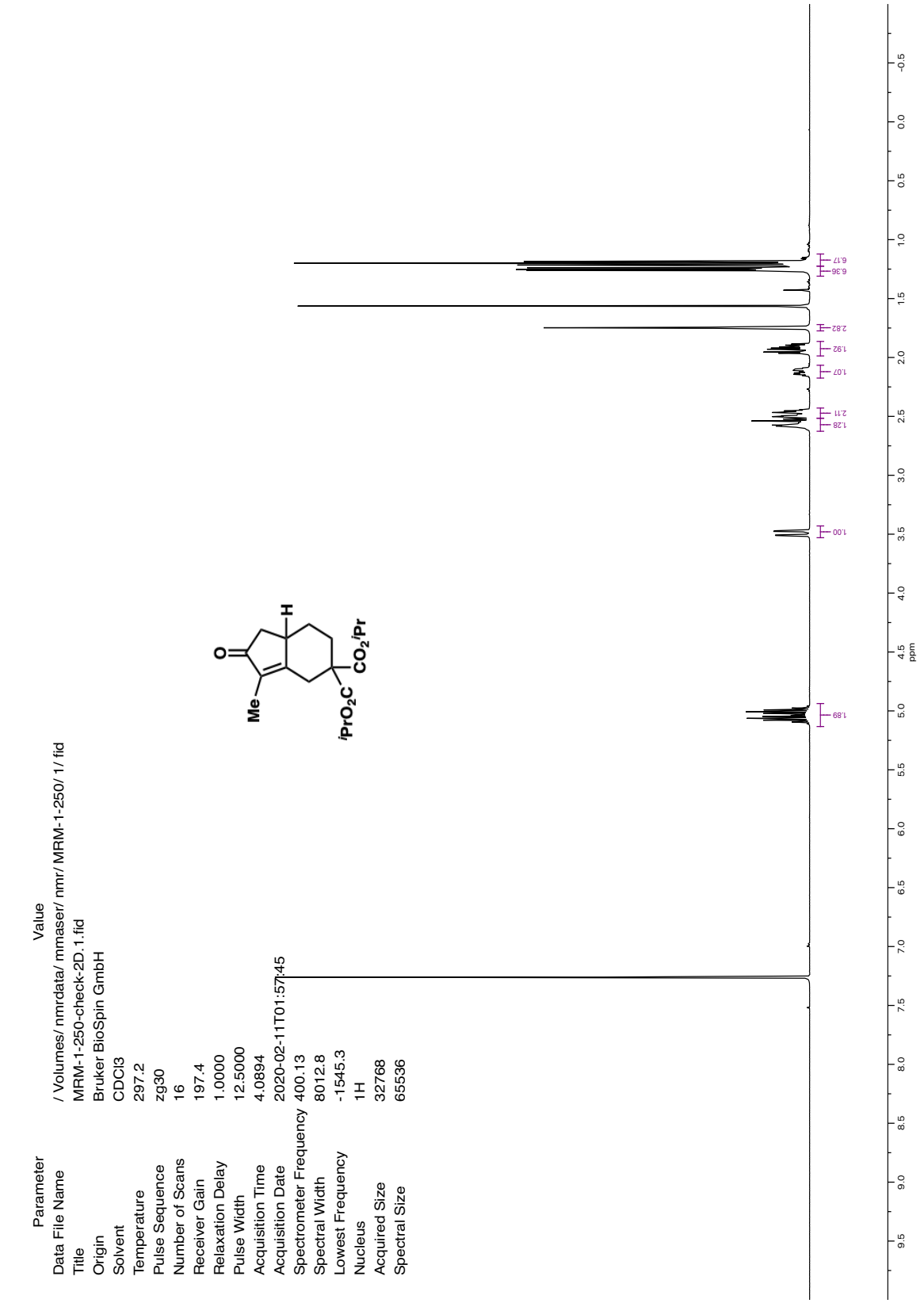

Parameter Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-106/ 2/ fid
Title SED-2-106.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 3000
Receiver Gain 43.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-10-31T23:07:49
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
Me
OH
B
O
O
Me
Me
Me
25.03
31.60
65.81
82.45
84.11
95.13
114.80
125.54
130.96
134.64
ppm

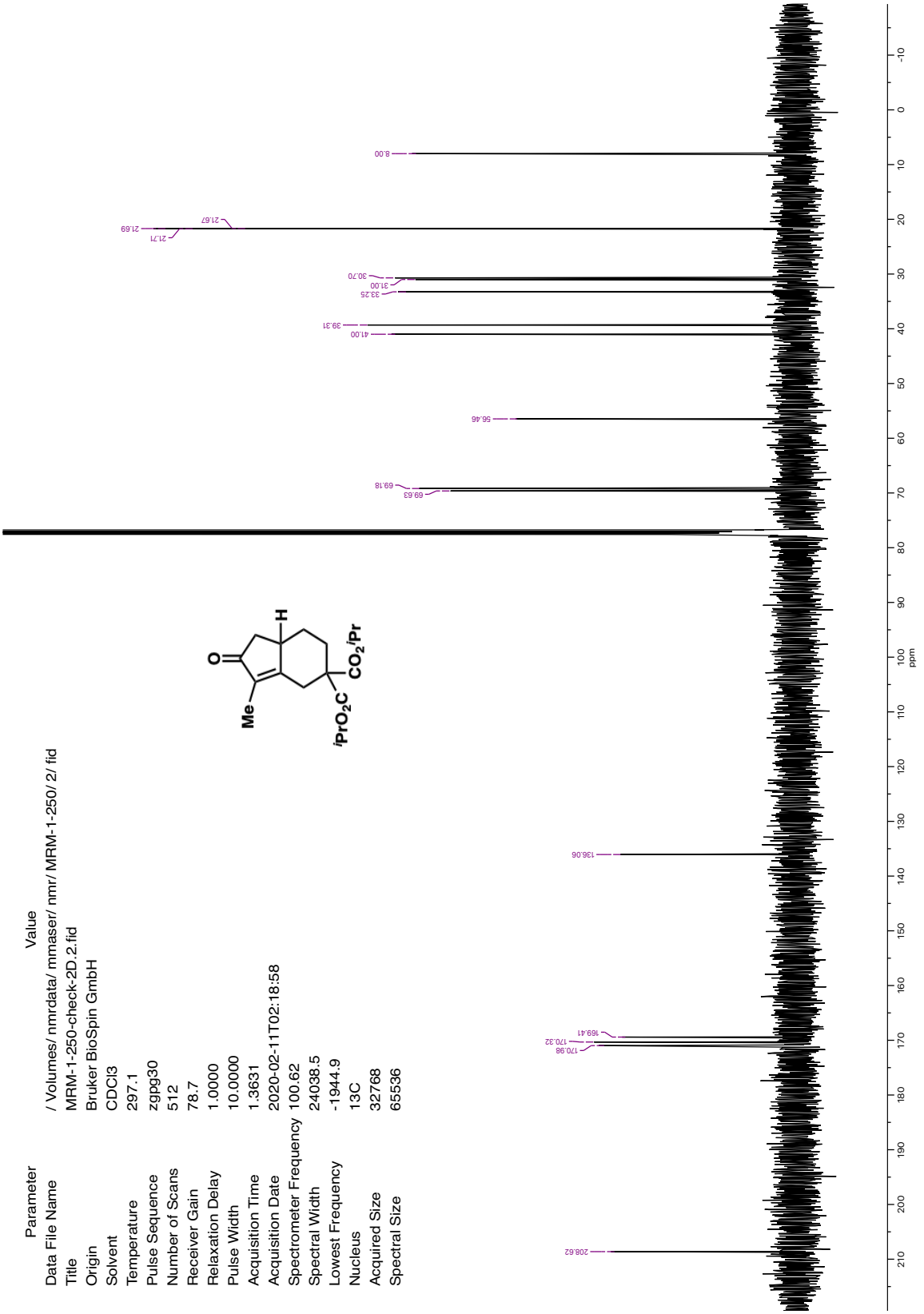

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-112-column1/ 1/ fid
Title SED-2-112-column1.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 8.7000
Acquisition Time 4.0894
Acquisition Date 2019-10-02T15:08:22
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
Me
O
Me
O
B
O
Me
Me
Me
Me
12.00
5.94
2.99
2.02
1.09
1.12
2.05
2.14
ppm

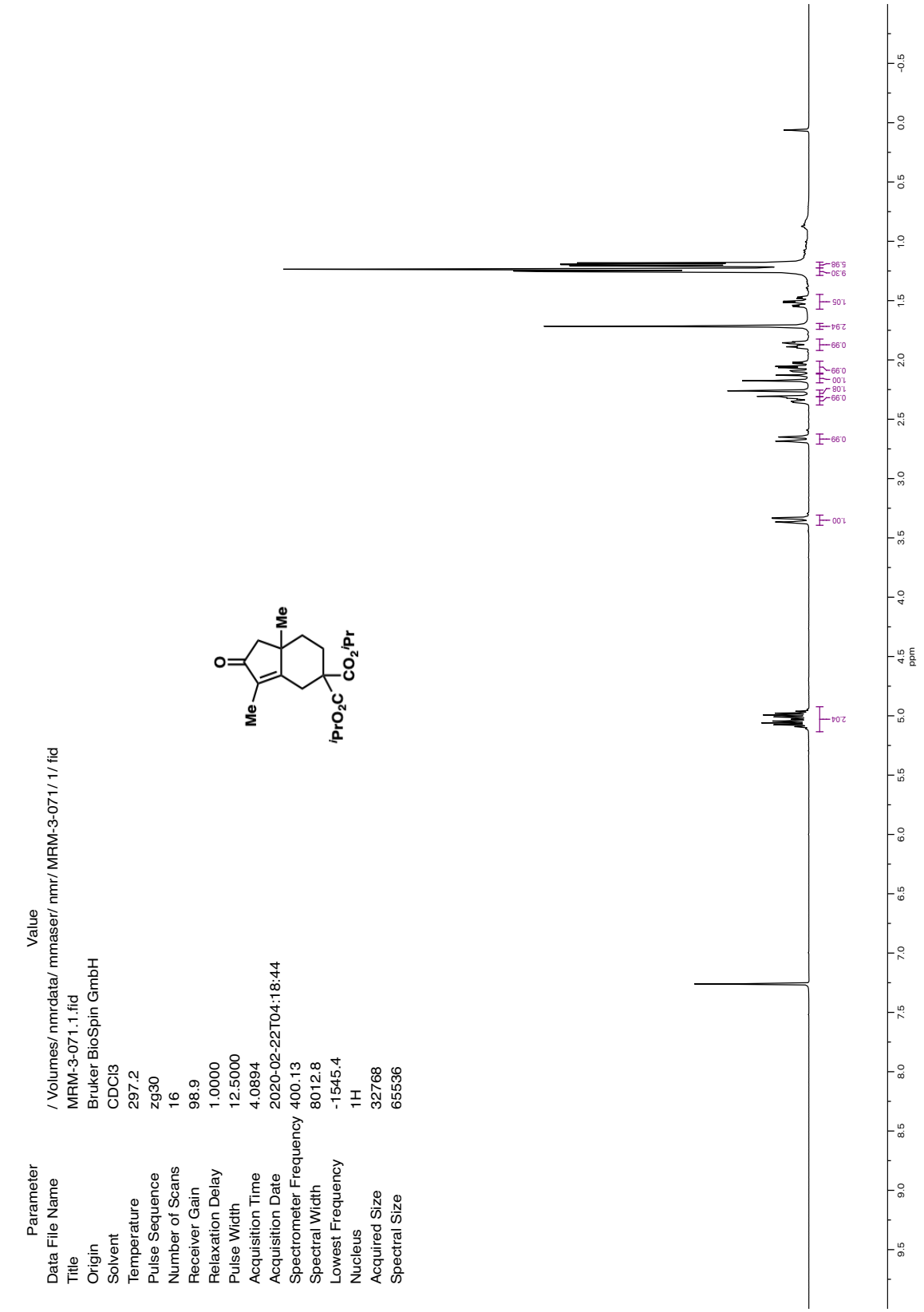

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-2-112-column1/ 3/ fid
Title SED-2-112-column1.3.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 2048
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-10-02T23:48:06
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.5
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
Me
O
Me
B
O
O
Me
Me
Me
Me
19.94
25.02
29.02
68.61
70.98
84.10
84.34
93.02
111.72
125.77
130.96
134.65
143.14
ppm

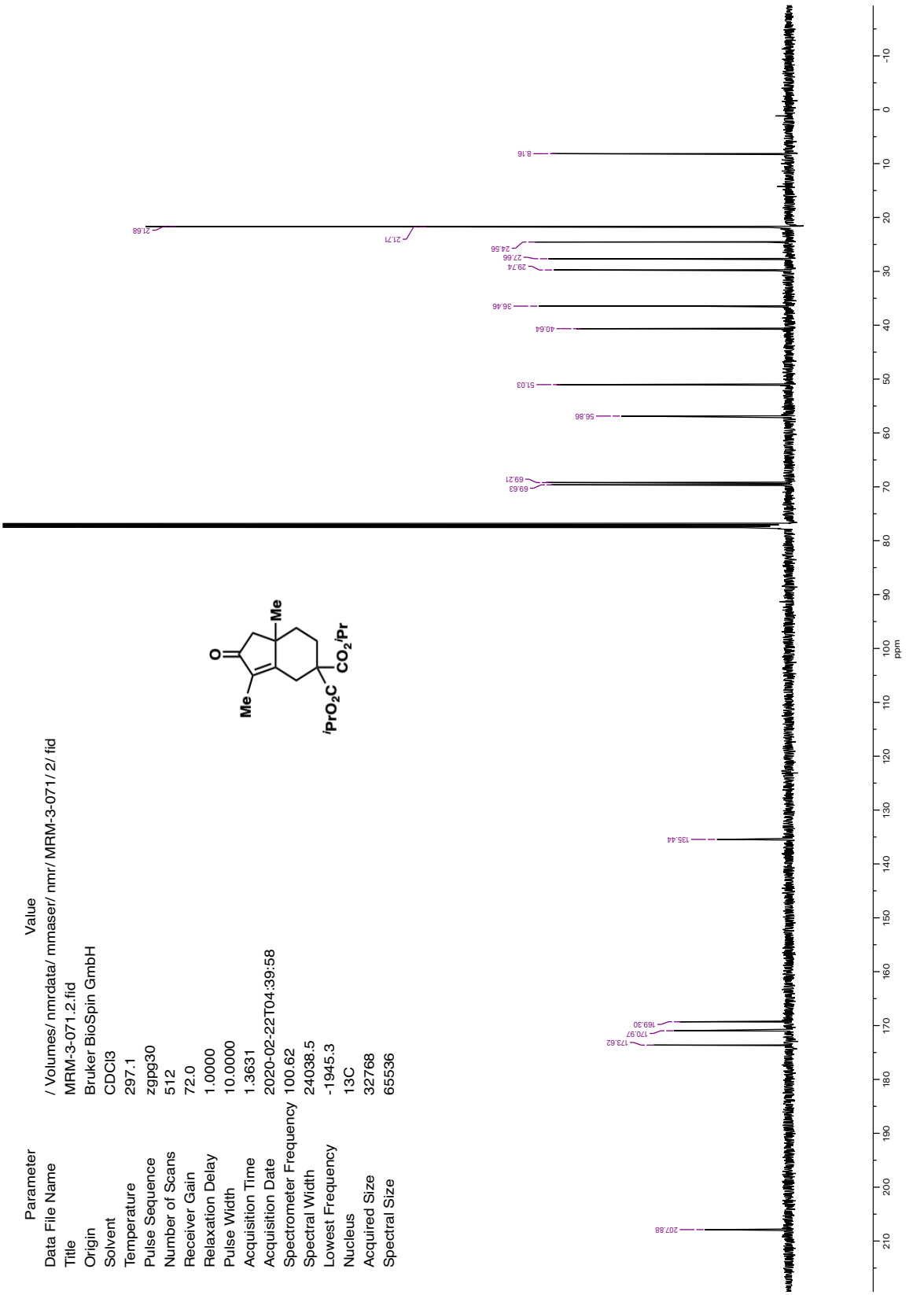

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-245/ 1/ fid
Title MRM-1-245-crude.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2018-05-23T20:12:20
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
O
O
iPr
iPrO
Me
13.58
3.22
2.07
1.82
1.95
0.92
3.18
1.00
ppm

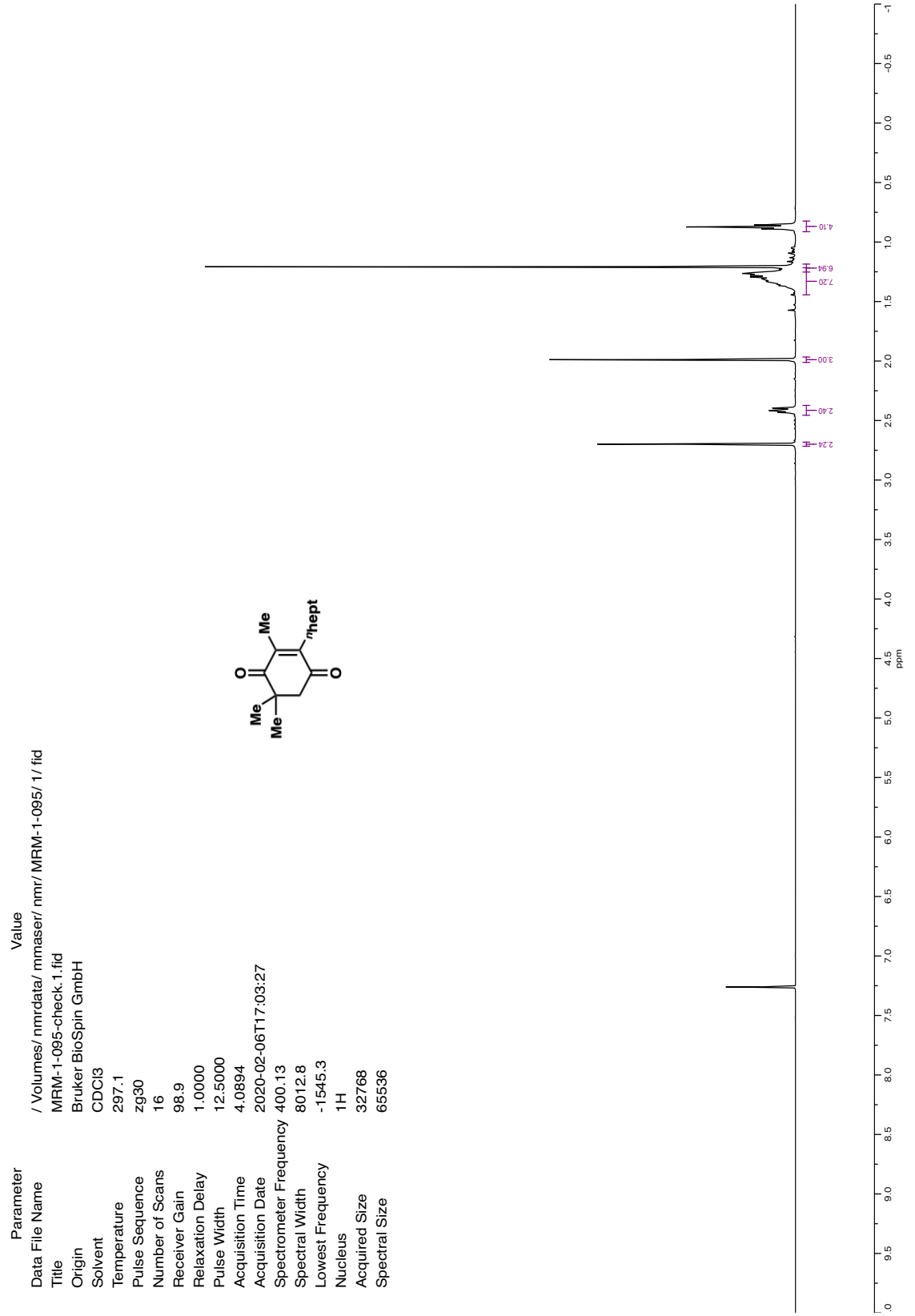

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-245/ 2/ fid
Title MRM-1-245-crude.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 256
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2018-05-23T20:23:11
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
O
O
O
iPrO
OiPr
Me
170.15
137.89
115.10
78.60
73.63
68.92
56.91
31.80
31.19
28.48
23.14
21.71
21.67
3.61
ppm


Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-250/ 1/ fid
Title MRM-1-250-check-2D.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-02-11T01:57:45
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
H
Me
iPrO2C
CO2iPr
1.89
1.00
2.11
1.28
1.07
1.92
2.82
6.36
6.17
ppm

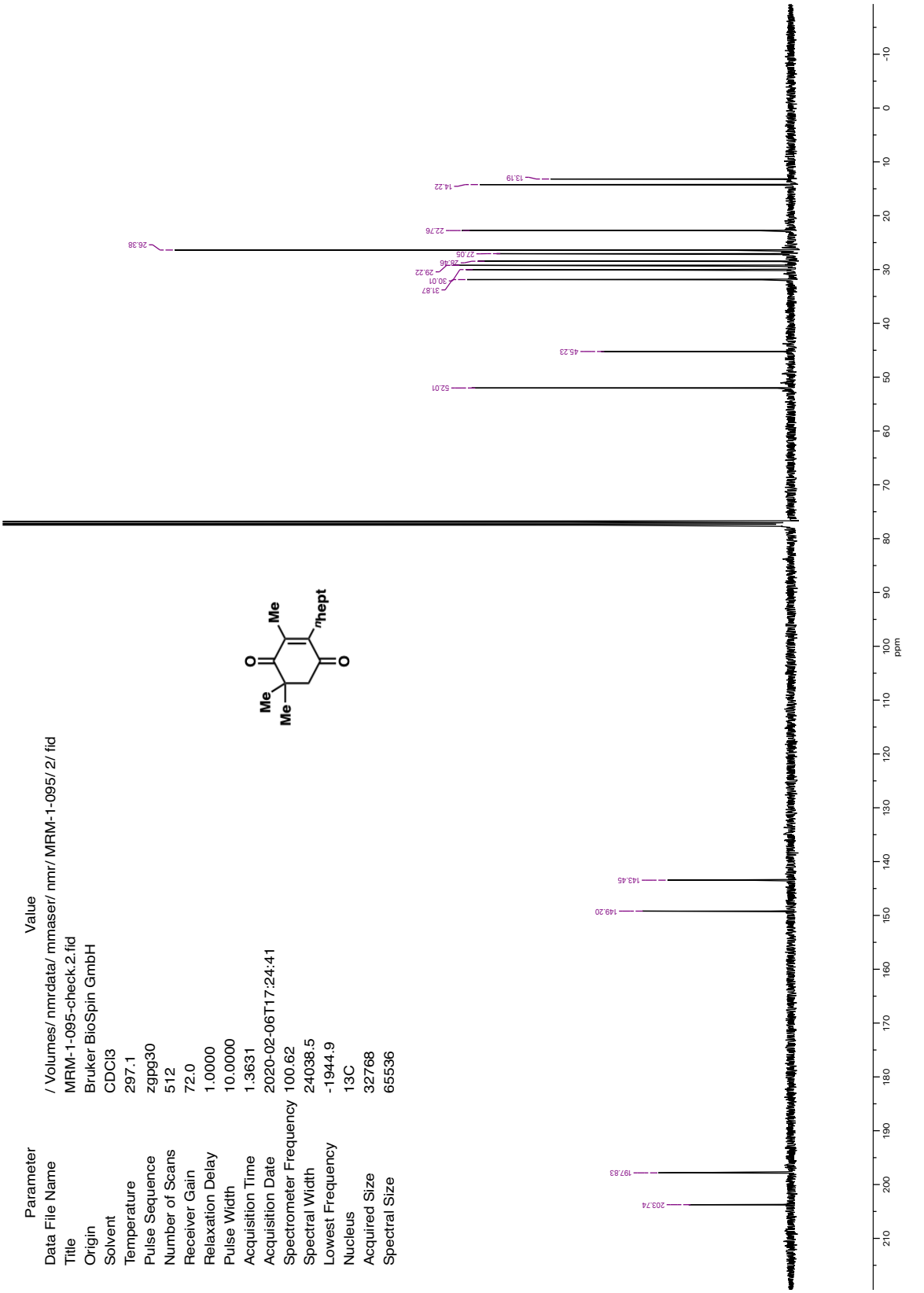

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-250/ 2/ fid
Title MRM-1-250-check-2D.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-02-11T02:18:58
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
O
H
Me
iPrO2C
CO2iPr
208.62
169.41
170.98
170.32
136.06
69.63
69.18
56.46
41.00
39.31
33.25
31.00
30.70
21.67
21.71
21.69
8.00
ppm

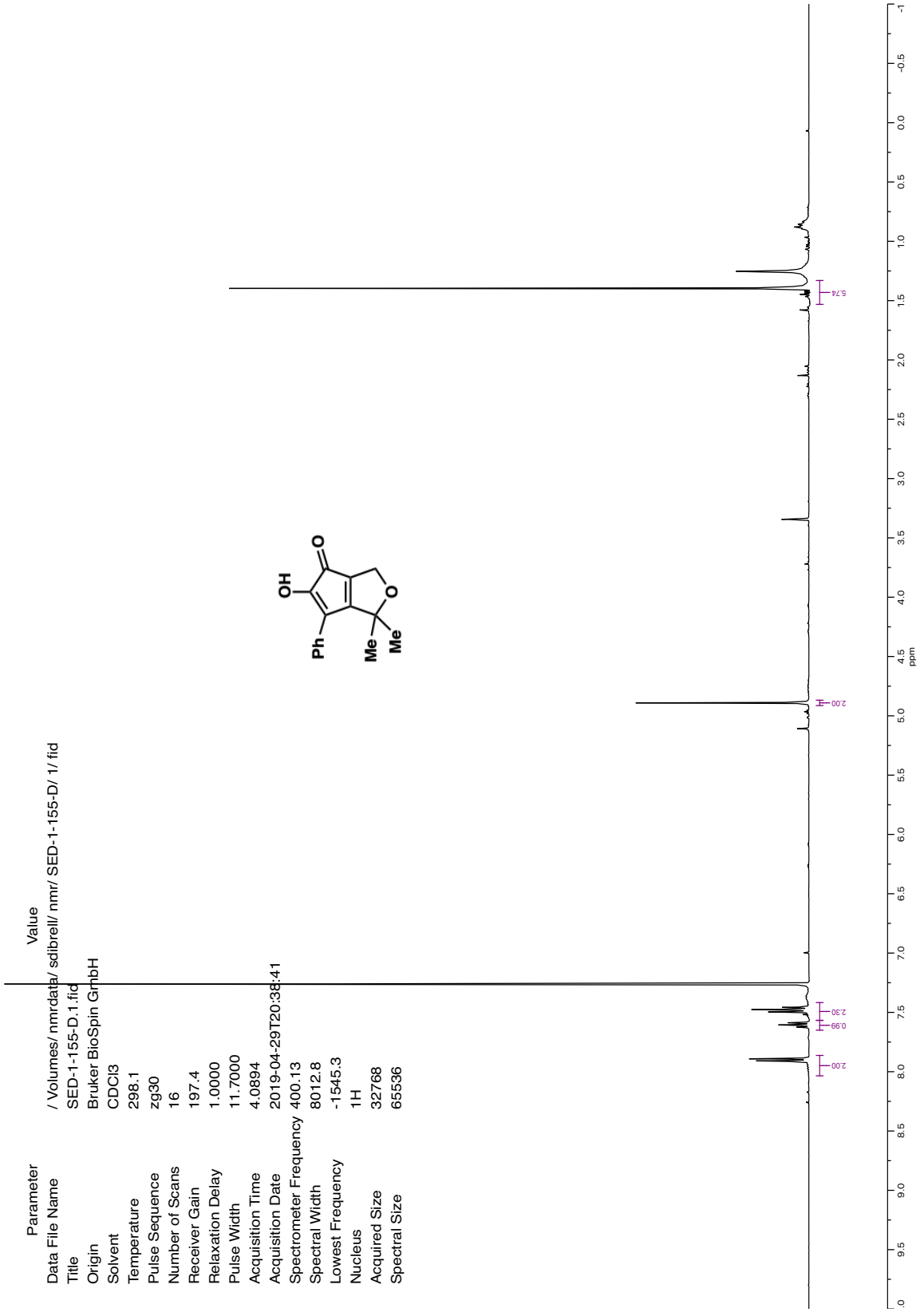

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-071/ 1/ fid
Title MRM-3-071.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 98.9
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-02-22T04:18:44
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
Me
Me
iPrO2C
CO2iPr
2.04
1.00
0.99
1.08
1.00
0.99
0.99
2.94
1.05
9.30
5.98
ppm

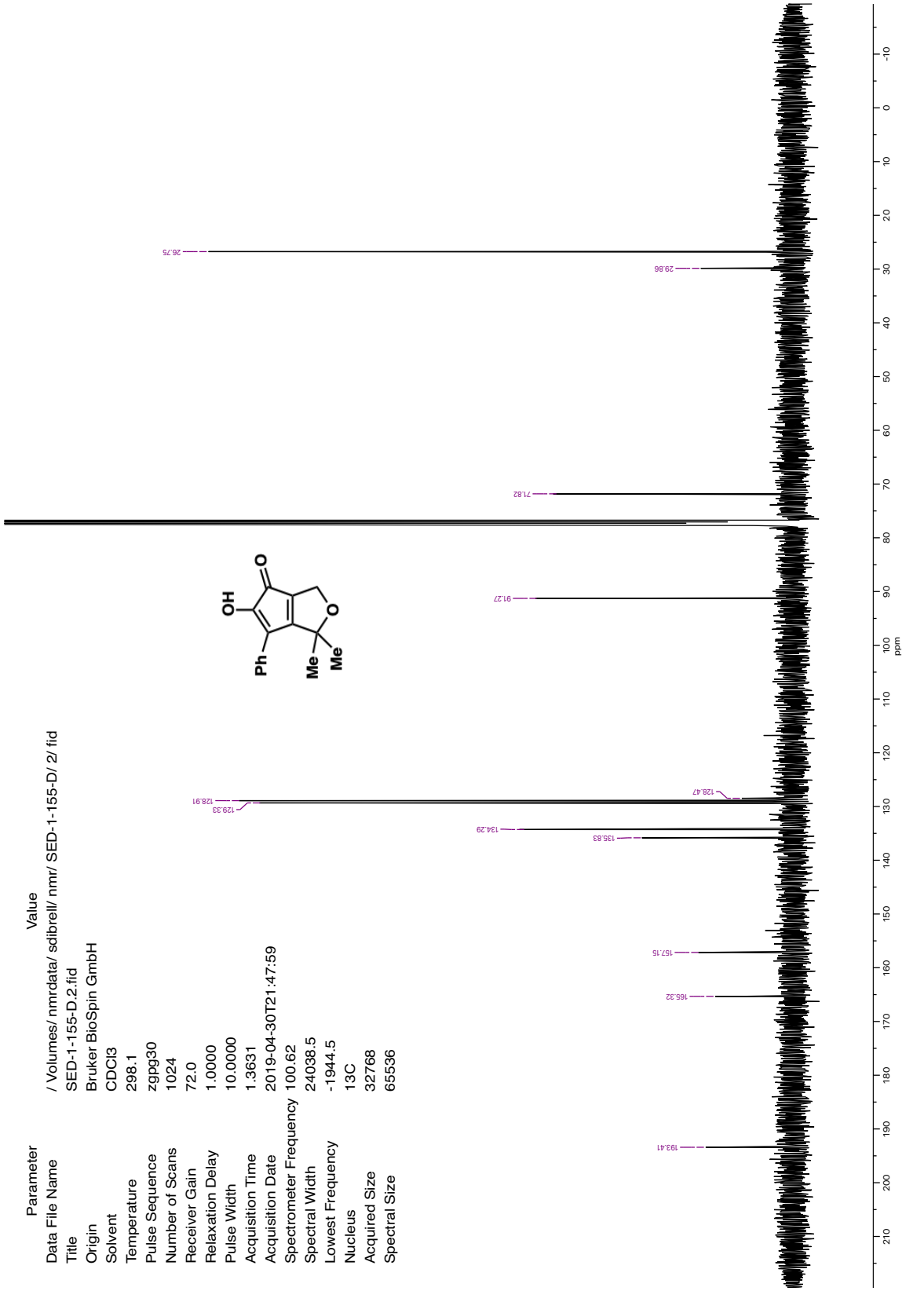

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-071/ 2/ fid
Title MRM-3-071.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-02-22T04:39:58
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
O
Me
Me
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
207.88
173.62
170.97
169.30
135.44
69.63
69.21
56.86
51.03
40.64
36.46
29.74
27.66
24.56
21.71
21.68
8.16
ppm

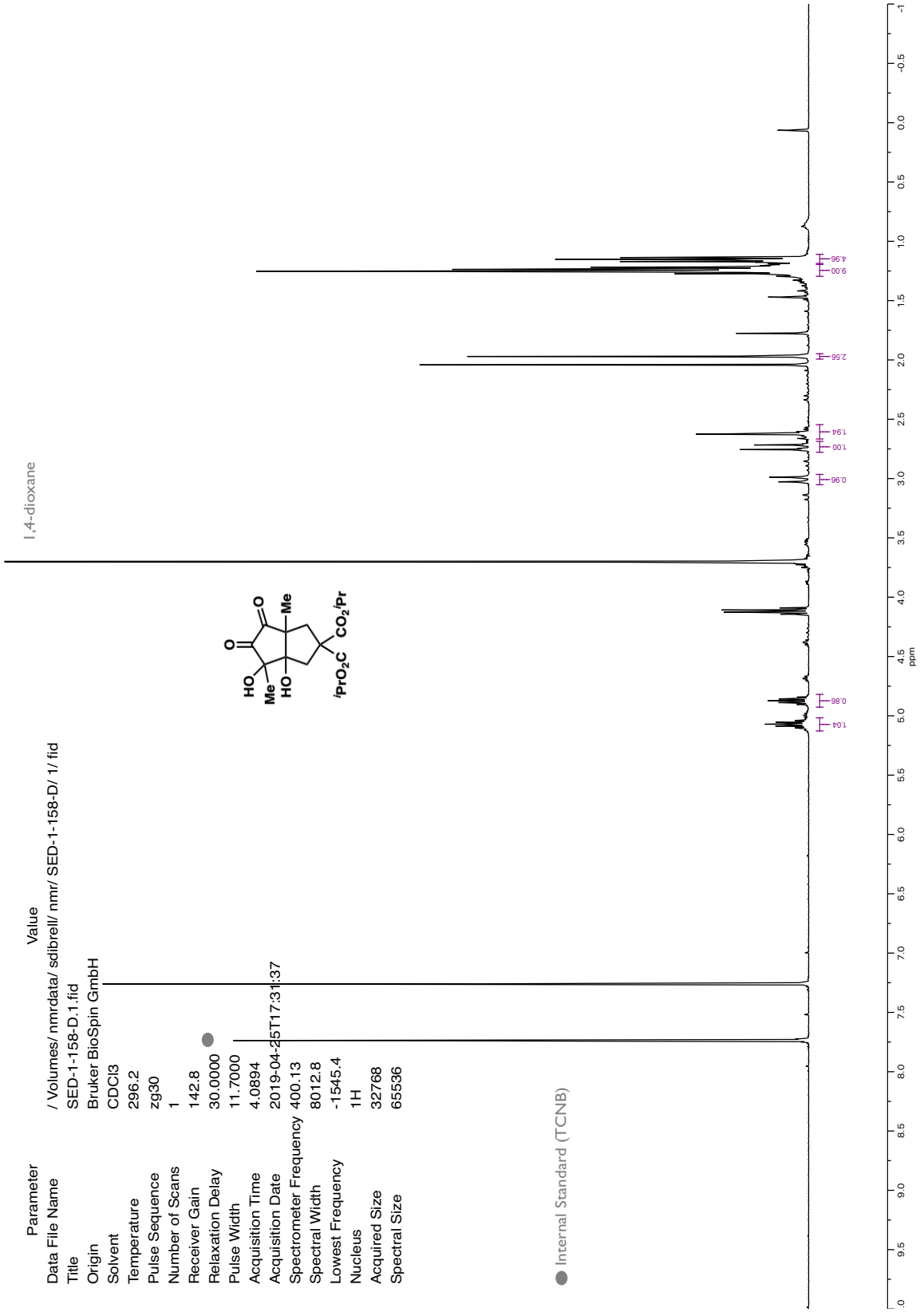


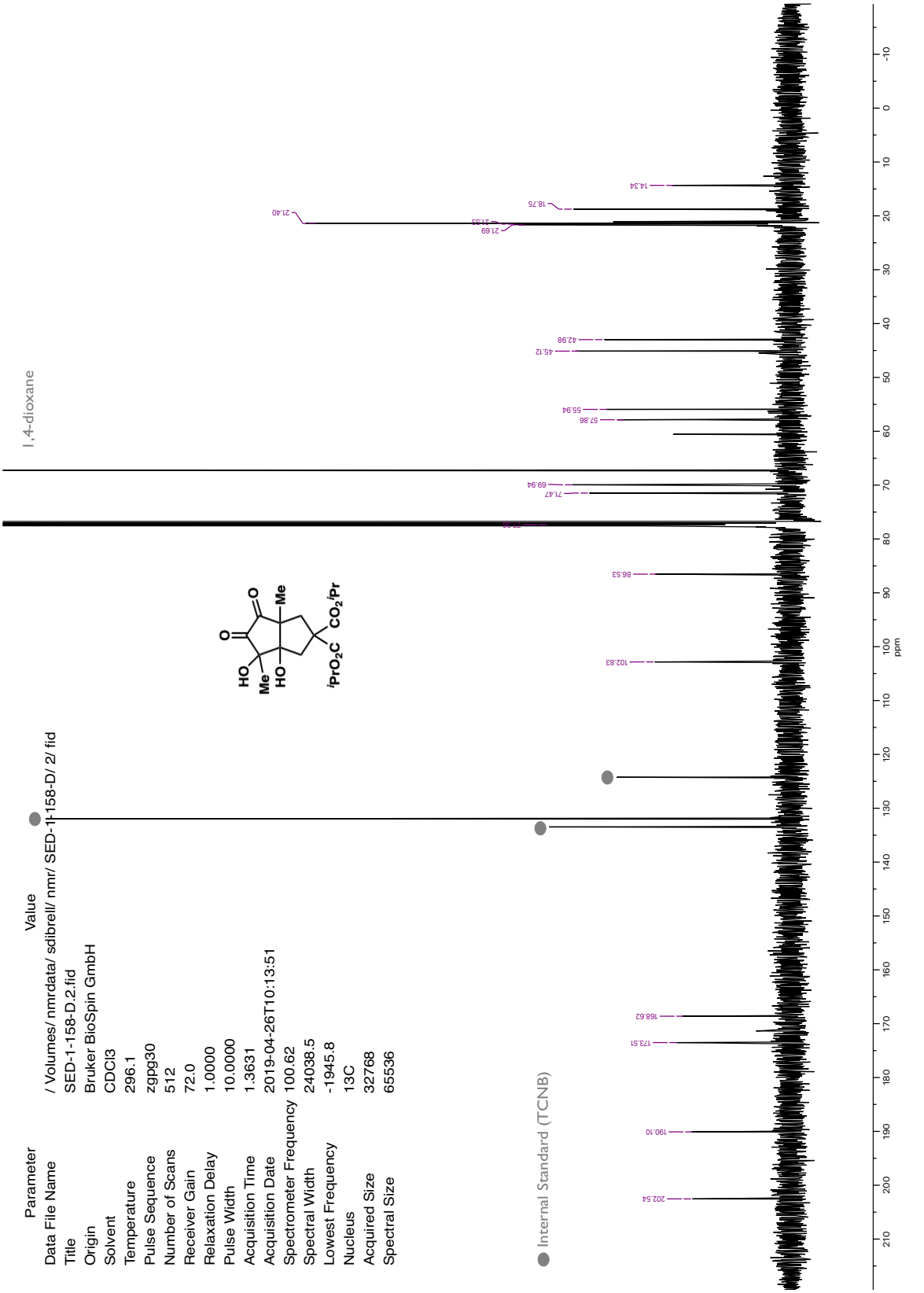
Parameter	Value
Data File Name	/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-095/ 1/ fid
Title	MRM-1-095-check.1.fid
Origin	Bruker BioSpin GmbH
Solvent	CDCl3
Temperature	297.1
Pulse Sequence	zg30
Number of Scans	16
Receiver Gain	98.9
Relaxation Delay	1.0000
Pulse Width	12.5000
Acquisition Time	4.0894
Acquisition Date	2020-02-06T17:03:27
Spectrometer Frequency	400.13
Spectral Width	8012.8
Lowest Frequency	-1545.3
Nucleus	1H
Acquired Size	32768
Spectral Size	65536

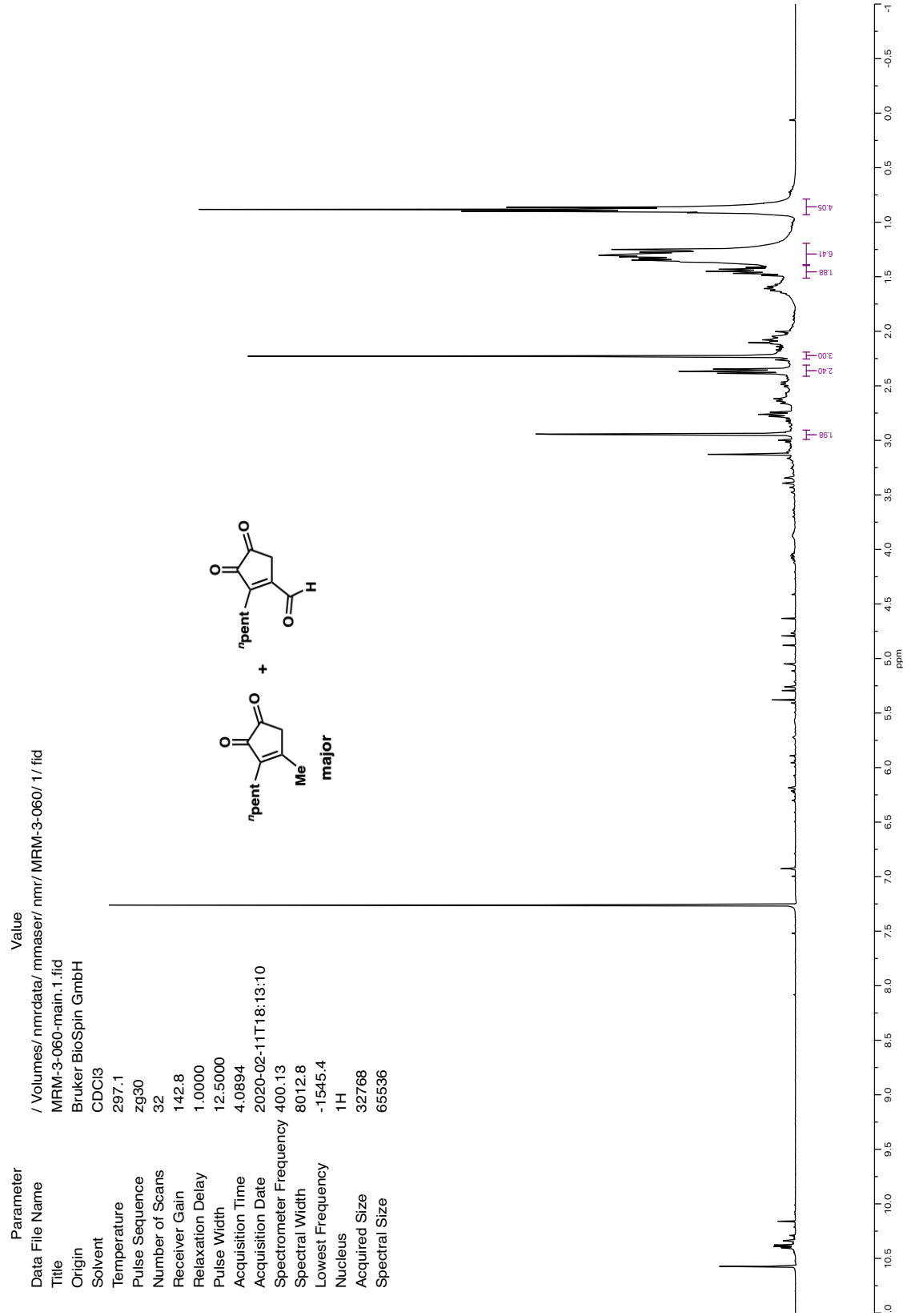

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-1-095/ 2/ fid
Title
MRM-1-095-check.2.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
zgpg30
Number of Scans
512
Receiver Gain
72.0
Relaxation Delay
1.0000
Pulse Width
10.0000
Acquisition Time
1.3631
Acquisition Date
2020-02-06T17:24:41
Spectrometer Frequency
100.62
Spectral Width
24038.5
Lowest Frequency
-1944.9
Nucleus
13C
Acquired Size
32768
Spectral Size
65536
Me
nept
Me
Me
O
O
203.74
197.83
149.20
143.45
52.01
45.23
31.87
30.01
29.22
28.46
27.05
26.38
22.76
14.22
13.19
ppm

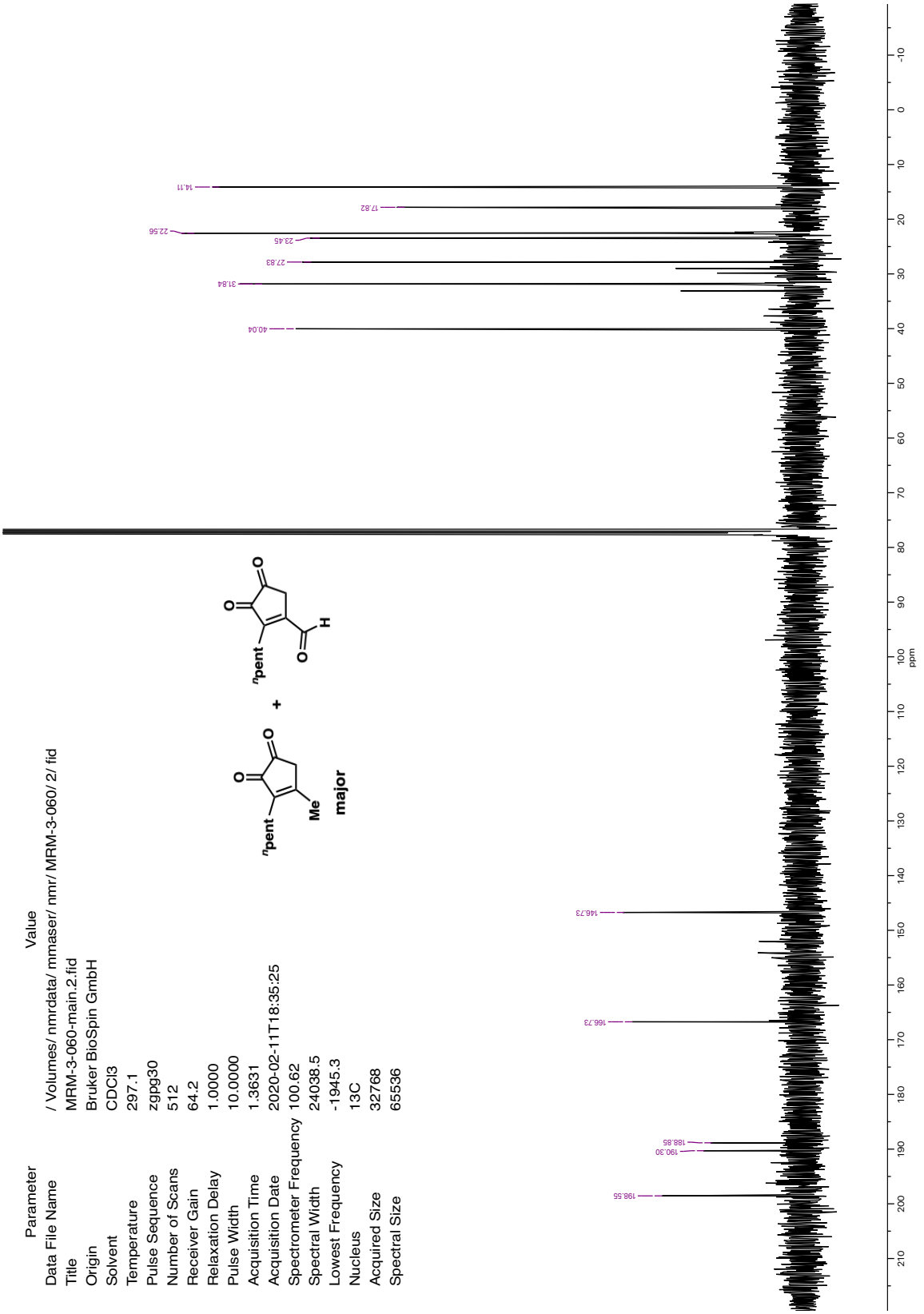

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-155-D/ 1/ fid
Title SED-1-155-D.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 298.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-04-29T20:38:41
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
O
Ph
Me
Me
O
2.00
0.99
2.30
2.00
5.74
2.00
ppm

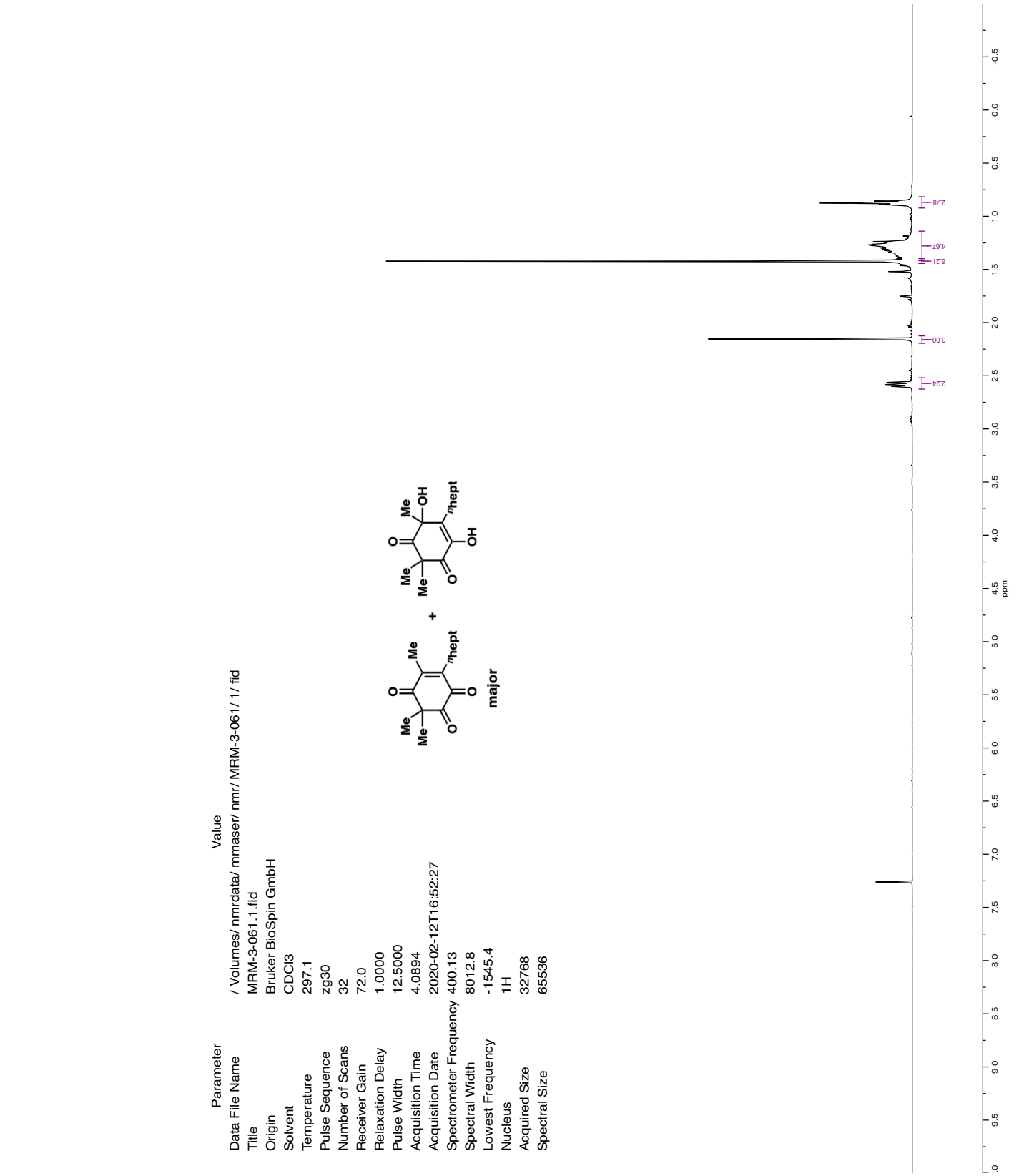

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-155-D/ 2/ fid
Title
SED-1-155-D.2.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
298.1
Pulse Sequence
zgpg30
Number of Scans
1024
Receiver Gain
72.0
Relaxation Delay
1.0000
Pulse Width
10.0000
Acquisition Time
1.3631
Acquisition Date
2019-04-30T21:47:59
Spectrometer Frequency
100.62
Spectral Width
24038.5
Lowest Frequency
-1944.5
Nucleus
13C
Acquired Size
32768
Spectral Size
65536
OH
O
Ph
Me
Me
O
193.41
165.32
157.15
135.83
134.29
129.33
128.91
128.47
91.27
71.82
29.86
26.75
ppm

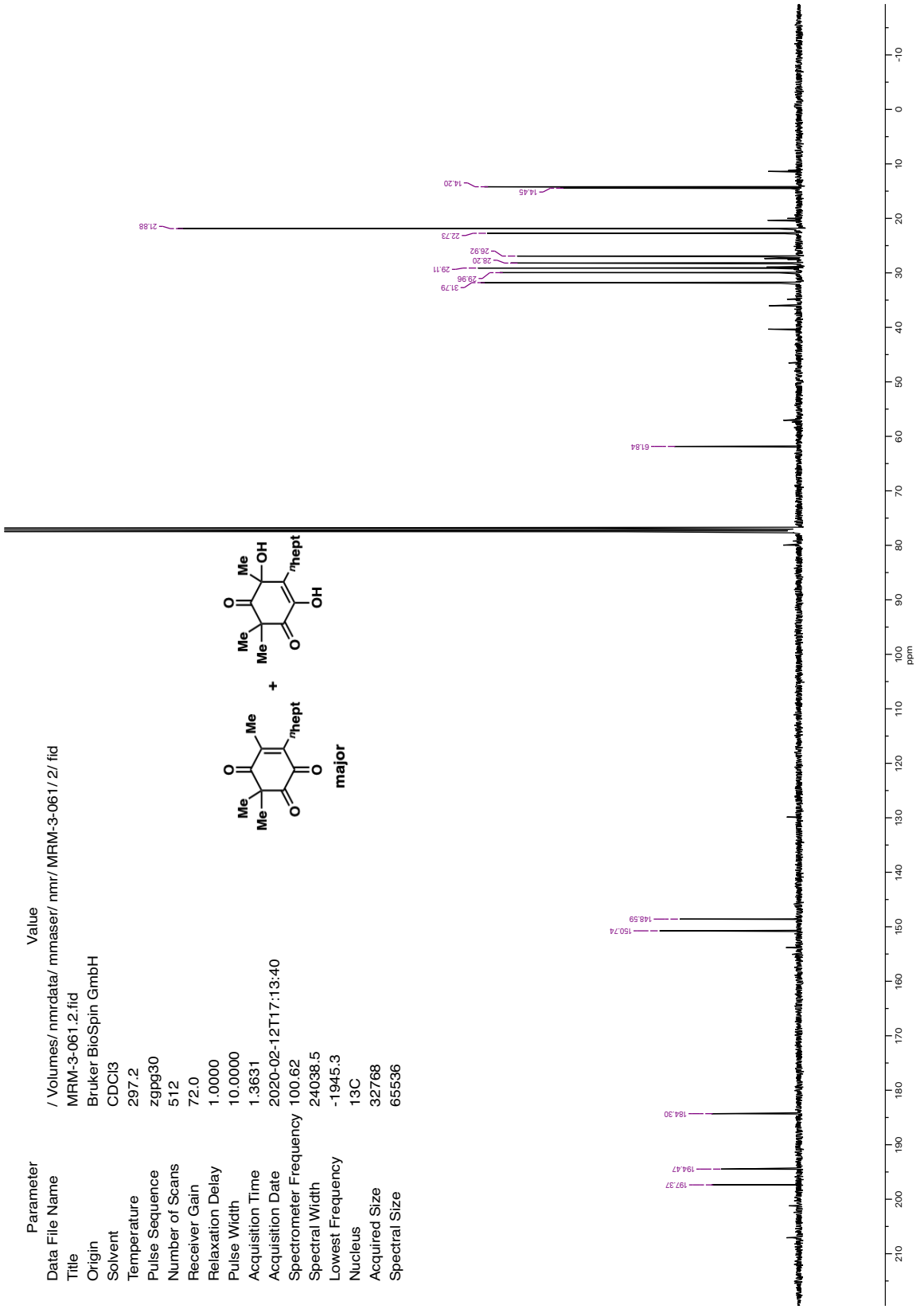

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-158-D/ 1/ fid
Title SED-1-158-D.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.2
Pulse Sequence zg30
Number of Scans 1
Receiver Gain 142.8
Relaxation Delay 30.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-04-25T17:31:37
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Internal Standard (TCNB)
1,4-dioxane
HO
Me
HO
O
O
Me
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
4.96
9.00
2.56
1.94
1.00
0.96
0.86
1.04
ppm

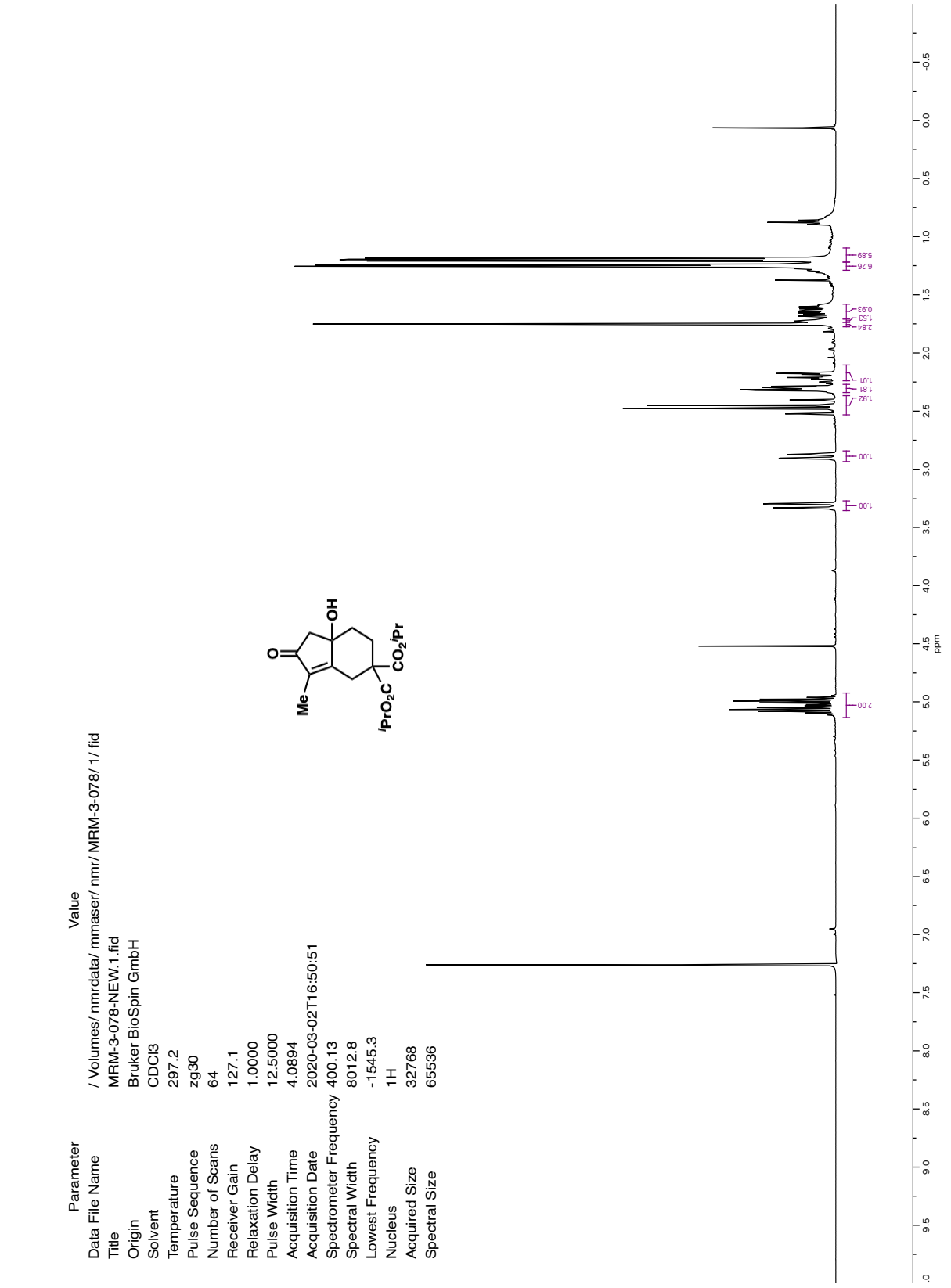

Parameter
Value
Data File Name / Volumes/ nmrdata/ sdibrell/ nmr/ SED-1-158-D/ 2/ fid
Title SED-1-158-D.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-04-26T10:13:51
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.8
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
1,4-dioxane
Internal Standard (TCNB)
HO
Me
HO
O
O
Me
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
14.34
18.75
21.40
21.89
21.53
42.98
45.12
55.94
57.86
69.94
71.47
86.53
102.83
168.62
173.51
190.10
202.54
ppm

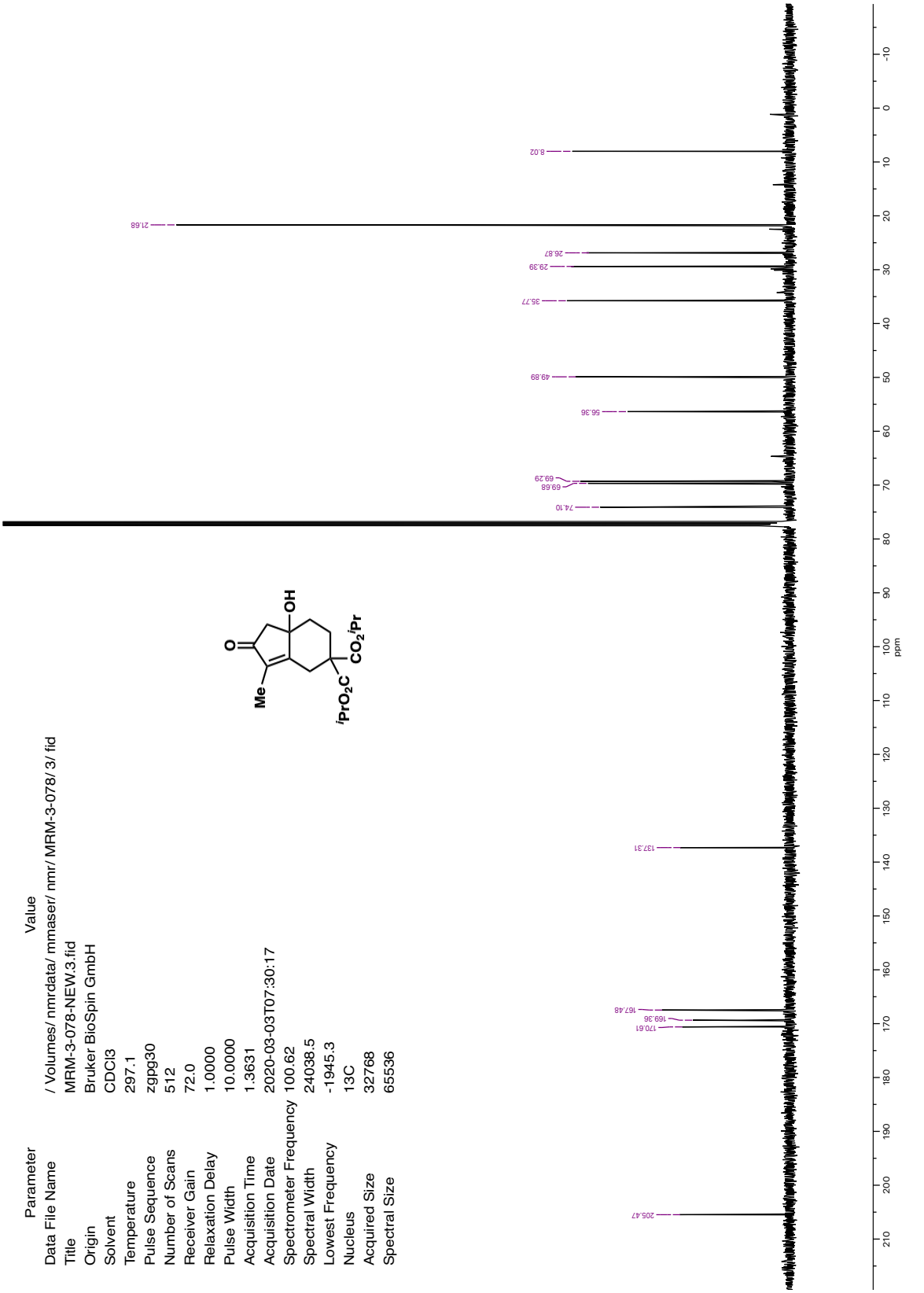

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-060/ 1/ fid
Title MRM-3-060-main.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 142.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-02-11T18:13:10
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
nPent
Me
major
+
nPent
H
4.05
6.41
1.88
3.00
2.40
1.98
ppm

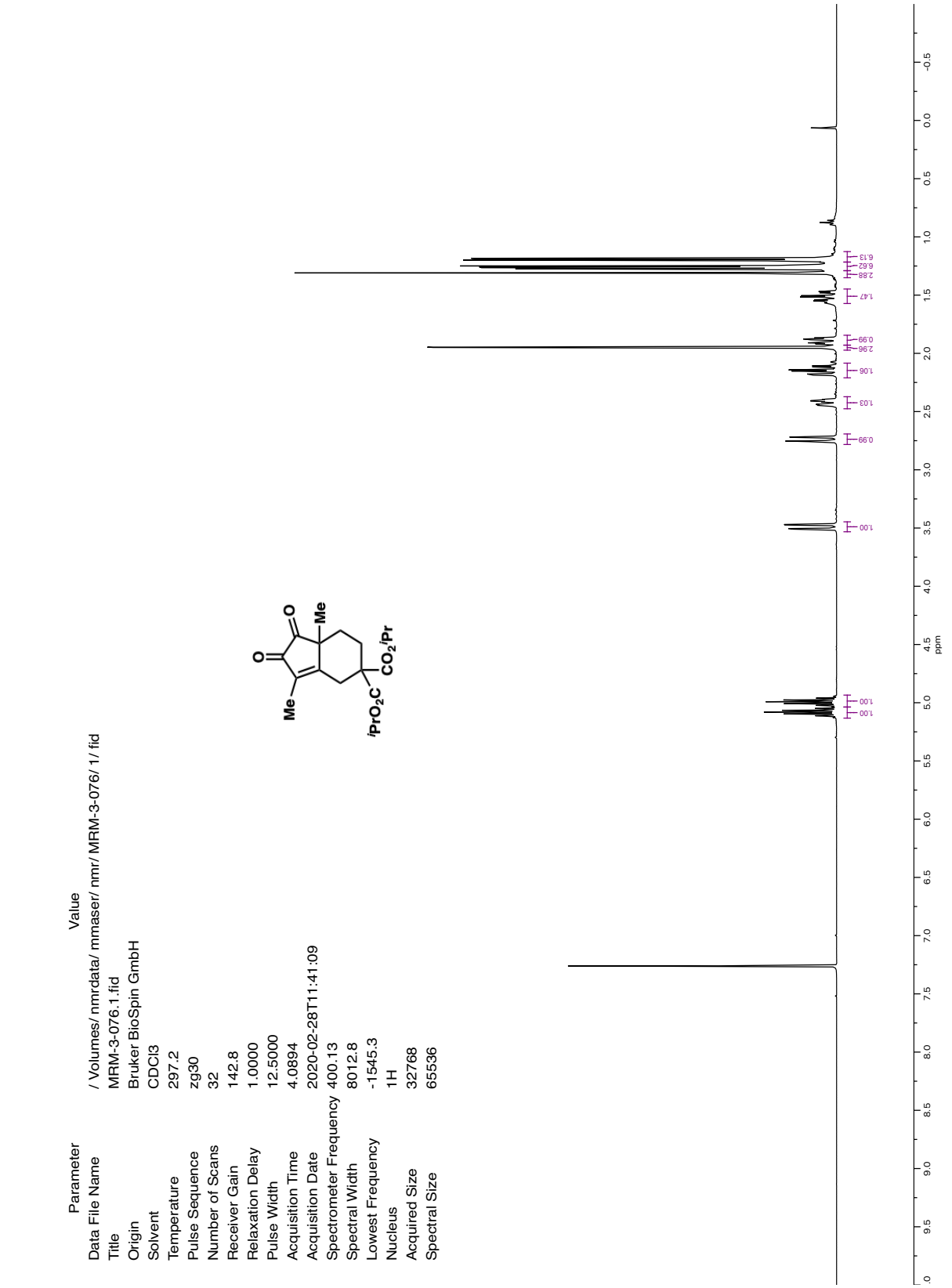

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-060/ 2/ fid
Title MRM-3-060-main.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-02-11T18:35:25
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
nPent
Me
major
+
nPent
H
14.11
17.82
22.56
23.45
27.83
31.84
40.04
146.73
166.73
188.55
190.30
198.55
ppm

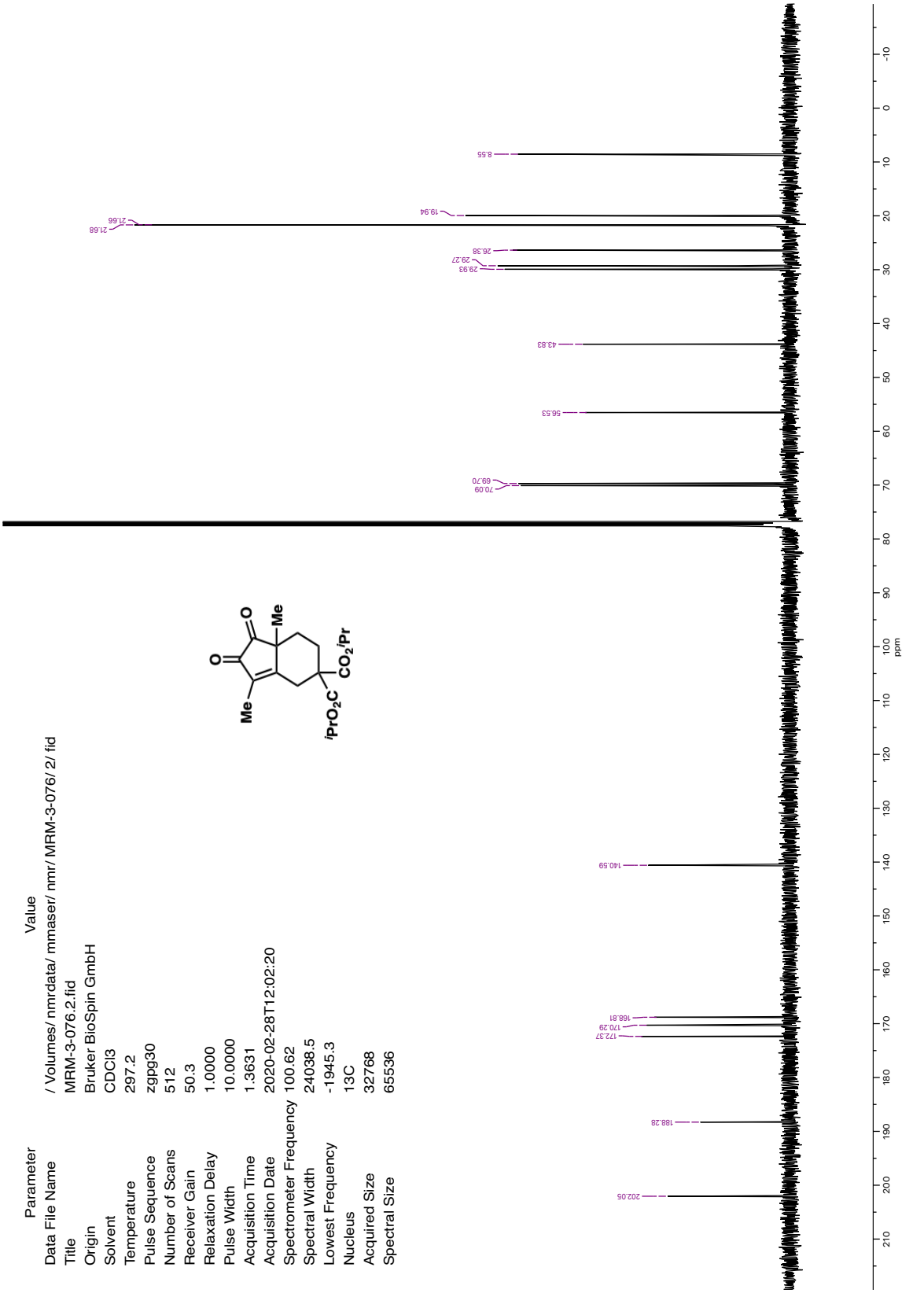

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-061/ 1/ fid
Title MRM-3-061.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-02-12T16:52:27
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
major
2.24
3.00
6.21
4.67
2.78
ppm


Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-061/ 2/ fid
Title MRM-3-061.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-02-12T17:13:40
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
major
+
14.20
14.45
21.88
22.73
26.92
28.20
29.11
29.96
31.79
61.84
148.59
150.74
184.30
194.47
197.37
ppm


Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-078/ 1/ fid
Title MRM-3-078-NEW.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 64
Receiver Gain 127.1
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-03-02T16:50:51
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
OH
O
iPrO2C
CO2iPr
2.00
1.00
1.00
1.92
1.81
1.01
2.84
1.53
0.93
6.26
5.89
ppm


Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-078/ 3/ fid
Title MRM-3-078-NEW.3.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-03-03T07:30:17
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OH
O
$^{i}PrO_2C$
$CO_2{}^{i}Pr$
205.47
170.61
169.36
167.48
137.31
74.10
69.68
69.29
56.36
49.89
35.77
29.39
26.87
21.68
8.02
ppm


Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-076/ 1/ fid
Title MRM-3-076.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 142.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-02-28T11:41:09
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
O
O
Me
Me
iPrO2C
CO2iPr
1.00
1.00
1.00
0.99
1.03
1.06
2.96
0.99
1.47
2.88
6.62
6.13
ppm


Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-076/ 2/ fid
Title MRM-3-076.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-02-28T12:02:20
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
O
O
Me
Me
iPrO2C
CO2iPr
202.05
188.28
172.37
170.29
168.81
140.59
70.09
69.70
56.53
43.83
29.93
29.27
26.38
19.94
21.68
21.66
8.55
ppm

Chapter 2

Total Synthesis of Ritterazine B and Analogs for the Treatment of Cancer[1]

2.1 INTRODUCTION

Ritterazine B (**46**) is a bis-steroidal pyrazine (BSP) natural product that was isolated in 1995 from the marine tunicate *Riterella tokioka* off of Japan's Izu Peninsula.[1] The BSPs include some of the most potent anticancer compounds discovered to date,[2] and **46** in particular has been described as "among the most potent growth inhibitors ever tested" by the National Cancer Institute (NCI).[3-5] It possesses sub-nanomolar activity against P388 leukemia cells (0.17 nM IC_{50})[6] and an average GI_{50} of 3.2 nM in the NCI-60 cell line screen.[5,7] Given that the BSPs display distinct activity patterns in NCI-60 COMPARE analyses, they are proposed to act by a distinct mode of action from existing chemotherapies.[4,7,8]

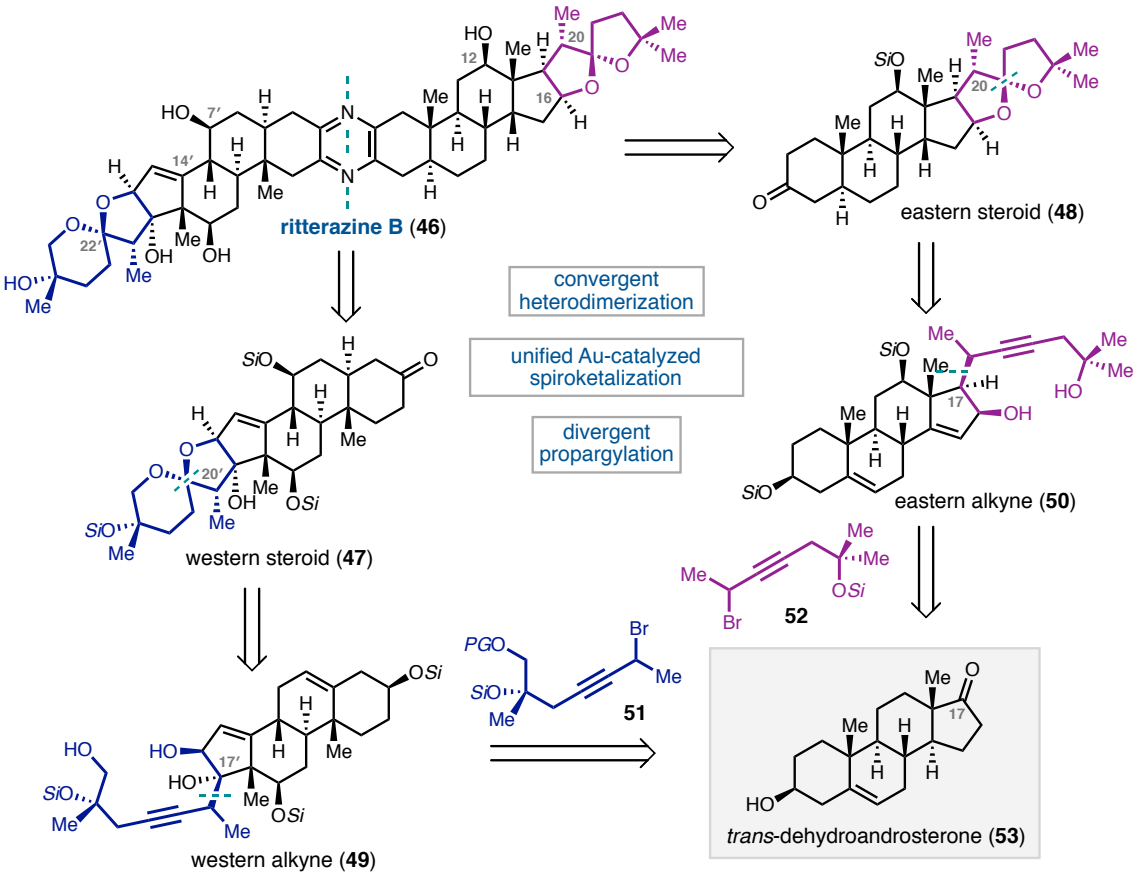
Although the BSPs are known to induce apoptosis,[3] a lack of natural material has hampered translational investigations of **46** and related compounds. Landmark studies from Shair and co-workers implicated BSPs as high-affinity ligands for oxysterol binding proteins,[4] while more recent evidence indicates that the endoplasmic reticulum-specific heat-shock protein GRP78 may be their efficacious target.[9,10] Given these promising

[1] Portions of this chapter were adapted from the following communications: Nakayama*, Y.; Maser*, M. R.; Okita, T.; Dubrovskiy, A. V.; Campbell, T. L.; Reisman, S. E. Total Synthesis of Ritterazine B. *J. Am. Chem. Soc.* **2021**, *143*, 4187–4192. DOI: 10.1021/jacs.1c01372. Copyright © 2021 American Chemical Society. The research discussed in this chapter was completed in collaboration with former members of the Reisman Lab: Postdoc Dr. Yasuaki Nakayama, visiting scholar Tatsuya Okita, postdoc Prof. Dr. Anton V. Dubrovskiy, and graduate student Taryn L. Campbell, as well as scientists from the City of Hope Comprehensive Cancer Center: Min Lin and David Horne.

foundational studies, improved synthetic access to **46** is required to fully evaluate its potential as a chemotherapeutic.[2,8] In this communication, we report the first total synthesis of **46**. Our approach uses a common strategy to prepare both of the steroid spiroketals from *trans*-dehydroandrosterone, a commercially available and inexpensive steroid.

2.2 RETROSYNTHETIC ANALYSIS

Figure 2.1 *Retrosynthetic analysis*



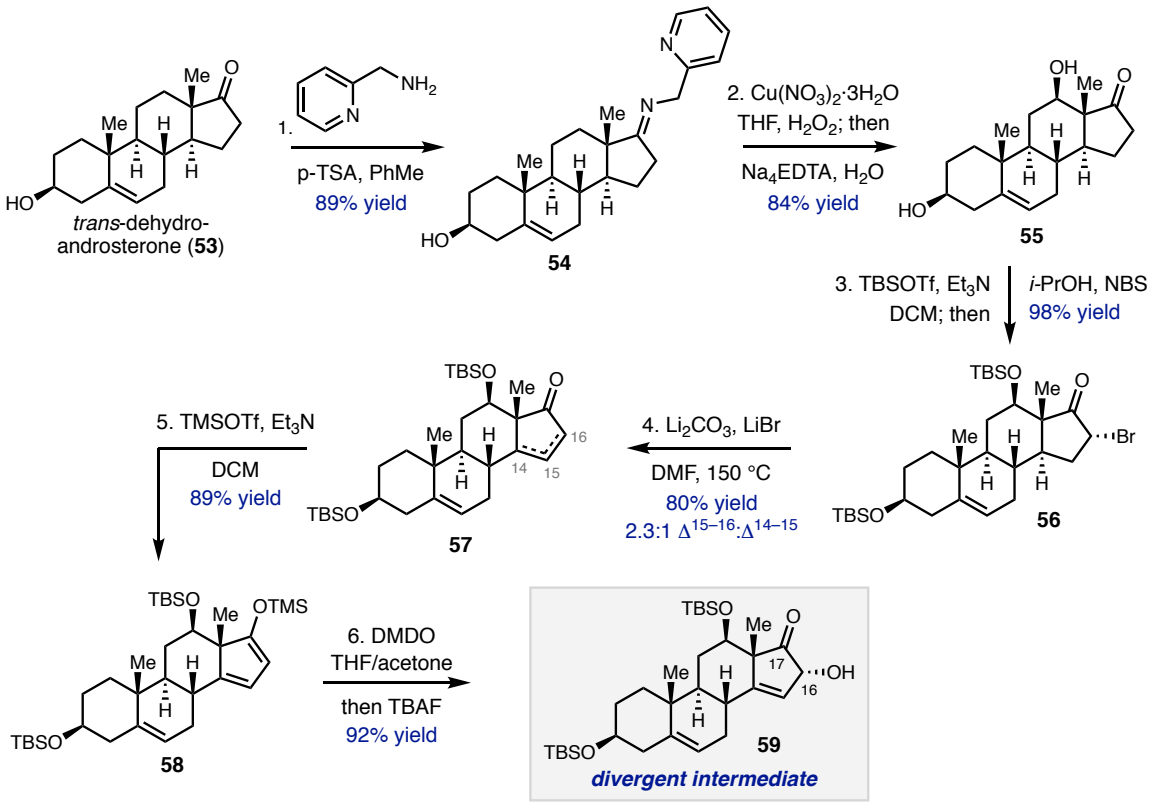
In line with prior efforts to the BSPs,[5,11] our retrosynthetic analysis began with scission of the central pyrazine ring, revealing the 'western' and 'eastern' steroids **47** and **48**, respectively (**Figure 2.1**). To streamline our route development, we sought to prepare both **47** and **48** from a common starting material, using the same general tactics for C–C

bond formation and spiroketalization. In this vein, steroids **47** and **48** were simplified to the corresponding alkynes **49** and **50**, where transition metal-catalyzed cycloisomerization[12] would be used to form the respective spiroketals. This retrosynthetic step shifted the synthetic challenge to the union of differentiated alkyne fragments with a common steroid core. We envisioned preparing alkynes **49** and **50** by 1,2-addition of the propargyl-metal species derived from **51** or **52** to an α-hydroxy ketone accessible from *trans*-dehydroandrosterone (TDEA, **53**). The choice of **53** as the starting material was seen as strategic: the C5–C6 alkene would provide a handle for late-stage B-ring oxidation. This tactic has yet to be utilized in synthetic approaches to the BSPs,[11] which could be why BSPs with C7/C7′ oxidation have not previously been synthesized.[7]

2.3 FORWARD SYNTHESIS

2.3.1 ACCESSING THE DIVERGENT INTERMEDIATE

Scheme 2.1 *Preparation of the divergent intermediate*

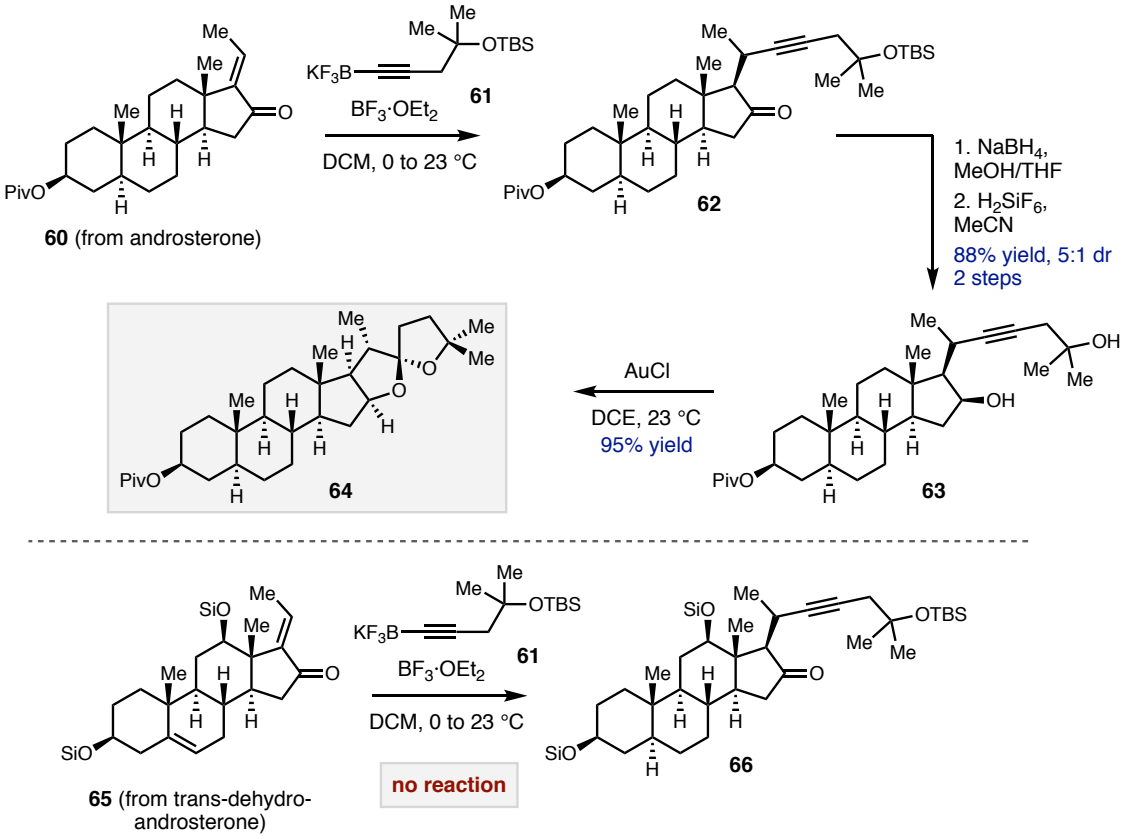


In the forward sense, known steroid **55** was prepared in two steps from TDEA by copper-mediated C–H oxidation at C12 directed by the pyridyl imine in **54** (**Scheme 2.1**).[13] **55** was treated with excess *tert*-butyldimethylsilyl triflate (TBSOTf) and triethylamine (Et_3N) to protect the C3 and C12 alcohols and form of a silyl enol ether at C17. Direct addition of isopropanol and *N*-bromosuccinimide to the reaction mixture afforded α-bromo-ketone **56** in quantitative yield in one pot. Elimination of the C16 bromide under basic conditions gave an inconsequential mixture of isomeric enones ($\Delta^{14–15}$ and $\Delta^{15–16}$, **57**), which converged to silyl-dienol ether **58** on treatment with Et_3N and trimethylsilyl

triflate (TMSOTf). Selective epoxidation of the C16–C17 alkene with dimethyldioxirane and subsequent addition of tetrabutylammonium fluoride (TBAF) provided α-hydroxy-ketone **59** in 92% yield, which would serve as our divergent intermediate.

2.3.2 INITIAL ATTEMPTS FOR ALKYNE INSTALLATION

Scheme 2.2 *First-generation approach to alkyne installation and spirocyclization*



In first-generation efforts in the group, the installation of the alkynes was attempted by conjugate addition of alkynyl boron species such as **61** to exocyclic enones such as **60** (**Scheme 2.2**, top). This strategy proved successful for under-functionalized steroid precursors like **60**. Addition product **62** was able to be carried in three additional steps to spirocycle **64** in excellent yield, demonstrating the efficacy of the gold-catalyzed spiroketalization. Unfortunately, the same procedure failed when more oxidized steroid

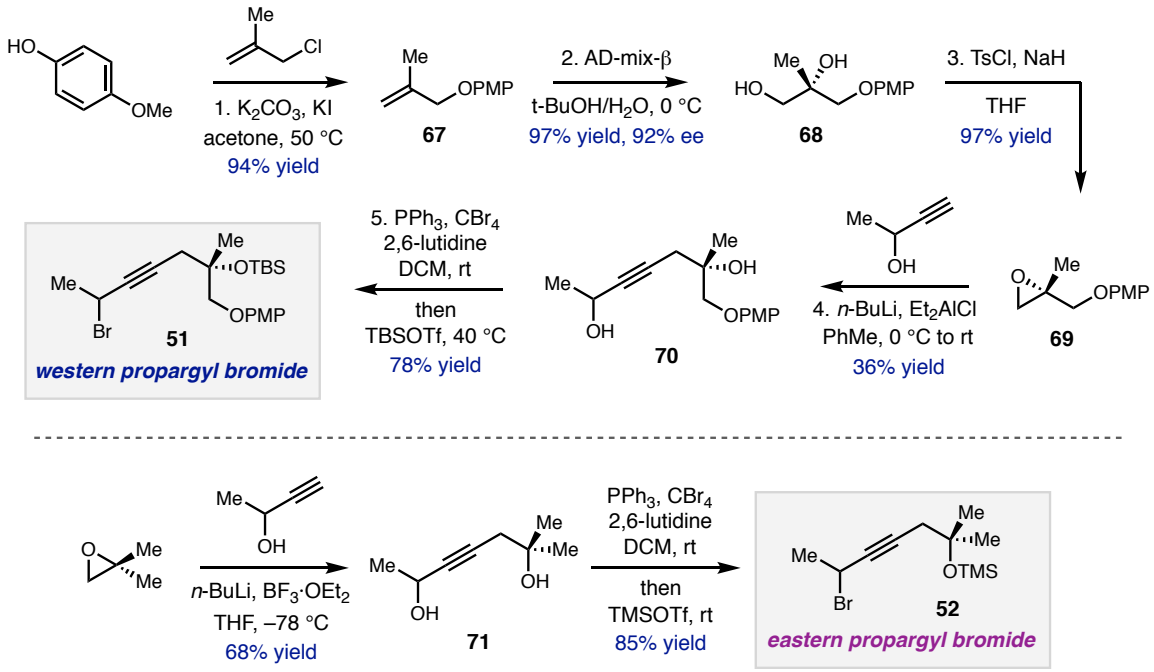
derivatives such as **65** were employed (**Scheme 2.2**, bottom), ultimately necessitating a revised approach.

2.3.3 SECOND-GENERATION APPROACH

2.3.3.1 Preparing the Propargyl Bromides

To facilitate the second-generation 1,2-addition strategy (see retrosynthesis in **Figure 2.1**), we sought the preparation of properly functionalized propargyl bromides such as **51** and **52**. Each could be prepared in short order using known methods and readily available materials (**Scheme 2.3**). For the western alkyne (**Scheme 2.3**, top), p-anisole was alkylated with methallyl chloride and treated under Sharpless asymmetric dihydroxylation conditions to afford diol **68** in excellent yield and ee.[14] Epoxide formation was effected by treatment with tosyl chloride and base in 97% yield. The epoxide in **69** could then be opened by addition of lithiated butyn-2-ol in the presence of diethyl aluminum chloride in an unoptimized 36% yield. Appel bromination[15] and tertiary alcohol protection was then performed in the same pot to access the desired propargyl bromide in 78% yield. The eastern propargyl bromide was prepared in two steps (**Scheme 2.3**, bottom). Lithium alkyne opening of isopropylene oxide proceeded smoothly with $BF_3 \cdot OEt_2$, as did tandem Appel bromination and TMS protection to give **52**.

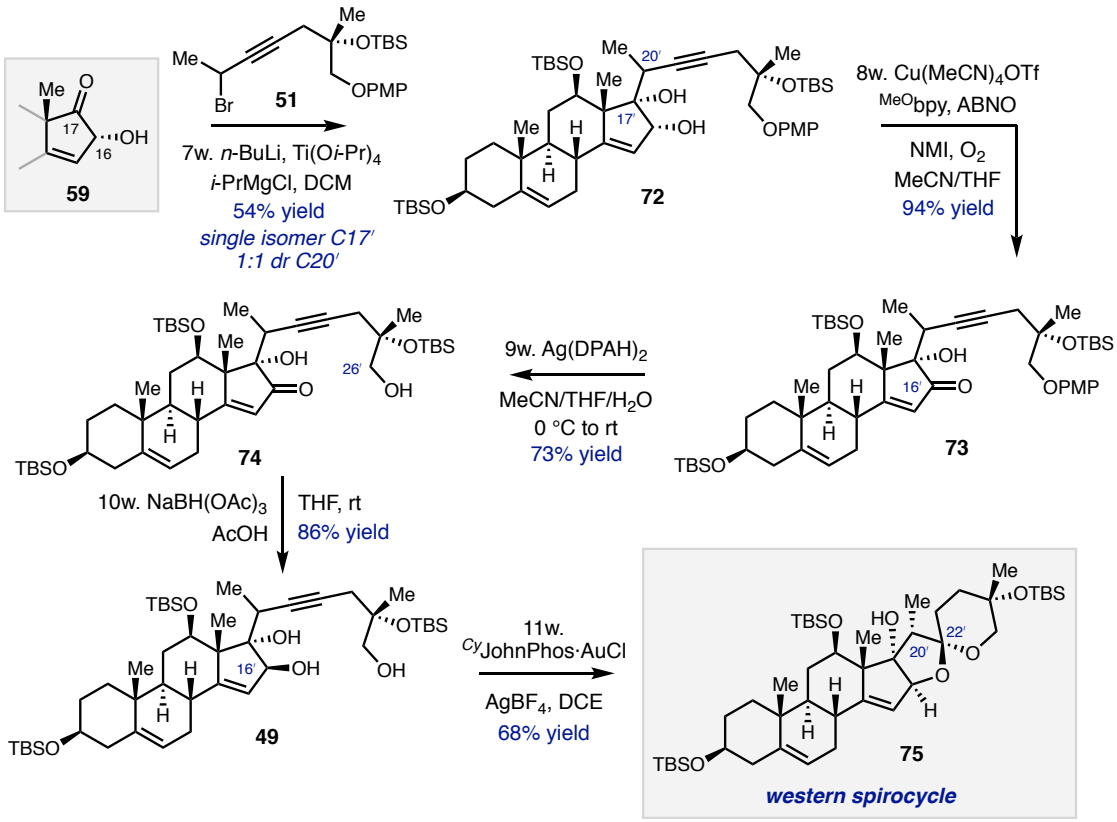
Scheme 2.3 *Preparation of the western (top) and eastern (bottom) propargyl bromides*



2.3.3.2 Preparing the Western Coupling Partner

At this stage, we turned our attention to preparing the distinct spiroketals found in the western and eastern steroids **47** and **48**, respectively. To this end, titanium-mediated propargylations based on conditions reported by Sato and co-workers proved uniquely effective (**Scheme 2.4**).[16] Deprotonation of the C16 alcohol in **59** by treatment with *n*-butyllithium followed by addition of the organotitanium species derived from propargyl bromide **51** resulted in 1,2-addition to give alkyne **72** in 54% yield. The addition occurred with exclusive β-face selectivity despite the axial methyl group at C13′,[17] possibly due to the formation of an α-disposed cyclic chelate between the C16′ and C17′ oxygens. While excellent diastereoselectivity was obtained at C17′, **72** was formed as a 1:1 mixture of epimers at C20′ (*vide infra*).

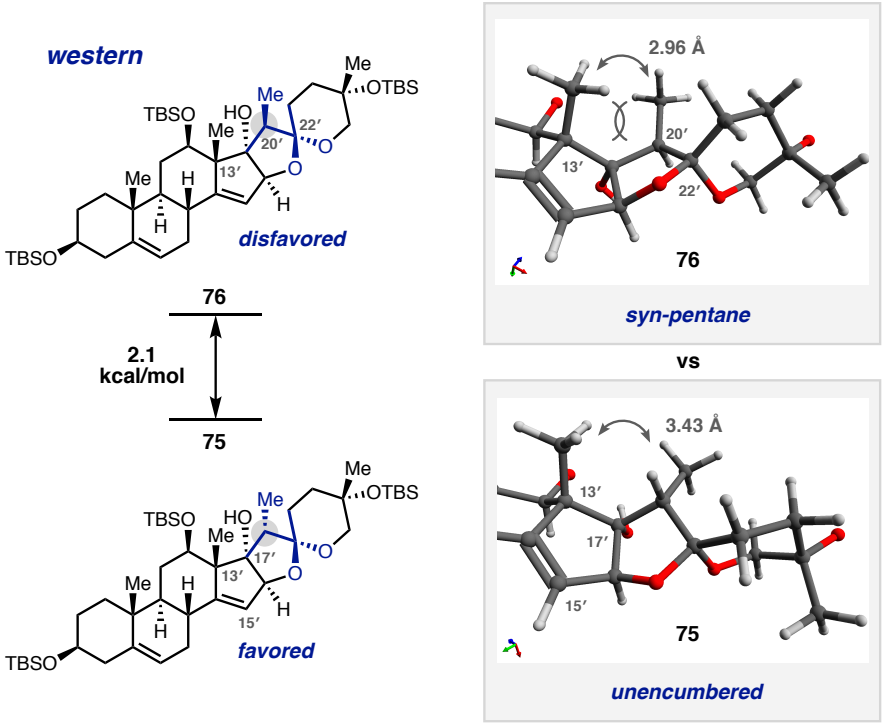
Scheme 2.4 *Preparation of the western spirocycle*



While the α-stereoisomer of the C16 alcohol was crucial for imparting the desired stereocontrol in the propargylation reaction, we required the β-alcohol to elaborate to the required spiroketal. For the preparation of the western steroid (**47**), stereoinversion was accomplished by oxidation to the enone followed by cleavage of the *p*-methoxyphenyl ether[18] and hydroxyl-directed 1,2-reduction, which furnished spiroketal precursor **49** in 86% yield. After extensive experimentation,[19] treatment of diol **49** with catalytic ^{Cy}JohnPhos·AuCl (10 mol %) and $AgBF_4$ (5 mol %) provided spiroketal **75** in 68% yield as a single diastereomer.[12] To our delight, this reaction not only provided the correct configuration at the spiroketal, but also proceeded with convergence of the C20′ epimers, furnishing **75** with the required α-disposed methyl group (*vide infra,* **Figure 2.2**). The

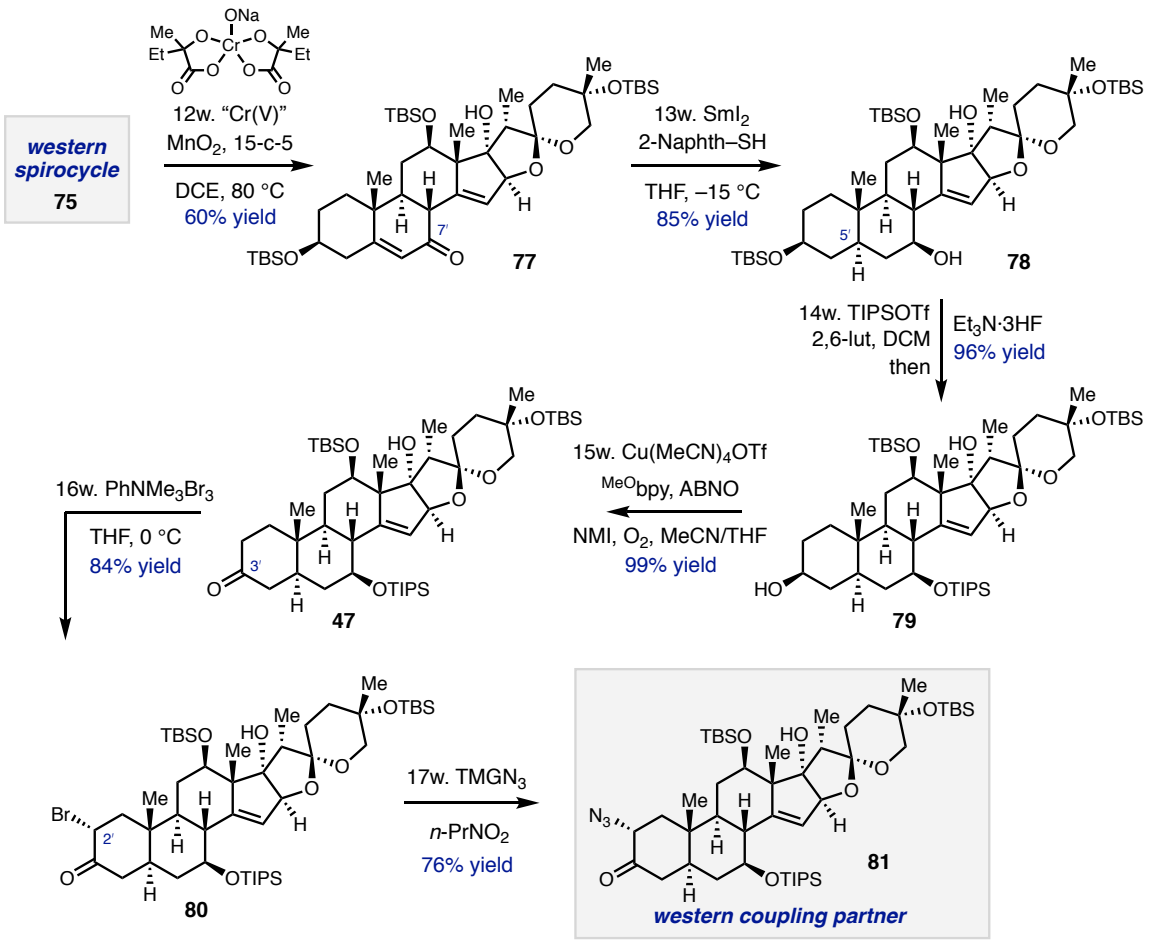
overall selectivity was found to be dependent on the choice of dichloroethane solvent, ^{Cy}JohnPhos ligand, and tetrafluoroborate counterion.

Figure 2.2 *Thermodynamic evaluation of the western spiroketal C20′ epimers*



Given the observed epimerization at C20′, we hypothesize that the process occurs through an initial stereoablative mono-cyclization/isomerization, followed by a doubly-diastereoselective ketal formation.[5,12a] To probe the preferential formation of the C20′ stereocenter, ground-state energies were calculated for DFT-optimized **75** and its epimer **76** (**Figure 2.2**). Indeed, the observed isomer was found to be 2.1 kcal/mol lower in energy than the un-natural C20′ epimer.[20] Conformational analysis of the disfavored species **76** revealed the existence of a *syn*-pentane-like interaction between the C20′-β-Me and the axial C13′-Me group (**Figure 2.2**, top-right).[5,21] These interactions are not present with the α-disposed C20-Me group, which appears less sterically encumbered (**Figure 2.2**, bottom-right).

Scheme 2.5 *Preparation of the western coupling partner*



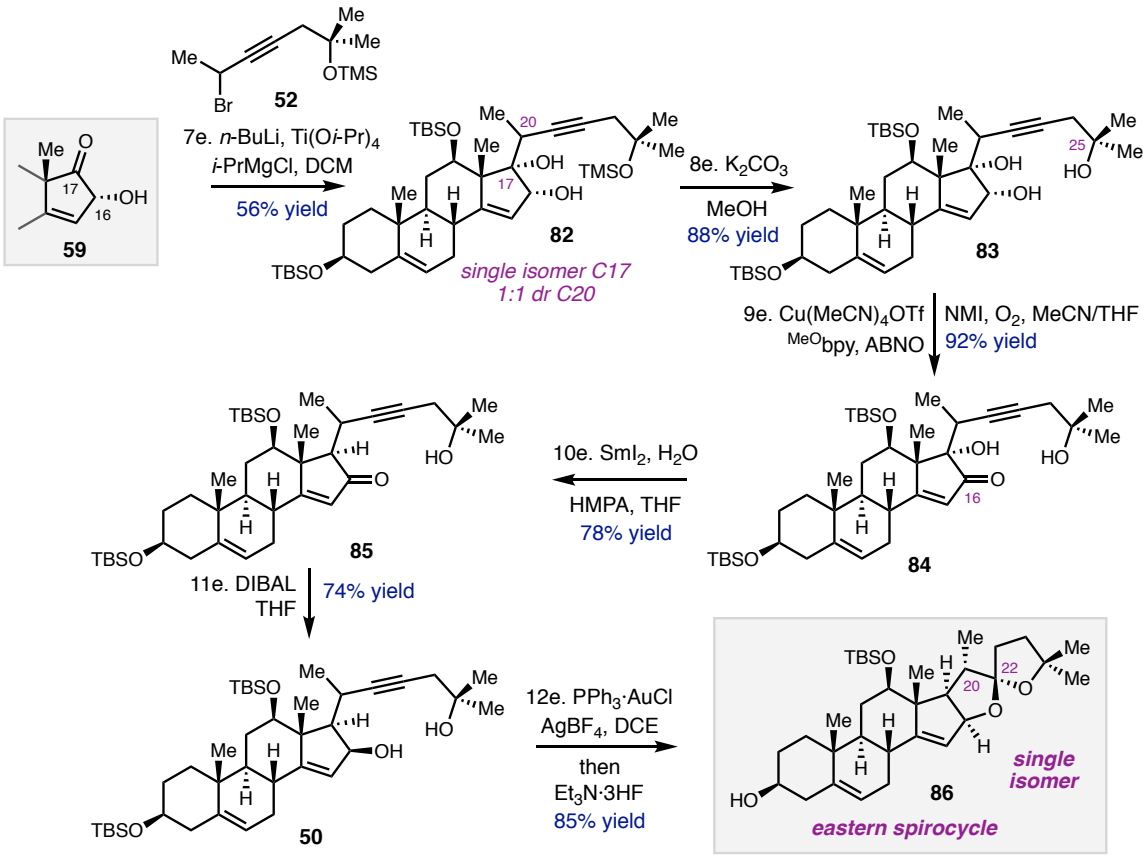
With the western spirocycle in hand, allylic oxidation of **75** at C7′ was achieved with oxochromate (**Cr(V)**, **Scheme 2.5**).[22] To obtain the fully saturated B′ ring, the intermediate enone **77** was reduced with SmI_2 and 2-naphtalene thiol to furnish **78** in 85% yield as a single diastereomer; low temperatures were crucial for preventing over-reduction of the C14–C15 alkene.[19] Protection of the C7′ alcohol of **78** as the tri-*iso*-propylsilyl ether was followed by addition of $Et_3N{\cdot}3HF$ to selectively reveal the C3′ alcohol in the same pot.

Oxidation of **79** to the ketone and two-step α-bromination and azidation at C2′ using procedures developed by Shair[11i] and Fuchs[11d] provided the western fragment as

keto-azide **81** in 63% overall yield. The use of nitropropane as solvent for the azidation instead of the traditional nitromethane was found necessary for solubilization of the intermediate bromide.[11d,i] This also significantly increased the yield by preventing undesired elimination of N_2 from **81** (*vide infra*).[11c]

2.3.3.3 Preparing the Eastern Coupling Partner

Scheme 2.6 *Preparation of the eastern spirocycle*

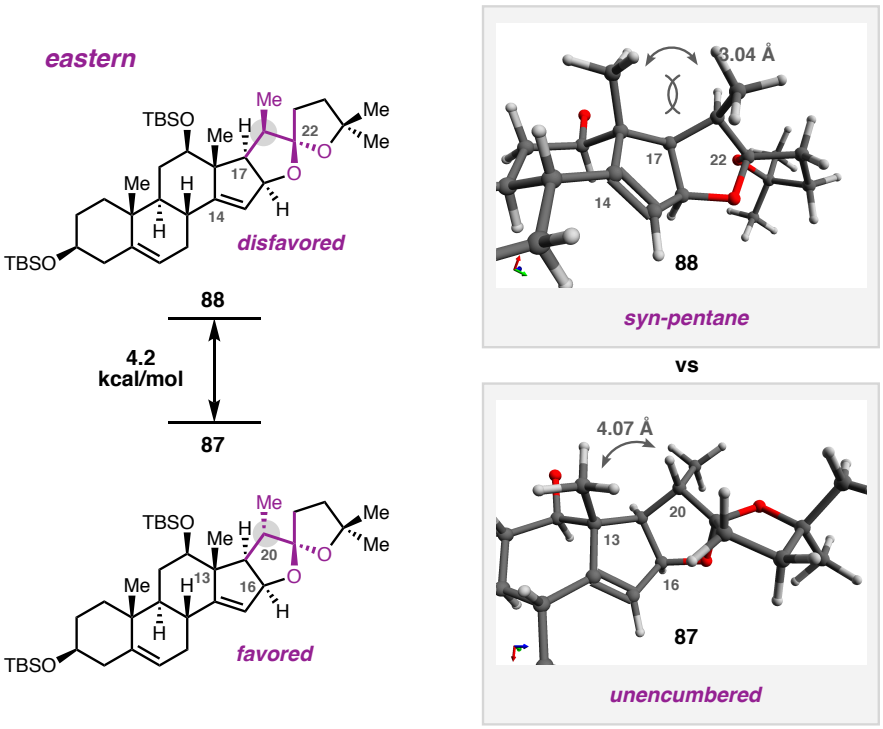


The synthesis of the eastern steroid followed a similar sequence to that described above but was modified slightly to prepare diol **50** (**Scheme 2.6**). Similar to the western fragment, **82** was prepared in 56% yield by addition of propargyl bromide **52** to deprotonated **59**, again as a single diastereomer at C17 but a 1:1 mixture at C20. Following deprotection of the trimethylsilyl ether and Stahl oxidation, the C17 alcohol was removed

via ketol deoxygenation with SmI_2 and H_2O to give **85**.[23] Diastereoselective 1,2-reduction was effected by treatment with di-*iso*-butylaluminum hydride to give **50**; the four electron reduction of **84** directly to give **50** could be accomplished with excess SmI_2, though the yield and diastereoselectivity was greatly reduced relative to the two-step procedure.

As observed for the western fragment, Au(I)-catalyzed spirocyclization proceeded smoothly to construct the 5/5 ring system, again as a single isomer. Here, it was found that direct addition of $Et_3N{\cdot}3HF$ to the reaction mixture resulted in selective deprotection of the C3 silyl ether, ultimately providing **86** in 85% yield. In this case, the thermodynamic preference for the observed α-disposed C20 center was even greater than that for the western steroid. Single-point DFT evaluations revealed a 4.2 kcal/mol difference between the ground-state energies of observed isomer **87** and epimer **88** (**Figure 2.3**). Again, a *syn*-pentane interaction between C19 and C21 was observed in the optimized geometry of **88**, with the C21-Me group disposed in a local concave region of the DE bicycle (**Figure 2.3**, top-right). This interaction is alleviated in **87**, where the C21-Me group is in an open 3D space (**Figure 2.3**, bottom-right)

Figure 2.3 *Thermodynamic evaluation of the eastern spiroketal C20 epimers*



The remaining two stereocenters required for the eastern fragment were installed in a single step *via* hydrogen atom transfer (HAT) reduction of the C5–C6 and C14–C15 alkenes under conditions developed by Shenvi and co-workers (**Scheme 2.7**).[24] The fully saturated product was obtained as a single isomer with *trans*-fusion at the A/B ring-junction and *cis*-fusion at the C/D ring-junction. This reaction proved critical for accessing late-stage material in the correct oxidation state, as typical alkene hydrogenation conditions were unable to reduce the C14–C15 double bond. To complete the eastern coupling partner, the same oxidation/bromination/azidation protocol as described for the western fragment was followed by ketone condensation with $MeONH_2 \cdot HCl$ to give **92** and Staudinger reduction[25] to furnish amino-methoxime **93** in quantitative yield.[11d]

Scheme 2.7 *Preparation of the eastern coupling partner*

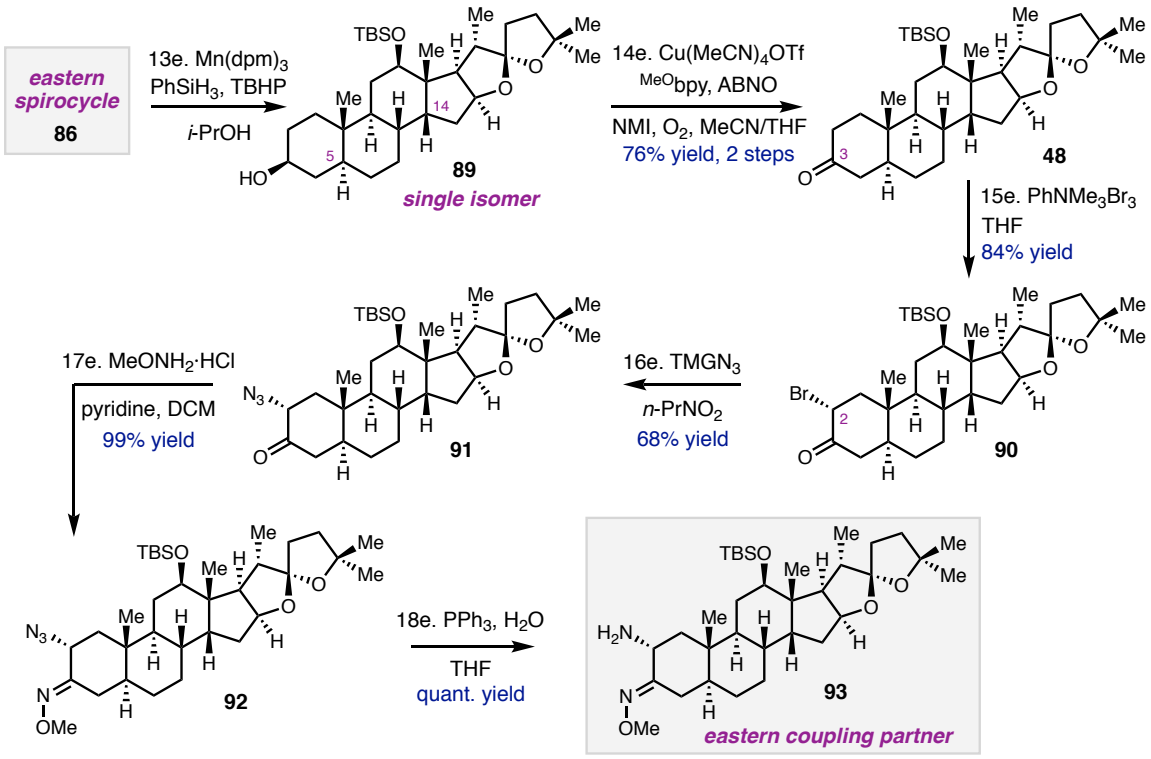
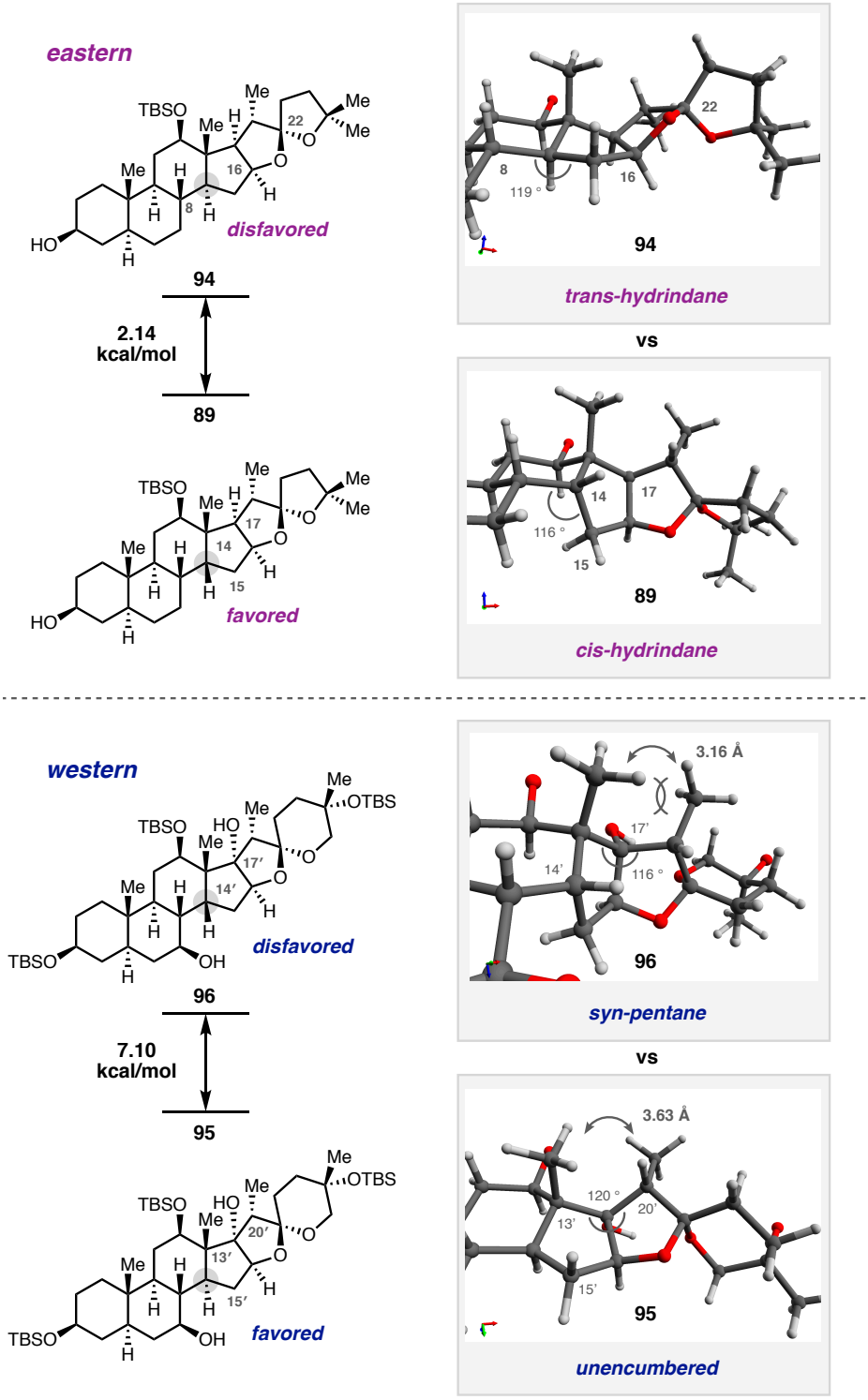


Figure 2.4 *Thermodynamic evaluation of the C14–C15/C14′–C15′ reduction products in both the eastern (top) and western (bottom) fragments*

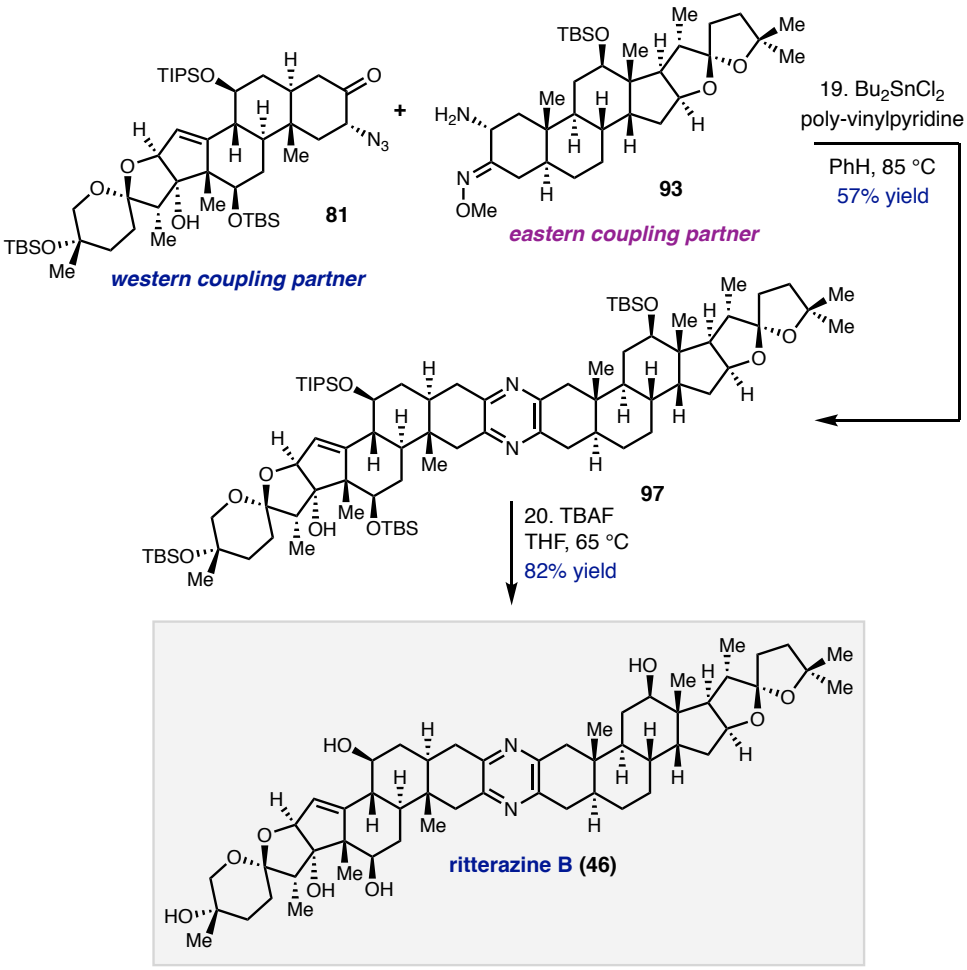


Intuitively, DFT studies supported the thermodynamic preference for the observed stereochemistry at C14 (**Figure 2.4**, top).[19,24] The ground-state geometry of the *cis*-hydrindane **89** was calculated to be 2.14 kcal/mol lower in energy than the *trans*-hydrindane **94**, in which the C8-14-15 bond angle was distorted to nearly 120°. Interestingly, this relationship was reversed in the fully saturated western fragment **95** (**Figure 2.4**, bottom), a byproduct of the SmI_2 double reduction (see Scheme 2.5 and Experimental Section). In this reaction, only the *trans*-hydrindane **95** is observed. Here, the *cis*-hydrindane **96** appears to cause a puckering of the E′ ring, where the C20′ methyl group actually creates a syn-pentane with that at C13′, even though the C20′-Me is α-disposed. This results in a massive 7.10 kcal/mol difference in ground state energy between **95** and **96**.

2.3.3.4 Completion of the Synthesis

To complete the synthesis, heterodimerization of fragments **81** and **93** under Lewis-acid catalysis as originally reported by Fuchs[11d] provided the desired pyrazine **97** (**Scheme 2.8**). Global deprotection with TBAF[11i] delivered ritterazine B in 82% yield, representing its first total synthesis. Spectroscopic characterization data matched that reported for the natural material.[1]

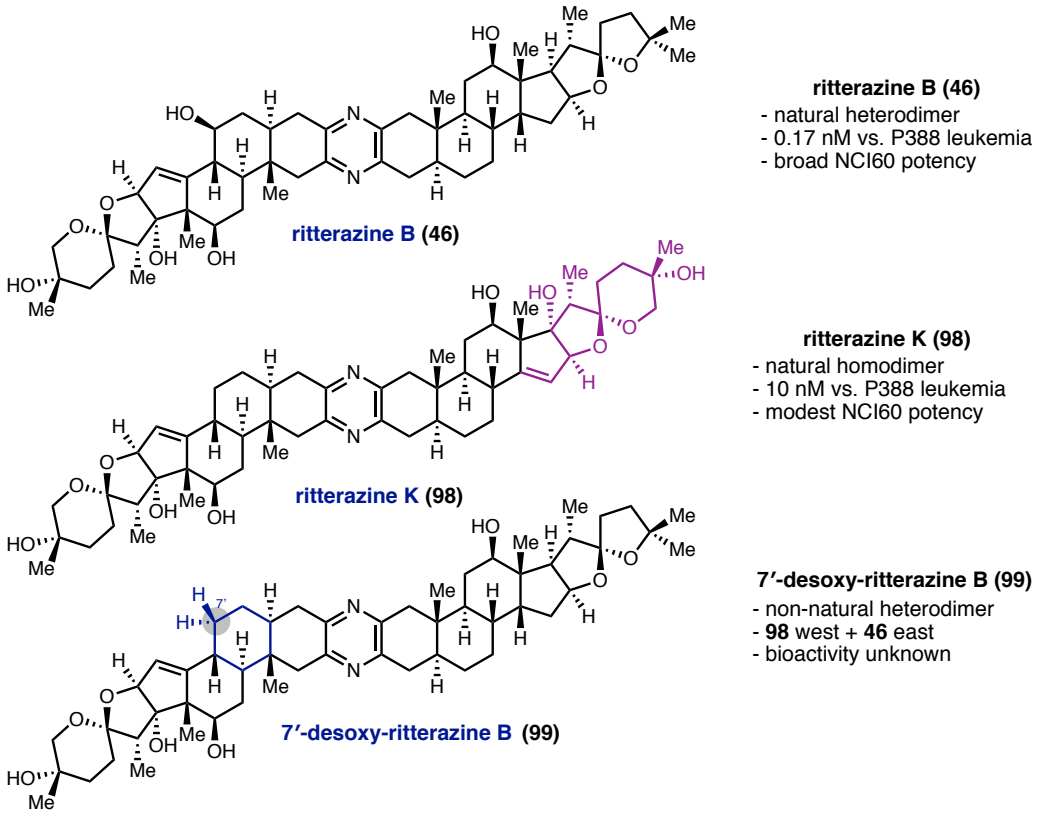
Scheme 2.8 *Completion of the synthesis of ritterazine B (**46**)*



2.4 ADDITIONAL RITTERAZINES

2.4.1 IDENTIFYING PROMISING TARGETS

Figure 2.5 *Additional targets ritterazine K (**98**) and 7′-desoxy-ritterazine B (**99**)*

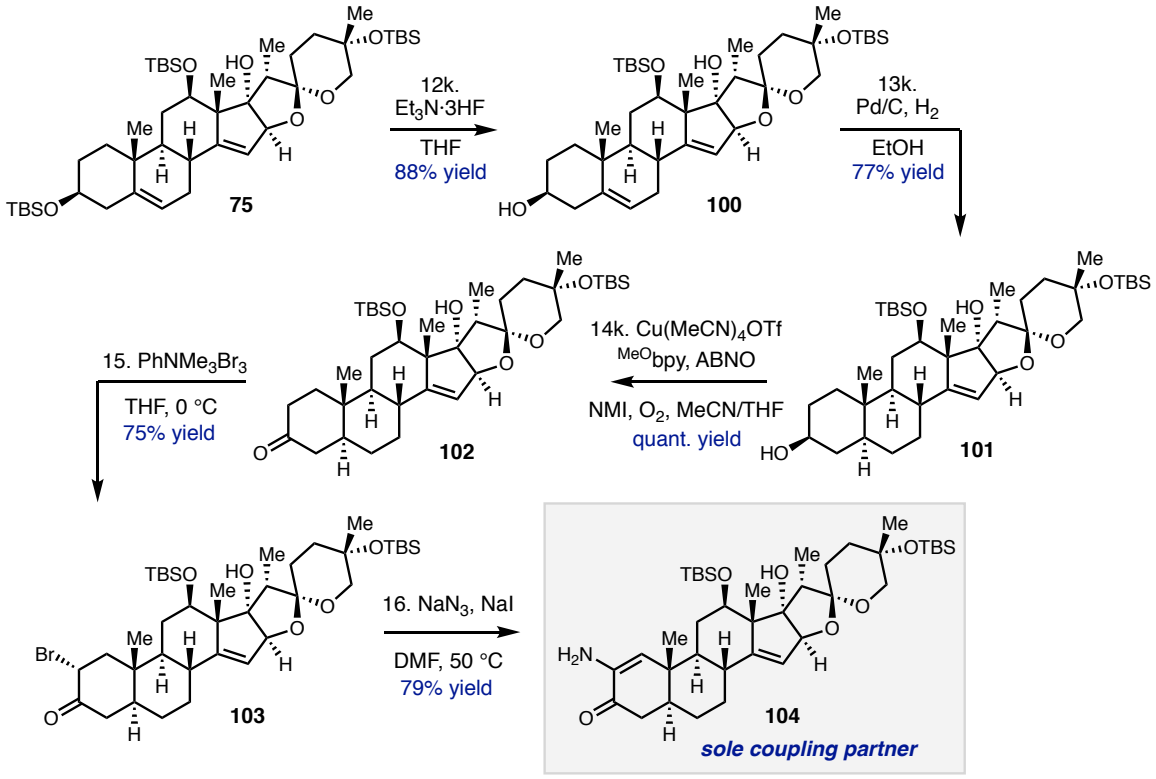


Having prepared **46**, we next sought to utilize the modular route to access other members of the ritterazine family, including non-natural compounds, that would allow for deeper biological interrogation.[2] We identified ritterazine K (**98**) and 7′-desoxy-ritterazine B (**99**) as exciting targets (**Figure 2.5**). **98** is a homodimer of the 7′-desoxy-western fragment of **46** that has shown much weaker activity in biological studies.[26] It was reported to have been prepared previously by Fuchs and coworkers, though its structure was not confirmed by NMR due to insufficient quantities.[11c] We hypothesized that **98** could serve as a potency baseline for the novel compound **99**, which combines the western steroid of

98 and the eastern steroid of **46**. Further, preparing **99** would allow us to interrogate the structure-activity relationship (SAR) around the C7′ alcohol that was challenging to access synthetically; no compound bearing C7(′) oxygenation had been synthesized prior to our studies. Finally, obtaining sufficient quantities of each compound would enable deeper MOA studies than have been possible to date.[1-4,6,10,26]

2.4.2 SYNTHESIS OF RITTERAZINE K

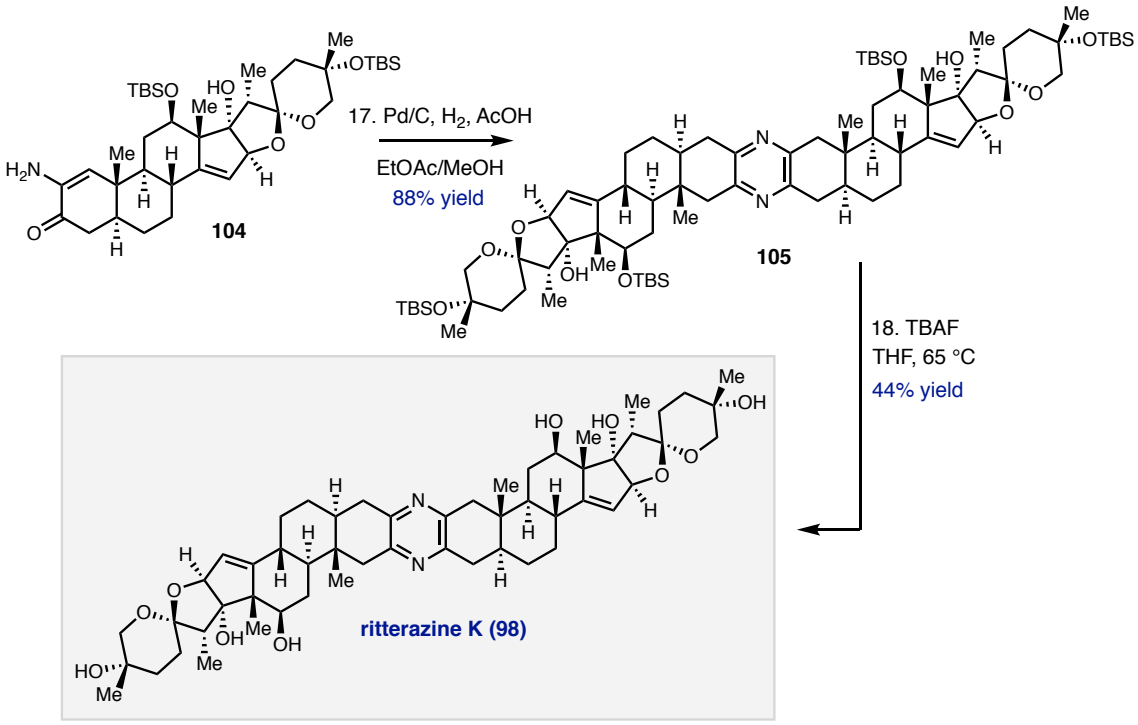
Scheme 2.9 *Preparation of the ritterazine K coupling partner* ***104***



To access the 7′-desoxy steroid of **98** and **99**, we retraced the western fragment synthesis of **46** to intermediate **75**, prior to the C7′ allylic oxidation (**Scheme 2.9**). **75** was treated with Et3N·3HF to remove the C3′–OTBS group and give **100**, the C5′–C6′ alkene of which was hydrogenated with Pd/C and H_2 in good yields. The typical Stahl oxidation/α-bromination sequence provided α-bromo-ketone **103** in 75% yield over the two steps from

101. Instead of azidation with $TMGN_3$, this material was treated with NaN_3 and NaI in DMF, which effected azide substitution and elimination of N_2 to give enamino-ketone **104** as the lone coupling partner for dimerization.[11c,27]

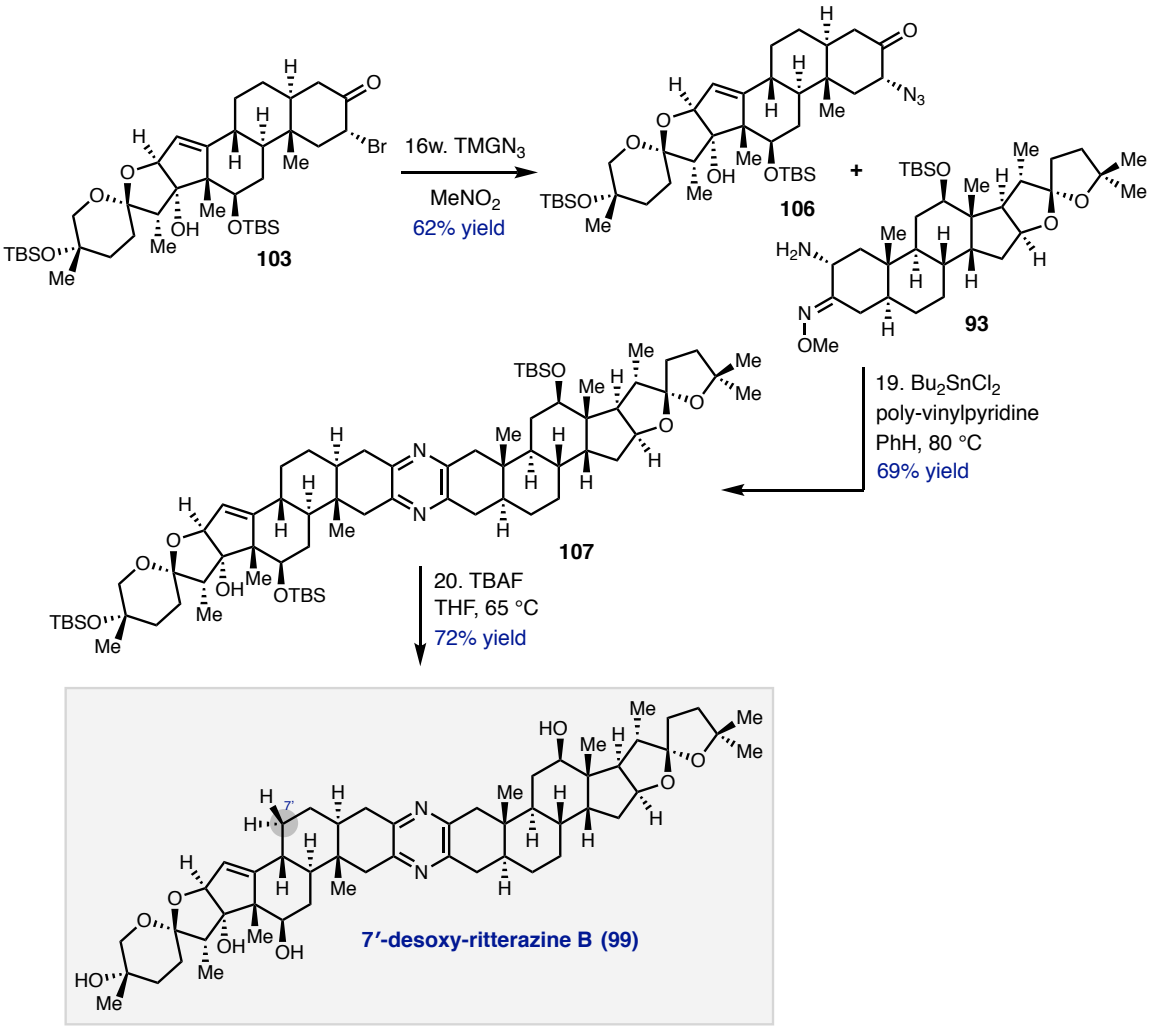
Scheme 2.10 *Completion of the synthesis of ritterazine K (**98**)*



Following Winterfeldt and coworkers,[27] the enamine in **104** was hydrogenated to an intermediate amino-ketone with Pd/C, H_2, and acetic acid in a mixture of ethyl acetate and methanol. *In-situ* dimerization afforded pyrazine **105** in excellent yield (**Scheme 2.10**). Global deprotection was effected with TBAF as in **46** to provide ritterazine K in an unoptimized 44% yield, completing its synthesis in 18 steps in the longest linear sequence. To our knowledge, this represents the first confirmed synthesis of **98**, with NMR spectral data matching that of the natural material as isolated in ref. 26.

2.4.3 SYNTHESIS OF 7′-DESOXY-RITTERAZINE B

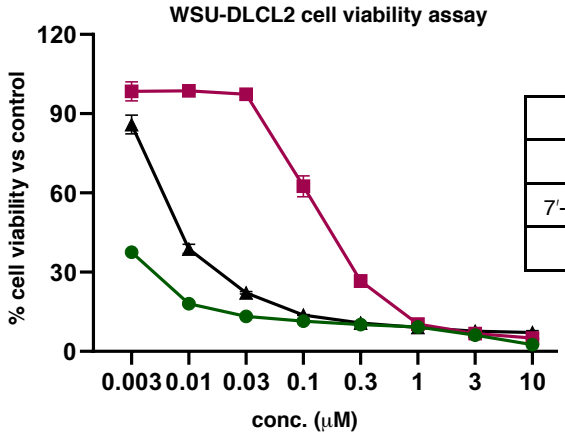
Scheme 2.11 *Completion of the synthesis of 7′-desoxy-ritterazine B (**99**)*



For the synthesis of **99** (**Scheme 2.11**), we needed only to treat α-bromo-ketone **103** from the synthesis of **98** with the canonical azidation conditions[11c] to prevent elimination of N_2 and obtain azido-ketone **106** as the western steroid coupling partner. This, combined with the natural eastern coupling partner **93** under the usual Lewis-acidic conditions, provided pyrazine **107** in 69% yield. Global deprotection then furnished 7′-desoxy-ritterazine B in 72% yield, representing its first synthesis.

2.5 BIOLOGICAL EVALUATION

Figure 2.6 *In vitro cytotoxicity of BSPs against triple-hit lymphoma*



compound	LC_{50} vs WSU-DLCL2
ritterazine B (**46**)	< 3 nM
7′-desoxy-ritterazine B (**99**)	9.3 nM
ritterazine K (**98**)	154 nM

Having obtained ample quantities of all three BSPs,[19] we initiated a collaboration with the City of Hope Comprehensive Cancer Center (CoH CCC) to determine their anticancer activities. The CoH CCC has recently begun investigations to develop treatments for highly aggressive triple-hit lymphoma,[28] using a cell line dubbed WSU-DLCL2 for in vitro studies. To our delight, **46**, **98**, and **99** all exhibited substantial cytotoxicity in this cell line (**Figure 2.6**). **46** again displayed the strongest potency (LC_{50}), below the assay detection limit of 3 nM. **99** exhibited slightly weaker, though still significant activity at roughly 9 nM. Finally, the homodimer **98** showed the weakest activity at 154 nM, a roughly 50-100-fold decrease from **46** and 16-fold decrease from **99**. These results lend further support to the hypothesis that non-identical steroid fragments are required for strong BSP activity.[2] They further lend insight to the SAR around the western fragment of **46**, demonstrating experimentally for the first time the deleterious effect of removing the C7′ alcohol.

2.6 CONCLUSION

In summary, the first total synthesis of ritterazine B has been completed starting from the simple steroid *trans*-dehydroandrosterone and using a unified approach to both steroid fragments. Key features of the strategy include titanium-mediated propargylations to access differentiated alkynes as well as gold-catalyzed, diastereoselective spirocyclizations to forge the spiroketals. The approach has allowed access to two additional bis-steroidal pyrazines to date, the homodimer ritterazine K and the chimeric 7′-desoxy-ritterazine B.

Investigations into the biological activity of the prepared compounds are currently underway and ritterazine B has demonstrated potent activity against triple-hit lymphoma in vitro. It is worth noting that several multi-milligram batches of these materials have been prepared thus far, and though we have elected to perform the final steps on small scale for safety, synthesis of the coupling fragments has proven to be scalable. We expect that our developed route will provide ample material for activity and mechanism of action studies, enabling further investigation of the BSPs as anticancer therapeutics.

2.7 EXPERIMENTAL SECTION

2.7.1 MATERIALS AND METHODS

Unless otherwise stated, reactions were performed under a nitrogen atmosphere using freshly dried solvents utilizing standard Schlenk techniques. Tetrahydrofuran (THF), methylene chloride (CH_2Cl_2/DCM), diethyl ether (Et_2O), benzene (PhH) and toluene (PhMe) were dried by passing through activated alumina columns. Dichloroethane ($(CH_2Cl)_2$/DCE), triethylamine (Et_3N), and isopropanol (*i*-PrOH) were distilled over calcium hydride prior to use. 2,6-lutidine and 3-butyn-2-ol were distilled by short-path distillation at ambient pressure prior to use. Unless otherwise stated, chemicals and reagents were purchased from Sigma-Aldrich and used as received. Glassware was oven-dried at 120 °C for a minimum of four hours, or flame-dried utilizing a Bunsen burner under high vacuum. All reactions were monitored by thin-layer chromatography using EMD/Merck silica gel 60 F254 pre-coated plates (0.25 mm) and were visualized by UV (254 nm), *p*-anisaldehyde staining. Flash column chromatography was performed using silica gel (SiliaFlash® P60, particle size 40-63 microns [230 to 400 mesh]) purchased from Silicycle. ^1H and ^{13}C NMR spectra were recorded on a Bruker Avance III HD with Prodigy Cryoprobe (at 400 MHz and 101 MHz, respectively) or a Varian Inova 500 (at 500 MHz and 101 MHz, respectively) and are reported relative to internal $CHCl_3$ (^1H, $\delta = 7.26$) and $CDCl_3$ (^{13}C, $\delta = 77.0$), C_6H_6 (^1H, $\delta = 7.16$) and C_6D_6 (^{13}C, $\delta = 128$), or C_5H_5N (^1H, $\delta = 8.74, 7.58, 7.22$) and C_5D_5N (^{13}C, $\delta = 150.35, 135.91, 123.87$). Data for ^1H NMR spectra are reported as follows: chemical shift (δ ppm) (multiplicity, coupling constant (Hz), integration). Multiplicity and qualifier abbreviations are as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet. IR spectra were recorded on a Perkin Elmer Paragon

1000 spectrometer and are reported in frequency of absorption (cm^{-1}). HRMS were acquired from the Caltech Mass Spectral Facility using fast-atom bombardment (FAB), electrospray ionization (ES+-TOF) or electron impact (EI). Optical rotations were measured on a Jasco P-2000 polarimeter using a 100 mm path-length cell at 589 nm. Analytical SFC was performed with a Mettler SFC supercritical CO_2 analytical chromatography system with a Chiralcel AD-H column (4.6 mm x 25 cm).

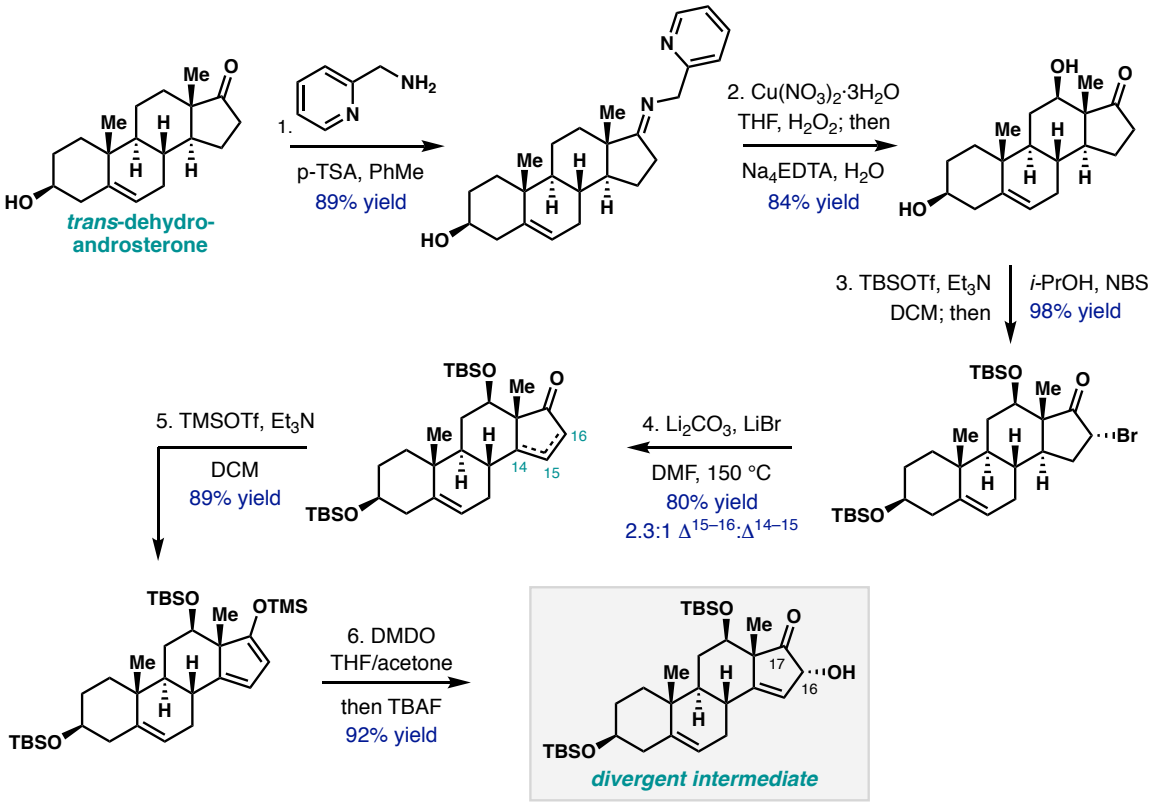
Computational details: DFT optimizations and energy evaluations were performed using the Entos Qcore[20] software package v0.8.14 on Entos clusters hosted on their academic JupyterHub servers. We thank Entos for providing access to these servers. Optimized structure visualizations in the main text and in this document were generated using Avogadro from the .xyz coordinate files obtained from Entos outputs. Si protecting groups were removed from these renderings for clarity but were present in the calculations. All optimizations and energy evaluations were conducted at the BLYP/6-31G** level of theory.

Abbreviations used: Et_2O – diethyl ether; CH_2Cl_2 – methylene chloride; PhMe – toluene; EtOAc – ethyl acetate; THF – tetrahydrofuran; *t*-BuOH – *tert*-butanol; $CHCl_3$ – chloroform; MeCN – acetonitrile; DMF – *N,N*-dimethylformamide; DMA – *N,N*-dimethylacetamide; dtbbpy – 4,4'-di*tert*butylbipyridine; dme – 1,2-dimethoxyethane; TBME – *tert*-butyl methyl ether

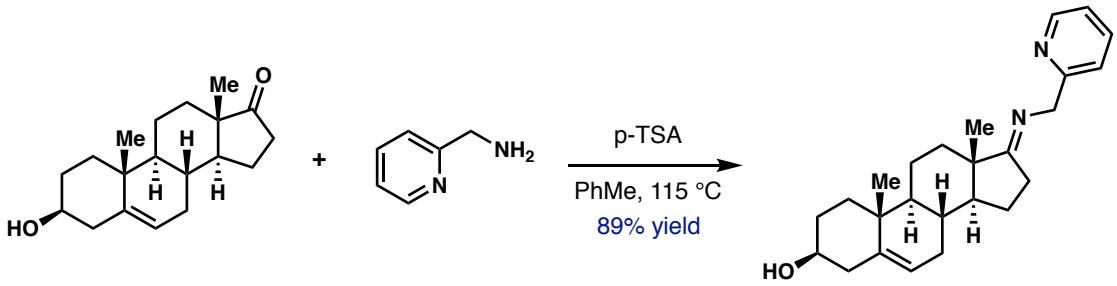
2.7.2 SYNTHETIC PROCEDURES

2.7.2.1 Ritterazine B

Synthesis of α-hydroxy ketone 59



Preparation of imine 54[13b]:



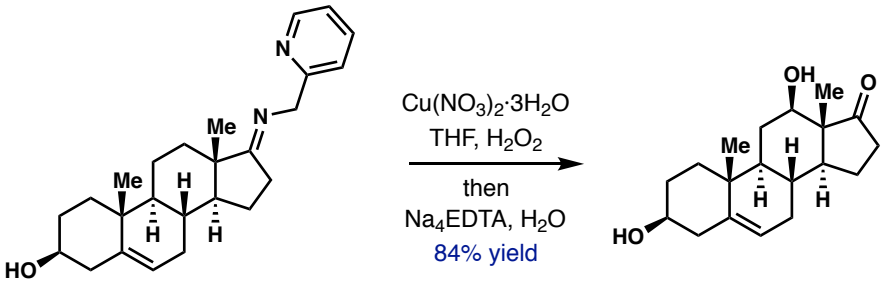
A 250 mL round-bottom flask with a stir bar was equipped with a Dean–Stark apparatus and reflux condenser and was charged with *trans*-dehydroandrosterone (**53**, 10.0 g, 34.7

mmol), 2-aminomethyl pyridine (10.7 mL, 104 mmol, 3.0 equiv), *p*-TsOH (150 mg, 867 μmol, 2.5 mol %) and toluene (87 mL, 0.4 M). The resulting solution was heated to reflux for 8 h. After cooling to ambient temperature, the reaction was diluted with $CHCl_3$ (80 mL), quenched with st. $NaHCO_3$ aq. (40 mL). The layers were separated, and the aqueous layer was extracted with $CHCl_3$ (3 × 20 mL). The combined organic layers were washed with brine (50 mL) and dried over anhydrous Na_2SO_4, filtered, and concentrated. The crude mixture was disolved in a minimal amount of acetone and diluted with hexane to precipitate out the product. The resulting solid was collected by filtration to afford imine **54** (11.6 g, 89% yield) as a white solid. Spectroscopic data were identical to those reported in the literature.[13a]

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.51 (m, 1H), 7.65 (td, *J* = 7.7, 1.8 Hz, 1H), 7.43 (d, *J* = 7.9 Hz, 1H), 7.13 (m, 1H), 5.37 (d, *J* = 5.1 Hz, 1H), 4.60 (q, *J* = 16.5 Hz, 2H), 3.52 (m, 1H), 2.53 – 2.43 (m, 1H), 2.36 – 2.29 (m, 2H), 2.26 (m, 1H), 2.07 (m, 2H), 1.92 – 1.87 (m, 1H), 1.87 – 1.81 (m, 2H), 1.73 – 1.60 (m, 3H), 1.60 – 1.53 (m, 1H), 1.53 – 1.44 (m, 2H), 1.41 (dd, *J* = 12.6, 3.9 Hz, 1H), 1.19 (m, 1H), 1.11 (dd, *J* = 13.3, 3.8 Hz, 1H), 1.04 (s, 3H), 1.03 – 0.98 (m, 1H), 0.92 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 149.08, 141.22, 136.74, 121.78, 121.75, 121.24, 71.76, 58.01, 53.41, 50.62, 45.81, 42.40, 37.37, 36.81, 34.15, 31.77, 31.67, 31.42, 28.25, 23.48, 20.88, 19.59, 16.36.

Preparation of diol 55[13b]:



A 500 mL round-bottom flask with a stir bar was charged with imine **54** (10.0 g, 26.4 mmol), $Cu(NO_3)_2$•$3H_2O$ (7.02 g, 29.1 mmol, 1.1 equiv) and THF (130 mL, 0.2 M). The heterogeneous mixture was stirred vigorously for 45 min at room temperature, and then 50 °C for 30 min to give a pale blue precipitate. After cooling to room temperature, hydrogen peroxide (13.5 mL, 132 mmol, 5.0 equiv, 30 wt % in H_2O) was then added to the reaction mixture dropwise which resulted in the dissolution of the precipitate to give a deep blue solution. This was accompanied by a slight exotherm and formation of gas. The reaction mixture was then stirred for 3 h at room temperature. Then, EtOAc (100 mL) and saturated aqueous Na_4EDTA (22.1 g, 58.1 mmol, 2.2 equiv in 100 mL of H_2O) were added, and the reaction mixture was stirred for 21 h. The layers were separated. The aqueous layer was further extracted with EtOAc (3 × 200 mL), combined organic layers were washed with 1 N HCl aq. (2 × 100 mL), st. $NaHCO_3$ aq. (100 mL), brine (2 × 100 mL), dried over anhydrous Na_2SO_4, filtered, and concentrated. The crude product was triturated with Et_2O (50 mL) and the resulting solid was collected by filtration, washed with Et_2O (50 mL), dried under high vacuum to give diol **55** (5.28 g, 17.3 mmol, 66% yield, two steps) as a white solid. The Et_2O residue was concentrated in vacuo, and subjected to flash column chromatography (50 to 75% EtOAc/hexanes) to afford diol **55** (800 mg, 2.62 mmol, 10% yield, two steps) as a white solid. This material was combined with that from trituration to

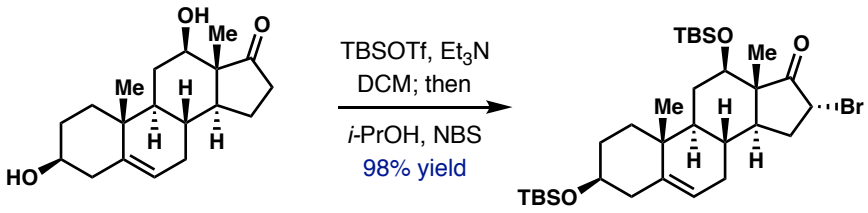
give **55** (6.08 g, 19.9 mmol, 84% yield, single step). Spectroscopic data were identical to those reported in the literature.[13a]

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.38 (dt, *J* = 5.0, 1.8 Hz, 1H), 3.81 (ddd, *J* = 11.3, 4.7, 1.3 Hz, 1H), 3.51 (m, 1H), 3.08 (d, *J* = 1.4 Hz, 1H), 2.54 – 2.42 (m, 1H), 2.33 (ddd, *J* = 13.1, 5.1, 2.2 Hz, 1H), 2.24 (m, 1H), 2.18 – 2.06 (m, 2H), 2.04 – 1.95 (m, 1H), 1.87 (m, 1H), 1.85 – 1.82 (m, 1H), 1.80 (t, *J* = 4.7 Hz, 1H), 1.72 – 1.65 (m, 1H), 1.64 (m, 1H), 1.61 (m, 1H), 1.57 – 1.54 (m, 1H), 1.53 – 1.48 (m, 1H), 1.47 – 1.40 (m, 1H), 1.29 – 1.21 (m, 1H), 1.16 – 1.08 (m, 2H), 1.04 (s, 3H), 0.96 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 141.14, 120.90, 72.81, 71.68, 51.53, 49.67, 49.25, 42.23, 37.28, 36.91, 35.90, 31.69, 30.71, 30.51, 28.33, 21.84, 19.49, 8.20.

Note: The C17 ketone ^{13}C signal is known to appear at 222 ppm.[13a] The scan width of the ^{13}C experiment does not reach this chemical shift and thus the peak is absent; the remainder of the spectrum is identical to that reported in ref 13a.

Preparation of α-bromo ketone 56:



A 1 L round-bottom flask with a stir bar was charged with diol **55** (14.0 g, 46.0 mmol), Et_3N (48.1 mL, 345 mmol, 7.5 equiv) and CH_2Cl_2 (230 mL, 0.2 M). TBSOTf (52.8 mL, 230 mmol, 5.0 equiv) was added at room temperature. After stirring at room temperature

for 30 min, the reaction was cooled to 0 °C stirred at this temperature for 15 min. Thereafter, *i*-PrOH (10.6 mL, 138 mmol, 3.0 equiv) was added. After stirring at 0 °C for 30 min, NBS (16.4 g, 92.0 mmol, 2.0 equiv) was added in one portion and stirred at 0 °C for 30 min. The reaction was diluted with hexane (200 mL), quenched with H_2O (200 mL). The layers were separated. The aqueous layer was further extracted with hexane (3 × 200 mL), combined organic layers were washed with st. $NaHCO_3$ aq. (200 mL), brine (2 × 200 mL), dried over anhydrous Na_2SO_4, filtered, and concentrated. The crude product was purified by flash column chromatography (5 to 7 to 10% Et_2O/hexanes) to afford α-bromo ketone **56** (27.3 g, 44.7 mmol, 98% yield) as a colorless foam.

TLC (10% Et_2O/hexanes): R_f = 0.50 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.33 (dt, *J* = 5.5, 1.8 Hz, 1H), 4.50 (dd, *J* = 6.9, 1.7 Hz, 1H), 3.80 (dd, *J* = 10.7, 4.9 Hz, 1H), 3.48 (m, 1H), 2.31 – 2.26 (m, 1H), 2.24 (m, 1H), 2.23 – 2.17 (m, 2H), 2.04 (ddd, *J* = 12.4, 5.2, 2.6 Hz, 1H), 1.82 – 1.73 (m, 2H), 1.71 (t, *J* = 4.9 Hz, 1H), 1.68 – 1.65 (m, 1H), 1.65 – 1.58 (m, 2H), 1.44 (m, 1H), 1.15 – 1.05 (m, 2H), 1.02 (s, 3H), 0.95 (s, 3H), 0.94 (s, 1H), 0.90 (d, *J* = 5.2 Hz, 18H), 0.15 (s, 3H), 0.06 (s, 9H).

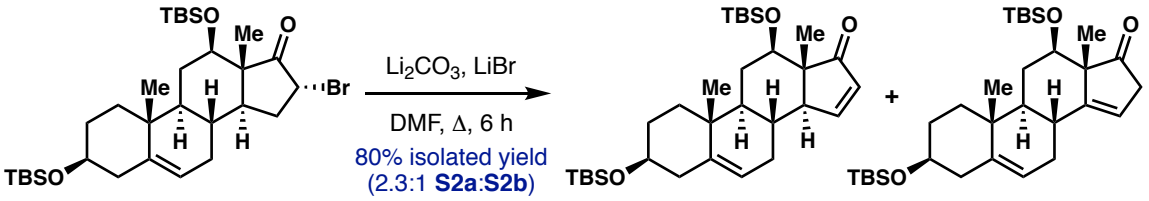
^{13}C NMR (101 MHz, $CDCl_3$): δ 210.62, 141.85, 120.21, 72.61, 72.50, 51.99, 49.12, 46.89, 46.78, 42.79, 37.28, 36.78, 33.50, 32.05, 31.38, 30.36, 29.90, 26.07, 19.50, 18.41, 18.28, 9.13, -4.14, -4.44, -4.46, -4.68.

FTIR (NaCl, thin film, cm^{-1}): 2929, 2855, 1755, 1747, 1733, 1470, 1462, 1362, 1250, 1092 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{31}H_{54}O_3Si_2Br$: 609.2795 $[M+H-H_2]^+$; found: 609.2778.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\boldsymbol{21}}$: –94° (*c* = 0.400, $CHCl_3$).

Preparation of unsaturated ketone 57:



A 1 L round-bottom flask with a stir bar was charged with α-bromo ketone **56** (27.3 g, 44.7 mmol), Li_2CO_3 (13.6 g, 224 mmol, 5.0 equiv), LiBr (14.6 g, 215 mmol, 4.8 equiv) and DMF (300 mL, 0.15 M). The resulting mixture was heated to 150 °C for 12 h. After cooling to room temperature, the reaction was filtered through a glass filter, washed with hexane (2 × 50 mL), the combined brown solution was diluted with hexane (100 mL) and H_2O (300 mL). The layers were separated. The aqueous layer was further extracted with hexane (3 × 150 mL), combined organic layers were washed with brine (2 × 150 mL), dried over anhydrous Na_2SO_4, filtered, and concentrated. The crude product was purified by flash column chromatography (5 to 7.5 to 10% Et_2O/hexanes) to afford unsaturated ketone **57** (19.0 g, 35.8 mmol, mixture of two regio isomers **57a** ($\Delta^{14–15}$) and **57b** ($\Delta^{15–16}$), **57a**:**57b** = 2.3:1, 80% yield) as a yellow foam. This mixture can be used directly in the next reaction, though it was purified further to obtain pure **57a** for analytical characterization.

Pure **57a**

TLC (8% Et_2O/hexanes): R_f = 0.26 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 7.40 (dd, *J* = 6.0, 1.8 Hz, 1H), 5.97 (dd, *J* = 6.0, 3.1 Hz, 1H), 5.36 (m, 1H), 4.01 (dd, *J* = 10.1, 5.4 Hz, 1H), 3.48 (tt, *J* = 10.5, 4.7 Hz, 1H), 2.26 (m, 3H), 2.15 (ddd, *J* = 11.3, 3.1, 1.8 Hz, 1H), 1.80 (m, 1H), 1.76 (m, 3H), 1.72 – 1.66 (m, 1H),

1.60 – 1.56 (m, 2H), 1.54 – 1.52 (m, 1H), 1.25 (s, 1H), 1.09 (s, 3H), 1.07 (s, 3H), 0.92 (s, 9H), 0.89 (s, 9H), 0.16 (s, 3H), 0.06 (d, J = 0.8 Hz, 9H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 210.60, 156.63, 142.29, 132.55, 119.87, 77.35, 77.03, 76.71, 72.39, 70.80, 55.18, 54.74, 50.71, 42.71, 37.07, 36.92, 31.91, 31.28, 30.38, 27.71, 25.94, 19.32, 18.28, 18.19, 15.10, -4.31, -4.57, -4.59, -4.75.

FTIR (NaCl, thin film, cm^{-1}): 2953, 2928, 2854, 1723, 1714, 1470, 1462, 1456, 1363, 1252, 1092 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{31}H_{55}O_3Si_2$: 531.3690 $[M+H]^+$; found: 531.3688.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: –116° ($c$ = 0.267, $CHCl_3$).

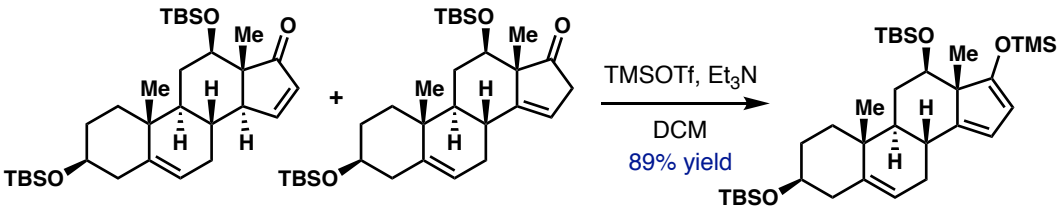
Mixture **57a**:**57b** (1:1):

^{1}H NMR (400 MHz, $CDCl_3$): δ 7.40 (dd, J = 6.0, 1.8 Hz, 1H), 5.97 (dd, J = 5.9, 3.1 Hz, 1H), 5.60 (q, J = 2.2 Hz, 1H), 5.38 (ddt, J = 14.9, 3.5, 2.0 Hz, 2H), 4.01 (dd, J = 10.1, 5.4 Hz, 1H), 3.57 – 3.51 (m, 1H), 3.48 (tt, J = 10.5, 4.8 Hz, 2H), 2.95 – 2.78 (m, 2H), 2.33 – 2.27 (m, 2H), 2.25 (s, 1H), 2.23 (dd, J = 5.3, 1.9 Hz, 2H), 2.20 (dd, J = 5.4, 2.1 Hz, 1H), 2.17 (dd, J = 3.1, 1.9 Hz, 1H), 2.15 – 2.09 (m, 2H), 1.81 (d, J = 4.7 Hz, 1H), 1.80 – 1.73 (m, 5H), 1.67 (dt, J = 13.4, 5.0 Hz, 2H), 1.62 – 1.59 (m, 1H), 1.57 (td, J = 6.6, 6.2, 2.6 Hz, 2H), 1.53 (d, J = 3.0 Hz, 1H), 1.11 (s, 3H), 1.08 (s, 3H), 1.07 (s, 3H), 1.06 (s, 3H), 1.05 – 1.02 (m, 2H), 0.92 (s, 9H), 0.91 (s, 9H), 0.89 (s, 18H), 0.16 (s, 3H), 0.06 (s, 15H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 218.90, 210.71, 156.75, 151.57, 142.40, 140.98, 132.67, 120.49, 120.00, 116.41, 72.55, 72.51, 72.02, 70.91, 56.53, 55.29, 54.85, 50.82, 48.74, 42.83, 42.76, 41.75, 37.26, 37.18, 37.11, 37.03, 32.03, 31.40, 30.93, 30.85, 30.50, 27.86,

27.82, 26.07, 26.03, 19.45, 19.32, 18.41, 18.31, 18.28, 15.23, 13.54, -4.18, -4.32, -4.45, -4.46, -4.62, -4.75.

Preparation of TMS enol ether 58:



A 1 L round-bottom flask with a stir bar was charged with unsaturated ketone **57** (16.7 g, 31.5 mmol, mixture of two regio isomers, **57a**:**57b** = 2.3:1), Et_3N (19.8 mL, 142 mmol, 4.5 equiv) and CH_2Cl_2 (315 mL, 0.1 M). TMSOTf (17.1 mL, 94.5 mmol, 3.0 equiv) was added to the reaction at room temperature. After stirring at room temperature for 30 min, the reaction was diluted with hexane (150 mL), quenched with H_2O (150 mL). The layers were separated. The aqueous layer was further extracted with hexane (3 × 150 mL), combined organic layers were washed with st. $NaHCO_3$ aq. (150 mL) and brine (2 × 150 mL), dried over anhydrous Na_2SO_4, filtered, and concentrated in vacuo. The crude product was dissolved in Et_2O (80 mL), and then MeCN (400 mL) was added. The resulting solid was collected by filtration, washed with MeCN (2 × 80 mL), dried under high vacuum to afford TMS enol ether **58** (17.2 g, 28.0 mmol, 89% yield) as a white solid.

TLC (2% Et_2O/hexanes): R_f = 0.38 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.76 (t, *J* = 2.2 Hz, 1H), 5.40 (dt, *J* = 4.2, 2.0 Hz, 1H), 5.13 (d, *J* = 2.4 Hz, 1H), 3.48 (m, 1H), 3.19 (dd, *J* = 10.6, 4.7 Hz, 1H), 2.33 – 2.24 (m, 1H), 2.22 (m, 1H), 2.19 (m, 1H), 2.15 (m, 1H), 2.11 – 2.04 (m, 1H) 1.80 – 1.70 (m, 2H), 1.63 –

1.57 (m, 2H), 1.54 – 1.49 (m, 1H), 1.07 (s, 3H), 1.01 (s, 3H), 0.90 (s, 9H), 0.89 (s, 9H), 0.86 – 0.80 (m, 1H), 0.22 (s, 9H), 0.06 (m, 7H), 0.02 (s, 3H), 0.01 (s, 3H).

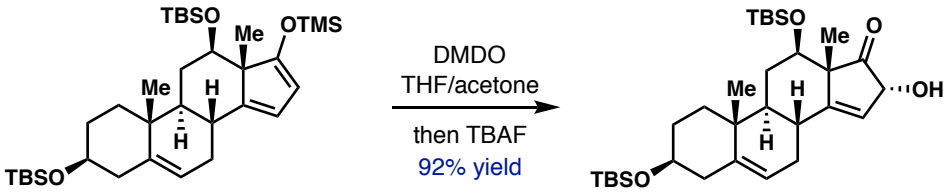
^{13}C NMR (101 MHz, $CDCl_3$): δ 170.35, 145.14, 140.92, 121.09, 119.48, 101.93, 76.31, 72.71, 56.13, 50.20, 42.83, 37.64, 37.11, 32.12, 31.76, 30.81, 28.49, 26.18, 26.09, 19.53, 18.42, 18.33, 12.39, 0.22, -4.42, -4.44, -4.53.

FTIR (NaCl, thin film, cm^{-1}): 2928, 2892, 2856, 1714, 1471, 1462, 1386, 1372, 1361, 1255, 1089 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{34}H_{62}O_3Si_3$: 602.4007 [M+•]$^-$; found: 602.4002.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: +113° (*c* = 0.200, $CHCl_3$).

Preparation of α-hydroxy ketone 59:



A 4 L flask with a stir bar was charged with TMS enol ether **58** (17.2 g, 28.0 mmol), Me_2C=CHMe (5.93 mL, 56.0 mmol, 2.0 equiv) and THF (1500 mL, 0.02 M). The solution was cooled to –78 °C and stirred at this temperature for 30 min. Thereafter, DMDO solution (0.060 M in acetone, 513 mL, 30.8 mmol, 1.1 equiv) was added dropwise by cannula. The reaction was stirred at –78 °C for 30 min, then warmed to room temperature. After stirring for 30 min, TBAF (1.0 M in THF, 28.0 mL, 28.0 mmol, 1.0 equiv) was added dropwise. After stirring for 30 min, the reaction was diluted with EtOAc (500 mL), quenched with H_2O (500 mL). The layers were separated. The aqueous layer was further extracted with EtOAc (3 × 150 mL), combined organic layers were washed with brine (2 × 150 mL), dried

over anhydrous Na_2SO_4, filtered, and concentrated. The crude product was purified by flash column chromatography (5 to 7.5% EtOAc/hexanes) to afford α-hydroxy ketone **59** (14.1 g, 25.8 mmol, 92% yield) as a colorless foam.

Note: Compound **59** decomposes within hours at room temperature under air and is stable for roughly two weeks at freezer temperature under inert atmosphere.

Preparation of DMDO solution in acetone: A three-necked 4 L flask was charged with $NaHCO_3$ (1.05 kg), acetone (1.1 L) and H_2O (1.7 L). The flask was fitted with a tall reflux condenser, which was connected to a distillation neck with a 1L receiving flask charged with activated 3Å MS pellets. The distillation neck was cooled with circulating ice water and the receiving flask was cooled in a dry ice/acetone bath. Oxone (2.5 kg) was added in several portions over 15 min at room temperature, the resulting DMDO was carefully distilled directly from the reaction mixture over 30 min under low pressure using a diaphragm pump. The concentration of DMDO solution was determined by ^{1}H NMR using thioanisole.

TLC (15% EtOAc/hexanes): R_f = 0.41 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.76 (t, *J* = 2.0 Hz, 1H), 5.39 (m, 1H), 4.76 (dd, *J* = 3.7, 1.9 Hz, 1H), 3.63 (dd, *J* = 11.2, 4.5 Hz, 1H), 3.54 – 3.43 (m, 1H), 2.53 (s, 1H), 2.33 – 2.26 (m, 1H), 2.23 (m, 2H), 2.17 (m, 1H), 2.06 (m, 1H), 1.81 (dt, *J* = 12.7, 3.3 Hz, 1H), 1.75 (m, 1H), 1.69 (dt, *J* = 13.3, 4.2 Hz, 1H), 1.63 – 1.49 (m, 3H), 1.20 (s, 3H), 1.16 – 1.10 (m,

1H), 1.08 (m, 1H), 1.04 (m, 4H), 0.92 (s, 9H), 0.91 (s, 3H), 0.89 (s, 9H), 0.11 (s, 3H), 0.10 (s, 2H), 0.06 (s, 6H), 0.03 (s, 3H).

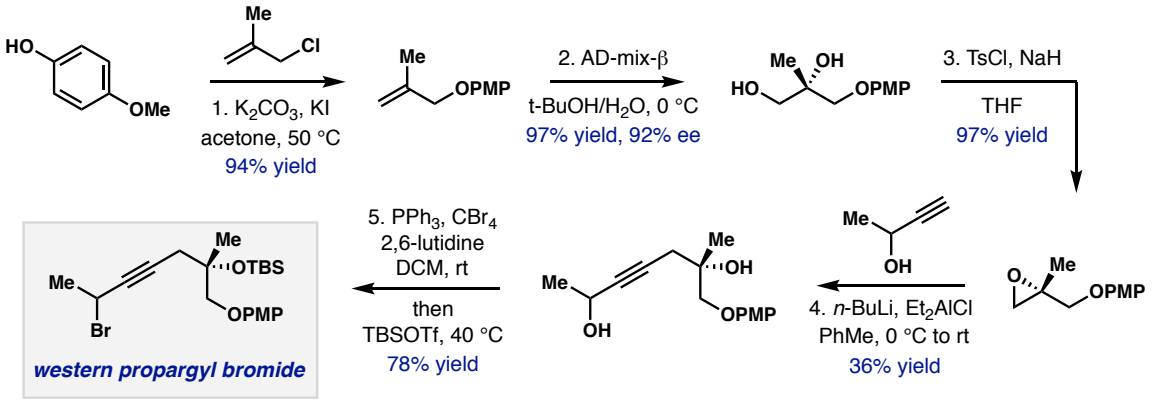
^{13}C NMR (101 MHz, $CDCl_3$): δ 217.10, 154.88, 140.87, 121.40, 120.18, 77.45, 77.14, 76.82, 74.14, 72.43, 71.67, 54.43, 47.31, 42.69, 37.17, 36.90, 31.96, 30.84, 30.73, 27.91, 26.05, 26.01, 25.77, 19.24, 18.38, 18.22, 14.39, -3.46, -4.12, -4.47, -4.74.

FTIR (NaCl, thin film, cm^{-1}): 3410, 2929, 2888, 2857, 1759, 1471, 1462, 1371, 1360, 1252, 1096 cm^{-1}.

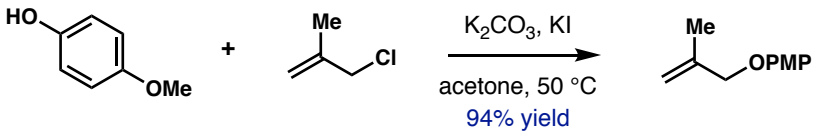
HRMS (FAB+, m/z): calc'd for $C_{31}H_{53}O_4Si_2$: 545.3482 $[M+H]^+$; found: 545.3461.

$[\alpha]_D^{21}$: –37° (*c* = 0.333, CH_2Cl_2).

Synthesis of western propargyl bromide 51



Preparation of PMP ether 67:



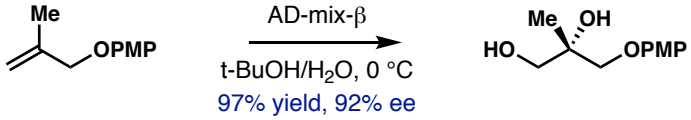
A 500 mL round-bottom flask with a stir bar was charged with *p*-methoxy phenol (20.0 g, 161 mmol), K_2CO_3 (24.5 g, 177 mmol, 1.1 equiv), KI (2.67 g, 16.1 mmol, 10 mol %), 2-methyl allylchloride (17.3 mL, 177 mmol, 1.1 equiv) and acetone (80 mL, 2.0 M). The

resulting mixture was heated to 50 °C and stirred for 12 h. After cooling to room temperature, the solid was removed by filtration and washed with EtOAc (2 × 50 mL), the resulting solution was diluted with H_2O (100 mL). The layers were separated. The aqueous layer was further extracted with EtOAc (3 × 50 mL), combined organic layers were washed with st. $Na_2S_2O_3$ aq. (50 mL) and brine (2 × 50 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (10 to 15% Et_2O/hexanes) to afford PMP ether **67** (27.0 g, 151 mmol, 94% yield) as colorless crystals. Spectroscopic data were identical to those reported in the literature.[29]

^{1}H NMR (400 MHz, $CDCl_3$): δ 6.91 – 6.79 (m, 4H), 5.09 (m, 1H), 4.98 (m, 1H), 4.39 (m, 2H), 3.77 (s, 3H), 1.83 (m, 3H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 153.95, 153.05, 141.35, 115.84, 114.68, 112.70, 72.55, 55.83, 19.56.

Preparation of diol 68:



A 1 L round-bottom flask with a stir bar was charged with AD-mix-β (70.0 g, 1.4 g/mmol), *t*-BuOH (230 mL, 0.1 M total) and H_2O (250 mL, 0.1 M total). The resulting mixture was stirred at room temperature until the solid was completely dissolved. The reaction was cooled to 0°C and stirred at this temperature for 1 h. Thereafter, PMP ether **67** (8.91 g, 50.0 mmol) was added, rinsed with *t*-BuOH (2 × 10 mL). The reaction was stirred at 0 °C for 24 h, quenched with Na_2SO_3 (75.0 g, 1.5 g/mmol), diluted with EtOAc (250 mL) and H_2O

(120 mL). The layers were separated. The aqueous layer was further extracted with EtOAc (3 × 150 mL), combined organic layers were washed with brine (2 × 150 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (65 to 100% EtOAc/hexanes) to afford diol **68** (10.3 g, 48.6 mmol, 97% yield, 92% ee) as a white solid. Spectroscopic data were identical to those reported in the literature.[30]

TLC (50% EtOAc/hexanes): R_f = 0.21 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 6.90 – 6.79 (m, 4H), 3.94 – 3.83 (m, 2H), 3.77 (s, 3H), 3.65 (dd, *J* = 61.1, 11.2 Hz, 2H), 1.28 (s, 3H).

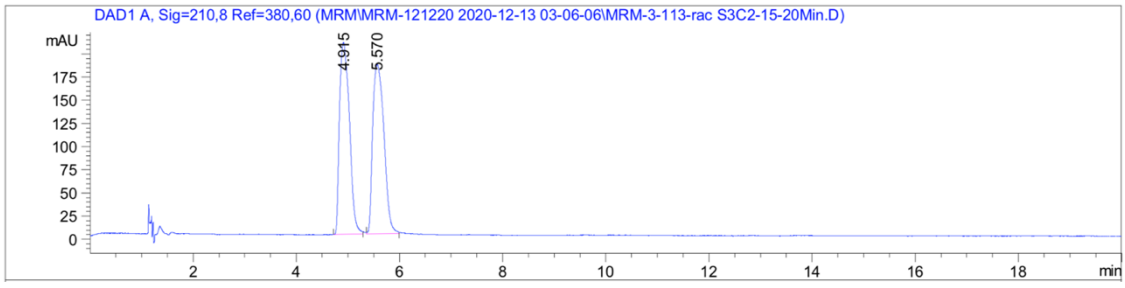
^{13}C NMR (101 MHz, $CDCl_3$): δ 154.39, 152.75, 115.70, 114.84, 73.81, 72.11, 67.95, 55.88, 21.61.

FTIR (NaCl, thin film, cm^{-1}): 3342, 3303, 2923, 2836, 1513, 1470, 1291, 1243, 1148, 1051 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{11}H_{16}O_4$: 212.1049 $[M+\bullet]^-$; found: 212.1022.

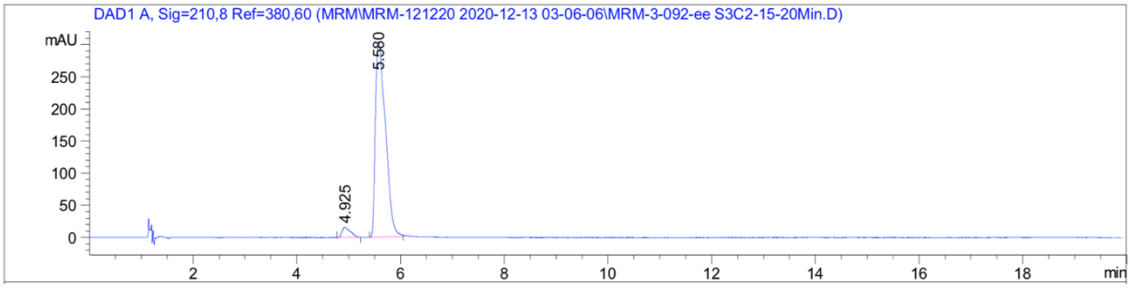
$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: +0.86° (*c* = 0.250, $CHCl_3$).

SFC data for racemic 68:



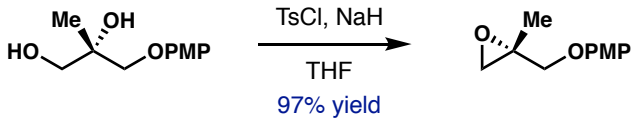
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	4.915	BB	0.1947	2767.92993	207.42258	49.9694
2	5.570	BB	0.2228	2771.32056	182.88441	50.0306

SFC data for enantioenriched 68:



Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	4.925	BB	0.1562	180.82196	15.71460	4.2452
2	5.580	BB	0.1944	4078.62842	304.27750	95.7548

Preparation of epoxide 69:



A 2 L round-bottom flask with a stir bar was charged with diol **68** (10.3 g, 48.6 mmol) and THF (500 mL, 0.1 M). NaH (60% in mineral oil, 5.83 g, 146 mmol, 3.0 equiv) was added in one portion at room temperature. This was accompanied by a slight exotherm and formation of gas. After stirring for 1 h, TsCl (10.2 g, 53.5 mmol, 1.1 equiv) was added in one portion at room temperature. After stirring for 24 h at room temperature, the reaction was poured into an ice-cooled mixture of Et_2O (500 mL), st. NH_4Cl aq. (250 mL) and H_2O (250 mL). The layers were separated. The aqueous layer was further extracted with Et_2O

(3 × 200 mL), combined organic layers were washed with brine (2 × 200 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (7 to 10% EtOAc/hexanes) to afford epoxide **69** (9.15 g, 47.1 mmol, 97% yield) as a white solid.

TLC (50% EtOAc/hexanes): R_f = 0.66 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 6.91 – 6.77 (m, 4H), 4.00 – 3.88 (m, 2H), 3.77 (s, 3H), 2.79 (dd, *J* = 55.2, 4.8 Hz, 2H), 1.48 (s, 3H).

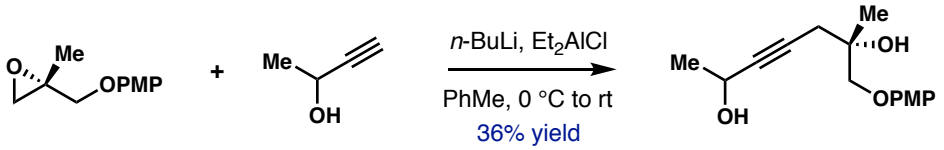
^{13}C NMR (101 MHz, $CDCl_3$): δ 154.25, 152.97, 115.82, 114.76, 72.41, 55.85, 55.77, 52.19, 18.65.

FTIR (NaCl, thin film, cm^{-1}): 3047, 2994, 2934, 2834, 1513, 1505, 1455, 1290, 1234, 1037 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{11}H_{14}O_3$: 194.0943 $[M+\bullet]^-$; found: 194.0969.

$[\alpha]_D^{21}$: –12° (*c* = 0.150, $CHCl_3$).

Preparation of propargyl alcohol 70:



A 200 mL round-bottom flask with a stir bar was charged with (±)-3-butyn-2-ol (392 µL, 5.0 mmol, 2.0 equiv) and PhMe (25 mL, 0.1 M). The resulting solution was cooled to 0 °C and stirred at this temperature for 1 h. Thereafter, *n*-BuLi (2.5 M in hexane, 4.0 mL, 10 mmol, 4.0 equiv) was added dropwise at 0 °C. After stirring at 0 °C for 1 h, Et_2AlCl (25

wt. % in PhMe, 2.7 mL, 5.0 mmol, 2.0 equiv) was added dropwise. After stirring at 0 ° C for 1 h, epoxide **69** (486 mg, 2.5 mmol) in PhMe (2.5 mL, rinsed with 2.5 mL) was added dropwise. The resulting mixture was stirred at room temperature for 18 h, then was diluted with Et_2O (20 mL), quenched with st. NH_4Cl aq. (10 mL), st. Na/K-tartrate aq. (10 mL) and H_2O (10 mL). After vigorously stirring for 1 h, the layers were separated. The aqueous layer was further extracted with Et_2O (3 × 20 mL), combined organic layers were washed with brine (2 × 20 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (50 to 75 to 100% EtOAc/hexanes) to afford propargyl alcohol **70** (238 mg, 0.9 mmol, 1:1 mixture of diastereomers, 36% yield) as a colorless oil.

TLC (50% EtOAc/hexanes): R_f = 0.19 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 6.90 – 6.78 (m, 4H), 4.56 – 4.46 (m, 1H), 3.87 (dd, *J* = 8.9, 1.1 Hz, 1H), 3.84 – 3.80 (m, 1H), 3.77 (s, 3H), 2.63 – 2.52 (m, 3H), 2.51 – 2.24 (m, 1H), 1.42 (dd, *J* = 6.6, 0.8 Hz, 3H), 1.38 (d, *J* = 0.7 Hz, 3H).

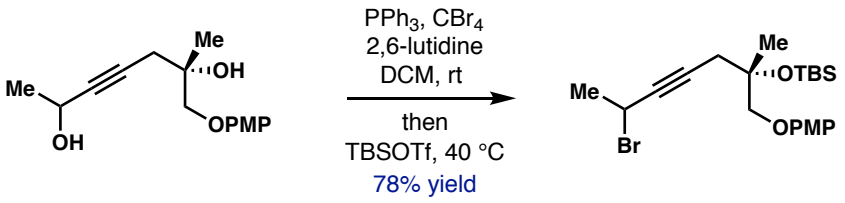
^{13}C NMR (101 MHz, $CDCl_3$): δ 154.30, 152.86, 115.77, 114.81, 85.30, 80.30, 74.45, 71.64, 58.52, 55.87, 29.76, 24.69, 23.87.

FTIR (NaCl, thin film, cm^{-1}): 3390, 2978, 2933, 2249, 2064, 1732, 1504, 1470, 1228, 1049 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{15}H_{20}O_4$: 264.1362 $[M+\bullet]^-$; found: 264.1368.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\boldsymbol{20}}$: +2.3° (*c* = 0.333, $CHCl_3$).

Preparation of propargyl bromide 51:



A 25 mL round-bottom flask was charged with propargyl alcohol **70** (160 mg, 0.60 mmol), CBr_4 (358 mg, 1.08 mmol, 1.8 equiv), 2,6-lutidine (280 μL, 2.4 mmol, 4.0 equiv) and CH_2Cl_2 (3.0 mL, 0.2 M). Ph_3P (262 mg, 1.0 mmol, 1.65 equiv) was added in one portion at room temperature. After stirring for 30 min, TBSOTf (207 μL, 0.9 mmol, 1.5 equiv) was added. After stirring for 30 min at room temperature, the reaction was diluted with hexane (3 mL), quenched with H_2O (3 mL). The layers were separated. The aqueous layer was further extracted with hexane (3 × 3 mL), combined organic layers were washed with st. $Na_2S_2O_3$ aq. (2 × 5 mL), st. $NaHCO_3$ aq. (5 mL) and brine (2 × 5 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (0 to 5 to 10% Et_2O/hexanes) to afford propargyl bromide **51** (207 mg, 0.47 mmol, 1:1 mixture of diastereomers, 78% yield) as a colorless oil.

TLC (10% EtOAc/hexanes): R_f = 0.65 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 6.84 (m, 4H), 4.62 (m, 1H), 3.81 (dd, *J* = 8.9, 1.8 Hz, 1H), 3.77 (m, 4H), 2.60 (ddd, *J* = 16.6, 2.2, 1.5 Hz, 1H), 2.53 (dd, *J* = 16.6, 2.2 Hz, 1H), 1.87 (dd, *J* = 6.8, 0.6 Hz, 3H), 1.39 (d, *J* = 1.0 Hz, 3H), 0.87 (s, 9H), 0.13 (s, 3H), 0.11 (s, 3H).

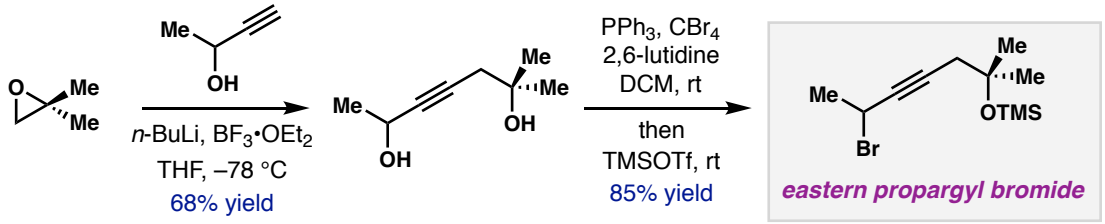
^{13}C NMR (101 MHz, $CDCl_3$): δ 154.04, 153.10, 115.61, 115.59, 114.79, 84.48, 82.60, 74.66, 74.62, 74.58, 55.90, 32.45, 31.02, 27.87, 25.92, 25.09, 18.36, -2.25.

FTIR (NaCl, thin film, cm^{-1}): 2953, 2930, 2857, 1510, 1463, 1287, 1231, 1168, 1062, 834 cm^{-1}.

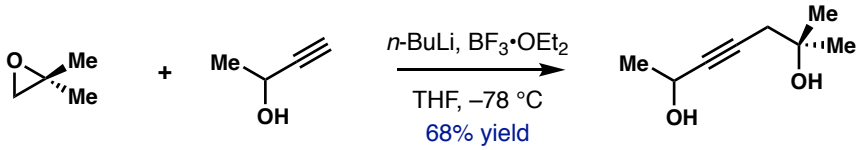
HRMS (FAB+, m/z): calc'd for $C_{21}H_{33}BrO_3Si$: 440.1382 $[M+\bullet]^-$; found: 440.1363.

$[\alpha]_D^{21}$: +10° (*c* = 0.100, $CHCl_3$).

Synthesis of eastern propargyl bromide 52



Preparation of propargyl alcohol 71:



A 1 L round-bottom flask with a stir bar was charged with (±)-3-butyn-2-ol (5.88 mL, 75.0 mmol, 2.0 equiv) and THF (375 mL, 0.1 M). The resulting solution was cooled to –78 °C and stirred at this temperature for 30 min. Thereafter, *n*-BuLi (2.5 M in hexane, 60 mL, 150 mmol, 4.0 equiv) was added via cannula and the resulting mixture was stirred at –78 °C for 5 min, warmed to room temperature, and stirred at this temperature for 30 min, then cooled back to –78 °C. After stirring at –78 ° C for 30 min, $BF_3 \cdot OEt_2$ (6.95 mL, 56.3 mmol, 1.5 equiv) was added dropwise, followed by isobutylene oxide (3.33 mL, 37.5 mmol). The reaction was stirred at –78 °C for 30 min, then warmed to room temperature. After stirring for 12 h, the reaction was diluted with Et_2O (150 mL), quenched with st. NH_4Cl aq. (150

mL) and H_2O (150 mL). The layers were separated. The aqueous layer was further extracted with Et_2O (10 × 200 mL), combined organic layers were washed with brine (2 × 200 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (50 to 75% EtOAc/hexanes) to afford propargyl alcohol **71** (3.61 g, 25.4 mmol, 68% yield) as a colorless oil.

TLC (50% EtOAc/hexanes): R_f = 0.16 (*p*-anisaldehyde)

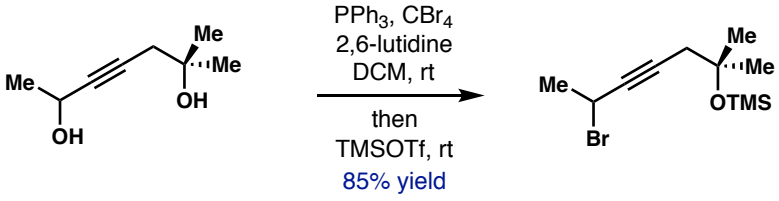
^{1}H NMR (400 MHz, $CDCl_3$): ^{1}H NMR (500 MHz, Chloroform-*d*) δ 4.55 (qt, *J* = 6.7, 2.0 Hz, 1H), 2.40 (d, *J* = 1.9 Hz, 2H), 1.82 (s, 2H), 1.46 (dd, *J* = 6.5, 0.5 Hz, 3H), 1.31 (s, 6H).

^{13}C NMR (101 MHz, $CDCl_3$): ^{13}C NMR (101 MHz, $CDCl_3$) δ 85.42, 81.00, 70.21, 58.44, 34.33, 28.78, 24.72.

FTIR (NaCl, thin film, cm^{-1}): 3356, 2973, 2930, 2249, 1666, 1462, 1366, 1154, 1080, 1006 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_8H_{13}O$: 125.0966 $[M–H_2O]^+$; found: 125.0994.

Preparation of propargyl bromide 52:



A 500 mL round-bottom flask with a stir bar was charged with propargyl alcohol **71** (2.93 g, 20.6 mmol), CBr_4 (12.3 g, 37.1 mmol, 1.8 equiv), 2,6-lutidine (9.60 mL, 82.4 mmol, 4.0 equiv) and CH_2Cl_2 (100 mL, 0.2 M). Ph_3P (8.92 g, 34.0 mmol, 1.65 equiv) was added in one portion at room temperature. After stirring for 30 min, TMSOTf (5.59 mL, 30.9 mmol,

1.5 equiv) was added. After stirring for 30 min at room temperature, the reaction was diluted with hexane (100 mL), quenched with H_2O (100 mL). The layers were separated. The aqueous layer was further extracted with hexane (3 × 100 mL), combined organic layers were washed with st. $Na_2S_2O_3$ aq. (2 × 100 mL), st. $NaHCO_3$ aq. (100 mL) and brine (2 × 100 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (0 to 5 to 10% Et_2O/hexanes) to afford propargyl bromide **52** (4.84 g, 17.5 mmol, 85% yield) as a colorless oil.

TLC (10% EtOAc/hexanes): R_f = 0.75 (*p*-anisaldehyde)

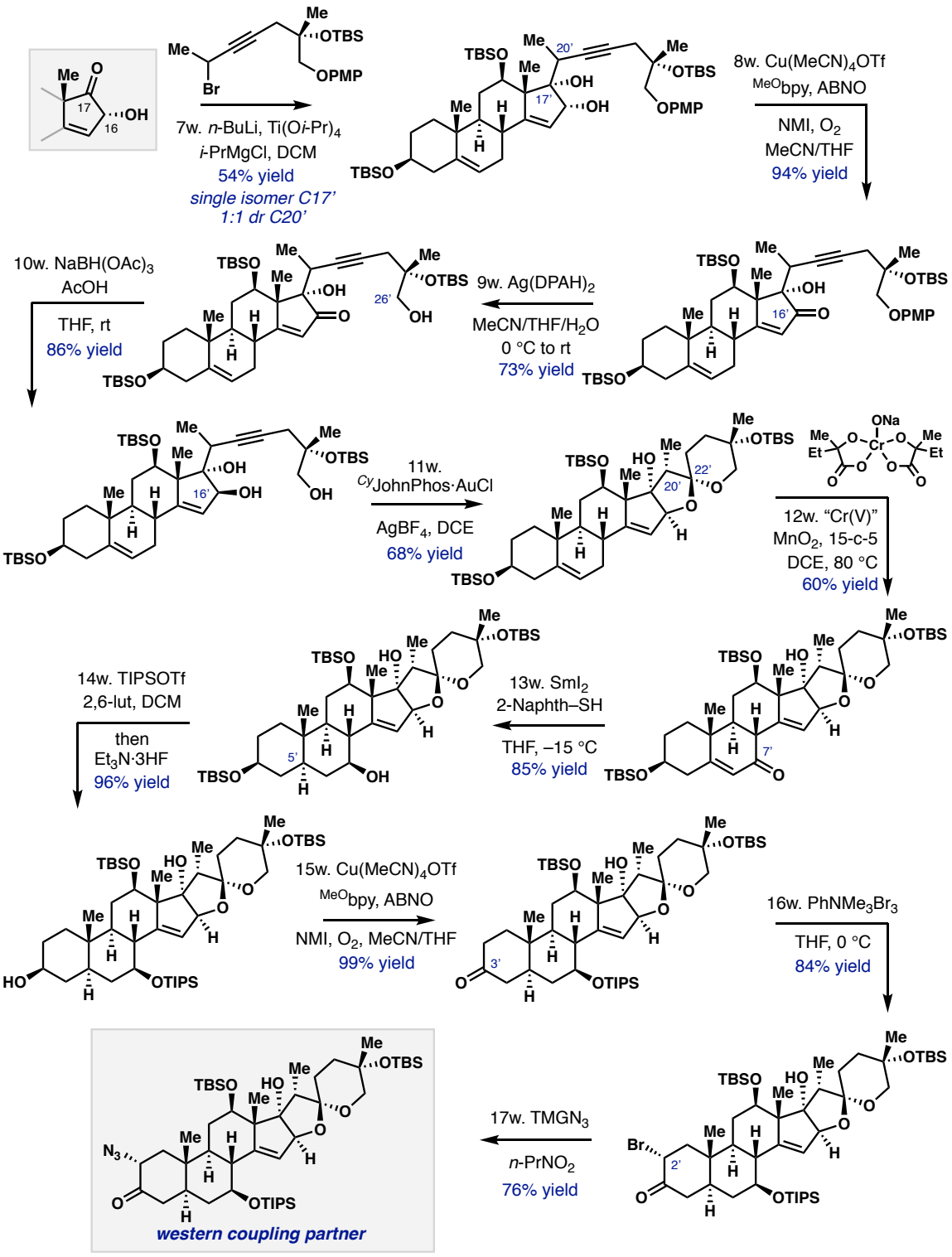
^{1}H NMR (400 MHz, $CDCl_3$): δ 4.65 (qt, *J* = 6.8, 2.2 Hz, 1H), 2.38 (d, *J* = 2.2 Hz, 2H), 1.90 (d, *J* = 6.8 Hz, 3H), 1.32 (s, 6H), 0.13 (s, 9H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 85.31, 82.32, 73.69, 35.49, 32.60, 29.56, 27.97, 2.64.

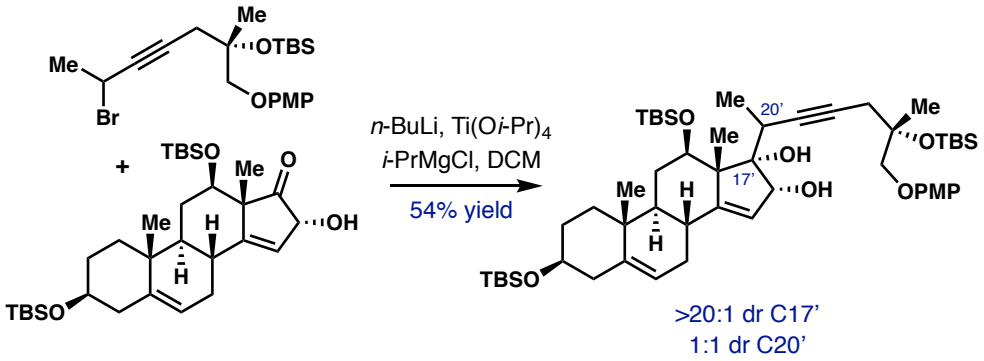
FTIR (NaCl, thin film, cm^{-1}): 2976, 2930, 2243, 1442, 1382, 1366, 1250, 1216, 1182, 1048 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_8H_{12}Br$: 187.0122 $[M–OTMS]^+$; found: 187.0142.

Synthesis of western azido ketone 81



Preparation of homopropargyl alcohol 72:



A 500 mL round-bottom flask with a stir bar was charged with α-hydroxy ketone **59** (3.45 g, 6.35 mmol, 1.0 equiv), western propargyl bromide **6** (5.59 g, 12.7 mmol, 2.0 equiv) and CH_2Cl_2 (125 mL, 0.05 M). The resulting solution was cooled to –78 °C and stirred at this temperature for 30 min. Thereafter, *n*-BuLi (2.5 M in hexane, 2.67 mL, 6.67 mmol, 1.05 equiv) was added dropwise at –78 °C. After stirring for 15 min, Ti(O*i*-Pr)$_4$ (3.76 mL, 12.7 mmol, 2.0 equiv) was added dropwise at –78 °C. After stirring for 15 min, *i*-PrMgCl (2.0 M in Et_2O, 12.7 mL, 25.4 mmol, 4.0 equiv) was added dropwise at –78 °C. The reaction was stirred at –78 °C for 30 min, then warmed to room temperature. After stirring at room temperature for 12 h, the reaction was diluted with Et_2O (100 mL), quenched with st. NH_4Cl aq. (100 mL) and H_2O (100 mL). After vigorously stirring for 3 h, the resulting solid was removed by filtration, washed with Et_2O (2 × 100 mL). The layers were separated. The aqueous layer was further extracted with EtOAc (3 × 100 mL), combined organic layers were washed with brine (2 × 100 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (10 to 15 to 20% EtOAc/hexanes) to afford homopropargyl alcohol **72** (3.12 g, 3.43 mmol, 1:1 mixture of diastereomers at C20′, 54% yield) as a colorless foam.

TLC (15% EtOAc/hexanes): R_f = 0.26 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 6.82 (m, 8H), 5.38 – 5.32 (m, 2H), 5.26 (m, 1H), 5.13 (t, *J* = 1.8 Hz, 1H), 4.87 (s, 1H), 4.71 (s, 1H), 4.23 (dd, *J* = 11.5, 4.5 Hz, 1H), 4.17 (dd, *J* = 11.5, 4.4 Hz, 1H), 3.84 – 3.79 (m, 3H), 3.76 (t, *J* = 1.5 Hz, 8H), 3.65 (d, *J* = 3.3 Hz, 1H), 3.50 – 3.42 (m, 2H), 3.26 (s, 1H), 3.12 (qt, *J* = 7.1, 2.3 Hz, 1H), 2.92 – 2.83 (m, 1H), 2.56 (ddd, *J* = 20.0, 16.3, 2.2 Hz, 4H), 2.44 (dt, *J* = 16.4, 2.5 Hz, 2H), 2.31 – 2.19 (m, 6H), 2.12 – 2.03 (m, 3H), 2.00 (m, 1H), 1.97 – 1.95 (m, 1H), 1.94 – 1.87 (m, 1H), 1.80 – 1.70 (m, 6H), 1.67 (m, 2H), 1.61 – 1.47 (m, 8H), 1.36 (d, *J* = 4.2 Hz, 6H), 1.31 (s, 2H), 1.30 – 1.28 (m, 4H), 1.27 – 1.25 (m, 4H), 1.12 (s, 1H), 1.09 (d, *J* = 2.1 Hz, 6H), 1.06 (m, 2H), 1.05 – 1.03 (m, 2H), 1.02 (s, 1H), 1.00 (d, *J* = 3.8 Hz, 6H), 0.97 (s, 1H), 0.93 (s, 3H), 0.92 (d, *J* = 2.5 Hz, 3H), 0.89 (m, 36H), 0.86 (d, *J* = 3.3 Hz, 18H), 0.11 (s, 3H), 0.10 (s, 3H), 0.09 (s, 3H), 0.09 (s, 3H), 0.08 (s, 3H), 0.08 (s, 3H), 0.07 (s, 4H), 0.06 (s, 12H).

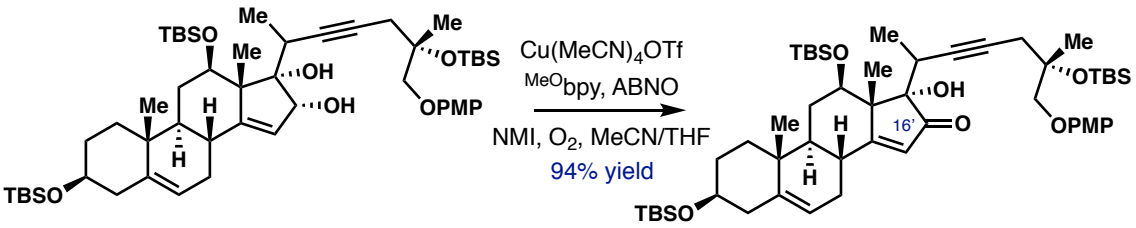
^{13}C NMR (101 MHz, $CDCl_3$): δ 154.23, 153.97, 153.91, 153.63, 153.21, 153.16, 140.70, 140.68, 122.95, 121.95, 120.82, 115.51, 115.48, 114.74, 114.72, 86.26, 84.22, 82.46, 82.35, 80.77, 79.42, 78.78, 76.75, 74.79, 74.75, 74.57, 74.38, 73.35, 72.56, 72.43, 57.94, 56.87, 55.86, 55.82, 47.83, 47.48, 42.73, 37.10, 37.01, 34.28, 33.81, 32.91, 32.08, 31.03, 30.63, 30.24, 30.18, 30.07, 29.85, 29.21, 28.97, 26.24, 26.17, 26.08, 25.92, 25.19, 25.05, 22.49, 19.30, 19.28, 18.96, 18.41, 18.33, 18.32, 18.25, 17.62, 14.80, 14.42, 14.22, -2.20, -2.21, -2.26, -3.34, -3.52, -3.57, -3.76, -4.43.

FTIR (NaCl, thin film, cm^{-1}): 3500, 2948, 2928, 2854, 1723, 1715, 1506, 1462, 1372, 1230, 1090 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{52}H_{87}O_7Si_3$: 907.5760 $[M+H-H_2]^+$; found: 907.5748.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: –68° (*c* = 0.200, $CHCl_3$).

Preparation of α-hydroxy enone 73:



A 500 mL round-bottom flask with a stir bar was charged with homopropargyl alcohol **72** (3.00 g, 3.30 mmol, 1:1 mixture of diastereomers at C20′), $Cu(MeCN)_4OTf$ (123 mg, 0.33 mmol, 10 mol %), ^{MeO}bpy (mg, 0.33 mmol, 10 mol %), ABNO (71.4 mg, 0.33 mmol, 10 mol %), NMI (52.6 μL, 0.66 mmol, 20 mol %) and MeCN/THF (3:1, 330 mL, 0.01 M). The resulting solution was sparged with O_2 for 5 min, then vigurously stirred (1200 rpm) at room temperature under O_2 atmosphere for 12 h. The reaction was passed through a short pad of SiO_2 using 50% EtOAc/hexanes as the eluent and concentrated in vacuo. The crude product was purified by flash column chromatography (15 to 20 to 25% EtOAc/hexanes) to afford α-hydroxy enone **73** (2.81 g, 3.10 mmol, 1:1 mixture of diastereomers at C20′, 94% yield) as a colorless foam.

TLC (15% EtOAc/hexanes): R_f = 0.31 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 6.83 – 6.80 (m, 8H), 5.83 (dd, *J* = 22.2, 1.5 Hz, 2H), 5.34 (m, 2H), 3.92 – 3.86 (m, 2H), 3.76 (s, 6H), 3.74 (d, *J* = 3.8 Hz, 2H), 3.72 (s, 1H), 3.71 (s, 1H), 3.49 (dd, *J* = 10.5, 5.2 Hz, 2H), 3.43 (d, *J* = 15.9 Hz, 2H), 3.01 – 2.91 (m, 1H), 2.74 (m, 1H), 2.51 (ddd, *J* = 16.2, 12.9, 2.2 Hz, 3H), 2.41 – 2.30 (m, 4H), 2.30 – 2.18 (m, 6H), 2.04 – 1.93 (m, 4H), 1.80 – 1.70 (m, 8H), 1.69 – 1.64 (m, 2H), 1.64 – 1.59 (m, 2H), 1.58 –

1.52 (m, 2H), 1.42 (s, 3H), 1.39 (s, 3H), 1.35 (m, 4H), 1.31 (s, 3H), 1.29 (s, 1H), 1.28 (s, 3H), 1.27 (s, 1H), 1.11 (d, J = 10.8 Hz, 2H), 1.08 (d, J = 2.0 Hz, 1H), 1.06 (s, 3H), 1.05 (s, 4H), 1.03 (d, J = 5.9 Hz, 2H), 1.00 (d, J = 3.0 Hz, 1H), 0.98 (d, J = 4.1 Hz, 1H), 0.94 (d, J = 2.1 Hz, 1H), 0.93 (d, J = 1.4 Hz, 2H), 0.92 (s, 1H), 0.91 (s, 2H), 0.90 (s, 6H), 0.89 (s, 15H), 0.89 (s, 3H), 0.88 (s, 8H), 0.86 (m, 4H), 0.83 (s, 6H), 0.83 (s, 8H), 0.08 (s, 6H), 0.07 (s, 3H), 0.06 (s, 15H), 0.06 (s, 6H).

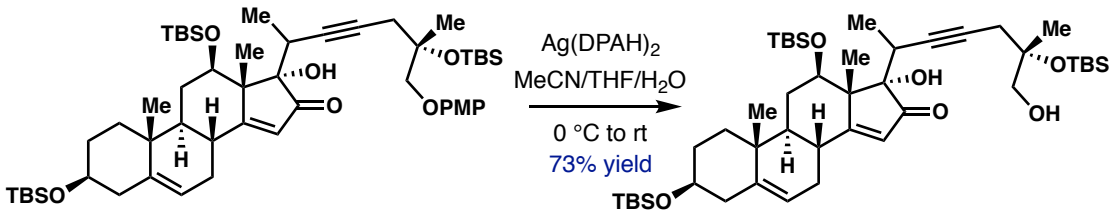
^{13}C NMR (101 MHz, $CDCl_3$): δ 206.72, 186.78, 185.46, 153.87, 153.10, 140.80, 140.72, 125.28, 124.93, 120.12, 120.07, 115.47, 115.33, 114.77, 114.74, 114.72, 84.18, 83.05, 82.55, 82.47, 81.50, 81.20, 76.52, 75.58, 74.75, 74.69, 74.12, 74.04, 72.31, 55.99, 55.80, 55.02, 48.42, 48.15, 47.79, 42.66, 37.73, 37.21, 33.38, 32.69, 32.17, 32.02, 31.98, 30.95, 30.88, 30.27, 30.03, 28.43, 28.13, 26.06, 26.03, 25.93, 25.89, 24.98, 19.38, 19.32, 18.40, 18.32, 18.28, 18.24, 18.19, 16.28, 15.78, 13.21, 12.09, -2.31, -2.92, -3.19, -4.16, -4.45, -4.47.

FTIR (NaCl, thin film, cm^{-1}): 3467, 2953, 2931, 2856, 1704, 1608, 1507, 1463, 1372, 1231, 1092 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{52}H_{85}O_7Si_3$: 905.5603 $[M+H-H_2]^+$; found: 905.5611.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\boldsymbol{21}}$: +57° ($c$ = 0.133, $CHCl_3$).

Preparation of primary alcohol 74:



A 500 mL round-bottom flask with a stir bar was charged with α-hydroxy enone **73** (3.81 g, 4.21 mmol, 1:1 mixture of diastereomers at C20′) and MeCN/THF/H_2O (2:2:1, 135 mL, 0.03 M). The resulting solution was cooled to 0 °C and stirred at this temperature for 15 min. Thereafter, $Ag(DPAH)_2 \cdot H_2O$ (7.38 g, 16.8 mmol, 4.0 equiv) was added in one portion. After stirring for 30 min at 0 °C, the reaction was warmed to room temperature stirred for 1 h, then diluted with EtOAc (100 mL), quenched with st. $Na_2S_2O_3$ aq. (100 mL) and H_2O (100 mL). The layers were separated. The aqueous layer was further extracted with EtOAc (3 × 100 mL), combined organic layers were washed with brine (2 × 100 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (10 to 15 to 20% EtOAc/hexanes) to afford primary alcohol **74** (2.46 g, 3.07 mmol, 1:1 mixture of diastereomers at C20′, 73% yield) as a colorless foam.

Preparation of $Ag(DPAH)_2 \cdot H_2O$: A 2 L round-bottom flask with a stir bar was charged with $AgNO_3$ (3.0 g, 17.4 mmol), 2,6-pyridinedicarboxylic acid (4.28 g, 34.8 mmol, 2.0 equiv) and H_2O (870 mL, 0.02 M). $K_2S_2O_8$ (23.5 g, 87.1 mmol, 5.0 equiv) was added in one portion and the resulting solution was stirred at room temperature for 24 h. The resulting solid was collected by filtration, washed with H_2O (2 × 50 mL), dried under high vacuum to afford $Ag(DPAH)_2 \cdot H_2O$ (6.79 g, 15.5 mmol, 89% yield) as a black solid.[31]

TLC (20% EtOAc/hexanes): R_f = 0.46 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.94 (dd, *J* = 9.4, 1.5 Hz, 2H), 5.40 (dd, *J* = 5.1, 2.3 Hz, 2H), 3.94 (dd, *J* = 11.0, 5.0 Hz, 2H), 3.49 (m, 4H), 3.46 – 3.39 (m, 2H), 3.33 (dd, *J* = 13.2, 10.6 Hz, 2H), 2.96 (qt, *J* = 7.1, 2.4 Hz, 1H), 2.75 (qt, *J* = 7.1, 2.3 Hz, 1H), 2.60 – 2.49 (m,

2H), 2.44 (ddd, *J* = 18.4, 16.2, 2.2 Hz, 2H), 2.30 (dd, *J* = 13.1, 2.6 Hz, 2H), 2.26 (s, 1H), 2.25 – 2.22 (m, 2H), 2.21 – 2.16 (m, 2H), 2.15 (d, *J* = 2.8 Hz, 1H), 2.13 – 2.07 (m, 1H), 1.84 – 1.79 (m, 2H), 1.77 (m, 3H), 1.74 (s, 3H), 1.70 (d, *J* = 12.7 Hz, 2H), 1.61 – 1.54 (m, 2H), 1.47 (s, 2H), 1.45 (s, 3H), 1.37 (d, *J* = 7.0 Hz, 3H), 1.29 (d, *J* = 6.2 Hz, 4H), 1.25 (d, *J* = 5.1 Hz, 5H), 1.21 – 1.20 (m, 3H), 1.16 (dd, *J* = 11.8, 6.7 Hz, 2H), 1.12 (s, 3H), 1.11 (s, 3H), 1.10 – 1.06 (m, 2H), 1.06 – 0.98 (m, 2H), 0.93 (d, *J* = 4.7 Hz, 1H), 0.91 (s, 6H), 0.89 (d, *J* = 0.9 Hz, 24H), 0.86 (m, 3H), 0.85 (s, 6H), 0.85 (s, 7H), 0.10 (m, 6H), 0.09 (m, 9H), 0.07 (s, 3H), 0.06 (s, 12H).

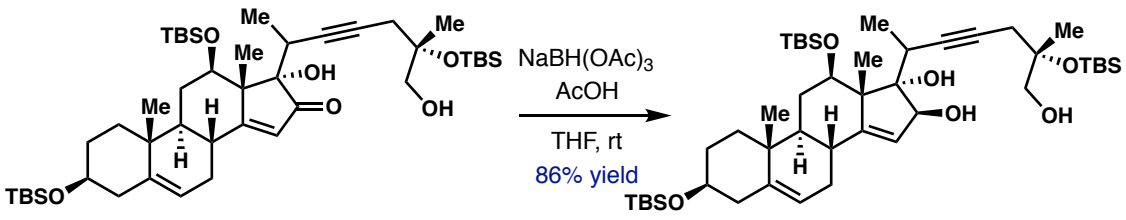
^{13}C NMR (101 MHz, $CDCl_3$): δ 206.69, 206.65, 140.94, 140.89, 125.41, 125.38, 124.98, 124.97, 120.08, 120.03, 84.41, 82.51, 81.04, 76.53, 76.06, 76.02, 72.31, 72.30, 69.73, 69.59, 55.94, 55.08, 48.47, 48.19, 42.66, 37.80, 37.78, 37.22, 37.20, 33.42, 33.36, 33.28, 33.11, 33.01, 32.69, 32.32, 32.13, 32.03, 30.32, 30.11, 29.96, 29.86, 28.52, 28.24, 26.08, 26.06, 26.03, 25.88, 23.97, 23.93, 19.49, 18.40, 18.26, 18.21, 18.19, 15.79, 13.30, 12.18, -2.09, -2.12, -2.91, -3.22, -4.15, -4.42, -4.47.

FTIR (NaCl, thin film, cm^{-1}): 3470, 2951, 2930, 2856, 1704, 1608, 1472. 1462, 1254, 1092 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{45}H_{81}O_6Si_3$: 801.5341 $[M+H]^+$; found: 801.5359.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\boldsymbol{21}}$: +29° (*c* = 0.208, $CHCl_3$).

Preparation of allylic alcohol 49:



A 2 L round-bottom flask with a stir bar was charged with $NaBH_4$ (8.70 g, 230 mmol, 100 equiv) and THF (460 mL, 0.005 M). The resulting mixture was cooled to 0 °C and stirred at this temperature for 15 min. Thereafter, AcOH (52.5 mL, 0.92 mol, 400 equiv) was added dropwise at 0 °C, then warmed to room temperature. This was accompanied by a slight exotherm and formation of gas. The remaining AcOH (52.5 mL, 0.92 mol, 400 equiv) was then added. After stirring for 30 min at room temperature, primary alcohol **74** (1.84 g, 2.30 mmol, 1.0 equiv) in THF (20 mL, further rinsed with 2 × 5.0 mL) was added. After stirring for 12 h at room temperature, the reaction was quenched with st. $NaHCO_3$ aq. (200 mL). The layers were separated. The aqueous layer was further extracted with EtOAc (3 × 200 mL), combined organic layers were washed with brine (2 × 100 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (20 to 25 to 30% EtOAc/hexanes) to afford allylic alcohol **49** (1.59 g, 1.98 mmol, 1:1 mixture of diastereomers at C20′, 86% yield) as a colorless foam.

TLC (20% EtOAc/hexanes): R_f = 0.25 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.45 (t, *J* = 2.4 Hz, 1H), 5.42 (t, *J* = 2.2 Hz, 1H), 5.37 (m, 2H), 4.57 (m, 1H), 4.34 (d, *J* = 2.8 Hz, 1H), 4.05 (ddd, *J* = 11.6, 5.9, 4.2 Hz, 2H), 3.68 – 3.57 (m, 1H), 3.55 (d, *J* = 10.8 Hz, 1H), 3.51 (d, *J* = 10.7 Hz, 1H), 3.49 – 3.45 (m, 2H), 3.45 – 3.39 (m, 2H), 3.14 (qt, *J* = 7.0, 2.4 Hz, 1H), 3.04 (qt, *J* = 7.1, 2.0 Hz, 1H), 2.53 (dd, *J* = 11.1, 2.3 Hz, 1H), 2.49 (dd, *J* = 11.0, 2.3 Hz, 1H), 2.39 (dd, *J* = 16.3, 2.2 Hz, 1H), 2.33 (dd, *J* = 16.3, 2.3 Hz, 1H), 2.29 – 2.24 (m, 2H), 2.24 – 2.19 (m, 4H), 2.18 – 2.14 (m, 2H), 2.12 – 2.08 (m, 1H), 2.06 – 1.92 (m, 4H), 1.76 (m, 4H), 1.72 (d, *J* = 4.5 Hz, 2H), 1.70 (t, *J* = 4.5 Hz, 1H), 1.62 (d, *J* = 12.4 Hz, 1H), 1.58 (m, 1H), 1.57 (m, 1H), 1.55 – 1.49 (m, 2H),

1.40 (d, J = 7.1 Hz, 3H), 1.35 (d, J = 7.1 Hz, 3H), 1.31 (d, J = 3.0 Hz, 6H), 1.29 (s, 3H), 1.25 (s, 6H), 1.10 (s, 3H), 1.06 (d, J = 3.9 Hz, 1H), 1.05 (s, 3H), 1.03 (s, 1H), 1.01 (m, 4H), 0.92 (s, 1H), 0.92 (s, 6H), 0.91 (s, 9H), 0.90 (m, 2H), 0.89 (s, 15H), 0.88 (s, 2H), 0.87 (s, 9H), 0.87 (s, 1H), 0.86 (s, 6H), 0.85 (s, 1H), 0.14 (s, 3H), 0.13 (s, 3H), 0.12 (d, J = 1.4 Hz, 5H), 0.11 (d, J = 1.7 Hz, 5H), 0.10 (s, 3H), 0.08 (s, 3H), 0.07 (s, 3H), 0.06 (s, 9H).

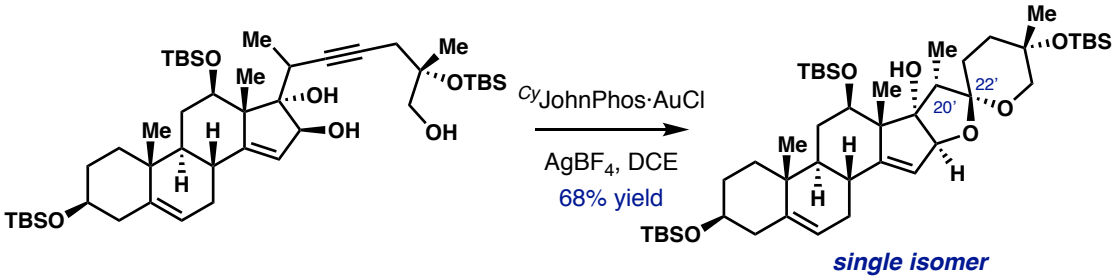
^{13}C NMR (101 MHz, $CDCl_3$): δ 158.37, 154.51, 140.71, 140.60, 122.74, 120.91, 120.78, 86.16, 86.04, 85.92, 85.58, 85.41, 85.14, 79.80, 79.19, 76.04, 75.85, 74.92, 72.53, 72.14, 70.03, 57.47, 57.36, 48.95, 47.59, 42.66, 37.28, 36.94, 32.73, 32.10, 31.07, 30.83, 30.53, 30.31, 30.24, 30.06, 29.85, 29.54, 28.81, 26.45, 26.22, 26.08, 25.91, 24.32, 24.17, 22.85, 22.49, 19.39, 19.32, 18.60, 18.41, 18.34, 18.23, 17.10, 15.89, 14.93, 1.17, -2.07, -2.43, -3.13, -3.70, -4.44.

FTIR (NaCl, thin film, cm^{-1}): 3437, 2957, 2927, 2854, 1471, 1463, 1372, 1256, 1090 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{45}H_{81}O_6Si_3$: 801.5341 $[M+H-H_2]^+$; found: 801.5333.

$[\alpha]_D^{21}$: –9.3° (c = 0.125, $CHCl_3$).

Preparation of spiroketal 75:



In an N_2-filled glovebox, a 100 mL round-bottom flask with a stir bar was charged with ^{Cy}JohnPhos·AuCl (36.7 mg, 0.06 mmol, 10 mol %), $AgBF_4$ (6.13 mg, 0.03 mmol, 5.0 mol

%) and $(CH_2Cl)_2$ (21 mL, 0.03 M). The flask was removed from the glovebox and stirred at room temperature for 10 min. Allylic alcohol **49** (503 mg, 0.63 mmol, 1.0 equiv) in $(CH_2Cl)_2$ (2 mL, rinsed with 2 × 1.0 mL) was added. After stirring for 20 h, the reaction was passed through a short pad of SiO_2 using 20% EtOAc/hexanes as the eluent and concentrated in vacuo. The crude product was purified by flash column chromatography (0 to 5 to 10% EtOAc/hexanes) to afford spiroketal **75** (344 mg, 0.43 mmol, 68% yield) as a colorless foam.

TLC (8% EtOAc/hexanes): R_f = 0.44 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.41 (t, *J* = 1.9 Hz, 1H), 5.37 (m, 1H), 4.64 (d, *J* = 2.3 Hz, 1H), 4.08 (dd, *J* = 11.3, 4.8 Hz, 1H), 3.66 (d, *J* = 11.6 Hz, 1H), 3.52 – 3.42 (m, 2H), 3.22 (s, 1H), 2.31 – 2.19 (m, 2H), 2.15 (dd, *J* = 13.5, 4.7 Hz, 1H), 2.10 – 2.04 (m, 2H), 1.99 (q, *J* = 6.9 Hz, 1H), 1.81 (ddd, *J* = 11.8, 9.6, 4.0 Hz, 2H), 1.77 – 1.71 (m, 2H), 1.64 – 1.60 (m, 1H), 1.59 (d, *J* = 3.0 Hz, 1H), 1.54 (m, 2H), 1.33 (m, 1H), 1.25 (s, 2H), 1.12 (s, 3H), 1.08 (s, 3H), 1.06 (s, 1H), 1.05 (s, 3H), 1.03 (d, *J* = 6.9 Hz, 3H), 0.91 (s, 6H), 0.89 (s, 9H), 0.88 (s, 6H), 0.10 (s, 9H), 0.08 (s, 3H), 0.06 (s, 6H).

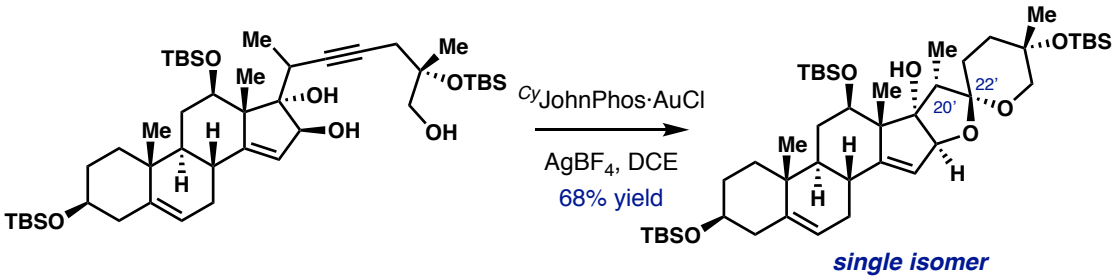
^{13}C NMR (101 MHz, $CDCl_3$): δ 155.72, 140.58, 120.91, 119.70, 107.28, 92.95, 91.64, 74.95, 72.53, 69.85, 69.38, 56.63, 48.88, 48.37, 42.73, 37.36, 37.31, 34.10, 32.11, 30.98, 30.42, 29.86, 28.65, 28.01, 26.57, 26.18, 26.07, 26.00, 19.48, 18.41, 18.34, 18.27, 14.88, 8.06, -2.04, -2.19, -2.77, -4.05, -4.45.

FTIR (NaCl, thin film, cm^{-1}): 3532, 2953, 2931, 2857, 1471, 1462, 1366, 1252, 1089, 1059 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{45}H_{81}O_6Si_3$: 801.5341 $[M+H-H_2]^+$; found: 801.5351.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: –104° (*c* = 0.133, $CHCl_3$).

Table 2.1. Optimization studies for western spiroketalization

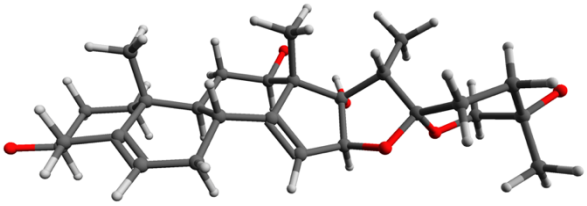


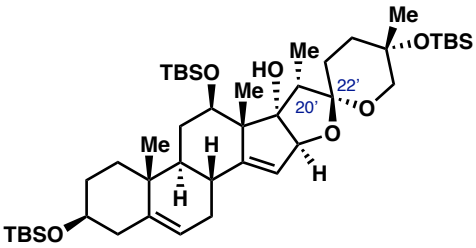
entry	cat.	X	MY	Z	solvent	20′S/22′S	20′R/22′S	20′S22′R	20′R/22′R
1	$PdCl_2(PhCN)_2$	10	–	–	MeCN/THF	33%	–	–	–
2	$PtCl_2$	10	–	–	DCE	37%	–	17%	–
3	$[Pt(C_2H_4)Cl_2]_2$	10	–	–	DCE	44%[a]	–	12%[a]	–
4	$HgCl_2$	10	–	–	DCE	–	6%	–	38%
5	$Ph_3PAuNTf_2$	10	–	–	DCE	41%	20%	5%	3%
6	AuCl	20	–	–	DCE	62%	–	–	–
7	AuCl	10	–	–	DCE	45%	–	–	–
8	AuCl	10	–	–	PhMe	27%	–	–	–
9	AuCl	10	–	–	THF	–	–	–	–
10	AuCl	10	–	–	Et_2O	4%	–	–	–
11	AuCl	5	–	–	DCE	52%	–	–	–
12	JohnPhos·Au(MeCN)SbF_6	10	–	–	DCE	46%	19%	–	–
13	JohnPhos·AuCl	10	–	–	DCE	10%	–	7%	–
15	JohnPhos·AuCl	10	–	–	MeOH	20%	2%	9%	–
16	JohnPhos·AuCl	10	AgOTf	10	DCE	44%[a]	–	22%[a]	–
17	–	–	AgOTf	10	DCE	25%[a]	–	14%[a]	–
18	JohnPhos·AuCl	10	$AgSbF_6$	10	DCE	44%[a]	–	11%[a]	–
19	JohnPhos·AuCl	10	$AgPF_6$	10	DCE	52%[a]	–	4%[a]	–
20	JohnPhos·AuCl	10	$AgBF_4$	10	DCE	58%[a]	–	–	–
21	JohnPhos·AuCl	10	$AgBF_4$	10	DCE	58%[a]	–	–	–
22	JohnPhos·AuCl	10	$AgNTf_2$	10	DCE	32%[a]	10%[a]	–	–
23	JohnPhos·AuCl	10	$AgNTf_2$	10	THF	40%	10%	–	–
24	JohnPhos·AuCl	20	$AgBF_4$	10	DCE	76%	5%	2%	–
25	JohnPhos·AuCl	10	$AgBF_4$	20	DCE	77%[b]	–	–	–
26	CyJohnPhos·AuCl	20	$AgBF_4$	10	DCE	63%	3%	–	–
27	IPr·AuCl	20	$AgBF_4$	10	DCE	56%	4%	–	–
28	PMe_3·AuCl	20	$AgBF_4$	10	DCE	44%	6%	–	–
29	PPh_3·AuCl	20	$AgBF_4$	10	DCE	54%	5%	–	–
30	CyJohnPhos·AuCl	10	$AgBF_4$	5	DCE	72%[c]	–	–	–

[a]C12′–OTBS deprotected. [b]Mix of 64% desired with 13% C12′–OH. [c]Isolated yield.

Minimized Structures and energies:

75





Single-point energy = -3082.95785963937 (Hartree)
N_atoms = 136
C 0.200633 -3.678912 -0.916686
C 0.347653 -2.467007 -1.7092325
C 0.744005 -1.267871 -0.9148065
O 1.5687955 -1.6099555 0.111838
C 1.5381655 -0.4872035 -1.886369
C 0.8853585 0.7681405 -2.004716
C -0.134173 0.8741265 -1.245642
C -0.9347195 2.041061 -1.0049915
C -0.550866 3.188451 -1.822703
C -1.1196715 4.4141595 -1.3622155
C -1.67093 4.5922305 -0.226109
C -2.2308525 5.874307 0.1204845
C -1.7521545 6.378206 1.407227
O -2.416175 7.5396265 1.654909
Si -2.193256 8.7149525 2.7194515
C -3.369792 9.9550165 2.1758225
C -2.631524 8.1980255 4.3891065
C -0.4794105 9.37401 2.6830895
C -0.057511 9.6060135 1.294181
C -0.4485845 10.654676 3.4086925
C 0.469698 8.452965 3.326004
C -1.9678685 5.3743205 2.441396
C -1.341937 4.1015075 2.083451
C -1.7745335 3.5379695 0.7786155
C -3.1834765 3.1043705 0.870883
C -0.8667405 2.4006145 0.4323645
C -1.0451705 1.224704 1.2984455
C -0.1742205 0.102592 0.953563
O -0.400336 -0.9649555 1.798664
Si 0.177218 -1.186496 3.293762
C 1.1767455 -2.6812695 3.24448
C 1.2268975 0.1815935 3.8051045
C -1.252707 -1.4177875 4.411917
C -2.0527165 -0.1888605 4.4947045
C -2.104404 -2.503452 3.9038915
C -0.780117 -1.7663525 5.7611415

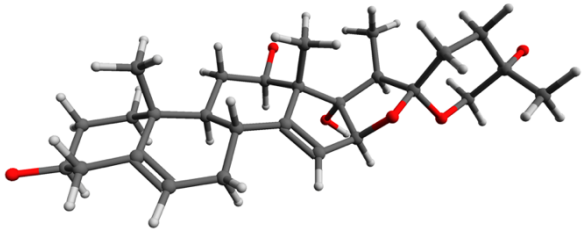
C	-0.3693745	-0.326353	-0.4546535
C	-1.7270495	-0.856267	-0.643668
O	1.516954	-1.172041	-3.082208
C	1.3343405	-2.4801045	-2.8226925
C	0.8965635	-3.0758575	-4.089428
C	1.2034795	-4.495635	-4.2434435
C	2.59362	-4.8038815	-3.899427
C	3.535642	-4.226312	-4.8495825
O	2.6738015	-6.169178	-3.84323
Si	3.828751	-7.2610465	-4.020438
C	4.356791	-7.3882135	-5.7379445
C	5.267732	-6.892519	-2.999212
C	3.021351	-8.80666	-3.4727875
C	3.963562	-9.9316975	-3.5671195
C	2.5736185	-8.6699195	-2.079305
C	1.8533555	-9.0907495	-4.319063
C	2.8174555	-4.2870855	-2.544209
O	2.5707185	-2.9522485	-2.4396645
H	-0.0354685	-4.4973335	-1.513305
H	-0.5695395	-3.571889	-0.223601
H	1.0722335	-3.8867675	-0.3939465
H	-0.559616	-2.273429	-2.1785325
H	1.0068005	-1.8650735	0.8023455
H	2.5260005	-0.4096075	-1.558762
H	1.241538	1.502042	-2.635373
H	-1.930436	1.8075585	-1.224319
H	-0.817191	3.0215385	-2.818541
H	0.4915975	3.281152	-1.8277885
H	-1.0611125	5.2165375	-2.0159165
H	-2.023655	6.571328	-0.6241065
H	-3.272348	5.8190685	0.1924215
H	-0.7249965	6.5667295	1.3288665
H	-3.112518	10.351381	1.248626
H	-3.4442735	10.743417	2.851786
H	-4.3149545	9.5286735	2.0816825
H	-2.0016375	7.4622235	4.769072
H	-2.597849	9.00649	5.0460905
H	-3.60241	7.821631	4.407674
H	-0.688731	10.2731345	0.802365
H	0.904936	10.012401	1.2670285
H	-0.0419655	8.7245005	0.7411235
H	-0.745387	10.552656	4.4027465
H	0.521309	11.0437835	3.417937
H	-1.063108	11.3685295	2.965006
H	0.222798	8.2671035	4.320774
H	1.432421	8.8599195	3.319292

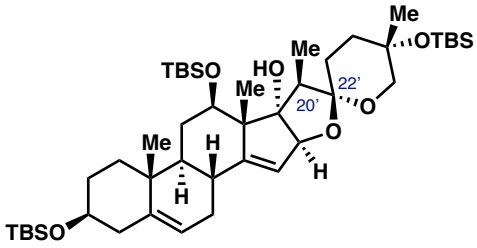
H 0.523974 7.5348435 2.8366325
H -1.576791 5.702751 3.349663
H -2.994751 5.261024 2.583684
H -1.525696 3.409129 2.8399645
H -0.3068075 4.2388395 2.052043
H -3.486621 2.6110635 0.006342
H -3.8195895 3.9164865 0.993063
H -3.3384915 2.470823 1.681554
H 0.108519 2.7465865 0.5895225
H -2.031601 0.897339 1.276368
H -0.8504935 1.4908785 2.2850325
H 0.8192595 0.4012695 1.0488865
H 2.004668 -2.559225 2.624193
H 1.533895 -2.924774 4.1921985
H 0.637573 -3.499436 2.8930565
H 0.714481 1.080494 3.9113595
H 1.667517 -0.0369235 4.7235845
H 1.999328 0.3327775 3.123251
H -2.4198215 0.0961915 3.56365
H -1.5063225 0.6096635 4.8824725
H -2.8752985 -0.330423 5.1234745
H -1.5946035 -3.410834 3.8564485
H -2.4663005 -2.292582 2.9509405
H -2.927649 -2.644344 4.5317575
H -0.246348 -2.660652 5.7736295
H -1.5923365 -1.884892 6.40739
H -0.1712025 -1.025016 6.1684595
H -1.915217 -1.0654455 -1.6432615
H -2.4498455 -0.174448 -0.3416375
H -1.8678045 -1.721456 -0.0866115
H -0.125357 -2.9141755 -4.209206
H 1.368668 -2.5277345 -4.8379175
H 1.0056965 -4.799895 -5.2206445
H 0.5990975 -5.075383 -3.6231865
H 3.4966585 -3.1888485 -4.808089
H 3.3232815 -4.531692 -5.821197
H 4.5119095 -4.505318 -4.6212635
H 5.0151 -8.185368 -5.868691
H 3.5492735 -7.5373515 -6.3775105
H 4.8490335 -6.5256285 -6.0484215
H 5.0117955 -6.7852515 -1.9960325
H 5.736298 -6.012381 -3.299708
H 5.9756795 -7.6545725 -3.064279
H 4.299072 -10.0831905 -4.541992
H 4.8005445 -9.798429 -2.960506
H 3.5025625 -10.817981 -3.2607675

H 3.371185 -8.5100575 -1.427831
H 2.08862 -9.540556 -1.76464
H 1.9085715 -7.877375 -1.970667
H 1.3763895 -9.9628125 -3.996303
H 1.1579775 -8.317215 -4.2863815
H 2.1195475 -9.236512 -5.316067
H 2.2319535 -4.8481715 -1.8880245
H 3.8131925 -4.3983815 -2.2613525

76

Single-point energy = -3082.954550249199 (Hartree)
N_atoms = 136
C 1.466518 -0.078627 2.535555
C 2.19979 0.4857655 1.405724
C 1.4772195 0.747566 0.120291
O 1.0333855 2.04133 0.071605
C 2.5366685 0.52288 -0.8813455
C 1.87228 0.078004 -2.0502105
C 0.6605685 -0.2386645 -1.810049
C -0.3076485 -0.7350785 -2.7491685
C 0.1198105 -0.572783 -4.137453
C -0.9352625 -0.8010155 -5.0716795
C -2.1752095 -0.8755835 -4.782652
C -3.1507085 -1.1325815 -5.8131405
C -4.230389 -0.1455775 -5.8201545
O -5.1781105 -0.5116205 -6.7324445
Si -5.592588 0.327334 -8.034784
C -6.0891305 1.9976615 -7.577062
C -4.2013805 0.4443155 -9.17241
C -6.9787615 -0.623924 -8.739758
C -8.0743475 -0.6941745 -7.7616195
C -6.5346625 -1.9900345 -9.053416
C -7.466618 0.020027 -9.9683845
C -4.8314535 -0.0606015 -4.498142
C -3.8303275 0.2426255 -3.47692
C -2.6845885 -0.7034435 -3.425315
C -3.142154 -2.0116545 -2.9157065
C -1.6312605 -0.112882 -2.542791
C -1.9893465 -0.106712 -1.120438
C -0.991572 0.5602855 -0.284477
O -1.4266855 0.5698545 1.012926
Si -2.170915 1.781596 1.7639375
C -0.973065 2.7680325 2.67507
C -2.9941975 2.8743405 0.5907905



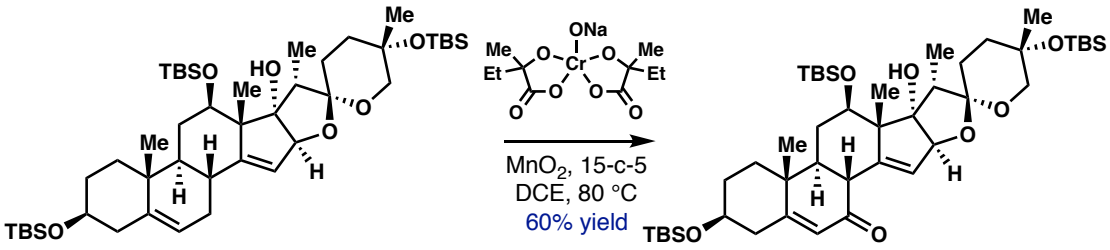


C	-3.367006	0.976803	2.8887475
C	-4.3861205	0.258627	2.110545
C	-4.026739	1.9862645	3.7294495
C	-2.665452	0.0140755	3.749849
C	0.326677	-0.103918	-0.3896065
C	0.2814815	-1.455387	0.1937305
O	3.3212245	-0.469598	-0.331235
C	3.472911	-0.18686	0.97625
C	3.8981425	-1.424368	1.6220795
C	4.5598235	-1.233146	2.9146775
C	5.5931785	-0.1905565	2.877388
C	6.772911	-0.59721	2.1235295
O	5.8999815	0.197467	4.1530365
Si	6.876111	-0.3632395	5.291283
C	6.73248	-2.147931	5.497326
C	8.597207	0.032705	4.93857
C	6.3021925	0.5145445	6.78837
C	6.4673835	1.964903	6.613335
C	7.074198	0.081323	7.9625975
C	4.879162	0.2255335	7.018542
C	4.9878265	0.997386	2.262963
O	4.510152	0.734011	1.0198685
H	1.947742	0.1491135	3.429675
H	1.3654005	-1.109448	2.4887865
H	0.512841	0.328653	2.571502
H	2.487029	1.4417225	1.7058585
H	1.7762115	2.5801855	0.1766575
H	3.1251815	1.3767525	-1.024033
H	2.344166	0.024001	-2.9649205
H	-0.417811	-1.761423	-2.57171
H	0.499697	0.3934815	-4.274068
H	0.9178055	-1.213964	-4.3438245
H	-0.648425	-0.9095705	-6.062091
H	-2.6888345	-1.15926	-6.7458725
H	-3.6017495	-2.0656125	-5.676875
H	-3.825998	0.7888955	-6.0686785
H	-5.3141905	2.5028805	-7.0975435
H	-6.3453995	2.558475	-8.416982
H	-6.9074265	1.996987	-6.934066
H	-3.3953445	0.929922	-8.724661
H	-3.8704745	-0.4943935	-9.476614
H	-4.4505125	0.986862	-10.0263875
H	-7.7628285	-1.1380995	-6.8736445
H	-8.452477	0.2491805	-7.531567
H	-8.8669335	-1.260789	-8.139844
H	-5.7759805	-1.999414	-9.7678665

H -6.181386 -2.480139 -8.2063215
H -7.3276355 -2.550921 -9.4393015
H -8.2656585 -0.522144 -10.367046
H -7.809189 0.9889135 -9.794761
H -6.7303195 0.0686705 -10.704214
H -5.572933 0.671895 -4.4945335
H -5.30844 -0.9663165 -4.2990155
H -4.2931475 0.277962 -2.5442485
H -3.4535675 1.2003445 -3.6560175
H -2.3408095 -2.6595055 -2.7713965
H -3.7913715 -2.4767615 -3.580541
H -3.6493895 -1.912255 -2.012912
H -1.5311955 0.886232 -2.840107
H -2.135121 -1.0749185 -0.767614
H -2.8923145 0.3915635 -0.9832625
H -0.8749575 1.534371 -0.628433
H -0.18119 3.0037385 2.042972
H -0.581076 2.261118 3.495131
H -1.3931775 3.656205 3.021597
H -2.3074385 3.3153475 -0.056305
H -3.472369 3.647818 1.099501
H -3.7106795 2.391008 0.011399
H -3.9571555 -0.460984 1.492742
H -5.0552825 -0.230614 2.7469725
H -4.945519 0.902513 1.5117715
H -3.3480425 2.5065615 4.3248005
H -4.7139595 1.530281 4.370891
H -4.552486 2.685614 3.162802
H -2.1905045 -0.7220735 3.1882895
H -1.946143 0.4746075 4.3469845
H -3.3389625 -0.462543 4.391708
H -0.166554 -1.4483625 1.127851
H -0.2700045 -2.098183 -0.40907
H 1.23801 -1.8521055 0.263237
H 3.085748 -2.0656365 1.730352
H 4.5474145 -1.8829445 0.949816
H 4.9698845 -2.1378105 3.229648
H 3.866087 -0.952934 3.6390505
H 7.507049 0.1380305 2.1757775
H 7.171414 -1.489808 2.4816065
H 6.527108 -0.72846 1.122432
H 7.0429335 -2.661504 4.6461755
H 5.755501 -2.4400055 5.704429
H 7.3326345 -2.47287 6.2850005
H 9.2090485 -0.2289295 5.740734
H 8.729372 1.049486 4.759984

H 8.94556 -0.481895 4.1037705
H 6.1110725 2.4757195 7.4526615
H 7.467138 2.2334345 6.4932255
H 5.9427515 2.3137805 5.784898
H 8.0921585 0.2815845 7.8626695
H 6.9674485 -0.938543 8.1471255
H 6.7452725 0.5844045 8.817094
H 4.5362905 0.733402 7.8650435
H 4.296164 0.5241045 6.210456
H 4.708171 -0.7900275 7.1788475
H 4.2393145 1.357188 2.9018865
H 5.6971995 1.7499495 2.147168

Preparation of enone 77:



In an N_2-filled glovebox, a 500 mL round-bottom flask with a stir bar was charged with spiroketal **75** (1.70 g, 2.12 mmol, 1.0 equiv), "Cr (V)" (6.85 g, 21.2 mmol, 10 equiv), MnO_2 (3.69 g, 42.4 mmol, 20 equiv) and $(CH_2Cl)_2$/15-crown-5 (49:1, 140 mL, 0.015 M). The flask was removed from the glovebox and the resulting mixture was heated to 80 °C and stirred for 48 h. After cooling to room temperature, the reaction was passed through a short pad of SiO_2 using 50% hexane/EtOAc as the eluent and concentrated in vacuo. The crude product was purified by flash column chromatography (0 to 5 to 10 to 15 to 20% EtOAc/hexanes) to afford starting spiroketal **75** (357 mg, 0.45 mmol, 21% yield) as a colorless foam and enone **77** (1.04 g, 1.27 mmol, 60% yield) as a colorless foam.

Preparation of "Cr (V)": A 100 mL round-bottom flask was charged with $Na_2Cr_2O_7•H_2O$ (ca. 8 g), heated to 100 °C for 2 h under high vacuum to afford anhydrous $Na_2Cr_2O_7$ (ca. 7.5 g). A 1 L flask was charged with anhydrous $Na_2Cr_2O_7$ (6.29 g, 24.0 mmol), 2-hydroxy-2-methylbutyric acid (17.0 g, 144 mmol, 6.0 equiv) and acetone (32 mL, 0.75 M). The resulting mixture was stirred at room temperature for 6 h. The resulting brown solid was collected by filtration, washed with acetone (2 × 10 mL). The crude product was recrystalized with acetone (20 mL), collected by filtration, washed with acetone (2 × 10 mL), dried under high vacuum to afford "Cr (V)" (7.45 g, 23.0 mmol, 96% yield) as a brown solid.[22,32]

TLC (15% EtOAc/hexanes): R_f = 0.55 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.82 (d, *J* = 1.4 Hz, 1H), 5.80 (dd, *J* = 2.1, 1.5 Hz, 1H), 4.65 (dd, *J* = 2.1, 1.0 Hz, 1H), 4.07 – 3.99 (m, 1H), 3.68 (d, *J* = 11.6 Hz, 1H), 3.66 – 3.56 (m, 1H), 3.42 (dd, *J* = 11.6, 2.7 Hz, 1H), 3.07 (s, 1H), 2.82 – 2.75 (m, 1H), 2.46 – 2.43 (m, 1H), 2.42 (s, 1H), 2.12 (td, *J* = 13.2, 4.5 Hz, 1H), 1.96 (q, *J* = 6.9 Hz, 1H), 1.88 (m, 1H), 1.86 – 1.76 (m, 3H), 1.70 – 1.63 (m, 1H), 1.63 – 1.55 (m, 6H), 1.37 (ddd, *J* = 13.0, 4.3, 2.4 Hz, 1H), 1.30 (m, 1H), 1.25 (s, 2H), 1.21 (m, 1H), 1.16 (s, 3H), 1.11 (s, 3H), 1.09 (s, 3H), 1.04 (d, *J* = 7.0 Hz, 3H), 0.90 – 0.88 (m, 27H), 0.10 (d, *J* = 0.7 Hz, 6H), 0.08 (d, *J* = 0.9 Hz, 6H), 0.06 (s, 6H).

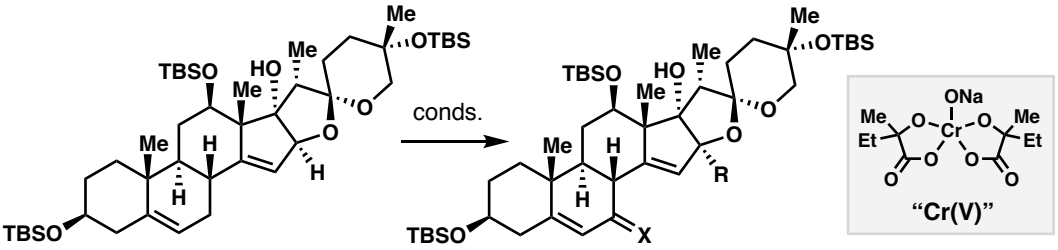
^{13}C NMR (101 MHz, $CDCl_3$): δ 196.72, 167.34, 145.79, 126.82, 123.69, 107.37, 94.10, 90.89, 73.81, 71.06, 69.90, 69.43, 56.70, 49.43, 48.11, 43.39, 42.70, 37.96, 36.14, 33.87, 31.76, 29.85, 27.50, 26.53, 26.18, 26.00, 25.97, 18.34, 18.30, 18.03, 13.76, 7.98, -2.03, -2.17, -3.33, -4.13, -4.51, -4.55.

FTIR (NaCl, thin film, cm^{-1}): 3522, 2949, 2931, 2856, 1681, 1470, 1462, 1373, 1252, 1086 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{45}H_{81}O_7Si_3$: 817.5290 $[M+H]^+$; found: 817.5270.

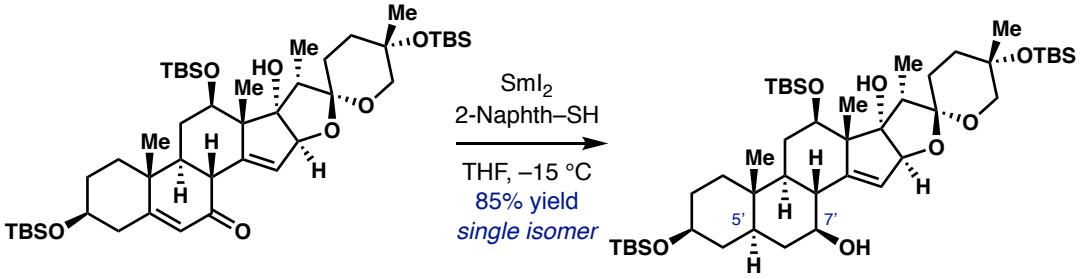
$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: +17° (*c* = 0.100, $CHCl_3$).

Table 2.2. Other attempted allylic oxidations



entry	conds.	X = O R = H	X = OH, H R = H	X = O R = OH	X = H, H R = H
1	$Mn(OAc)_3$ (10 mol %), TBHP (10 equiv) 3Å MS, EtOAc, 22 °C	11%	–	–	–
2	$Rh_2(cap)_4$ (1 mol %), TBHP (10 equiv) K_2CO_3 (0.5 equiv), DCM, 22 °C	17%	–	–	–
3	$Rh_2(cap)_4$ (1 mol %), TBHP (10 equiv) DCE, 22 °C	17%	8%	1%	11%
4	CuI (70 mol %), TBAB (10 mol %) TBHP (30 equiv), DCm, 45 °C	10%	3%	2%	15%
5	$Na_2Cr_2O_7$ (1.5 equiv), Cl_4NHPI (2.2 equiv) acetone/py (9:1), 22 °C	3%	–	–	80%
6	$Na_2Cr_2O_7$ (1.5 equiv), Cl_4NHPI (2.2 equiv) acetone/py (19:1), 22 °C	37%	–	–	50%
7	$Na_2Cr_2O_7$ (1.5 equiv), Cl_4NHPI (2.2 equiv) acetone, 22 °C	22%	2%	2%	23%
8	PCC (30 equiv), NaOAc (30 equiv) celite (100 wt % PCC), PhH, 85 °C	1%	2%	2%	–
9	PDC (1.5 equiv), NHPI (1.5 equiv) MeCN/DCM/py/H_2O (9:9:1:1), 23 °C	29%	–	15%	–
10	PDC (1.5 equiv), NHPI (5.0 equiv) acetone/py (19:1), 23 °C	43%	–	–	17%
11	NHPI (20 mol %), $Co(OAc)_2$ (10 mol %) $Mn(OAc)_2$ (10 mol %), O_2, acetone, 22 °C	3%	10%	10%	0%
12	$PhI(OAc)_2$ (3.0 equiv), TBHP (4.0 equiv) K_2CO_3 (0.5 equiv), *n*-$PrCO_2Bu$/hexane, 0 °C	10%	1%	1%	4%
13	$NaClO_2$ (1.5 equiv), NHPI (10 mol %) MeCN/THF/H_2O, 50 °C	–	–	–	78%
14	"Cr(V)" (2.0 equiv), NHPI (5.0 equiv) acetone/py (19:1), 23 °C	29%	–	15%	18%
15	"Cr(V)" (5.0 equiv), MnO_2 (50 equiv) $PhCF_3$/15-c-5 (49:1), 80 °C	37%	–	1%	35%
16	"Cr(V)" (10 equiv), MnO_2 (20 equiv) $PhCF_3$/15-c-5 (49:1), 80 °C	41%	–	2%	22%
17	"Cr(V)" (10 equiv), MnO_2 (20 equiv) EtOAc/15-c-5 (49:1), 80 °C	23%	–	1%	44%
18	"Cr(V)" (10 equiv), MnO_2 (20 equiv) dioxane/15-c-5 (49:1), 80 °C	13%	–	–	61%
19	"Cr(V)" (10 equiv), MnO_2 (20 equiv) DCE/15-c-5 (49:1), 80 °C	60%	–	2%	5%

Preparation of C7 alcohol 78:



In a N_2-filled glovebox, a 200 mL round-bottom flask with a stir bar was charged with enone **77** (147 mg, 0.18 mmol, 1.0 equiv), 2-naphthalene thiol (577 mg, 3.60 mmol, 20.0 equiv) and THF (3.6 mL, 0.05 M). Also in the glovebox, a 24 mL syringe was used to withdraw a premade SmI_2 solution (0.05 M in THF (6.1 mL 0.1M SmI_2 in THF + 6.1 mL THF), 36.0 mL, 1.8 mmol, 10.0 equiv). The flask and syringe were sealed and removed from the glovebox. The reaction flask was placed under N_2, stirred, and brought to –10 °C by lowering into a NaCl/ice water bath (~1:3 by mass). The SmI_2 solution was then added dropwise over 1 h via syringe pump, maintaining an ice bath temperature of –10 °C. Upon completion of the addition, the ice bath was removed and the reaction was allowed to come to room temperature and stirred for an additional 30 min. The reaction was quenched by addition of sat. aq. Rochelle's salt (30 mL) and 1M aq. NaOH (30 mL). The layers were separated, and the aqueous layer was extracted with EtOAc (3 × 30 mL). Combined organics were washed with brine (30 mL), dried over anhydrous Na_2SO_4, filtered, and concentrated in vacuo. The crude product was purified by flash column chromatography (10 to 15 to 20% EtOAc/hexanes) to afford C7′ alcohol **78** (126 mg, 0.15 mmol, 85% yield, single isomer C5′ and C7′) as a colorless foam.

Preparation of SmI_2 solution in THF: A 125 mL sep-funnel was charged with $(CH_2I)_2$ (ca. 4 g), Et_2O (50 mL) and sat. aq. $Na_2S_2O_3$ (10 mL). The layers were separated, and the organic layer was further washed with sat. aq. $Na_2S_2O_3$ (4 × 10 mL) and H_2O (5 × 10 mL). Combined organic layers were dried over anhydrous Na_2SO_4, filtered, and concentrated in vacuo. The resulting material was dried under high vacuum in the dark for 15 min to afford purified $(CH_2I)_2$ (ca. 3.2 g) as a colorless crystal. In a N_2-filled glovebox, a 200 mL flask was charged with freshly purified $(CH_2I)_2$ (2.81 g, 10.0 mmol, 1.0 equiv), Sm (300 mesh, 3.01 g, 20.0 mmol, 2.0 equiv) and THF (100 mL). The resulting deep-green mixture was stirred at room temperature in the dark for 12 h to afford deep-blue SmI_2 solution (0.1 M in THF), which is usable for several months.

TLC (15% EtOAc/hexanes): R_f = 0.26 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.54 (t, *J* = 1.5 Hz, 1H), 4.67 (d, *J* = 2.3 Hz, 1H), 4.00 (dd, *J* = 11.1, 5.0 Hz, 1H), 3.74 (ddd, *J* = 10.9, 9.4, 5.1 Hz, 1H), 3.67 (d, *J* = 11.6 Hz, 1H), 3.54 (h, *J* = 5.3 Hz, 1H), 3.45 (dd, *J* = 11.6, 2.7 Hz, 1H), 3.16 (s, 1H), 2.20 – 2.11 (m, 1H), 1.99 (t, *J* = 9.4 Hz, 1H), 1.91 (q, *J* = 6.9 Hz, 1H), 1.83 – 1.57 (m, 7H), 1.54 – 1.44 (m, 2H), 1.44 – 1.37 (m, 3H), 1.35 – 1.32 (m, 1H), 1.32 – 1.29 (m, 1H), 1.27 – 1.24 (m, 1H), 1.17 (dt, *J* = 11.5, 2.5 Hz, 1H), 1.12 (s, 3H), 1.06 (s, 3H), 1.03 (d, *J* = 6.9 Hz, 3H), 0.97 – 0.91 (m, 1H), 0.89 (s, 9H), 0.88 (s, 9H), 0.88 (s, 9H), 0.86 (s, 3H), 0.84 – 0.78 (m, 1H), 0.78 – 0.72 (m, 1H), 0.10 (s, 3H), 0.08 (d, *J* = 1.6 Hz, 6H), 0.07 (s, 3H), 0.05 (s, 6H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 152.81, 119.21, 107.31, 93.61, 91.29, 74.76, 71.87, 70.47, 69.90, 69.35, 56.66, 50.52, 48.36, 43.00, 42.01, 38.19, 36.85, 35.99, 35.60, 34.03, 31.85,

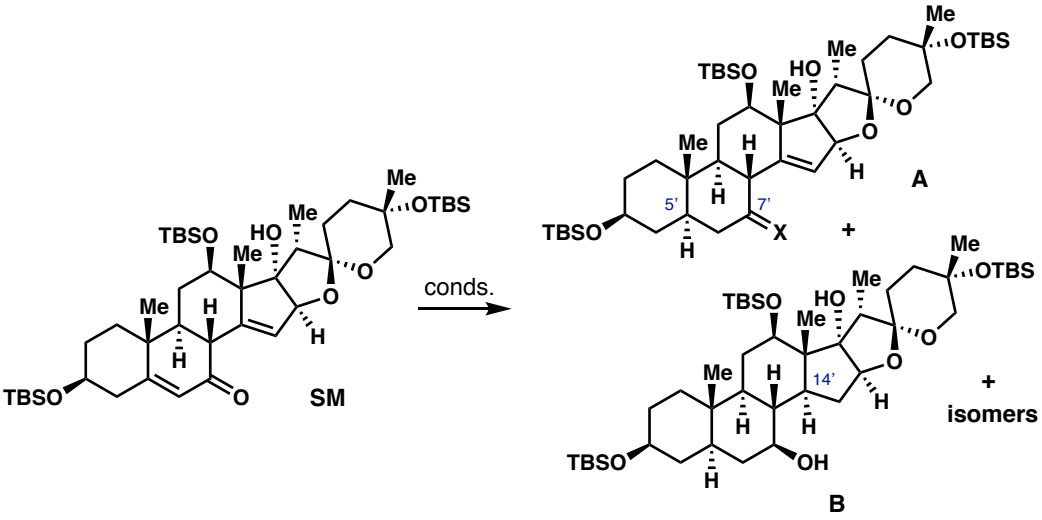
31.28, 27.74, 26.55, 26.16, 26.06, 25.99, 18.38, 18.33, 18.25, 14.50, 12.50, 8.01, -2.03, -2.19, -3.02, -4.19, -4.45.

FTIR (NaCl, thin film, cm^{-1}): 3522, 2953, 2926, 2853, 2243, 1471, 1462, 1372, 1251, 1091, 1061 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{45}H_{83}O_7Si_3$: 819.5447 $[M+H-H_2]^+$; found: 819.5470.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: +45° (*c* = 0.167, $CHCl_3$).

Table 2.3. Optimization of SmI_2 double reduction



entry	conds.	A X = β-OH, H	A X = α-OH, H	A X = O	B	SM
1	Pt/C (250 wt %), H_2 (balloon) EtOH, 22 °C	36%	–	11%	–	–
2	Pt/C (250 wt %), H_2 (balloon) EtOAc, 22 °C	–	–	–	45%	–
3	Pt/C (250 wt %), H_2 (balloon) hexane, 22 °C	–	–	–	51%	–
4	Pd/C (250 wt %), H_2 (balloon) *i*-PrOH, 22 °C	–	–	–	38%	–
5	Pd/C (20 wt %), H_2 (balloon) EtOH, 22 °C	17%	9%	25%	–	–
6	Li(0) (100 equiv), $FeCl_2{\cdot}4H_2O$ (10 equiv) THF, 22 °C	–	–	–	–	–
7	SmI_2 (10 equiv), t-BuOH (20 equiv) THF, –78 to 22 °C	–	–	–	–	36%
8	SmI_2 (10 equiv), PhSH (20 equiv) THF, –78 to 22 °C	–	–	60%	–	–
9	SmI_2 (10 equiv), PhSH (20 equiv) THF, 22 °C	75%	–	5%	–	–
10	SmI_2 (10 equiv), PhSH (20 equiv) Et_3N (20 equiv), THF, 22 °C	44%	–	21%	–	–
11	SmI_2 (10 equiv), TripSH (20 equiv) THF, 22 °C	74%	–	21%	–	–
12	SmI_2 (10 equiv), PhSH (20 equiv) THF, 28 °C[a]	77%	–	–	11%	–
13	SmI_2 (10 equiv), 2-NaphthSH (20 equiv) THF, 28 °C[a]	78%	–	–	10%	–
14	SmI_2 (10 equiv), 2-NaphthSH (20 equiv) THF, 0 °C	83%	–	–	5%	–
15	SmI_2 (10 equiv), 2-NaphthSH (20 equiv) THF, –15 °C	85%	–	–	–	–

[a]Glove box temperature.

Characterization data for **95**:

TLC (15% EtOAc/hexanes): R_f = 0.22 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 4.01 – 3.92 (m, 2H), 3.82 (q, *J* = 2.7 Hz, 1H), 3.59 (dd, *J* = 11.4, 5.5 Hz, 2H), 3.38 – 3.32 (m, 1H), 2.16 – 2.11 (m, 1H), 2.10 – 2.06 (m, 1H), 2.02 (d, *J* = 7.2 Hz, 1H), 1.99 – 1.94 (m, 1H), 1.94 – 1.88 (m, 1H), 1.78 (dd, *J* = 13.3, 4.4 Hz, 1H), 1.75 – 1.71 (m, 2H), 1.71 – 1.67 (m, 1H), 1.65 – 1.60 (m, 2H), 1.60 – 1.58 (m, 2H), 1.56 – 1.53 (m, 1H), 1.53 – 1.50 (m, 1H), 1.48 (d, *J* = 2.4 Hz, 1H), 1.46 (t, *J* = 2.7 Hz, 1H), 1.44 – 1.41 (m, 2H), 1.41 – 1.37 (m, 2H), 1.37 – 1.31 (m, 3H), 1.30 – 1.26 (m, 2H), 1.25 (s, 2H), 1.24 – 1.20 (m, 2H), 1.12 (d, *J* = 1.2 Hz, 1H), 1.11 (s, 1H) 1.07 – 0.98 (m, 2H), 0.98 – 0.94 (m, 1H), 0.94 (d, *J* = 7.2 Hz, 3H), 0.90 (m, 2H), 0.89 (s, 9H), 0.88 (s, 9H), 0.88 (s, 9H), 0.87 (d, *J* = 2.5 Hz, 2H), 0.81 (s, 3H), 0.77 (s, 3H), 0.10 (s, 3H), 0.08 (s, 3H), 0.07 (s, 3H), 0.05 (s, 3H), 0.04 (s, 3H), 0.03 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 110.15, 90.42, 90.39, 72.04, 72.00, 69.39, 69.36, 67.74, 50.24, 45.99, 45.18, 44.58, 38.62, 38.19, 37.34, 36.94, 36.57, 35.67, 33.85, 31.98, 30.64, 30.50, 27.13, 26.52, 26.39, 26.10, 26.07, 25.98, 18.41, 18.33, 11.40, 11.37, 8.09, -2.02, -2.17, -2.98, -3.38, -4.43.

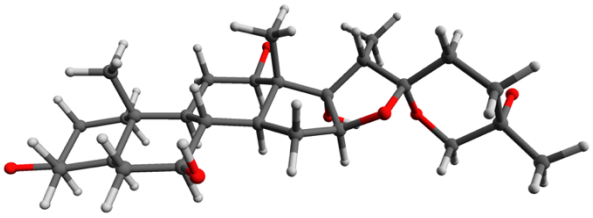
FTIR (NaCl, thin film, cm^{-1}): 3470, 2952, 2929, 2856, 1472, 1462, 1372, 1252, 1076, 1055 cm^{-1}.

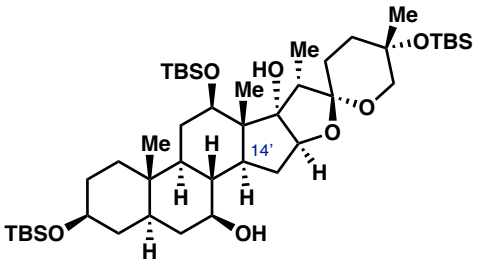
HRMS (FAB+, m/z): calc'd for $C_{45}H_{85}O_7Si_3$: 821.5603 $[M+H-H_2]^+$; found: 821.5611.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: –103° (*c* = 0.208, $CHCl_3$).

Minimized Structures and energies (see also Figure 2.4):

95





Single-point energy = -3160.596228982013 (Hartree)
N_atoms = 141
C 3.6828335 -1.566169 -0.698298
C 3.2484025 -0.2333645 -1.080993
C 2.0359535 0.277253 -0.393421
O 1.9947235 -0.132346 0.908973
C 2.238595 1.7602055 -0.439994
C 1.1002695 2.34322 -1.1341905
C 0.0457595 1.388837 -0.7640805
C -1.3193675 1.517233 -1.2832235
C -1.882646 2.861582 -1.0906235
O -1.2040675 3.826385 -1.78369
C -3.305516 2.920781 -1.4064395
C -4.0722185 1.9130655 -0.682034
C -5.5190885 2.0748215 -0.8491255
C -6.2655465 1.147861 -0.0032745
O -7.601078 1.211832 -0.275681
Si -8.7478235 2.012233 0.503491
C -8.5791755 1.790282 2.283262
C -8.657534 3.774689 0.1410285
C -10.2983405 1.286401 -0.1275045
C -10.3314755 1.3857205 -1.594424
C -11.456058 2.000169 0.4313535
C -10.3747555 -0.13092 0.2546505
C -5.82524 -0.223664 -0.2230865
C -4.3738155 -0.378722 -0.104425
C -3.597996 0.5334955 -0.973908
C -3.803509 0.1571735 -2.3811575
C -2.1494425 0.4823885 -0.608029
C -1.5511685 -0.8551325 -0.752543
C -0.147479 -0.9577595 -0.3338695
O 0.336949 -2.205157 -0.61305
Si 0.049404 -3.652758 0.01298
C -1.5724555 -4.2650935 -0.4935145
C 1.2857045 -4.6858865 -0.7802245
C 0.2094215 -3.722632 1.8408735
C -0.850646 -2.952039 2.507906
C 0.092221 -5.129185 2.2618995
C 1.51949 -3.2074715 2.263612
C 0.6795835 0.0570365 -1.006641

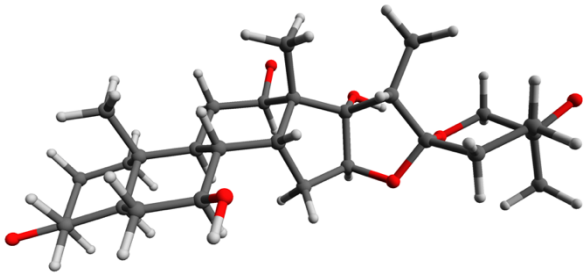
C	0.781053	-0.24966	-2.4386995
O	3.4531985	1.9866045	-1.0342765
C	4.196928	0.863041	-0.8500925
C	5.3609245	0.9375775	-1.714233
C	6.266658	2.008159	-1.3029195
C	6.628472	1.9001905	0.116312
C	7.375933	3.077253	0.5483615
O	7.3649115	0.755355	0.2395915
Si	8.4066525	0.170329	1.30255
C	7.8614825	0.392704	3.0070025
C	10.018068	0.953831	1.1133865
C	8.4986015	-1.6042745	0.8797605
C	9.5210035	-2.2728475	1.6985345
C	8.844512	-1.7611275	-0.5401485
C	7.1961475	-2.2405485	1.1256135
C	5.3889305	1.7822705	0.8931155
O	4.596287	0.752226	0.469909
H	2.928299	-2.251797	-0.8945395
H	3.911267	-1.6223715	0.3115955
H	4.528986	-1.8465035	-1.2363785
H	3.0836515	-0.221722	-2.1068175
H	2.826817	0.0969995	1.2506865
H	2.2654125	2.1066575	0.544933
H	1.2713075	2.3584915	-2.1606155
H	0.891314	3.314646	-0.8406255
H	-0.046879	1.4691115	0.2771535
H	-1.318746	1.3359065	-2.31331
H	-1.75222	3.135953	-0.0906495
H	-1.2828675	3.6135705	-2.6796
H	-3.432851	2.801452	-2.437559
H	-3.6547145	3.8765285	-1.1796285
H	-3.873371	2.064578	0.336452
H	-5.7870505	3.0551385	-0.61471
H	-5.8073055	1.920804	-1.8395275
H	-6.091825	1.3990745	0.9995695
H	-8.587303	0.7856655	2.555014
H	-9.358894	2.260918	2.789491
H	-7.6923955	2.2059555	2.638571
H	-9.4199605	4.3016735	0.6172295
H	-7.754346	4.1786375	0.4676905
H	-8.7321045	3.9619535	-0.880077
H	-11.207162	0.954411	-1.9674045
H	-9.5239335	0.8924515	-2.0265745
H	-10.311789	2.376156	-1.9177685
H	-11.4854725	1.948658	1.4720535
H	-11.462123	3.006599	0.162309

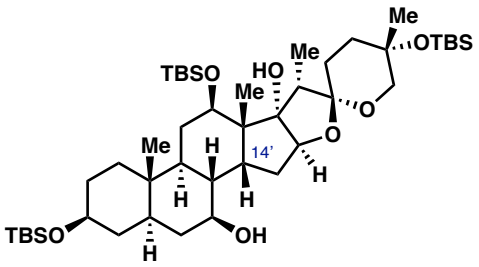
H -12.3459425 1.5817835 0.078238
H -10.392 -0.2550755 1.289199
H -9.5667555 -0.6712915 -0.1173365
H -11.2475465 -0.5656565 -0.12156
H -6.305253 -0.852386 0.4556595
H -6.1501415 -0.5137035 -1.170239
H -4.092626 -0.208108 0.8882815
H -4.129945 -1.3700105 -0.310593
H -4.794469 0.266709 -2.673356
H -3.551462 -0.840437 -2.534365
H -3.2286975 0.724568 -3.0356635
H -2.1128785 0.7171475 0.4132145
H -1.6182505 -1.1777195 -1.7397145
H -2.095934 -1.536146 -0.1866165
H -0.0783775 -0.784618 0.6899275
H -1.69053 -4.153117 -1.5222925
H -1.668044 -5.280266 -0.2781215
H -2.3653885 -3.774284 -0.0317115
H 2.2509625 -4.455701 -0.4683135
H 1.249422 -4.5547545 -1.8126515
H 1.1198655 -5.695511 -0.5862835
H -0.7687645 -3.047389 3.545706
H -0.7960145 -1.935201 2.2913295
H -1.802668 -3.2846705 2.2452255
H 0.8368525 -5.7254105 1.84365
H 0.1848445 -5.204775 3.300184
H -0.8293345 -5.5433055 2.006018
H 1.6050965 -3.248823 3.3049755
H 2.299844 -3.776776 1.8720845
H 1.6694305 -2.222179 1.966015
H 1.197245 -1.1907855 -2.572741
H 1.362461 0.4414205 -2.952041
H -0.1507885 -0.2599995 -2.897451
H 5.856665 0.025125 -1.668616
H 5.0448685 1.085375 -2.695004
H 7.1370825 1.9885455 -1.874111
H 5.803766 2.928113 -1.4583945
H 8.252565 3.1690095 -0.003512
H 7.628757 3.019981 1.556715
H 6.8155775 3.9436675 0.4113105
H 6.9201075 -0.018 3.1745825
H 7.8229145 1.3977605 3.27751
H 8.526055 -0.0712715 3.662486
H 10.727732 0.4988735 1.7258495
H 9.986601 1.9618145 1.3709605
H 10.3607115 0.8900185 0.1324935

H 10.4736615 -1.8853055 1.532169
H 9.3196265 -2.197909 2.718495
H 9.5583245 -3.291376 1.4690075
H 8.146178 -1.29769 -1.1564295
H 9.7800275 -1.3586555 -0.7624205
H 8.8747295 -2.773318 -0.7988585
H 6.4337535 -1.771656 0.59619
H 7.2209575 -3.2433735 0.8329915
H 6.9383645 -2.2212295 2.1351375
H 5.5976565 1.5805665 1.890745
H 4.862702 2.6823135 0.830351

96

Single-point energy = -3160.584909522271 (Hartree)
N_atoms = 141
C 3.894561 0.672822 -1.6333795
C 3.0514845 0.6588015 -0.4447055
C 1.805253 -0.1598375 -0.551061
O 2.065222 -1.312574 -1.225026
C 1.496808 -0.4297535 0.8813245
C 0.322551 0.351028 1.240056
C 0.0989075 1.269184 0.109301
C -1.2650505 1.8143115 0.0079925
C -1.659226 2.4861505 1.2466325
O -0.760376 3.4944085 1.457682
C -3.0290855 2.9831515 1.1906125
C -3.9588335 1.9060305 0.858411
C -5.3642645 2.29335 1.001157
C -6.254003 1.1572205 0.789951
O -7.541191 1.5979085 0.893897
Si -8.835444 0.765445 1.3425365
C -8.448537 -0.2832585 2.7557565
C -9.4436295 -0.2773155 0.005934
C -10.0540525 2.0547065 1.7705935
C -10.3037475 2.9122065 0.6020385
C -9.5336975 2.887649 2.865099
C -11.3196175 1.4375 2.1934205
C -5.994839 0.5428 -0.5079255
C -4.5850655 0.174443 -0.664207
C -3.6461795 1.299142 -0.465423
C -3.801911 2.2643985 -1.566359
C -2.242783 0.7845005 -0.392121
C -1.775709 0.1393475 -1.6213385
C -0.4355005 -0.449057 -1.5276105



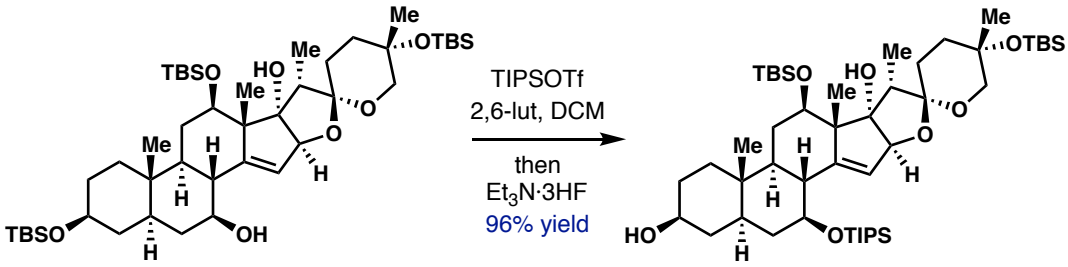


O	-0.091581	-0.955688	-2.7378
Si	-0.499229	-2.2499745	-3.5807775
C	0.859139	-2.426199	-4.7391405
C	-2.0082275	-1.9127475	-4.5127175
C	-0.716264	-3.7840825	-2.5895935
C	0.5447765	-4.15998	-1.9348415
C	-1.762785	-3.632841	-1.567124
C	-1.1267045	-4.870648	-3.495648
C	0.5935005	0.5390235	-1.096443
C	0.887561	1.4549655	-2.1989945
O	2.5906975	-0.019767	1.612237
C	3.645115	0.1226885	0.7971675
C	4.6086765	0.947766	1.524264
C	6.0133665	0.727771	1.1858705
C	6.365821	-0.6949045	1.171955
C	6.2960885	-1.305408	2.492896
O	7.588235	-0.832692	0.5817995
Si	9.1096655	-1.181226	0.926164
C	9.2818705	-2.571221	2.060259
C	9.769062	-1.6736155	-0.6693495
C	10.014791	0.278687	1.5740765
C	11.4483945	-0.044003	1.6661995
C	9.537444	0.6525985	2.912475
C	9.8466285	1.4223805	0.6663865
C	5.43655	-1.3521725	0.2452345
O	4.1242035	-1.179054	0.590347
H	4.0561015	-0.2864835	-1.9938125
H	3.4445215	1.21276	-2.396766
H	4.814849	1.1172805	-1.439443
H	2.783131	1.639347	-0.212307
H	2.7873295	-1.6801755	-0.768189
H	1.3616705	-1.45159	1.0347665
H	-0.4937785	-0.278896	1.3839045
H	0.482697	0.8591595	2.132435
H	0.7274045	2.087477	0.2505125
H	-1.2513605	2.550394	-0.732449
H	-1.6045575	1.790143	2.0300465
H	-0.9951035	3.901695	2.2520915
H	-3.081821	3.756454	0.4936445
H	-3.285442	3.3968615	2.116181
H	-3.7927875	1.14274	1.5586365
H	-5.619518	3.044935	0.325101
H	-5.528438	2.683299	1.954223
H	-6.061418	0.449357	1.5389195
H	-8.0534205	0.260621	3.550436
H	-9.303517	-0.769342	3.0995255

H -7.758126 -1.022507 2.506575
H -10.3373725 -0.743987 0.269272
H -9.6161645 0.27178 -0.861134
H -8.7608435 -1.028082 -0.2281585
H -10.703687 2.376246 -0.1971705
H -9.430369 3.3721985 0.272949
H -10.9896945 3.6624215 0.844305
H -10.213248 3.6463145 3.0981335
H -8.631854 3.3355625 2.6036205
H -9.3772325 2.3346885 3.7344795
H -11.7458175 0.8595735 1.438408
H -11.200748 0.829272 3.031573
H -12.016843 2.173253 2.4473995
H -6.5894585 -0.3044655 -0.628999
H -6.283946 1.216239 -1.24917
H -4.4507055 -0.242976 -1.608949
H -4.352798 -0.5777055 0.024193
H -3.0828455 3.013803 -1.539621
H -4.729034 2.732759 -1.5449845
H -3.7207495 1.789661 -2.488493
H -2.2538235 0.05597 0.359378
H -1.7725235 0.824724 -2.404849
H -2.429958 -0.6210095 -1.8963775
H -0.45674 -1.2168175 -0.820975
H 0.733201 -3.244081 -5.3713125
H 0.934591 -1.5757865 -5.3355655
H 1.7625055 -2.5366935 -4.2358505
H -2.172201 -2.634027 -5.246616
H -2.858181 -1.8799175 -3.912054
H -1.9294755 -0.9916055 -4.9924145
H 1.2867455 -4.3818175 -2.6319785
H 0.4097775 -5.0146495 -1.347471
H 0.9090615 -3.3994955 -1.3260835
H -1.501059 -2.9433395 -0.832852
H -2.673496 -3.332499 -1.9756275
H -1.9309565 -4.54384 -1.0823925
H -1.216834 -5.765597 -2.9633355
H -0.436543 -5.0367315 -4.2584275
H -2.0509185 -4.6828315 -3.938007
H 1.5371965 2.208697 -1.897375
H 1.301935 0.944324 -3.000605
H 0.0235205 1.9210535 -2.5420825
H 4.4559875 0.741033 2.532656
H 4.359519 1.9494925 1.387184
H 6.6143735 1.2413395 1.8648275
H 6.2345705 1.108404 0.241864

H 5.315227 -1.3035385 2.8359785
H 6.887611 -0.796147 3.180957
H 6.610466 -2.295298 2.453031
H 10.2827235 -2.855356 2.122061
H 8.955558 -2.3631185 3.0251405
H 8.748434 -3.3966105 1.715495
H 9.6551945 -0.9239755 -1.381818
H 10.781023 -1.9129465 -0.61507
H 9.2643415 -2.5148105 -1.0185305
H 11.8587835 -0.2759705 0.737468
H 11.632326 -0.8465515 2.304843
H 11.9789865 0.774797 2.0400865
H 8.5288705 0.9125505 2.9109565
H 9.668334 -0.116766 3.602687
H 10.065938 1.4802975 3.269849
H 10.2122255 1.217102 -0.287348
H 8.8477235 1.696837 0.5673995
H 10.368508 2.253559 1.0255505
H 5.638024 -1.000488 -0.7150055
H 5.585593 -2.382277 0.2528335

Preparation of C3 alcohol 79:



A 25 mL round-bottom flask with a stir bar was charged with charged with C7′ alcohol **78** (50 mg, 0.06 mmol), 2,6-lutidine (19.8 μL, 0.17 mmol, 3.0 equiv) and CH_2Cl_2 (5.8 mL, 0.01 M). TIPSOTf (23.4 μL, 0.087 mmol, 1.5 equiv) was added dropwise at room temperature. After stirring for 30 min, $Et_3N{\cdot}3HF$ (186 μL, 1.16 mmol, 20 equiv) was added dropwise and stirred at room temperature for 12 h. The reaction was diluted with hexane (6 mL), quenched with H_2O (6 mL). The layers were separated. The aqueous layer was further extracted with hexane (3 × 6 mL), combined organic layers were washed with st.

$NaHCO_3$ aq. (6 mL) and brine (6 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (10 to 15 to 20 % EtOAc/hexanes) to afford C3′ alcohol **79** (50.5 mg, 0.058 mmol, 96% yield) as a colorless foam.

TLC (20% EtOAc/hexanes): R_f = 0.31 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.47 (t, *J* = 1.7 Hz, 1H), 4.74 (d, *J* = 1.9 Hz, 1H), 3.91 (ddd, *J* = 22.1, 10.5, 5.1 Hz, 2H), 3.70 (d, *J* = 11.7 Hz, 1H), 3.58 (tt, *J* = 10.6, 4.6 Hz, 1H), 3.44 (dd, *J* = 11.7, 2.7 Hz, 1H), 3.34 (s, 1H), 2.15 – 2.05 (m, 1H), 1.96 (t, *J* = 10.5 Hz, 1H), 1.89 – 1.84 (m, 1H), 1.83 – 1.79 (m, 1H), 1.79 – 1.70 (m, 2H), 1.66 – 1.63 (m, 1H), 1.63 – 1.53 (m, 5H), 1.46 – 1.40 (m, 2H), 1.40 – 1.34 (m, 2H), 1.34 – 1.29 (m, 1H), 1.25 (s, 3H), 1.22 (dd, *J* = 4.1, 2.4 Hz, 1H), 1.21 – 1.13 (m, 2H), 1.11 (s, 3H), 1.05 (s, 4H), 1.03 (s, 3H), 1.03 – 1.03 (m, 6H), 1.02 (s, 6H), 1.01 (d, *J* = 1.5 Hz, 2H), 0.94 (dd, *J* = 13.5, 4.0 Hz, 1H), 0.89 (s, 9H), 0.88 (s, 9H), 0.86 (s, 3H), 0.82 – 0.77 (m, 1H), 0.77 – 0.71 (m, 1H), 0.10 (s, 3H), 0.08 (s, 3H), 0.07 (s, 3H), 0.07 (s, 3H).

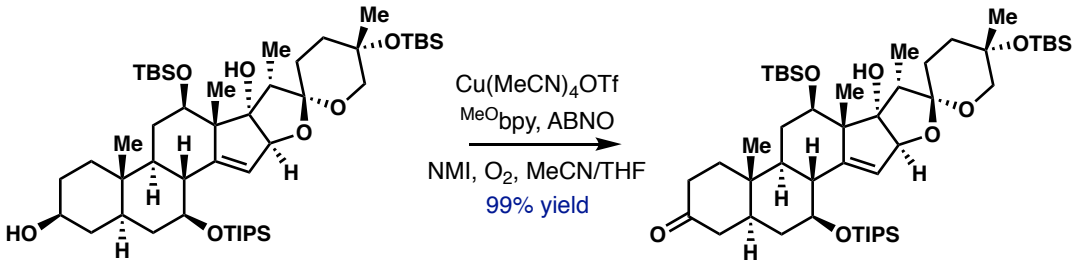
^{13}C NMR (101 MHz, $CDCl_3$): δ 150.67, 120.54, 107.03, 93.71, 91.87, 75.58, 71.31, 71.15, 69.91, 69.44, 56.48, 51.33, 48.41, 43.39, 41.89, 38.30, 38.00, 36.75, 35.36, 33.96, 31.66, 31.39, 29.85, 28.20, 26.58, 26.20, 26.01, 18.71, 18.69, 18.34, 18.28, 13.48, 13.25, 12.63, 8.10, -2.02, -2.19, -2.84, -4.25.

FTIR (NaCl, thin film, cm^{-1}): 3521, 2927, 2862, 1470, 1462, 1384, 1252, 1117, 1088, 1058 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{48}H_{90}O_7Si_3$: 862.5994 $[M+\bullet]^-$; found: 862.6001.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\boldsymbol{21}}$: +33° (*c* = 0.125, $CHCl_3$).

Preparation of ketone 47:



A 10 mL round-bottom flask with a stir bar was charged with C3′ alcohol **79** (29.0 mg, 0.035 mmol, 1.0 equiv), $Cu(MeCN)_4OTf$ (1.3 mg, 3.50 μmol, 10 mol %), ^{MeO}bpy (0.76 mg, 3.50 μmol, 10 mol %), ABNO (0.50 mg, 3.50 μmol, 10 mol %), NMI (0.56 μL, 7.00 μmol, 20 mol %) and MeCN/THF (3:1, 3.5 mL, 0.01 M). The resulting solution was sparged with O_2 for 5 min, then vigurously stirred (1200 rpm) at room temperature under O_2 atmosphere for 12 h. The reaction was passed through a short pad of SiO_2 using 50% EtOAc/hexanes as the eluent and concentrated in vacuo. The crude product was purified by flash column chromatography (10 to 15 to 20% EtOAc/hexanes) to afford western fragment **47** (28.9 mg, 0.034 mmol, 99% yield) as a colorless foam.

TLC (15% EtOAc/hexanes): R_f = 0.53 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.48 (t, *J* = 1.7 Hz, 1H), 4.73 (d, *J* = 2.0 Hz, 1H), 3.99 – 3.86 (m, 2H), 3.70 (d, *J* = 11.7 Hz, 1H), 3.43 (dd, *J* = 11.7, 2.7 Hz, 1H), 3.24 (s, 1H), 2.43 – 2.32 (m, 2H), 2.32 – 2.26 (m, 1H), 2.16 – 2.06 (m, 2H), 2.02 (t, *J* = 10.5 Hz, 1H), 1.91 (ddd, *J* = 14.6, 6.8, 3.1 Hz, 1H), 1.86 (q, *J* = 6.9 Hz, 1H), 1.82 – 1.76 (m, 1H), 1.76 – 1.72 (m, 1H), 1.67 (ddd, *J* = 12.3, 4.9, 2.3 Hz, 1H), 1.62 – 1.59 (m, 1H), 1.58 – 1.53 (m, 2H), 1.50 – 1.41 (m, 2H), 1.33 – 1.24 (m, 3H), 1.23 – 1.18 (m, 1H), 1.11 (s, 3H), 1.07 (s, 3H),

1.06 (s, 3H), 1.05 (d, J = 6.9 Hz, 3H), 1.03 (d, J = 1.3 Hz, 3H), 1.03 – 1.02 (m, 12H), 1.01 (d, J = 1.3 Hz, 3H), 0.89 (s, 9H), 0.88 (s, 9H), 0.87 – 0.84 (m, 1H), 0.10 (s, 3H), 0.09 (s, 3H), 0.08 (s, 3H), 0.07 (s, 3H).

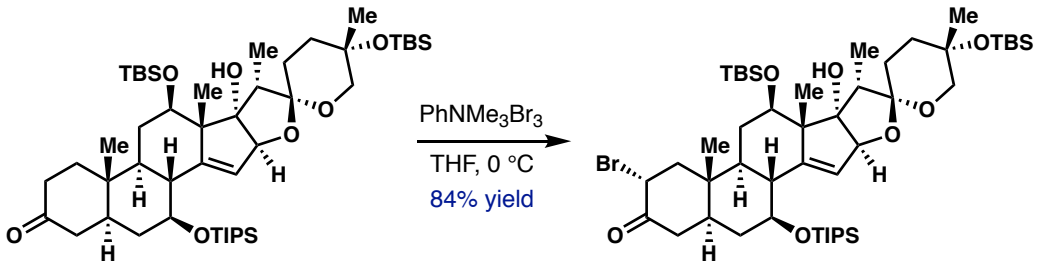
^{13}C NMR (101 MHz, $CDCl_3$): δ 210.91, 150.19, 120.90, 107.10, 93.69, 91.79, 75.06, 70.86, 69.90, 69.43, 56.43, 50.57, 48.33, 44.33, 43.71, 43.22, 38.35, 38.06, 37.99, 35.49, 33.91, 32.01, 29.86, 28.03, 26.57, 26.21, 26.01, 18.69, 18.67, 18.34, 18.29, 13.49, 13.22, 11.79, 8.08, -2.02, -2.19, -3.00, -4.24.

FTIR (NaCl, thin film, cm^{-1}): 3524, 2932, 2860, 1715, 1470, 1462, 1384, 1252, 1113, 1089, 1056 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{48}H_{87}O_7Si_3$: 859.5760 $[M+H-H_2]^+$; found: 859.5757.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: +70° ($c$ = 0.167, $CHCl_3$).

Preparation of α-bromo ketone 80:



A 25 mL round-bottom flask with a stir bar was charged with ketone **47** (27.0 mg, 0.031 mmol, 1.0 equiv) and THF (1.5 mL, 0.01 M total). The resulting solution was cooled to 0 °C and stirred at this temperature for 5 min. $PhNMe_3Br_3$ (12.2 mg, 0.33 mmol, 1.05 equiv) in THF (1.5 mL, 0.01 M total) was also cooled to 0 °C and was added by cannula in one portion, and the reaction was stirred at 0 °C for 30 min. The reaction was quenched with st. $Na_2S_2O_3$ aq. (1.5 mL) and H_2O (1.5 mL), diluted with hexane (3 mL). The layers were

separated. The aqueous layer was further extracted with hexane (3 × 3 mL), combined organic layers were washed with brine (3 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (5 to 10 to 15% EtOAc/hexanes) to afford α-bromo ketone **80** (24.4 mg, 0.026 mmol, 84% yield) as a colorless foam.[11i]

TLC (10% EtOAc/hexanes): R_f = 0.35 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.47 (t, *J* = 1.7 Hz, 1H), 4.77 – 4.65 (m, 2H), 3.92 (ddd, *J* = 18.4, 10.4, 5.1 Hz, 2H), 3.69 (d, *J* = 11.6 Hz, 1H), 3.43 (dd, *J* = 11.7, 2.7 Hz, 1H), 3.15 (s, 1H), 2.52 – 2.43 (m, 2H), 2.16 – 2.06 (m, 1H), 1.99 (t, *J* = 10.6 Hz, 1H), 1.85 (q, *J* = 6.9 Hz, 1H), 1.82 – 1.73 (m, 2H), 1.73 – 1.64 (m, 3H), 1.64 – 1.54 (m, 3H), 1.52 – 1.42 (m, 2H), 1.26 (s, 2H), 1.24 – 1.14 (m, 2H), 1.14 (s, 3H), 1.11 (s, 3H), 1.08 (d, *J* = 4.9 Hz, 1H), 1.07 (s, 3H), 1.05 (d, *J* = 6.9 Hz, 4H), 1.02 (s, 9H), 1.01 (s, 6H), 1.00 (s, 3H), 0.91 (d, *J* = 3.0 Hz, 1H), 0.90 (s, 6H), 0.89 (s, 1H), 0.88 (s, 9H), 0.10 (s, 6H), 0.08 (d, *J* = 1.5 Hz, 6H).

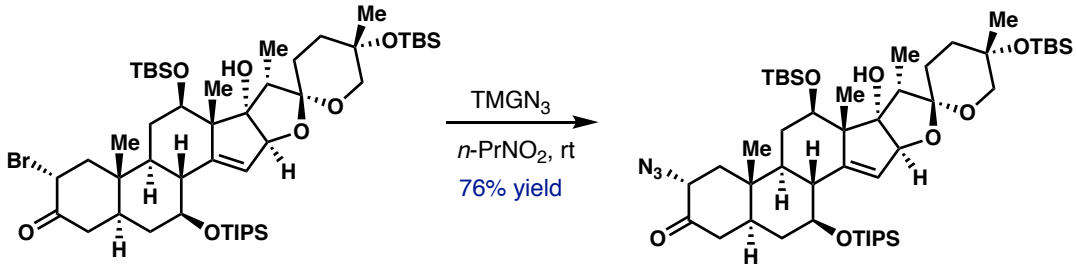
^{13}C NMR (101 MHz, $CDCl_3$): δ 200.38, 149.77, 121.05, 107.16, 93.68, 91.72, 74.54, 70.57, 69.90, 69.42, 56.38, 53.75, 50.81, 50.20, 48.24, 44.33, 43.52, 42.79, 38.66, 37.81, 33.87, 32.12, 29.85, 27.85, 26.55, 26.21, 26.01, 18.67, 18.65, 18.34, 18.30, 13.55, 13.18, 12.43, 8.07, -2.01, -2.18, -3.04, -4.24.

FTIR (NaCl, thin film, cm^{-1}): 3519, 2924, 2862, 1728, 1470, 1462, 1387, 1254, 1089, 1062 cm^{-1}.

HRMS (TOF-ESI, m/z): calc'd for $C_{48}H_{88}O_7BrSi_3$: 939.5021 $[M+H]^+$; found: 939.5021.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: +39° (*c* = 0.250, $CHCl_3$).

Preparation of western fragment α-azido ketone 81:



In an N_2-filled glovebox, a 10 mL round-bottom flask with a stir bar was charged with α-bromo ketone **80** (21 mg, 0.023 mmol) and *n*-$PrNO_2$ (2.3 mL, 0.01). $TMGN_3$ (14.3 mg, 0.09 mmol, 4.0 equiv) was added, and the reaction was stirred at room temperature for 4 h. Upon completion, the reaction was diluted with heptane (2.3 mL) and was concentrated in vacuo. The crude material was azeotroped with another 2.3 mL of heptane to remove *n*-$PrNO_2$. The crude product was purified by flash column chromatography (5 to 10 to 15% EtOAc/hexanes) to afford α-azido ketone **18** (15.8 mg, 0.017 mmol, 76% yield) as a colorless foam.[11i]

TLC (10% EtOAc/hexanes): R_f = 0.40 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.47 (t, *J* = 1.6 Hz, 1H), 4.71 (d, *J* = 2.0 Hz, 1H), 4.04 – 3.85 (m, 3H), 3.69 (d, *J* = 11.6 Hz, 1H), 3.43 (dd, *J* = 11.7, 2.7 Hz, 1H), 3.15 (s, 1H), 2.42 (t, *J* = 14.1 Hz, 1H), 2.29 (dd, *J* = 14.5, 3.8 Hz, 1H), 2.17 (dd, *J* = 12.7, 6.2 Hz, 1H), 2.12 – 2.06 (m, 1H), 2.00 (t, *J* = 10.5 Hz, 1H), 1.85 (q, *J* = 6.9 Hz, 1H), 1.79 (dd, *J* = 13.3, 4.2 Hz, 1H), 1.73 (t, *J* = 4.6 Hz, 1H), 1.70 (t, *J* = 5.0 Hz, 1H), 1.61 (m, 3H), 1.52 –1.38 (m, 3H), 1.29 (d, *J* = 12.9 Hz, 2H), 1.25 (s, 3H), 1.20 (ddd, *J* = 12.8, 4.2, 2.4 Hz, 2H), 1.13 (s, 3H), 1.11 (s, 3H), 1.07 (s, 3H), 1.05 (d, *J* = 6.9 Hz, 5H), 1.03 (t, *J* = 1.4 Hz, 9H), 1.02 (s,

6H), 1.00 (d, J = 1.4 Hz, 3H), 0.89 (s, 9H), 0.88 (s, 9H), 0.87 – 0.85 (m, 1H), 0.10 (d, J = 1.2 Hz, 6H), 0.07 (s, 6H).

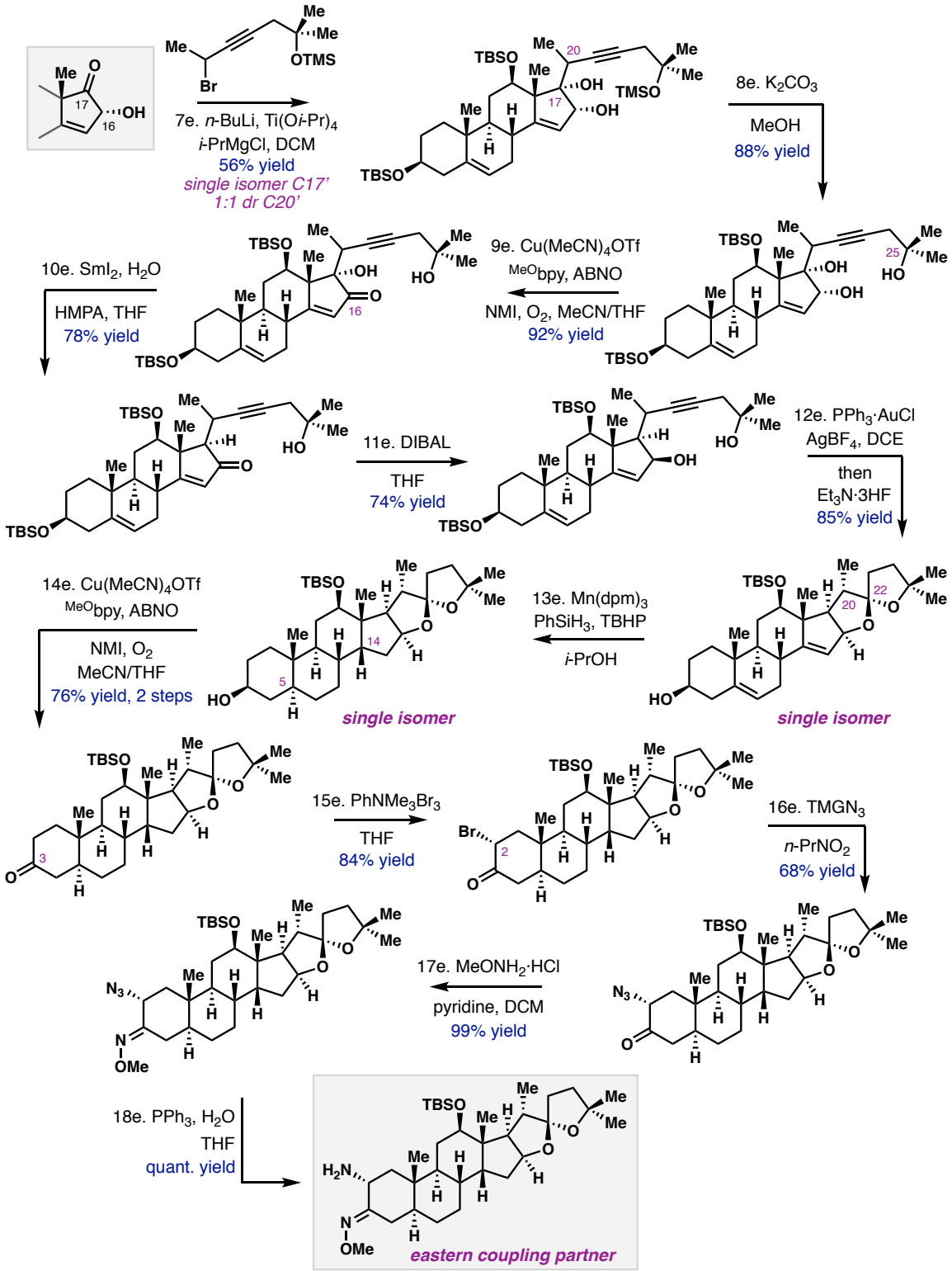
^{13}C NMR (101 MHz, $CDCl_3$): δ 204.52, 149.82, 121.03, 107.16, 93.68, 91.72, 74.55, 70.61, 69.90, 69.42, 63.63, 56.38, 50.29, 48.24, 44.69, 44.39, 43.40, 42.74, 37.80, 36.67, 33.87, 32.18, 29.86, 27.86, 26.55, 26.21, 26.01, 18.67, 18.66, 18.35, 18.30, 13.54, 13.18, 12.79, 8.07, 1.17, -2.01, -2.18, -3.07, -4.25.

FTIR (NaCl, thin film, cm^{-1}): 3526, 2928, 2857, 2106, 1732, 1463, 1386, 1252, 1090, 1056 cm^{-1}.

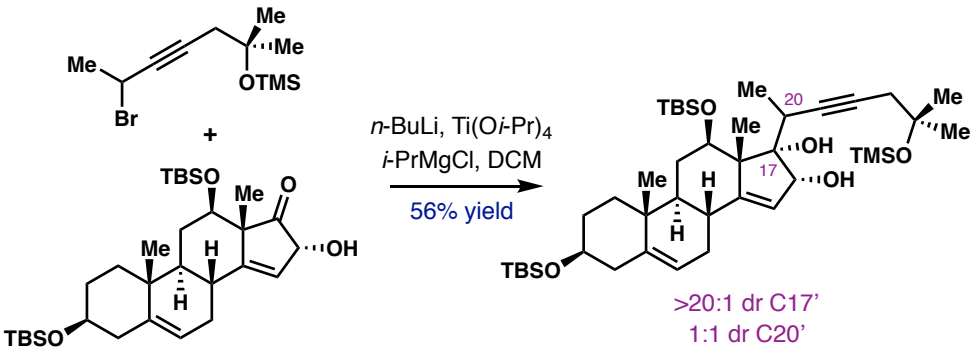
HRMS (TOF-ESI, m/z): calc'd for $C_{48}H_{88}N_3O_7Si_3$: 902.5930 $[M+H]^+$; found: 902.5936.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: +45° ($c$ = 0.083, $CHCl_3$).

Synthesis of eastern amino methoxime 93



Preparation of homopropargyl alcohol 82:



A 1 L round-bottom flask with a stir bar was charged with α-hydroxy ketone **59** (6.78 g, 12.4 mmol, 1.0 equiv), propargyl bromide **52** (5.13 g, 18.6 mmol, 1.5 equiv) and CH_2Cl_2 (250 mL, 0.05 M). The resulting solution was cooled to –78 °C and stirred at this temperature for 30 min. Thereafter, *n*-BuLi (2.5 M in hexane, 5.21 mL, 13.02 mmol, 1.05 equiv) was added dropwise at –78 °C. After stirring at –78 °C for 15 min, Ti(O*i*-Pr)$_4$ (7.34 mL, 24.8 mmol, 2.0 equiv) was added dropwise at –78 °C. After stirring at –78 °C for 15 min, *i*-PrMgCl (2.0 M in Et_2O, 24.8 mL, 49.6 mmol, 4.0 equiv) was added dropwise at –78 °C. The reaction was stirred at –78 °C for 30 min, then warmed to room temperature. After stirring at room temperature for 12 h, the reaction was diluted with Et_2O (150 mL), quenched with st. NH_4Cl aq. (150 mL) and H_2O (150 mL). After vigorously stirring for 3 h, the resulting solid was removed by filtration, washed with Et_2O (2 × 50 mL). The layers were separated. The aqueous layer was further extracted with EtOAc (3 × 150 mL), combined organic layers were washed with brine (2 × 150 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (5 to 10 to 15% EtOAc/hexanes) to afford homopropargyl alcohol **82** (5.17 g, 6.94 mmol, 1:1 mixture of diastereomers at C20, 56% yield) as a colorless foam.

TLC (10% EtOAc/hexanes): R_f = 0.48 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.37 – 5.32 (m, 2H), 5.26 (t, *J* = 1.9 Hz, 1H), 5.18 (t, *J* = 1.8 Hz, 1H), 4.79 (d, *J* = 13.2 Hz, 2H), 4.31 (dd, *J* = 11.5, 4.5 Hz, 1H), 4.20 (dd, *J* = 11.5, 4.2 Hz, 1H), 3.68 (s, 1H), 3.54 – 3.41 (m, 3H), 3.38 (s, 1H), 3.31 (d, *J* = 4.6 Hz, 1H), 3.21 (qt, *J* = 7.1, 2.4 Hz, 1H), 2.84 (qt, *J* = 7.0, 2.4 Hz, 1H), 2.64 (s, 1H), 2.31 (dd, *J* = 8.5, 2.3 Hz, 5H), 2.25 (d, *J* = 2.5 Hz, 1H), 2.23 (d, *J* = 2.4 Hz, 1H), 2.20 (d, *J* = 5.4 Hz, 2H), 2.18 – 2.13 (m, 2H), 2.13 – 2.07 (m, 2H), 2.05 – 2.02 (m, 1H), 2.02 – 1.98 (m, 1H), 1.98 – 1.93 (m, 1H), 1.81 – 1.78 (m, 1H), 1.76 (s, 3H), 1.74 – 1.69 (m, 3H), 1.67 (t, *J* = 4.3 Hz, 1H), 1.61 – 1.54 (m, 4H), 1.51 (dt, *J* = 7.5, 3.1 Hz, 3H), 1.47 (d, *J* = 8.6 Hz, 1H), 1.36 (d, *J* = 7.0 Hz, 3H), 1.33 (s, 3H), 1.31 (dd, *J* = 2.9, 1.8 Hz, 12H), 1.25 (s, 3H), 1.12 (d, *J* = 3.0 Hz, 4H), 1.08 (s, 4H), 1.06 – 1.04 (m, 2H), 1.02 (s, 3H), 1.01 (s, 3H), 0.93 – 0.92 (m, 2H), 0.91 (s, 9H), 0.90 (s, 9H), 0.89 (s, 18H), 0.15 (s, 9H), 0.12 (d, *J* = 1.2 Hz, 12H), 0.10 (s, 3H), 0.08 (s, 3H), 0.07 (s, 3H), 0.06 – 0.05 (m, 12H).

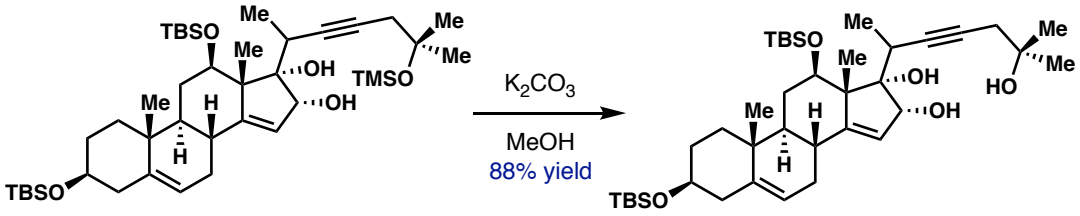
^{13}C NMR (101 MHz, $CDCl_3$): δ 154.03, 153.53, 140.69, 123.00, 121.95, 120.85, 120.82, 86.16, 83.93, 82.55, 82.24, 81.79, 79.72, 78.86, 76.51, 73.89, 73.86, 72.75, 72.57, 71.83, 58.17, 56.88, 47.67, 47.30, 42.73, 37.08, 37.05, 36.96, 35.55, 35.53, 33.54, 33.17, 32.09, 31.06, 30.25, 30.07, 29.86, 29.69, 29.62, 29.55, 29.51, 29.37, 29.05, 26.37, 26.19, 26.08, 19.32, 19.26, 18.46, 18.41, 18.24, 17.42, 15.08, 14.87, 2.66, -3.07, -3.38, -3.74, -3.80, -4.43.

FTIR (NaCl, thin film, cm^{-1}): 3472, 2927, 2854, 1748, 1722, 1470, 1463, 1371, 1248, 1049 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{42}H_{75}O_5Si_3$: 743.4922 $[M+H-H_2]^+$; found: 743.4940.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: –123° (*c* = 0.222, $CHCl_3$).

Preparation of triol 83:



A 500 mL round-bottom flask with a stir bar was charged with homopropargyl alcohol **82** (4.13 g, 5.55 mmol, 1:1 mixture of diastereomers at C20) and MeOH (185 mL, 0.03 M). K_2CO_3 (3.84 g, 27.8 mmol, 5.0 equiv) was added in one portion and stirred at room temperature for 12 h. The reaction was quenched with st. NH_4Cl aq. (50 mL) and H2O (50 mL). The resulting mixture was concentrated in vacuo to remove most of the MeOH. EtOAc (100 mL) was added, the layers were separated. The aqueous layer was further extracted with EtOAc (3 × 100 mL), combined organic layers were washed with brine (2 × 100 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (20 to 30 to 40% EtOAc/hexanes) to afford tertiary alcohol **83** (3.64 g, 4.88 mmol, 1:1 mixture of diastereomers at C20, 88% yield) as a colorless foam.

TLC (20% EtOAc/hexanes): R_f = 0.19 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.38 – 5.33 (m, 2H), 5.31 (t, *J* = 1.9 Hz, 1H), 5.23 (t, *J* = 1.9 Hz, 1H), 4.82 – 4.70 (m, 2H), 4.28 (dd, *J* = 11.4, 4.6 Hz, 1H), 4.16 (dd, *J* = 11.5, 4.3 Hz, 1H), 3.92 (s, 1H), 3.53 –3.43 (m, 3H), 3.23 – 3.12 (m, 1H), 2.85 (qt, *J* = 7.0, 2.4 Hz, 1H), 2.74 (s, 1H), 2.33 (dd, *J* = 6.6, 2.4 Hz, 4H), 2.30 – 2.27 (m, 1H), 2.27 – 2.24 (m, 1H), 2.24 – 2.18 (m, 3H), 2.18 – 2.13 (m, 1H), 2.13 – 2.07 (m, 2H), 2.06 – 2.03 (m, 1H), 2.00

(t, J = 2.7 Hz, 1H), 1.98 – 1.91 (m, 1H), 1.78 (q, J = 3.5 Hz, 1H), 1.76 – 1.74 (m, 2H), 1.72 (d, J = 4.1 Hz, 2H), 1.69 (t, J = 4.3 Hz, 1H), 1.61 – 1.55 (m, 5H), 1.55 – 1.46 (m, 3H), 1.34 (d, J = 2.9 Hz, 3H), 1.33 (d, J = 2.9 Hz, 3H), 1.30 – 1.28 (m, 9H), 1.28 (s, 4H), 1.25 (s, 2H), 1.13 (s, 3H), 1.11 (s, 1H), 1.10 (s, 3H), 1.09 – 1.05 (m, 2H), 1.05 – 1.04 (m, 1H), 1.02 (s, 3H), 1.01 (s, 3H), 0.91 (s, 9H), 0.90 (s, 9H), 0.89 (s, 18H), 0.12 (s, 3H), 0.11 (s, 3H), 0.09 (s, 3H), 0.08 (s, 3H), 0.06 (s, 12H).

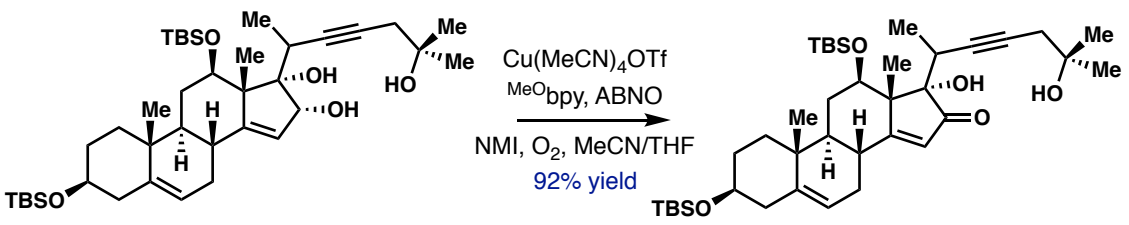
^{13}C NMR (101 MHz, $CDCl_3$): δ 154.39, 154.16, 140.67, 140.64, 122.89, 122.18, 120.82, 120.76, 86.75, 85.79, 82.71, 82.48, 80.21, 79.43, 78.05, 76.43, 73.57, 72.54, 72.52, 72.45, 70.35, 70.09, 58.06, 56.66, 47.89, 47.48, 42.70, 37.12, 37.09, 37.00, 34.62, 34.34, 33.82, 33.25, 32.07, 32.05, 30.96, 30.23, 30.17, 30.13, 29.24, 29.10, 28.98, 28.88, 28.82, 28.78, 26.33, 26.12, 26.08, 19.32, 19.31, 19.01, 18.44, 18.40, 18.21, 17.38, 14.88, 14.49, -3.14, -3.28, -3.65, -3.92, -4.44.

FTIR (NaCl, thin film, cm^{-1}): 3382, 2932, 2854, 1704, 1470, 1462, 1379, 1252, 1098, 1084 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{39}H_{67}O_5Si_2$: 671.4527 $[M+H-H_2]^+$; found: 671.4547.

$[\alpha]_D^{21}$: –91° (c = 0.200, $CHCl_3$).

Preparation of α-hydroxy enone 84:



A 500 mL round-bottom flask with a stir bar was charged with triol **83** (3.15 g, 4.69 mmol, 1:1 mixture of diastereomers at C20), $Cu(MeCN)_4OTf$ (177 mg, 0.47 mmol, 10 mol %),

^{MeO}bpy (102 mg, 0.47 mmol, 10 mol %), ABNO (65.9 mg, 0.47 mmol, 10 mol %), NMI (74.9 μL, 0.94 mmol, 20 mol %) and MeCN/THF (3:1, 160 mL, 0.03 M). The resulting solution was sparged with O_2 for 5 min, then vigorously stirred (1200 rpm) at room temperature under O_2 atmosphere for 12 h. The reaction was passed through a short pad of SiO_2 using 50% EtOAc/hexanes as the eluent and concentrated in vacuo. The crude product was purified by flash column chromatography (15 to 25 to 35% EtOAc/hexanes) to afford α-hydroxy enone **84** (2.89 g, 4.31 mmol, 1:1 mixture of diastereomers at C20, 92% yield) as a colorless foam.

TLC (30% EtOAc/hexanes): R_f = 0.47 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.95 (dd, *J* = 10.6, 1.5 Hz, 2H), 5.42 – 5.37 (m, 2H), 3.95 (ddd, *J* = 11.1, 5.1, 4.0 Hz, 2H), 3.58 (s, 1H), 3.55 – 3.41 (m, 4H), 3.01 (qt, *J* = 7.1, 2.4 Hz, 1H), 2.83 – 2.75 (m, 1H), 2.53 (qd, *J* = 11.3, 4.9 Hz, 2H), 2.39 – 2.31 (m, 2H), 2.29 (d, *J* = 2.4 Hz, 3H), 2.25 (q, *J* = 2.4 Hz, 5H), 2.21 (dd, *J* = 5.6, 1.9 Hz, 1H), 2.18 – 2.05 (m, 4H), 1.85 – 1.80 (m, 2H), 1.78 (d, *J* = 5.1 Hz, 3H), 1.76 – 1.71 (m, 5H), 1.71 – 1.68 (m, 2H), 1.67 – 1.63 (m, 3H), 1.61 (d, *J* = 3.6 Hz, 3H), 1.59 – 1.56 (m, 3H), 1.56 – 1.50 (m, 3H), 1.47 (s, 6H), 1.42 (d, *J* = 7.0 Hz, 4H), 1.32 – 1.29 (m, 6H), 1.28 – 1.27 (m, 3H), 1.25 (s, 9H), 1.24 (s, 6H), 1.21 (s, 4H), 1.20 (s, 4H), 1.18 – 1.16 (m, 2H), 1.15 – 1.12 (m, 2H), 1.11 (d, *J* = 1.5 Hz, 6H), 1.10 – 1.08 (m, 2H), 1.08 – 1.04 (m, 3H), 1.01 (d, *J* = 1.6 Hz, 1H), 0.91 (d, *J* = 1.4 Hz, 3H), 0.91 (s, 6H), 0.90 (s, 3H), 0.89 (s, 24H), 0.88 (d, *J* = 1.2 Hz, 3H), 0.10 (d, *J* = 1.7 Hz, 9H), 0.08 (s, 3H), 0.06 (d, *J* = 2.3 Hz, 15H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 206.85, 206.60, 187.25, 185.98, 140.79, 140.70, 125.25, 124.89, 119.90, 119.84, 85.54, 84.22, 82.25, 81.02, 80.90, 77.35, 77.03, 76.71, 76.20,

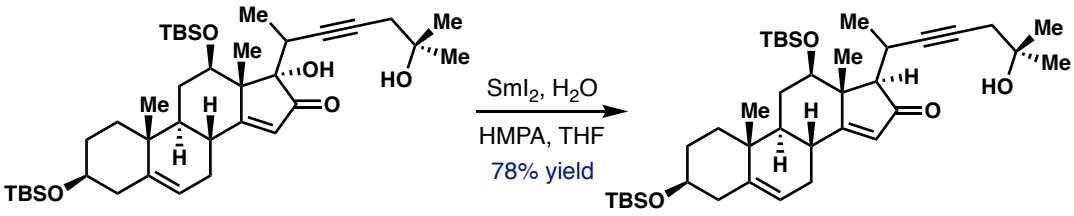
75.50, 72.16, 72.14, 70.02, 69.91, 55.74, 55.10, 48.26, 48.06, 42.50, 37.66, 37.07, 34.58, 33.31, 32.38, 32.24, 32.08, 31.89, 30.17, 30.02, 29.72, 28.61, 28.58, 28.55, 28.35, 28.12, 25.94, 25.92, 25.90, 19.34, 18.26, 18.13, 18.08, 15.84, 13.09, 12.19, -3.06, -3.37, -4.29, -4.53, -4.57, -4.60.

FTIR (NaCl, thin film, cm^{-1}): 3452, 2956, 2930, 2857, 1698, 1608, 1472, 1463, 1372, 1253, 1093 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{39}H_{67}O_5Si_2$: 671.4527 $[M+H]^+$; found: 671.4513.

$[\alpha]_D^{21}$: +24° (*c* = 0.222, $CHCl_3$).

Preparation of enone 85:



In a nitrogen-filled glovebox, a 500 mL round-bottom flask with a stir bar was charged with α-hydroxy enone **84** (1.90 g, 2.84 mmol, 1:1 mixture of diastereomers at C20), HMPA (4.94 mL, 28.4 mmol, 10 equiv), H_2O (0.1 M in THF, 28.4 mL, 2.84 mmol, 1.0 equiv) and THF (115 mL, 0.025 M). SmI_2 solution (0.1 M in THF, 56.8 mL, 5.68 mmol, 2.0 equiv, see preparation procedure in preparation of **78**) was added dropwise over 20 min at room temperature. After stirring an additional 30 min, the reaction was diluted with EtOAc (50 mL), quenched with st. $NaHCO_3$ aq. (50 mL), st. $Na_2S_2O_3$ aq. (50 mL), H_2O (50 mL). The layers were separated. The aqueous layer was further extracted with EtOAc (3 × 50 mL), combined organic layers were washed with brine (2 × 50 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column

chromatography (15 to 20 to 25% EtOAc/hexanes) to afford enone **85** (1.45 g, 2.22 mmol, 1:1 mixture of diastereomers at C20, 78% yield) as a colorless foam.

TLC (25% EtOAc/hexanes): R_f = 0.49 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.85 (d, *J* = 1.7 Hz, 1H), 5.81 (d, *J* = 1.8 Hz, 1H), 5.41 – 5.35 (m, 2H), 3.48 (tt, *J* = 10.7, 5.0 Hz, 3H), 3.36 – 3.29 (m, 1H), 3.29 – 3.24 (m, 1H), 3.12 – 3.05 (m, 1H), 2.83 – 2.77 (m, 1H), 2.77 – 2.63 (m, 1H), 2.54 – 2.46 (m, 2H), 2.43 (d, *J* = 2.9 Hz, 1H), 2.32 – 2.29 (m, 1H), 2.28 (s, 3H), 2.26 – 2.25 (m, 1H), 2.25 – 2.20 (m, 6H), 2.20 – 2.14 (m, 2H), 2.12 (d, *J* = 1.9 Hz, 1H), 2.09 – 2.03 (m, 1H), 2.03 – 1.98 (m, 1H), 1.80 (q, *J* = 3.4 Hz, 1H), 1.78 (q, *J* = 2.7, 1.6 Hz, 2H), 1.75 (s, 2H), 1.72 – 1.65 (m, 5H), 1.61 (dd, *J* = 13.3, 4.5 Hz, 1H), 1.57 (s, 1H), 1.54 (d, *J* = 7.2 Hz, 4H), 1.53 – 1.49 (m, 1H), 1.45 – 1.38 (m, 1H), 1.34 (s, 3H), 1.32 (t, *J* = 4.0 Hz, 1H), 1.30 (d, *J* = 1.7 Hz, 3H), 1.29 – 1.27 (m, 1H), 1.26 (s, 3H), 1.25 (d, *J* = 4.3 Hz, 5H), 1.24 (d, *J* = 3.1 Hz, 8H), 1.22 (s, 1H), 1.20 (d, *J* = 4.5 Hz, 1H), 1.19 (s, 3H), 1.17 (dd, *J* = 7.0, 4.6 Hz, 2H), 1.10 (s, 3H), 1.10 (s, 3H), 1.07 (d, *J* = 3.6 Hz, 1H), 1.06 (t, *J* = 2.5 Hz, 1H), 1.04 (d, *J* = 4.7 Hz, 1H), 1.02 (d, *J* = 7.0 Hz, 1H), 0.96 (d, *J* = 6.7 Hz, 1H), 0.92 – 0.90 (m, 13H), 0.89 (s, 9H), 0.89 (s, 21H), 0.07 (s, 3H), 0.05 (d, *J* = 1.7 Hz, 15H), 0.05 (s, 3H), 0.03 (s, 3H).

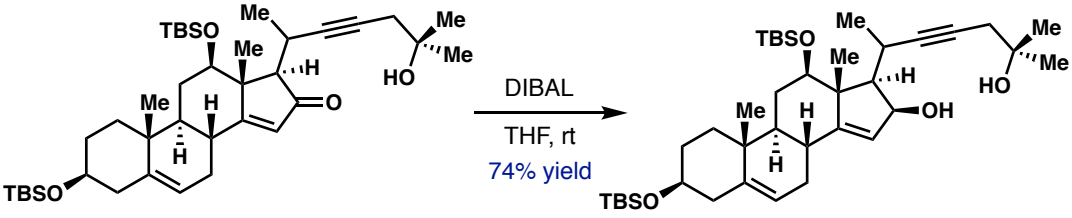
^{13}C NMR (101 MHz, $CDCl_3$): δ 208.86, 208.38, 188.89, 188.16, 141.07, 127.17, 127.02, 120.05, 87.41, 86.47, 80.33, 80.07, 78.89, 78.41, 72.40, 72.38, 70.07, 69.90, 61.07, 60.84, 53.25, 52.76, 48.45, 48.18, 42.66, 37.72, 37.64, 37.12, 34.64, 34.58, 32.13, 32.10, 32.02, 29.85, 29.83, 29.80, 29.15, 29.09, 28.69, 28.65, 28.63, 28.59, 26.06, 26.03, 25.90, 25.39, 25.33, 22.46, 19.46, 18.90, 18.40, 18.14, 18.04, 14.28, 13.33, 12.97, -3.50, -3.86, -4.20, -4.47, -4.59, -4.72.

FTIR (NaCl, thin film, cm^{-1}): 3439, 2949, 2927, 2853, 1694, 1471, 1462, 1374, 1252, 1092 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{39}H_{67}O_4Si_2$: 655.4578 $[M+H]^+$; found: 655.4552.

$[\alpha]_D^{21}$: +42° (*c* = 0.250, $CHCl_3$).

Preparation of allylic alcohol 50:



A 500 mL round-bottom flask with a stir bar was charged with enone **85** (1.82 g, 2.78 mmol, 1.0 equiv) and CH_2Cl_2 (140 mL, 0.02 M). The resulting solution was cooled to 0 °C and stirred at this temperature for 15 min. Thereafter, *i*-Bu_2AlH (1.0 M in CH_2Cl_2, 8.34 mL, 8.34 mmol, 3.0 equiv) was added dropwise. After stirring at 0 °C for 30 min, the reaction was diluted with Et_2O (100 mL), quenched with st. Na/K-tartrate aq. (100 mL) and vigorously stirred at room temperature for 1 h. The layers were separated. The aqueous layer was further extracted with Et_2O (3 × 50 mL), combined organic layers were washed with brine (2 × 50 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (20 to 30 to 40% EtOAc/hexanes) to afford allylic alcohol **50** (1.35 g, 2.06 mmol, 1:1 mixture of diastereomers at C20, 74% yield) as a colorless foam.

TLC (25% EtOAc/hexanes): R_f = 0.29 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.51 (t, *J* = 2.4 Hz, 1H), 5.48 (t, *J* = 2.3 Hz, 1H), 5.37 (dt, *J* = 4.0, 1.7 Hz, 2H), 4.85 (dd, *J* = 6.9, 2.5 Hz, 1H), 4.81 (dd, *J* = 6.5, 2.7 Hz, 1H), 3.47 (tt, *J* = 5.1, 4.7 Hz, 2H), 3.23 (ddd, *J* = 10.6, 8.4, 5.8 Hz, 3H), 3.10 – 3.02 (m, 1H), 2.34 (d, *J* = 2.4 Hz, 2H), 2.32 (d, *J* = 2.2 Hz, 2H), 2.27 – 2.23 (m, 3H), 2.23 – 2.19 (m, 3H), 2.19 – 2.14 (m, 2H), 2.08 (dd, *J* = 6.9, 4.4 Hz, 1H), 1.98 – 1.87 (m, 2H), 1.78 (d, *J* = 3.7 Hz, 1H), 1.75 (dd, *J* = 5.1, 2.9 Hz, 2H), 1.73 – 1.71 (m, 2H), 1.64 – 1.61 (m, 1H), 1.61 – 1.60 (m, 1H), 1.59 (s, 1H), 1.58 (s, 1H), 1.57 – 1.55 (m, 2H), 1.53 (q, *J* = 2.4 Hz, 1H), 1.52 – 1.48 (m, 1H), 1.44 (d, *J* = 7.1 Hz, 3H), 1.41 (d, *J* = 7.2 Hz, 3H), 1.33 – 1.29 (m, 2H), 1.28 – 1.26 (m, 12H), 1.25 (s, 6H), 1.04 (s, 3H), 1.03 (s, 3H), 1.02 (s, 1H), 0.99 (dd, *J* = 7.8, 3.1 Hz, 2H), 0.97 (q, *J* = 5.6 Hz, 3H), 0.91 (s, 9H), 0.91 (s, 9H), 0.89 (s, 18H), 0.09 (s, 3H), 0.05 (d, *J* = 1.7 Hz, 15H), 0.05 (s, 3H), 0.04 (s, 3H).

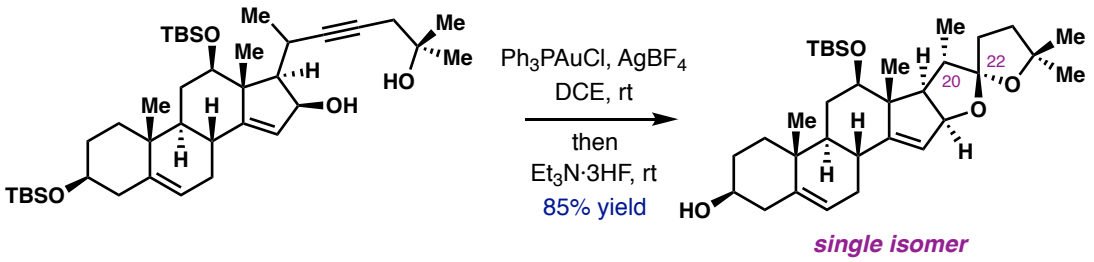
^{13}C NMR (101 MHz, $CDCl_3$): δ 160.33, 159.45, 155.87, 155.03, 140.82, 140.80, 125.73, 123.98, 121.10, 121.06, 89.60, 88.74, 81.86, 81.36, 76.29, 75.97, 72.63, 72.58, 70.08, 70.07, 59.32, 58.83, 56.79, 54.14, 50.11, 48.88, 42.80, 42.73, 37.79, 37.77, 37.24, 37.20, 34.79, 34.74, 32.13, 32.08, 31.72, 31.64, 30.79, 30.33, 29.86, 29.52, 28.93, 28.89, 28.80, 28.68, 28.63, 26.22, 26.20, 26.08, 26.04, 24.88, 22.85, 19.64, 19.35, 18.42, 18.35, 18.29, 14.28, 13.92, 13.20, 1.17, -3.33, -3.63, -3.75, -4.01, -4.26, -4.44.

FTIR (NaCl, thin film, cm^{-1}): 3383, 2957, 1930, 2856, 1714, 1470, 1463, 1372, 1256, 1090 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{39}H_{67}O_4Si_2$: 655.4578 $[M+H-H_2]^+$; found: 655.4571.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: +12° (*c* = 0.083, CH_2Cl_2).

Preparation of spiroketal 86:



A 250 mL round-bottom flask with a stir bar was charged with Ph_3PAuCl (98.9 mg, 0.20 mmol, 10 mol %), $AgBF_4$ (19.5 mg, 0.1 mmol, 5.0 mol %) and $(CH_2Cl)_2$ (65 mL, 0.03 M total) and stirred at room temperature for 10 min. Allylic alcohol **50** (1.30 g, 1.98 mmol, 1.0 equiv) in $(CH_2Cl)_2$ (5 mL, 0.03 M total) was added. After stirring for 20 h, $Et_3N{\cdot}3HF$ (6.33 mL, 39.6 mmol, 20 equiv) was added dropwise and stirred for 5 h at room temperature. The reaction was passed through a short pad of SiO_2 using 50% EtOAc/hexanes as the eluent, diluted with st. $NaHCO_3$ aq. (50 mL) and H_2O (50 mL). The layers were separated. The aqueous layer was further extracted with EtOAc (3 × 50 mL), combined organic layers were washed with brine (2 × 50 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (20 to 25 to 30% EtOAc/hexanes) to afford spiroketal **86** (911 mg, 1.68 mmol, 85% yield) as a colorless foam.

TLC (20% EtOAc/hexanes): R_f = 0.22 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.43 – 5.36 (m, 2H), 4.89 (dd, *J* = 8.2, 2.3 Hz, 1H), 3.51 (tt, *J* = 11.1, 4.7 Hz, 1H), 3.22 (dd, *J* = 10.6, 5.0 Hz, 1H), 2.47 (t, *J* = 8.2 Hz, 1H), 2.33 (ddd, *J* = 13.0, 5.1, 2.2 Hz, 1H), 2.25 (dt, *J* = 11.0, 2.3 Hz, 1H), 2.21 – 2.11 (m, 2H), 2.09 (td, *J* = 4.9, 2.4 Hz, 1H), 2.06 – 1.98 (m, 4H), 1.97 – 1.91 (m, 1H), 1.89 – 1.83 (m, 1H),

1.79 (dt, *J* = 13.1, 3.4 Hz, 1H), 1.69 (dd, *J* = 6.2, 3.8 Hz, 1H), 1.67 – 1.64 (m, 1H), 1.63 (dd, *J* = 10.2, 2.0 Hz, 1H), 1.59 – 1.57 (m, 1H), 1.54 – 1.48 (m, 1H), 1.36 (s, 3H), 1.17 (s, 3H), 1.10 (d, *J* = 3.3 Hz, 1H), 1.06 (s, 3H), 1.05 – 1.02 (m, 1H), 1.02 – 0.97 (m, 6H), 0.96 (d, *J* = 6.5 Hz, 1H), 0.91 (s, 9H), 0.08 (s, 3H), 0.05 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 157.02, 139.98, 121.54, 120.63, 117.51, 84.31, 81.98, 79.62, 71.76, 55.96, 53.61, 49.16, 42.25, 41.86, 37.48, 37.29, 37.13, 33.63, 31.64, 30.73, 30.33, 30.12, 29.08, 28.60, 26.13, 19.37, 18.20, 14.49, 13.88, -3.61, -4.05.

FTIR (NaCl, thin film, cm^{-1}): 3392, 2957, 2930, 2857, 1471, 1462, 1373, 1251, 1099, 1070 cm^{-1}.

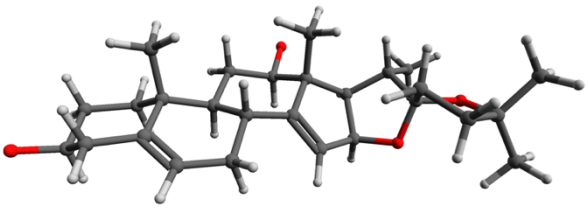
HRMS (FAB+, m/z): calc'd for $C_{33}H_{55}O_4Si$: 543.3870 $[M+H]^+$; found: 543.3893.

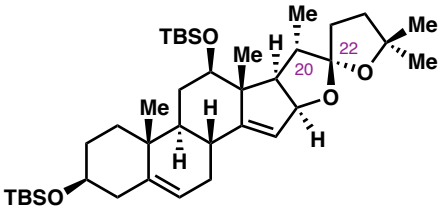
$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: –67° (*c* = 0.167, $CHCl_3$).

Minimized Structures and energies:

87

Single-point energy = -2406.101814191406 (Hartree)
N_atoms = 113
C -4.2612945 3.1439595 3.605928
C -3.1775145 2.161998 3.6497575
C -2.338009 2.237628 2.4354575
C -1.126178 2.968367 2.8920965
C -0.037664 2.0969355 2.5738055
C -0.4030685 1.035947 1.9668315
C 0.4228185 0.038453 1.350371
C 1.851007 0.2410315 1.577091
C 2.665246 -0.56034 0.721119
C 2.2599615 -1.1731 -0.321583
C 3.177214 -1.9667585 -1.1022185
C 3.162845 -1.604117 -2.52126
O 3.9388865 -2.449306 -3.2554915
Si 5.498797 -2.3265755 -3.601855





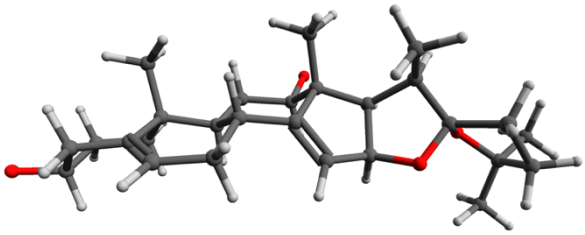
C	5.948927	-0.6085685	-3.9029365
C	6.507916	-2.9591545	-2.251198
C	5.7003365	-3.349371	-5.100168
C	4.913206	-2.7767915	-6.202072
C	7.115439	-3.397334	-5.4962875
C	5.2272305	-4.716723	-4.8364805
C	1.802126	-1.673919	-3.0297045
C	0.895728	-0.8317675	-2.2519365
C	0.879103	-1.1122685	-0.7921685
C	0.233898	-2.419202	-0.5490745
C	0.145802	-0.001614	-0.1075965
C	-1.2964885	0.0118315	-0.397269
C	-2.0391985	1.06258	0.3019835
O	-3.3672285	0.969084	-0.001853
Si	-4.156109	1.752503	-1.1604735
C	-4.154989	3.5223545	-0.828051
C	-3.391661	1.4974415	-2.772778
C	-5.8360395	1.0440185	-1.1016395
C	-6.441669	1.2780045	0.217749
C	-6.6797145	1.6665245	-2.1331015
C	-5.768712	-0.4049455	-1.3417285
C	-1.8455	0.982835	1.773467
C	-2.4439085	-0.250586	2.296286
O	-1.266547	3.235874	4.2289665
C	-2.2433055	2.421503	4.763641
C	-1.6318575	1.2687905	5.4323115
C	-1.155484	1.8676815	6.6755045
C	-2.1643135	2.907296	6.9664445
C	-1.567475	4.181009	7.3447035
C	-3.1626765	2.4701245	7.9341685
O	-2.840873	3.0931115	5.758497
H	-4.890132	2.9431275	2.8007305
H	-3.886739	4.1088405	3.4987295
H	-4.820111	3.119279	4.480334
H	-3.599564	1.2204305	3.7674535
H	-2.851029	2.7934	1.724306
H	-1.003927	3.899228	2.433875
H	0.9351425	2.3586905	2.796967
H	0.168064	-0.895751	1.746625
H	2.084343	0.04763	2.5765835
H	2.091115	1.248897	1.4282495
H	3.6665	-0.631848	0.980413
H	2.9352275	-2.9823115	-1.0481915
H	4.139768	-1.8719475	-0.716689
H	3.5105955	-0.6212135	-2.6202985
H	5.359258	-0.1718285	-4.641085

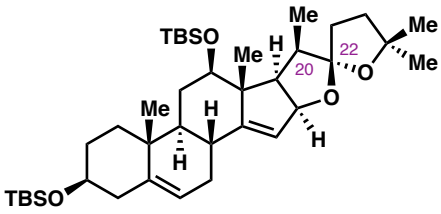
H 6.9415875 -0.540396 -4.2118435
H 5.8574045 -0.029555 -3.041828
H 6.3913755 -2.3850715 -1.3899405
H 7.520687 -2.945844 -2.496078
H 6.2527935 -3.939304 -2.011547
H 3.902499 -2.7207285 -5.9605395
H 5.2369505 -1.8184345 -6.4520035
H 5.003673 -3.367307 -7.0593855
H 7.715734 -3.824553 -4.759532
H 7.4977375 -2.450832 -5.7077085
H 7.2315895 -3.9743015 -6.359603
H 5.3201915 -5.304496 -5.6954565
H 4.227338 -4.72542 -4.5492845
H 5.7801645 -5.180051 -4.084546
H 1.4986315 -2.6715715 -3.0110535
H 1.789017 -1.378952 -4.0291855
H -0.0688435 -0.929708 -2.6329245
H 1.169036 0.1670185 -2.391283
H -0.689627 -2.4814635 -1.023815
H 0.0886605 -2.586492 0.467522
H 0.8150595 -3.2028295 -0.9059045
H 0.542161 0.8880405 -0.4913315
H -1.731757 -0.9002995 -0.1532465
H -1.4464755 0.137889 -1.418691
H -1.667219 1.9951145 0.00219
H -4.6748265 4.0431965 -1.565762
H -4.591445 3.749786 0.0887715
H -3.1841275 3.900505 -0.81366
H -3.9215535 1.9995125 -3.5164305
H -3.351356 0.491718 -3.036678
H -2.4181295 1.868029 -2.7903605
H -6.546272 2.293856 0.4245895
H -7.3960865 0.854645 0.2567
H -5.8759075 0.852813 0.9806505
H -6.781957 2.6936255 -1.9891665
H -6.3038965 1.514976 -3.0930615
H -7.640708 1.257107 -2.112223
H -5.1770425 -0.881394 -0.6307495
H -5.3792965 -0.6283295 -2.2823085
H -6.7233885 -0.8277365 -1.2963575
H -2.0575405 -1.0961895 1.832548
H -3.470172 -0.2522555 2.1344775
H -2.267392 -0.355883 3.3146135
H -0.8767095 0.8348725 4.8679075
H -2.3609855 0.5490025 5.625353
H -0.224557 2.304656 6.5219135

H -1.0683175 1.1826655 7.4529325
H -2.302247 4.902396 7.495953
H -1.017913 4.0862355 8.2253735
H -0.935361 4.510628 6.588847
H -3.5601575 1.5457125 7.6681095
H -2.7465385 2.386969 8.8853545
H -3.944365 3.15533 7.9786565

88

Single-point energy = -2406.095200694148 (Hartree)
N_atoms = 113
C 4.165177 3.0501625 1.713936
C 3.3047445 2.409058 2.7093075
C 2.1830125 1.5466105 2.235926
C 1.0977405 2.531394 2.0246775
C 0.934933 2.5771555 0.60687
C 1.518736 1.598232 0.0272995
C 1.3351735 1.1141125 -1.307094
C 0.677097 2.062186 -2.2000875
C 0.213455 1.45139 -3.4041515
C 0.0897705 0.1959845 -3.5930085
C -0.382162 -0.3063675 -4.859688
C -1.5199385 -1.2140155 -4.714311
O -1.83117 -1.7220925 -5.941544
Si -3.277002 -1.9114975 -6.607187
C -3.0118195 -3.1727675 -7.8537835
C -4.4705165 -2.4948565 -5.3899765
C -3.853194 -0.346217 -7.367844
C -5.15604 -0.555103 -8.0181545
C -2.881669 0.1035305 -8.375003
C -4.001034 0.7028065 -6.3483205
C -1.1770745 -2.2849475 -3.7886385
C -0.7502655 -1.747106 -2.4974415
C 0.393527 -0.799501 -2.56763
C 1.61898 -1.527351 -2.955478
C 0.542996 -0.1432985 -1.23022
C 1.0425765 -1.0552435 -0.1857525
C 1.3193565 -0.4342975 1.1150865
O 1.8814145 -1.36624 1.940363
Si 1.1637395 -2.1586965 3.137516
C -0.168945 -3.1988955 2.513372
C 0.4569945 -1.0198085 4.339836
C 2.5115755 -3.157736 3.8552445
C 3.6051505 -2.278235 4.292818

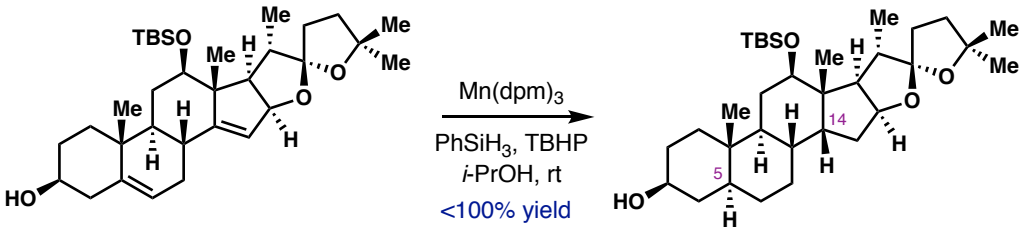




C 3.0326525 -4.0906515 2.8446695
C 2.0212995 -3.9237055 5.0105655
C 2.2185295 0.7470215 0.9795305
C 3.552609 0.306752 0.553624
O 1.498343 3.720257 2.594294
C 2.529238 3.424166 3.4400755
C 3.129114 4.6782045 3.8691695
C 2.2245685 5.107771 4.934098
C 1.720437 3.842531 5.5136085
C 0.2688195 3.8323085 5.6486895
C 2.3780685 3.499745 6.768162
O 2.084254 2.8579195 4.597344
H 4.8292765 2.3760075 1.291559
H 4.738239 3.7949275 2.1613515
H 3.6098065 3.488658 0.952997
H 3.8965665 1.88304 3.384562
H 1.956314 0.9195845 3.029925
H 0.202928 2.2480255 2.487883
H 0.3217055 3.2710085 0.15202
H 2.27009 0.881862 -1.713657
H 1.3247715 2.8507205 -2.4215785
H -0.1360585 2.501092 -1.7092
H -0.047243 2.095925 -4.173087
H 0.3650825 -0.847205 -5.3508105
H -0.6383655 0.4840945 -5.4867795
H -2.3289825 -0.677162 -4.3226655
H -2.2364565 -2.9116975 -8.496763
H -2.7549145 -4.075907 -7.4039465
H -3.863049 -3.3373845 -8.430687
H -5.4004875 -2.662843 -5.828346
H -4.61421 -1.8193755 -4.610655
H -4.156931 -3.3954845 -4.971456
H -5.89116 -0.846697 -7.3397065
H -5.484407 0.3312435 -8.46389
H -5.112386 -1.276808 -8.768593
H -2.7790195 -0.5839845 -9.1509985
H -1.9381825 0.253802 -7.961086
H -3.1879675 1.0070815 -8.801774
H -4.340848 1.59013 -6.783914
H -3.097942 0.9203335 -5.877459
H -4.6901845 0.4407325 -5.611719
H -0.425762 -2.863675 -4.221853
H -1.998064 -2.912836 -3.6545885
H -1.558779 -1.246481 -2.065037
H -0.511839 -2.533644 -1.8573405
H 1.762073 -2.3735705 -2.367685

H 1.5782045 -1.8440925 -3.943909
H 2.4611755 -0.9228625 -2.865755
H -0.417521 0.158847 -0.946046
H 0.3554715 -1.821242 -0.0271155
H 1.9183685 -1.5202885 -0.496489
H 0.420139 -0.090288 1.525379
H -0.9199715 -2.6263095 2.073017
H -0.6119725 -3.7410465 3.285346
H 0.16699 -3.8792175 1.8011975
H 0.008975 -1.54521 5.120425
H 1.1669925 -0.3805765 4.7519705
H -0.282121 -0.4255655 3.9086175
H 4.384925 -2.842694 4.699917
H 3.299193 -1.6042275 5.0265325
H 3.9928615 -1.7282095 3.499426
H 3.8275665 -4.6435005 3.237659
H 3.3824895 -3.5855545 2.004807
H 2.308498 -4.771177 2.5311465
H 1.6675595 -3.305478 5.7711555
H 1.2549685 -4.5817405 4.753097
H 2.7945525 -4.497624 5.4157755
H 3.511798 -0.459547 -0.14653
H 4.087601 1.0828045 0.1221405
H 4.0917335 -0.0548525 1.3650065
H 3.2054535 5.3632625 3.0936505
H 4.083182 4.4910825 4.242628
H 2.695679 5.6998515 5.6478295
H 1.44034 5.6592265 4.531202
H -0.1704975 4.0162405 4.725151
H -0.061072 2.9082135 5.995801
H -0.0482465 4.56342 6.320653
H 2.120763 2.53511 7.0616985
H 3.413302 3.539654 6.6681905
H 2.0969515 4.156803 7.5260445

Preparation of C3 alcohol 89:



A 100 mL round-bottom flask with a stir bar was charged with spiroketal **86** (147 mg, 0.27 mmol, 1.0 equiv) and adventitious water was removed via azeotropic drying with PhMe (3 x 10 mL) under high vacuum. Thereafter, *i*-PrOH (freshly distilled over CaH_2, 10.8 mL, 0.025 M), $PhSiH_3$ (200 μL, 1.62 mmol, 6.0 equiv), and TBHP solution (5.5 M in nonane, 197 μL, 1.08 mmol, 4.0 equiv) were added. The mixture was subjected to three freeze-pump-thaw cycles. Thereafter, $Mn(dpm)_3$ (82 mg, 0.135 mmol, 50 mol %) was added in one portion. After stirring at room temperature for 6 h, the reaction was passed through a short pad of SiO_2 using 50% EtOAc/hexanes as the eluent and concentrated in vacuo. The crude product was purified by flash column chromatography (15 to 25 to 35% EtOAc/hexanes) to afford C3 alcohol **89** (<147 mg, <0.27 mmol, <100 % yield) as a colorless foam.[24] This material contained an excess $PhSiH_3$ and could be carried through the next step where a two-step yield was recorded. It could be exhaustively purified by multiple rounds of chromatography for analytical characterization.

TLC (20% EtOAc/hexanes): R_f = 0.25 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 4.45 (t, *J* = 6.9 Hz, 1H), 3.58 (tt, *J* = 11.0, 4.7 Hz, 1H), 3.29 (dd, *J* = 11.5, 4.5 Hz, 1H), 2.40 (dd, *J* = 9.2, 7.4 Hz, 1H), 2.01 – 1.96 (m, 1H), 1.96 – 1.87 (m, 3H), 1.86 – 1.73 (m, 4H), 1.73 – 1.70 (m, 1H), 1.70 – 1.68 (m, 1H), 1.67 (d, *J* = 1.7 Hz, 1H), 1.66 – 1.61 (m, 2H), 1.57 – 1.54 (m, 2H), 1.50 (dd, *J* = 12.7, 3.5 Hz, 2H), 1.46 – 1.40 (m, 1H), 1.40 – 1.36 (m, 1H), 1.32 (s, 3H), 1.30 (d, *J* = 6.4 Hz, 1H), 1.28 (dd, *J* = 6.8, 2.3 Hz, 2H), 1.26 – 1.22 (m, 3H), 1.20 (d, *J* = 2.1 Hz, 1H), 1.14 (s, 3H), 1.10 – 1.03 (m, 2H), 1.01 (t, *J* = 3.9 Hz, 1H), 0.99 – 0.96 (m, 1H), 0.96 – 0.91 (m, 5H), 0.89 (d, *J* = 3.3 Hz, 14H), 0.80 (s, 3H), 0.04 (s, 3H), 0.02 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 116.66, 81.72, 79.61, 73.97, 71.47, 56.96, 48.70, 47.61, 45.52, 44.54, 41.93, 38.25, 37.50, 37.07, 35.51, 33.37, 32.56, 32.46, 32.07, 31.48, 30.56, 30.20, 29.00, 28.63, 26.11, 18.21, 14.50, 13.69, 12.33, -3.26, -4.65.

FTIR (NaCl, thin film, cm^{-1}): 3398, 2926, 2858, 1470, 1462, 1455, 1382, 1366, 1255, 1085 cm^{-1}.

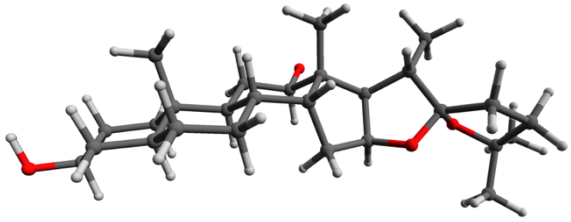
HRMS (FAB+, m/z): calc'd for $C_{33}H_{59}O_4Si$: 547.4183 $[M+H]^+$; found: 547.4181.

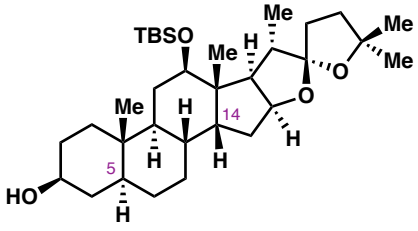
$[\alpha]_D^{21}$: –39° (*c* = 0.417, $CHCl_3$).

Minimized Structures and energies:

89

Single-point energy = -1882.063903502656 (Hartree)
N_atoms = 96
C 3.944346 0.312602 1.635718
C 3.307516 -0.4003555 0.539361
C 1.969351 0.0965405 0.162229
C 1.913441 -0.140255 -1.3051875
C 0.818017 -1.0723625 -1.5609195
C 0.5115085 -1.679696 -0.255265
C -0.803167 -2.328972 -0.118547
C -1.000564 -3.367881 -1.1331795
C -2.3053335 -4.0115085 -1.0236705
C -3.376423 -3.017425 -1.0604525
C -4.709759 -3.620338 -1.142073
C -5.7504075 -2.6129895 -1.3195475
O -6.9958975 -3.1614765 -1.4100405
C -5.672072 -1.592395 -0.2797915
C -4.329621 -1.0143335 -0.1661635
C -3.2571425 -2.0137195 0.035788
C -3.415332 -2.6232355 1.3660945
C -1.916405 -1.3584655 -0.0993295
C -1.654359 -0.3157245 0.9022815
C -0.3695275 0.3499545 0.6980205
O -0.1524575 1.31983 1.633433
Si -0.8056265 2.7625325 1.8804815



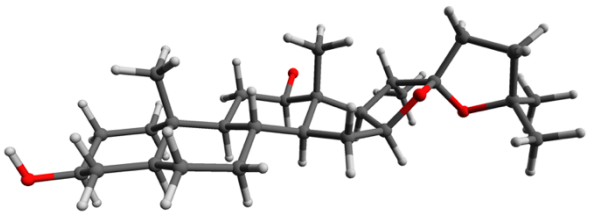


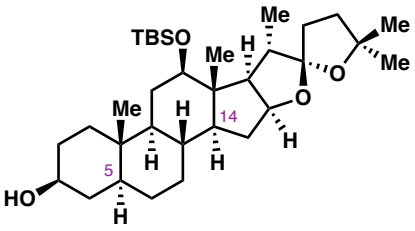
C	0.3526705	3.5221375	3.020832
C	-2.40027	2.6107335	2.7076945
C	-0.9769805	3.766546	0.351296
C	-1.323271	5.1447445	0.7401295
C	-2.043638	3.253464	-0.5209805
C	0.2863275	3.797662	-0.39945
C	0.776538	-0.5950615	0.728516
C	0.973571	-1.105086	2.084945
O	3.1066795	-0.7335845	-1.647204
C	4.0425055	-0.3553445	-0.7328265
C	5.217442	-1.2211035	-0.9083805
C	6.294254	-0.3244075	-1.3409265
C	5.602576	0.9116635	-1.747775
C	5.2159765	0.8949055	-3.1589645
C	6.340157	2.120509	-1.419949
O	4.47055	0.9175075	-0.948791
H	4.8918185	-0.0717345	1.8398205
H	4.054853	1.317962	1.3990485
H	3.3831535	0.237814	2.5078035
H	3.243239	-1.4133205	0.78023
H	1.919324	1.110536	0.384199
H	1.816006	0.7456915	-1.847319
H	1.1040865	-1.769905	-2.2766045
H	-0.0027755	-0.559142	-1.9462405
H	1.2289135	-2.422803	-0.09086
H	-0.7930175	-2.8068645	0.8115895
H	-0.235895	-4.075433	-1.0660275
H	-0.9217475	-2.9410185	-2.0829975
H	-2.4344075	-4.691833	-1.8042085
H	-2.3625865	-4.5644815	-0.1402655
H	-3.2470715	-2.475238	-1.9488365
H	-4.7449385	-4.292837	-1.9382685
H	-4.9110385	-4.1787525	-0.2805445
H	-5.6104955	-2.1545175	-2.247821
H	-7.158658	-3.5788125	-0.6015195
H	-5.9572725	-2.0202245	0.6307885
H	-6.368916	-0.8439225	-0.4799665
H	-4.328377	-0.3198375	0.610764
H	-4.125317	-0.4745325	-1.038241
H	-3.4165595	-1.893632	2.1079155
H	-2.6575535	-3.2963795	1.595842
H	-4.3130495	-3.1391765	1.4539025
H	-1.935744	-0.8885485	-1.0357095
H	-1.671601	-0.7163825	1.8629795
H	-2.4031345	0.4057935	0.8712275
H	-0.386264	0.8009835	-0.2450735

H 0.0098585 4.441291 3.369474
H 0.491802 2.903456 3.8467315
H 1.284919 3.6718015 2.583347
H -2.3023655 2.0202105 3.5600225
H -3.1341675 2.1787385 2.1099545
H -2.7542425 3.5422955 3.012745
H -1.4437575 5.740012 -0.110387
H -0.580194 5.585605 1.3211465
H -2.217193 5.192658 1.273684
H -2.966212 3.226555 -0.0381345
H -2.154418 3.86721 -1.359848
H -1.843366 2.293627 -0.8726485
H 0.5711675 2.8529575 -0.731216
H 1.0704655 4.178204 0.1710655
H 0.1918515 4.408824 -1.2418265
H 0.138821 -1.622803 2.4276005
H 1.774363 -1.7671955 2.1250255
H 1.149023 -0.3290195 2.7518925
H 5.454262 -1.712358 -0.023205
H 4.9899995 -1.94389 -1.6187815
H 6.939299 -0.119452 -0.549741
H 6.8644865 -0.724375 -2.113465
H 4.631229 1.72754 -3.376963
H 4.6596235 0.046946 -3.3816155
H 6.0582975 0.9177375 -3.7720445
H 7.231908 2.16526 -1.955958
H 5.780846 2.9664135 -1.653188
H 6.5635645 2.14666 -0.404464

94

Single-point energy = -1882.0604906961223 (Hartree)
N_atoms = 96
C 3.5722425 1.118087 1.4307455
C 2.952779 -0.1684775 1.1421925
C 2.074434 -0.176138 -0.0466905
C 2.6422945 -1.2551875 -0.8963405
C 1.5305195 -1.957912 -1.53796
C 0.391969 -1.0642045 -1.303132
C -0.963165 -1.5485285 -1.5538325
C -1.1134325 -2.102343 -2.9015645
C -2.4972815 -2.4472585 -3.213877
C -3.3855865 -1.306962 -3.0003935
C -4.752695 -1.551732 -3.468103

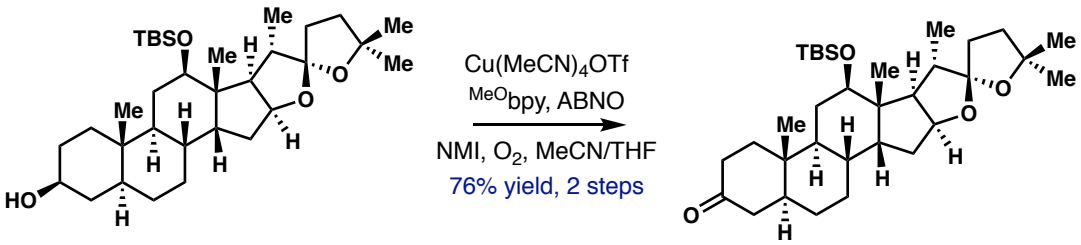




C -5.572809 -0.3470265 -3.3854935
O -6.8421005 -0.5468435 -3.8416765
C -5.5502265 0.202599 -2.034588
C -4.184652 0.413586 -1.5453305
C -3.3302465 -0.793485 -1.600991
C -3.821672 -1.775848 -0.6225175
C -1.9061645 -0.4290655 -1.3224595
C -1.6819995 0.1805835 -0.0021315
C -0.2957165 0.602557 0.2384105
O -0.191551 1.134971 1.493911
Si -0.2180235 2.659803 1.9929805
C 0.638165 3.728582 0.8219565
C 0.664457 2.5957165 3.555619
C -1.914292 3.3130315 2.2857885
C -2.7376265 2.3093695 2.9740575
C -1.816745 4.5088725 3.141301
C -2.574006 3.7091735 1.033025
C 0.628142 -0.531886 0.0686935
C 0.4134865 -1.541518 1.114901
O 3.313373 -2.0739425 -0.0231905
C 3.9327675 -1.249965 0.8733455
C 4.39893 -2.057566 1.989582
C 5.6610885 -2.600665 1.4932725
C 6.1482305 -1.5754445 0.542301
C 6.52762 -2.148384 -0.7431065
C 7.216124 -0.751898 1.095119
O 5.063463 -0.7222825 0.3401485
H 2.8487215 1.827705 1.660521
H 4.219029 1.051997 2.24449
H 4.1269075 1.458895 0.621165
H 2.407685 -0.4638755 1.976757
H 2.1435265 0.7366545 -0.546516
H 3.3228045 -0.8729375 -1.588984
H 1.393034 -2.883051 -1.080305
H 1.7032765 -2.1309035 -2.5486245
H 0.5272055 -0.2460775 -1.9463195
H -1.171694 -2.312882 -0.87365
H -0.502002 -2.9392445 -3.013498
H -0.776434 -1.397199 -3.597467
H -2.567375 -2.75951 -4.207298
H -2.8066845 -3.259257 -2.634604
H -3.008309 -0.529247 -3.594474
H -4.735727 -1.886766 -4.455404
H -5.2000925 -2.312752 -2.907023
H -5.1784265 0.364612 -4.040687
H -7.2250165 -1.17965 -3.2873265

H -6.078497 1.1007745 -2.0179755
H -6.06339 -0.4499315 -1.398691
H -3.7384465 1.164061 -2.120789
H -4.229812 0.780255 -0.571422
H -3.2158345 -2.617847 -0.5554155
H -4.778788 -2.1125805 -0.848506
H -3.868412 -1.353455 0.327448
H -1.662188 0.299177 -2.0369295
H -1.946957 -0.482376 0.7552895
H -2.2933075 1.0130665 0.1112715
H -0.0339425 1.3293425 -0.4692355
H 0.6282505 4.7163425 1.153779
H 0.1987725 3.717756 -0.121837
H 1.633979 3.4486105 0.7085145
H 0.8153305 3.5457725 3.953401
H 0.1426185 2.036359 4.261951
H 1.59652 2.151314 3.4324155
H -3.6848365 2.6964745 3.187111
H -2.8673855 1.452331 2.3988925
H -2.3142525 2.016396 3.8801035
H -1.223611 5.2525905 2.714623
H -2.7651 4.9220325 3.288575
H -1.4263455 4.29111 4.081263
H -2.684157 2.9132125 0.3710745
H -2.0486975 4.4598185 0.5378385
H -3.5315085 4.0794225 1.2284365
H -0.493894 -2.037716 1.0121505
H 1.164511 -2.2578245 1.1010615
H 0.4077855 -1.0890165 2.0501415
H 3.7079565 -2.7866765 2.2535695
H 4.5520975 -1.43972 2.81409
H 6.34852 -2.780232 2.252973
H 5.493048 -3.500268 0.99924
H 5.73072 -2.682459 -1.1426645
H 6.7939545 -1.3996465 -1.41537
H 7.341181 -2.791204 -0.636214
H 7.4259345 0.0407435 0.4547385
H 6.940649 -0.3562405 2.0173865
H 8.0870205 -1.3086895 1.222976

Preparation of ketone 48:



A 100 mL round-bottom flask with a stir bar was charged with C3 alcohol **89** (<147 mg, <0.27 mmol, 1.0 equiv), $Cu(MeCN)_4OTf$ (10.1 mg, 0.027 mmol, 10 mol %), ^{MeO}bpy (5.84 mg, 0.027 mmol, 10 mol %), ABNO (3.79 mg, 0.027 mmol, 10 mol %), NMI (4.30 μL, 0.054 mmol, 20 mol %) and MeCN/THF (3:1, 18.0 mL, 0.015 M). The resulting solution was sparged with O_2 for 5 min, then vigorously stirred (1200 rpm) at room temperature under O_2 atmosphere for 12 h. The reaction was passed through a short pad of SiO_2 using 50% EtOAc/hexanes as the eluent and concentrated in vacuo. The crude product was purified by flash column chromatography (10 to 15 to 20% EtOAc/hexanes) to afford eastern fragment ketone **48** (112 mg, 0.21 mmol, 76% yield over 2 steps) as a colorless foam.

TLC (20% EtOAc/hexanes): R_f = 0.42 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 4.45 (t, *J* = 6.9 Hz, 1H), 3.32 (dd, *J* = 11.4, 4.5 Hz, 1H), 2.44 – 2.35 (m, 2H), 2.34 – 2.31 (m, 1H), 2.30 – 2.21 (m, 1H), 2.10 (ddd, *J* = 15.2, 4.1, 2.1 Hz, 1H), 2.00 (ddt, *J* = 5.7, 4.1, 2.0 Hz, 1H), 1.98 – 1.96 (m, 1H), 1.96 – 1.88 (m, 3H), 1.85 (td, *J* = 7.9, 3.8 Hz, 1H), 1.80 – 1.75 (m, 1H), 1.69 (dd, *J* = 4.2, 3.1 Hz, 1H), 1.68 – 1.63 (m, 2H), 1.59 (d, *J* = 3.0 Hz, 1H), 1.58 – 1.52 (m, 3H), 1.50 – 1.42 (m, 1H), 1.37 (d, *J* = 4.1 Hz, 1H), 1.35 (s, 1H), 1.32 (d, *J* = 4.5 Hz, 5H), 1.30 – 1.24 (m, 2H), 1.14 (s, 3H), 1.12

(d, J = 6.4 Hz, 1H), 1.10 – 1.04 (m, 2H), 1.00 (s, 3H), 0.96 – 0.93 (m, 4H), 0.92 (s, 3H), 0.90 (d, J = 1.5 Hz, 10H), 0.05 (s, 3H), 0.02 (s, 3H).

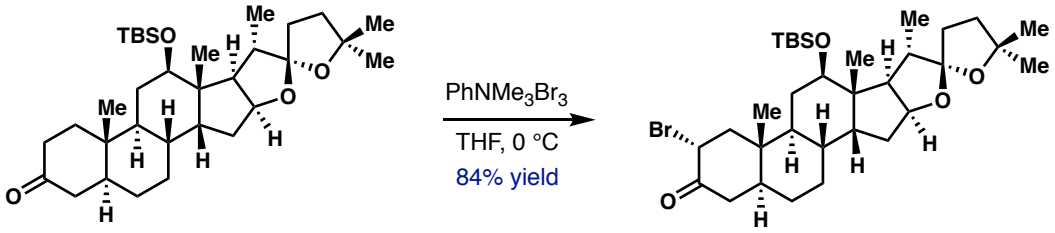
^{13}C NMR (101 MHz, $CDCl_3$): δ 211.90, 116.62, 81.75, 79.46, 73.72, 56.95, 48.74, 47.46, 46.26, 45.11, 44.72, 41.94, 38.42, 38.14, 37.50, 35.66, 33.35, 32.46, 32.42, 31.71, 30.70, 30.19, 29.23, 28.64, 26.09, 18.20, 14.49, 13.67, 11.50, -3.26, -4.63.

FTIR (NaCl, thin film, cm^{-1}): 2955, 2931, 2857, 1715, 1462, 1252, 1116, 1088, 1004, 836 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{33}H_{57}O_4Si$: 545.4026 $[M+H]^+$; found: 545.4041.

$[\alpha]_D^{21}$: –21° (c = 0.133, $CHCl_3$).

Preparation of α-bromo ketone 90:



In an N_2-filled glovebox, a 50 mL round-bottom flask with a stir bar was charged with ketone **48** (80 mg, 0.15 mmol) and THF (7.5 mL, 0.01 M total). The flask was removed from the glovebox and the resulting solution was cooled to 0 °C and stirred at this temperature for 5 min. A solution of $PhNMe_3Br_3$ (62 mg, 0.165 mmol, 1.1 equiv) in THF (7.5 mL, 0.01 M total) was also cooled to 0 °C, added by cannula, and the reaction stirred at 0 °C for 30 min. The reaction was quenched with st. $Na_2S_2O_3$ aq. (7.5 mL) and H_2O (7.5mL) and was diluted with hexane (7.5 mL). The layers were separated. The aqueous layer was further extracted with hexane (3 × 7.5 mL), combined organic layers were

washed with brine (1 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (10 to 15 to 20% EtOAc/hexanes) to afford α-bromo ketone **90** (78.5 mg, 0.13 mmol, 84% yield) as an off-yellow foam.[11i]

TLC (15% EtOAc/hexanes): R_f = 0.27 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 4.74 (dd, *J* = 13.3, 6.2 Hz, 1H), 4.49 – 4.41 (m, 1H), 3.31 (dd, *J* = 11.3, 4.5 Hz, 1H), 2.58 (dd, *J* = 12.8, 6.3 Hz, 1H), 2.45 – 2.38 (m, 3H), 2.01 – 1.96 (m, 1H), 1.96 – 1.91 (m, 2H), 1.89 – 1.81 (m, 2H), 1.78 (dd, *J* = 9.2, 3.2 Hz, 1H), 1.75 – 1.70 (m, 1H), 1.70 – 1.64 (m, 2H), 1.63 (dd, *J* = 6.6, 1.6 Hz, 1H), 1.59 (s, 1H), 1.57 (s, 1H), 1.55 – 1.49 (m, 2H), 1.43 – 1.38 (m, 1H), 1.38 – 1.36 (m, 1H), 1.34 (d, *J* = 8.7 Hz, 1H), 1.32 (s, 3H), 1.31 – 1.28 (m, 1H), 1.28 – 1.24 (m, 2H), 1.21 (d, *J* = 11.2 Hz, 1H), 1.19 – 1.15 (m, 1H), 1.14 (s, 3H), 1.13 – 1.09 (m, 1H), 1.08 (s, 3H), 1.02 – 0.96 (m, 1H), 0.94 (d, *J* = 6.7 Hz, 4H), 0.92 (s, 3H), 0.90 (s, 10H), 0.06 (s, 3H), 0.04 (s, 3H).

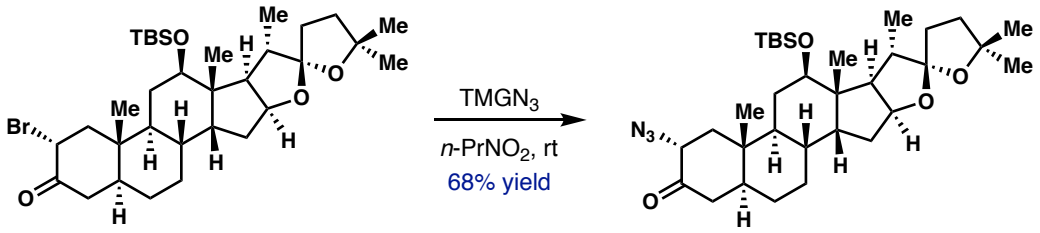
^{13}C NMR (101 MHz, $CDCl_3$): δ 201.03, 116.59, 81.78, 79.32, 73.42, 56.90, 54.28, 51.46, 48.73, 47.29, 47.06, 44.89, 43.95, 41.92, 38.99, 37.49, 33.31, 32.37, 32.02, 31.51, 30.71, 30.19, 28.70, 28.63, 26.08, 18.18, 14.48, 13.69, 12.14, -3.19, -4.64.

FTIR (NaCl, thin film, cm^{-1}): 2954, 2930, 2857, 1731, 1462, 1456, 1363, 1253, 1185, 1089 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{33}H_{56}O_4SiBr$: 623.3131 $[M+H]^+$; found: 623.3108.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: –0.22° (*c* = 0.208, $CHCl_3$).

Preparation of α-azido ketone 91:



In an N_2-filled glovebox, a 25 mL round-bottom flask with a stir bar was charged with α-bromo ketone **90** (31 mg, 0.05 mmol, 1.0 equiv) and *n*-$PrNO_2$ (5.0 mL, 0.01 M). $TMGN_3$ (31.4 mg, 0.20 mmol, 4.0 equiv) was added, and the reaction was stirred at room temperature for 4 h. Upon completion, the reaction was diluted with heptane (5 mL) and was concentrated in vacuo. The crude material was azeotroped with another 5 mL of heptane to remove *n*-$PrNO_2$. The crude product was purified by flash column chromatography (5 to 10 to 15 to 20 to 25% EtOAc/hexanes) to afford α-azido ketone **91** (19.7 mg, 0.034 mmol, 68% yield) as a colorless foam.[11i]

TLC (20% EtOAc/hexanes): R_f = 0.52 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 4.45 (t, *J* = 6.1 Hz, 1H), 3.98 (dd, *J* = 13.1, 6.3 Hz, 1H), 3.32 (dd, *J* = 11.3, 4.5 Hz, 1H), 2.44 – 2.38 (m, 1H), 2.38 – 2.32 (m, 1H), 2.28 (d, *J* = 4.0 Hz, 1H), 2.27 – 2.25 (m, 1H), 2.25 – 2.21 (m, 1H), 2.00 – 1.96 (m, 1H), 1.95 (s, 1H), 1.94 (d, *J* = 1.7 Hz, 1H), 1.92 (d, *J* = 1.7 Hz, 1H), 1.89 (dd, *J* = 7.3, 5.3 Hz, 1H), 1.86 – 1.82 (m, 1H), 1.77 (dq, *J* = 9.4, 6.7 Hz, 2H), 1.70 (d, *J* = 4.4 Hz, 1H), 1.67 (t, *J* = 1.7 Hz, 1H), 1.66 – 1.62 (m, 2H), 1.59 (s, 1H), 1.56 (s, 1H), 1.54 (d, *J* = 4.0 Hz, 1H), 1.52 (d, *J* = 4.0 Hz, 1H), 1.51 – 1.47 (m, 1H), 1.44 – 1.40 (m, 1H), 1.39 (s, 1H), 1.38 (d, *J* = 1.8 Hz, 1H), 1.36 (s, 1H), 1.34 – 1.33 (m, 1H), 1.32 (s, 3H), 1.30 (d, *J* = 1.7 Hz, 1H), 1.28 (d, *J* = 4.0 Hz,

1H), 1.25 (s, 3H), 1.20 – 1.17 (m, 1H), 1.14 (s, 4H), 1.13 – 1.10 (m, 1H), 1.08 (s, 4H), 0.94 (d, J = 6.6 Hz, 4H), 0.92 (s, 3H), 0.90 (s, 10H), 0.88 (d, J = 1.6 Hz, 1H), 0.87 – 0.82 (m, 2H), 0.06 (s, 3H), 0.03 (s, 3H).

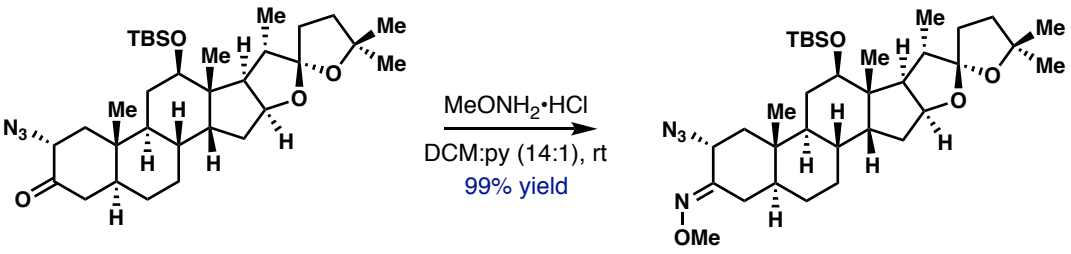
^{13}C NMR (101 MHz, $CDCl_3$): δ 205.11, 116.60, 81.79, 79.34, 73.45, 63.81, 56.92, 48.74, 47.31, 47.17, 45.26, 44.93, 43.79, 41.93, 37.49, 36.93, 33.32, 32.37, 31.95, 31.56, 30.77, 30.19, 29.85, 28.71, 28.63, 26.08, 18.19, 14.48, 13.69, 12.53, -3.19, -4.64.

FTIR (NaCl, thin film, cm^{-1}): 2924, 2858, 2103, 1728, 1557, 1462, 1456, 1363, 1258, 1090 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{33}H_{56}N_3O_4Si$: 586.4040 $[M+H]^+$; found: 586.4060.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: +21° ($c$ = 0.250, $CHCl_3$).

Preparation of α-azido methoxime 92:



A 10 mL round-bottom flask with a stir bar was charged with α-azido ketone **91** (16.0 mg, 0.027 mmol) and CH_2Cl_2/py. (14:1, 2.7 mL, 0.01 M). $MeONH_2$·HCl (6.84 mg, 0.082 mmol, 3.0 equiv) was added, and the reaction was stirred at room temperature for 2 h, then concentrated in vacuo. The crude product was purified by flash column chromatography (5 to 10 to 15% EtOAc/hexanes) to afford α-azido methoxime **92** (16.6 mg, 0.027 mmol, 99% yield) as a colorless foam.[11d]

TLC (10% EtOAc/hexanes): R_f = 0.43 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 4.45 (t, *J* = 6.8 Hz, 1H), 3.97 (dd, *J* = 12.0, 5.4 Hz, 1H), 3.90 (s, 3H), 3.31 (dd, *J* = 11.2, 4.5 Hz, 1H), 3.01 (dd, *J* = 15.5, 3.3 Hz, 1H), 2.41 (dd, *J* = 9.2, 7.4 Hz, 1H), 2.17 – 2.11 (m, 1H), 2.01 – 1.93 (m, 2H), 1.93 – 1.89 (m, 1H), 1.88 – 1.80 (m, 1H), 1.79 – 1.74 (m, 1H), 1.73 – 1.68 (m, 1H), 1.67 (t, *J* = 1.8 Hz, 1H), 1.65 – 1.60 (m, 2H), 1.59 (d, *J* = 5.0 Hz, 1H), 1.57 – 1.52 (m, 2H), 1.51 (t, *J* = 3.8 Hz, 1H), 1.45 – 1.39 (m, 1H), 1.32 (s, 3H), 1.31 – 1.29 (m, 1H), 1.27 (s, 1H), 1.26 – 1.23 (m, 3H), 1.20 (d, *J* = 5.4 Hz, 1H), 1.14 (s, 3H), 1.13 – 1.09 (m, 1H), 1.08 – 1.03 (m, 1H), 0.94 (d, *J* = 6.6 Hz, 3H), 0.91 (s, 3H), 0.90 (s, 11H), 0.89 – 0.86 (m, 2H), 0.06 (s, 3H), 0.04 (s, 3H).

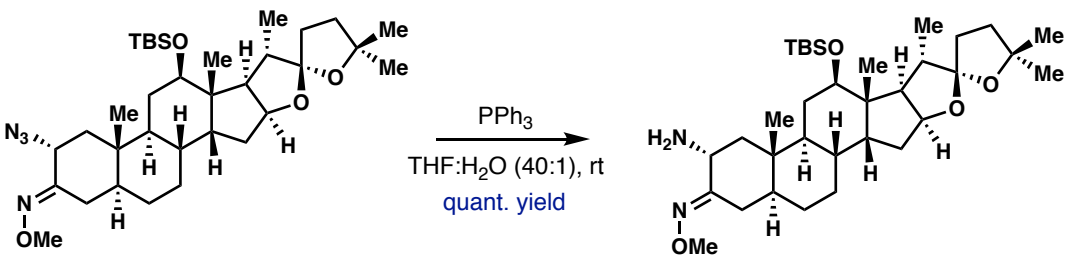
^{13}C NMR (101 MHz, $CDCl_3$): δ 155.54, 116.62, 81.76, 79.42, 73.62, 62.11, 57.77, 56.94, 48.69, 47.43, 45.10, 44.86, 44.48, 41.93, 37.50, 36.54, 33.34, 32.39, 32.03, 31.58, 30.53, 30.20, 28.64, 28.44, 27.58, 26.10, 18.21, 14.49, 13.71, 12.46, -3.18, -4.62.

FTIR (NaCl, thin film, cm^{-1}): 3146, 2930, 1858, 2106, 1675, 1462, 1455, 1381, 1251, 1088, 1048 cm^{-1}.

HRMS (FAB+, m/z): calc'd for $C_{34}H_{59}N_4O_4Si$: 615.4306 $[M+H]^+$; found: 615.4302.

$[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\mathbf{21}}$: +14° (*c* = 0.250, $CHCl_3$).

Preparation of eastern fragment α-amino methoxime 93:



A 10 mL round-bottom flask with a stir bar was charged with α-azido methoxime **92** (11.0 mg, 0.018 mmol), THF/H_2O (39:1, 1.8 mL, 0.01 M). Ph_3P (18.8 mg, 0.072 mmol, 4.0 equiv) was added, and the reaction was stirred at room temperature for 24 h, then concentrated in vacuo. The crude product was purified by flash column chromatography (0 to 3 to 7% MeOH/DCM) to afford α-amino methoxime **93** (10.5 mg, 0.018 mmol, quant. yield) as a colorless foam.[11d]

Note: The product α-amino methoxime is very sensitive to acidic conditions (e.g. $CDCl_3$, impurity in EtOAc).

TLC (5% MeOH/DCM): R_f = 0.36 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, C_6D_6): δ 4.94 (t, *J* = 7.0 Hz, 1H), 4.38 (s, 1H), 3.99 (s, 3H), 3.47 (dd, *J* = 11.5, 4.5 Hz, 1H), 3.43 – 3.35 (m, 1H), 3.18 (dd, *J* = 14.8, 3.7 Hz, 1H), 2.81 (dd, *J* = 9.4, 7.4 Hz, 1H), 2.26 (dd, *J* = 12.4, 4.8 Hz, 1H), 2.20 – 2.09 (m, 2H), 2.05 – 1.97 (m, 3H), 1.91 (dd, *J* = 13.8, 7.7 Hz, 1H), 1.74 (dt, *J* = 11.6, 5.8 Hz, 2H), 1.65 – 1.57 (m, 2H), 1.55 (s, 3H), 1.46 (d, *J* = 3.0 Hz, 2H), 1.43 – 1.33 (m, 6H), 1.29 (d, *J* = 6.6 Hz, 3H), 1.22 (s, 3H), 1.19 – 1.15 (m, 4H), 1.11 (s, 9H), 1.02 (dt, *J* = 10.1, 3.4 Hz, 2H), 0.99 – 0.94 (m, 1H), 0.94 – 0.87 (m, 3H), 0.66 (s, 3H), 0.39 (s, 1H), 0.20 (s, 3H), 0.18 (s, 3H).

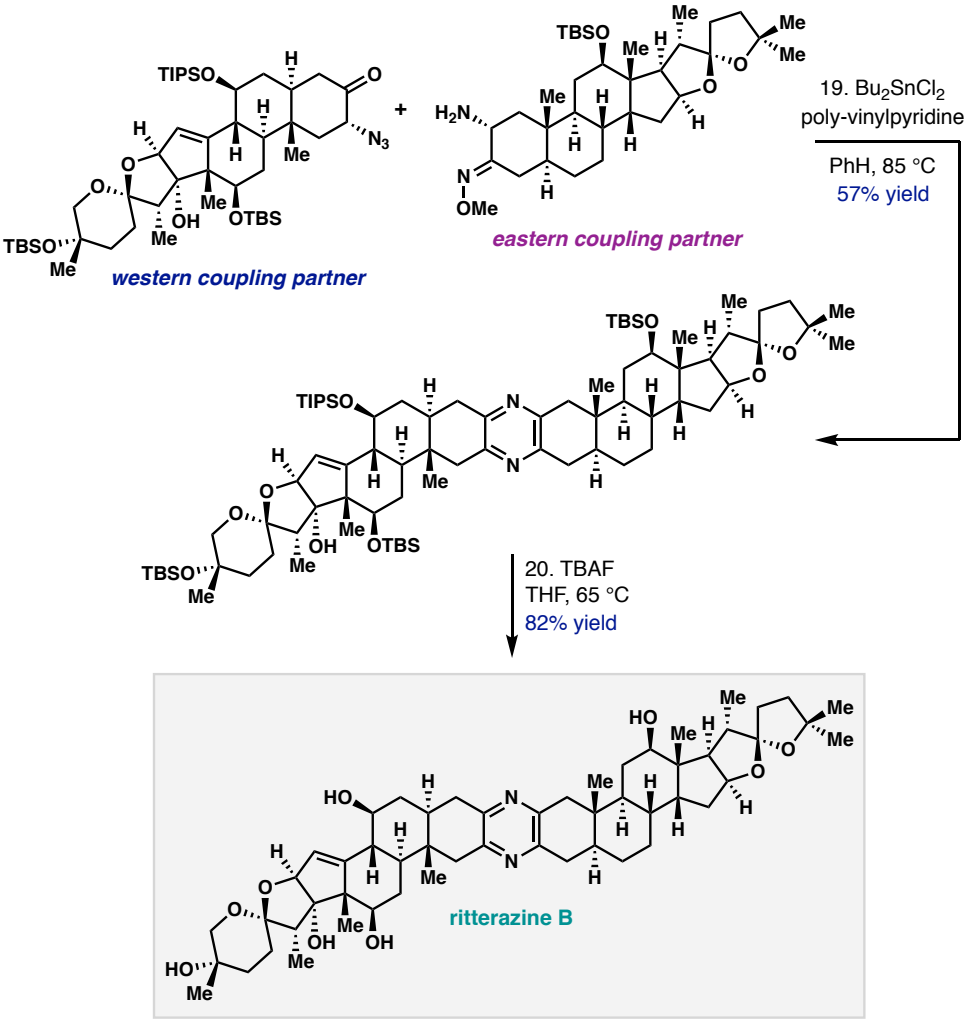
^{13}C NMR (101 MHz, C_6D_6): δ 115.64, 80.43, 78.58, 72.91, 60.26, 56.38, 47.86, 46.52, 44.05, 44.02, 41.19, 36.63, 35.05, 32.37, 31.65, 30.98, 30.62, 29.50, 29.16, 29.04, 27.57, 27.36, 26.47, 25.04, 17.13, 13.68, 12.74, 10.97, 0.22, -4.41, -5.92.

FTIR (NaCl, thin film, cm^{-1}): 3365, 2934, 2858, 1602, 1462, 1456, 1381, 1361, 1251, 1088, 1047 cm^{-1}.

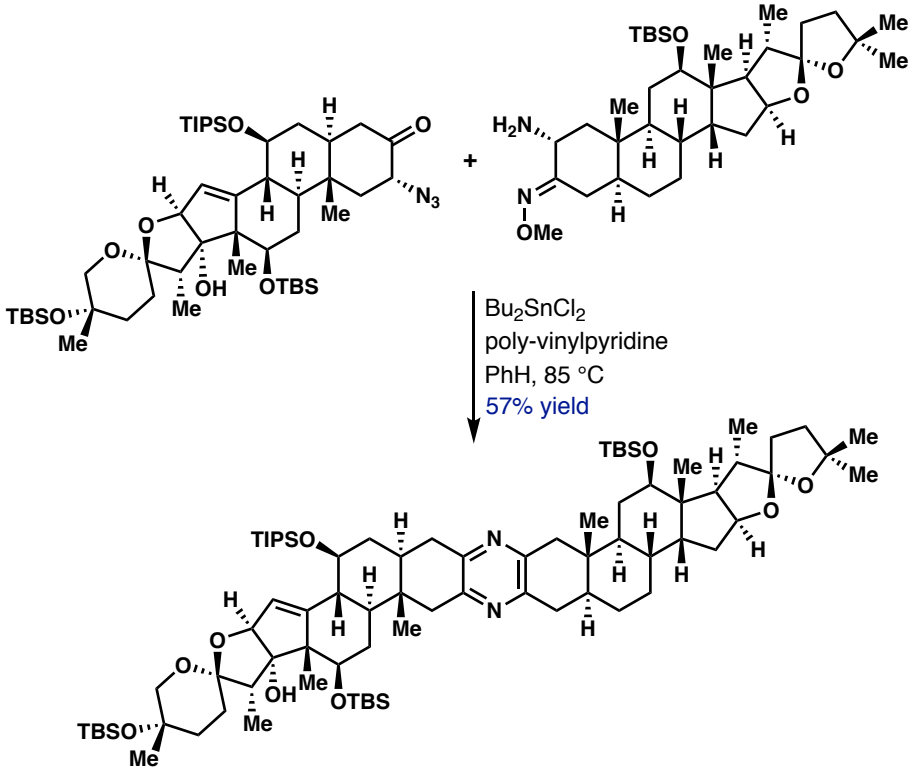
HRMS (FAB+, m/z): calc'd for $C_{34}H_{61}N_2O_4Si$: 589.4401 $[M+H]^+$; found: 589.4404.

$[\alpha]_D^{21}$: +16° (*c* = 0.083, CH_2Cl_2).

Synthesis of ritterazine B (46)



Preparation of pyrazine 97:



In an N_2-filled glovebox, a 10 mL round-bottom flask with a stir bar equipped with a Dean–Stark apparatus, reflux condenser was charged with α-azido ketone **81** (14.0 mg, 0.016 mmol, 1.0 equiv), α-amino methoxime **93** (9.14 mg, 0.016 mmol, 1.0 equiv), Bu_2SnCl_2 (1.0 mg, 0.032 mmol, 20 mol %), polyvinylpyridine (28.0 mg, 200 wt %) and PhH (3.2 mL, 0.005 M). The flask was removed from the glovebox, equipped with a reflux condenser, placed in an oil bath preheated to 85 °C, and the resulting mixture was heated to reflux for 6 h. After cooling to room temperature, the reaction was filtered through a cotton plug and was concentrated in vacuo. The crude product was purified by flash column chromatography (20 to 30 to 40% EtOAc/hexanes) to afford pyrazine **97** (12.7 mg, 0.01 mmol, 57% yield) as a colorless foam.[11d]

TLC (40% Et_2O/hexanes): R_f = 0.42 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, C_6D_6): δ 5.77 (t, *J* = 1.6 Hz, 1H), 5.12 – 5.08 (m, 1H), 4.83 (t, *J* = 7.0 Hz, 1H), 4.21 (dd, *J* = 10.7, 5.2 Hz, 1H), 3.97 (td, *J* = 10.3, 4.9 Hz, 1H), 3.76 (d, *J* = 11.6 Hz, 1H), 3.43 (dd, *J* = 11.4, 4.4 Hz, 1H), 3.32 – 3.27 (m, 2H), 3.24 – 3.12 (m, 3H), 2.97 (ddd, *J* = 18.1, 8.7, 5.5 Hz, 2H), 2.74 (dt, *J* = 15.2, 10.2 Hz, 3H), 2.67 – 2.53 (m, 4H), 2.31 (td, *J* = 13.0, 6.7 Hz, 2H), 2.24 (t, *J* = 7.4 Hz, 2H), 2.15 – 2.07 (m, 5H), 2.03 – 1.90 (m, 10H), 1.86 – 1.73 (m, 6H), 1.69 – 1.59 (m, 9H), 1.57 – 1.54 (m, 3H), 1.53 (d, *J* = 3.2 Hz, 2H), 1.50 (d, *J* = 3.1 Hz, 2H), 1.48 (d, *J* = 3.4 Hz, 2H), 1.47 – 1.44 (m, 7H), 1.43 – 1.41 (m, 3H), 1.41 – 1.38 (m, 4H), 1.37 – 1.36 (m, 6H), 1.35 (s, 3H), 1.34 (s, 6H), 1.33 – 1.31 (m, 7H), 1.31 – 1.29 (m, 7H), 1.28 – 1.26 (m, 6H), 1.22 (s, 6H), 1.20 (d, *J* = 2.7 Hz, 6H), 1.18 (s, 9H), 1.17 (d, *J* = 2.1 Hz, 9H), 1.15 (s, 3H), 1.13 (s, 6H), 1.08 (s, 4H), 1.06 (s, 9H), 1.03 (s, 12H), 1.01 (s, 9H), 1.00 (d, *J* = 1.2 Hz, 2H), 0.98 (s, 1H), 0.97 – 0.95 (m, 4H), 0.95 – 0.93 (m, 3H), 0.92 (s, 3H), 0.92 – 0.89 (m, 4H), 0.87 (s, 4H), 0.83 (dd, *J* = 11.7, 3.6 Hz, 3H), 0.80 – 0.77 (m, 1H), 0.73 (d, *J* = 5.2 Hz, 1H), 0.68 (s, 3H), 0.61 (s, 3H), 0.30 (s, 6H), 0.24 (s, 3H), 0.12 (s, 3H), 0.09 – 0.05 (m, 9H).

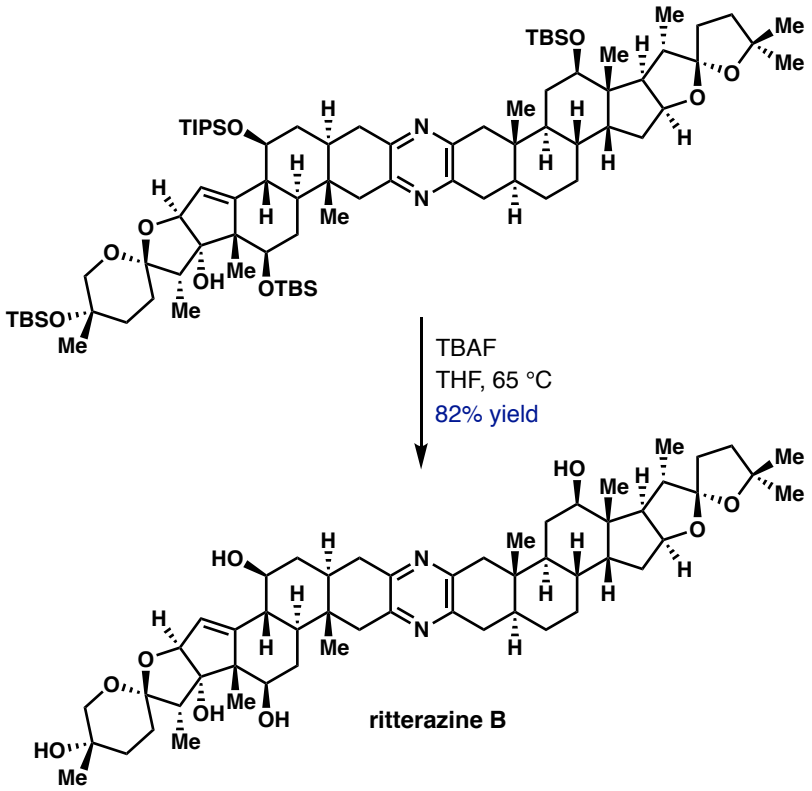
^{13}C NMR (101 MHz, C_6D_6): δ 151.28, 149.31, 149.16, 148.58, 148.28, 121.09, 116.86, 107.66, 94.79, 92.07, 81.65, 79.77, 74.56, 74.04, 71.59, 69.76, 69.68, 57.57, 56.46, 50.53, 49.01, 48.54, 47.71, 46.33, 45.89, 45.37, 43.39, 42.44, 41.54, 39.50, 38.65, 37.83, 35.74, 35.45, 34.25, 34.15, 33.60, 32.85, 32.50, 32.43, 32.37, 31.65, 30.82, 30.36, 30.23, 30.16, 29.86, 29.80, 29.58, 28.88, 28.78, 27.91, 26.44, 26.41, 26.23, 26.18, 23.15, 18.93, 18.90, 18.56, 18.50, 18.33, 14.89, 14.40, 13.84, 13.82, 13.60, 11.77, 11.76, 8.39, 1.42, -1.98, -2.14, -3.32, -3.51, -4.05, -4.69.

FTIR (NaCl, thin film, cm^{-1}): 3522, 2953, 2925, 1858, 1743, 1470, 1462, 1456, 1385, 1252, 1090 cm^{-1}.

HRMS (TOF-ESI, m/z): calc'd for $C_{81}H_{141}N_2O_9Si_4$: 1397.9714 $[M+H]^+$; found: 1397.9714.

$[\alpha]_D^{21}$: +28° (*c* = 0.083, CH_2Cl_2).

Preparation of ritterazine B (46):



A 10 mL round-bottom flask with a stir bar was charged with pyrazine **97** (8.0 mg, 5.7 μmol, 1.0 equiv) and THF (2.3 mL 0.0025 M). The flask was equipped with a reflux condenser and TBAF (1.0 M in THF, 114 μL, 0.11 mmol, 20 equiv) was added. The reaction was placed in an oil bath pre-heated to 65 °C and was stirred at for 12 h. After cooling to room temperature, the reaction was concentrated, and the resulting residue taken

up in EtOAc (1 mL) and washed with H_2O (8 × 1 mL). The organic layer was dried over anhydrous Na_2SO_4, concentrated in vacuo, and the crude product was purified by flash column chromatography (4 to 8 to 12% MeOH/DCM) to afford ritterazine B (**46**) (4.2 mg, 4.7 μmol, 82% yield) as a colorless glassy solid.[11i]

TLC (7% MeOH/DCM): R_f = 0.36 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, C_5D_5N): δ 6.16 (t, *J* = 1.7 Hz, 1H), 6.12 (d, *J* = 4.9 Hz, 1H), 5.87 (d, *J* = 5.0 Hz, 1H), 5.70 (s, 1H), 5.28 (d, *J* = 1.9 Hz, 1H), 4.81 (t, *J* = 6.8 Hz, 1H), 4.71 (d, *J* = 2.0 Hz, 1H), 4.27 – 4.19 (m, 1H), 4.13 – 4.06 (m, 1H), 4.04 (d, *J* = 11.4 Hz, 1H), 3.69 – 3.61 (m, 2H), 3.21 (d, *J* = 6.7 Hz, 1H), 3.19 – 3.14 (m, 2H), 2.98 (td, *J* = 18.0, 5.4 Hz, 2H), 2.85 – 2.77 (m, 1H), 2.77 – 2.71 (m, 2H), 2.71 – 2.64 (m, 1H), 2.57 – 2.49 (m, 1H), 2.47 – 2.41 (m, 1H), 2.27 – 2.16 (m, 4H), 2.14 (t, *J* = 3.6 Hz, 1H), 2.12 (s, 1H), 2.11 – 2.09 (m, 1H), 2.08 (s, 1H), 2.05 – 2.03 (m, 1H), 1.91 (s, 1H), 1.90 – 1.87 (m, 2H), 1.86 – 1.84 (m, 1H), 1.84 – 1.82 (m, 1H), 1.79 – 1.77 (m, 1H), 1.76 – 1.72 (m, 1H), 1.72 – 1.68 (m, 2H), 1.62 – 1.56 (m, 1H), 1.53 – 1.50 (m, 1H), 1.50 – 1.46 (m, 2H), 1.45 (s, 3H), 1.41 – 1.39 (s, 1H), 1.37 (s, 2H), 1.35 (s, 3H), 1.32 (d, *J* = 3.3 Hz, 1H), 1.29 (d, *J* = 7.0 Hz, 3H), 1.29 (s, 3H), 1.26 (s, 3H), 1.21 (s, 3H), 1.21 (d, *J* = 6.8 Hz, 3H), 1.18 – 1.16 (m, 1H), 1.15 – 1.12 (m, 1H), 1.11 – 1.07 (m, 1H), 0.87 (s, 3H), 0.77 (s, 3H).

^{13}C NMR (101 MHz, C_5D_5N): δ 152.01, 149.55, 149.23, 149.09, 148.74, 121.58, 117.43, 108.34, 94.39, 93.64, 81.81, 80.39, 76.17, 72.15, 70.57, 69.82, 66.15, 57.92, 56.51, 51.57, 49.01, 48.64, 48.18, 46.70, 46.30, 45.88, 43.24, 42.40, 41.94, 40.40, 38.83, 38.17, 36.31, 36.25, 36.03, 33.62, 33.61, 33.20, 33.03, 32.16, 31.15, 30.68, 29.60, 29.40, 29.16, 27.87, 27.38, 15.10, 14.13, 12.94, 12.31, 12.20, 8.49.

FTIR (NaCl, thin film, cm^{-1}): 3444, 2920, 2854, 2352, 2333, 1667, 1455, 1397, 1262, 1045, 1034 cm^{-1}.

HRMS (TOF-ESI, m/z): calc'd for $C_{54}H_{79}N_2O_9$: 899.5786 $[M+H]^+$; found: 899.5781.

$[\alpha]_D^{21}$: +31° (*c* = 0.250, CH_2Cl_2).

Note: NMR chemical shift references for C_5D_5N vary by source. There is no information on the reference shift used in the isolation report from Fusetani et al.[1] Herein, the reference shifts provided by Cambridge Isotope Laboratories, Inc., from whom our C_5D_5N was purchased, are used. For ^1H NMR: 8.74, 7.58, 7.22. For ^{13}C NMR: 150.35, 135.91, 123.87. These values lead to a consistent shift of 0.04 ppm in ^1H NMR and 0.4 ppm in ^{13}C NMR between our chemical shifts (at 297K) and those from Fusetani (at 303K). Raw Δδ and those normalized for this shift (norm) are presented below.

Table 2.4. Comparison of spectroscopic data for natural (Fusetani)[1] and synthetic (Reisman) ritterazine B (46)

^1H western half				
proton	natural (Fusetani)	synthetic (Reisman)	Δδ	norm. Δδ (–0.04)
1α'	2.68	2.71	0.03	-0.01
1β'	3.15	3.18	0.03	-0.01
4α'	2.98	3.01	0.03	-0.01
4β'	2.77	2.81	0.04	0.00
5'	1.84	1.86	0.02	-0.02
6α'	2.18	2.21	0.03	-0.01
6β'	1.76	1.78	0.02	-0.02
7'	4.06	4.09	0.03	-0.01
8'	2.41	2.45	0.04	0.00
9'	1.16	1.18	0.02	-0.02

^1H eastern half				
proton	natural (Fusetani)	synthetic (Reisman)	Δδ	norm. Δδ (–0.04)
1α	2.71	2.75	0.04	0.00
1β	3.17	3.21	0.04	0.00
4α	2.94	2.98	0.04	0.00
4β	2.68	2.72	0.04	0.00
5	1.57	1.60	0.03	-0.01
6α	1.48	1.49	0.01	-0.03
6β	1.28	1.32	0.04	0.00
7α	1.10	1.14	0.04	0.00
7β	1.49	1.50	0.01	-0.03
8	1.68	1.71	0.03	-0.01

11α'	2.17	2.20	0.03	-0.01
11β'	1.88	1.91	0.03	-0.01
12'	4.20	4.24	0.04	0.00
15'	6.13	6.16	0.03	-0.01
16'	5.25	5.28	0.03	-0.01
18'	1.33	1.36	0.03	-0.01
19'	0.85	0.87	0.02	-0.02
20'	2.21	2.24	0.03	-0.01
21'	1.26	1.29	0.03	-0.01
23α'	2.50	2.54	0.04	0.00
23β'	1.49	1.52	0.03	-0.01
24α'	1.87	1.90	0.03	-0.01
24β'	2.16	2.19	0.03	-0.01
26α'	3.61	3.64	0.03	-0.01
26β'	4.02	4.04	0.02	-0.02
27'	1.22	1.26	0.04	0.00
7'-OH	–	6.12	–	–
12'-OH	4.67	4.71	0.04	0.00
17'-OH	–	5.70	–	–
25'-OH	3.69	3.70	0.01	-0.03

9	1.36	1.39	0.03	-0.01
11α	2.04	2.08	0.04	0.00
11β	1.67	1.68	0.01	-0.03
12	3.64	3.67	0.03	-0.01
14	2.08	2.12	0.04	0.00
15α	1.80	1.82	0.02	-0.02
15β	1.83	1.84	0.01	-0.03
16	4.78	4.81	0.03	-0.01
17	3.15	3.18	0.03	-0.01
18	1.26	1.29	0.03	-0.01
19	0.75	0.77	0.02	-0.02
20	2.01	2.04	0.03	-0.01
21	1.18	1.21	0.03	-0.01
23α	1.85	1.88	0.03	-0.01
23β	2.12	2.14	0.02	-0.02
24α	1.68	1.70	0.02	-0.02
24β	2.04	2.10	0.06	0.02
26	1.18	1.21	0.03	-0.01
27	1.43	1.45	0.02	-0.02
12-OH	5.80	5.87	0.07	0.03

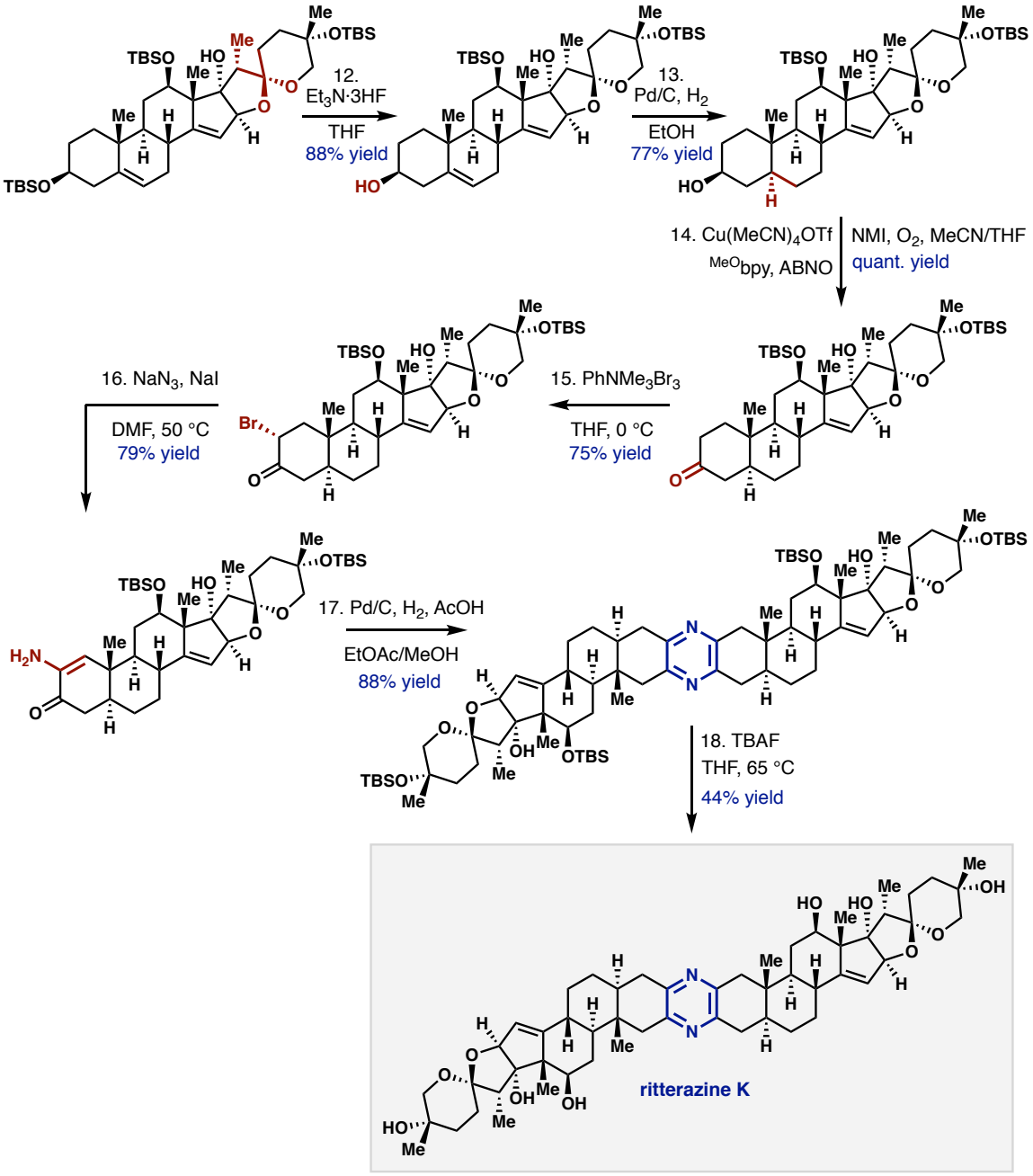
^{13}C western half				
carbon	natural (Fusetani)	synthetic (Reisman)	Δδ	norm. Δδ (–0.4)
1'	45.9	46.3	0.4	0.0
2'	148.6	149.1	0.5	0.1
3'	148.1	148.7	0.6	0.2
4'	35.6	36.0	0.4	0.0
5'	40.0	40.4	0.4	0.0
6'	38.4	38.8	0.4	0.0
7'	69.4	69.8	0.4	0.0
8'	42.8	43.2	0.4	0.0
9'	51.2	51.6	0.4	0.0
10'	35.9	36.2	0.3	-0.1
11'	29.2	29.6	0.4	0.0
12'	75.6	76.2	0.6	0.2
13'	56.1	56.5	0.4	0.0

^{13}C eastern half				
carbon	natural (Fusetani)	synthetic (Reisman)	Δδ	norm. Δδ (–0.4)
1	46.3	46.7	0.4	0.0
2	149.3	149.6	0.3	-0.1
3	149.0	149.2	0.2	-0.2
4	35.9	36.3	0.4	0.0
5	41.5	41.9	0.4	0.0
6	29.0	29.4	0.4	0.0
7	31.7	32.2	0.5	0.1
8	32.6	33.0	0.4	0.0
9	45.5	45.9	0.4	0.0
10	35.9	36.3	0.4	0.0
11	30.7	31.2	0.5	0.1
12	71.8	72.2	0.4	0.0
13	48.6	49.0	0.4	0.0

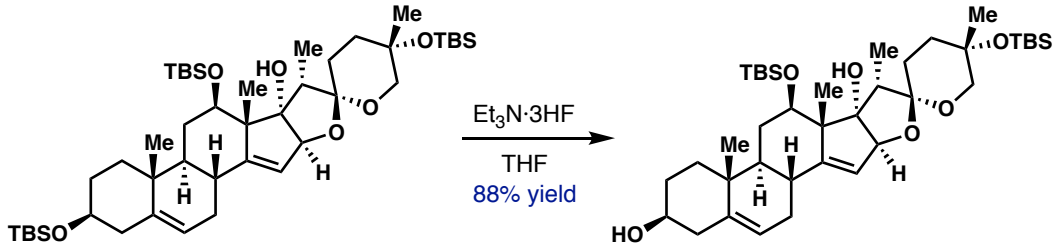
14'	151.6	152.0	0.4	0.0	14	47.8	48.2	0.4	0.0
15'	121.1	121.6	0.5	0.1	15	32.8	33.2	0.4	0.0
16'	94.0	94.4	0.4	0.0	16	80.0	80.4	0.4	0.0
17'	93.3	93.6	0.3	-0.1	17	57.5	57.9	0.4	0.0
18'	12.5	12.9	0.4	0.0	18	13.7	14.1	0.4	0.0
19'	11.5	12.2	0.7	0.3	19	11.9	12.3	0.4	0.0
20'	48.2	48.6	0.4	0.0	20	42.0	42.4	0.4	0.0
21'	8.1	8.5	0.4	0.0	21	14.7	15.1	0.4	0.0
22'	107.9	108.3	0.4	0.0	22	117.0	117.4	0.4	0.0
23'	27.5	27.9	0.4	0.0	23	33.2	33.6	0.4	0.0
24'	33.2	33.6	0.4	0.0	24	37.3	38.2	0.9	0.5
25'	65.8	66.2	0.4	0.0	25	81.4	81.8	0.4	0.0
26'	70.2	70.6	0.4	0.0	26	28.8	29.2	0.4	0.0
27'	27.0	27.4	0.4	0.0	27	30.3	30.7	0.4	0.0

2.7.2.2 Ritterazine K

Synthesis of ritterazine K (98)



Preparation of C3′ alcohol 100:



A 25 mL round-bottom flask with a stir bar was charged with charged with silyl ether **75** (220 mg, 0.27 mmol) and taken up in THF (5.5 mL, 0.05 M). $Et_3N{\cdot}3HF$ (895 μL, 5.50 mmol, 20 equiv) was added dropwise and the reaction stirred at room temperature for 24 h. The reaction was poured into a sep funnel with EtOAc (6 mL) and st. $NaHCO_3$ aq. (6 mL). The layers were separated. The aqueous layer was further extracted with EtOAc (3 × 3 mL), combined organic layers were washed with brine (6 mL), dried over anhydrous Na_2SO_4, and concentrated in vacuo. The crude product was purified by flash column chromatography (20 to 25 to 30% EtOAc/hexanes) to afford C3′ alcohol **100** (165 mg, 0.241 mmol, 88% yield) as a colorless foam.

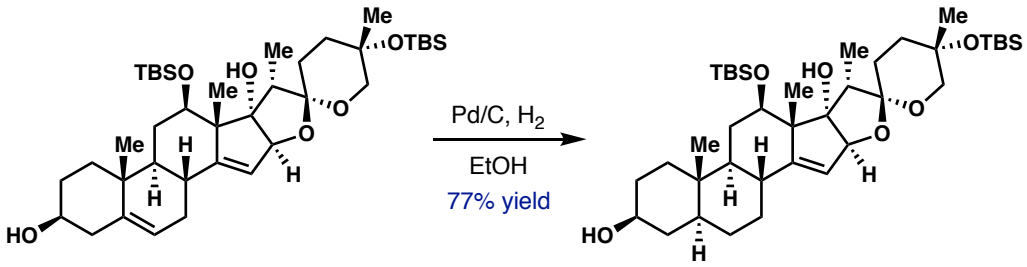
TLC (30% EtOAc/hexanes): R_f = 0.38 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.41 (q, *J* = 3.0, 2.4 Hz, 2H), 4.63 (dd, *J* = 2.3, 0.9 Hz, 1H), 4.09 (dd, *J* = 11.2, 4.9 Hz, 1H), 3.66 (d, *J* = 11.7 Hz, 1H), 3.53 (d, *J* = 4.4 Hz, 1H), 3.47 (dd, *J* = 11.6, 2.7 Hz, 1H), 3.20 (s, 1H), 2.33 (ddd, *J* = 13.1, 5.1, 2.2 Hz, 1H), 2.25 (ddd, *J* = 13.1, 7.8, 5.7 Hz, 1H), 2.16 (td, *J* = 13.4, 4.6 Hz, 1H), 2.12 – 2.05 (m, 2H), 1.99 (q, *J* = 7.0 Hz, 1H), 1.90 – 1.84 (m, 1H), 1.84 – 1.79 (m, 2H), 1.77 (t, *J* = 4.6 Hz, 1H), 1.66 – 1.58 (m, 2H), 1.52 (dd, *J* = 11.1, 2.8 Hz, 1H), 1.48 (s, 1H), 1.33 (ddd, *J* = 13.0, 4.2, 2.5

Hz, 1H), 1.12 (s, 3H), 1.07 (d, J = 10.0 Hz, 7H), 1.03 (d, J = 7.0 Hz, 3H), 0.91 (s, 9H), 0.88 (s, 9H), 0.10 (d, J = 1.5 Hz, 9H), 0.08 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 155.60, 139.81, 121.44, 119.77, 107.28, 92.94, 91.62, 74.85, 71.68, 69.84, 69.37, 56.61, 48.75, 48.37, 42.18, 37.24, 37.20, 34.09, 31.62, 31.00, 30.40, 28.64, 27.99, 26.56, 26.17, 25.99, 19.44, 18.34, 18.27, 14.90, 8.07, -2.04, -2.20, -2.77, -4.06.

Preparation of saturated B ring 101:

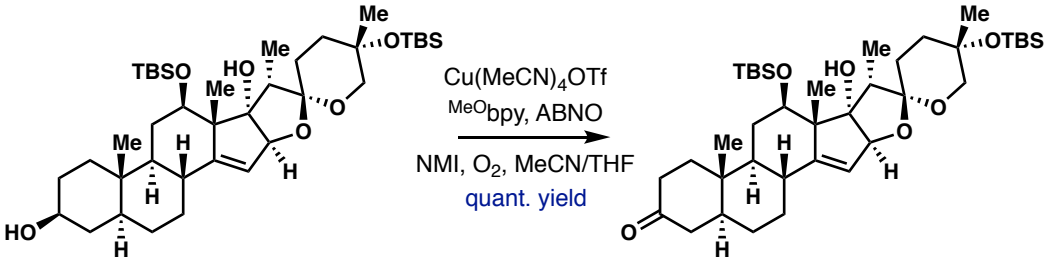


A dry 10 mL round-bottom flask with a stir bar was charged with C3′ alcohol **100** (105 mg, 0.15 mmol, 1.0 equiv) and Pd/C (10% loading, 21.1 mg, 45.7 μmol). The solids were taken up in EtOH (3.1 mL, 0.05 M) and stirred at room temperature. The resulting mixture was sparged with H_2 for 10 min by balloon, then stirred at room temperature under H_2 atmosphere for 24 h. The reaction was opened to air and passed through a short pad of Celite using EtOAc as the eluent and concentrated in vacuo. The crude product was purified by flash column chromatography (20 to 25 to 30% EtOAc/hexanes) to afford saturated B ring **101** (80.8 mg, 0.118 mmol, 77% yield) as a colorless foam.

TLC (30% EtOAc/hexanes): R_f = 0.35 (*p*-anisaldehyde)

^{1}H NMR (500 MHz, $CDCl_3$): δ 5.36 (t, *J* = 1.9 Hz, 1H), 4.62 (dd, *J* = 2.3, 0.9 Hz, 1H), 4.03 (dd, *J* = 11.2, 4.9 Hz, 1H), 3.66 (d, *J* = 11.6 Hz, 1H), 3.63 – 3.55 (m, 1H), 3.46 (dd, *J* = 11.6, 2.8 Hz, 1H), 3.11 (s, 1H), 2.16 (td, *J* = 13.3, 4.6 Hz, 1H), 1.96 (p, *J* = 6.8, 6.2 Hz, 2H), 1.86 – 1.83 (m, 1H), 1.81 (dq, *J* = 4.1, 2.4, 1.9 Hz, 2H), 1.80 – 1.76 (m, 1H), 1.68 – 1.64 (m, 1H), 1.62 (dt, *J* = 4.9, 2.6 Hz, 1H), 1.59 (dd, *J* = 4.7, 2.5 Hz, 1H), 1.47 – 1.40 (m, 2H), 1.39 – 1.35 (m, 1H), 1.35 – 1.32 (m, 2H), 1.31 – 1.28 (m, 2H), 1.12 (s, 3H), 1.06 (s, 3H), 1.03 (d, *J* = 7.0 Hz, 3H), 1.00 – 0.94 (m, 1H), 0.90 (s, 9H), 0.89 (s, 9H), 0.85 (s, 3H), 0.76 – 0.69 (m, 1H), 0.11 (s, 3H), 0.09 (s, 3H), 0.09 (s, 3H), 0.08 (s, 3H).

Preparation of ketone 102:

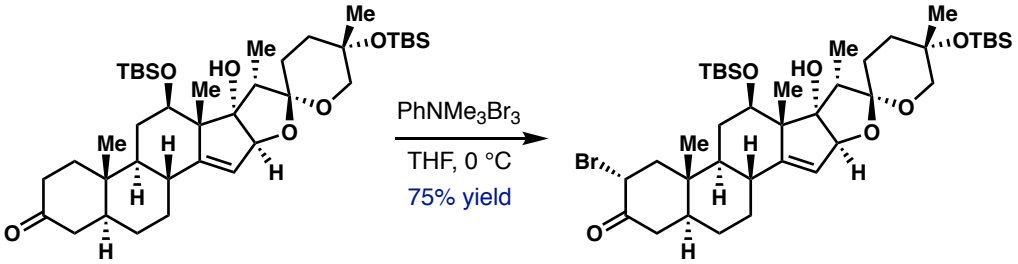


A 10 mL round-bottom flask with a stir bar was charged with C3′ alcohol **101** (77.6 mg, 0.112 mmol, 1.0 equiv), $Cu(MeCN)_4OTf$ (4.2 mg, 11.2 μmol, 10 mol %), ^{MeO}bpy (2.4 mg, 11.2 μmol, 10 mol %), ABNO (1.6 mg, 11.2 μmol, 10 mol %), NMI (1.8 μL, 22.4 μmol, 20 mol %) and MeCN/THF (3:1, 11.2 mL, 0.01 M). The resulting solution was sparged with O_2 for 5 min, then vigorously stirred (1200 rpm) at room temperature under O_2 atmosphere for 12 h. The reaction was passed through a short pad of SiO_2 using 50% EtOAc/hexanes as the eluent and concentrated in vacuo. The crude product was purified by flash column chromatography (10 to 15 to 20% EtOAc/hexanes) to afford western fragment **2** (28.9 mg, 0.034 mmol, 99% yield) as a colorless foam.

TLC (20% EtOAc/hexanes): R_f = 0.39 (*p*-anisaldehyde)

^{1}H NMR (500 MHz, $CDCl_3$): δ 5.38 (t, *J* = 2.0 Hz, 1H), 4.62 (dd, *J* = 2.3, 1.0 Hz, 1H), 4.06 (dd, *J* = 11.1, 4.9 Hz, 1H), 3.67 (d, *J* = 11.6 Hz, 1H), 3.46 (dd, *J* = 11.6, 2.7 Hz, 1H), 3.03 (s, 1H), 2.40 (td, *J* = 14.5, 13.5, 6.4 Hz, 1H), 2.36 – 2.32 (m, 1H), 2.31 – 2.24 (m, 1H), 2.18 (dd, *J* = 13.4, 4.5 Hz, 1H), 2.16 – 2.12 (m, 1H), 2.05 – 1.99 (m, 1H), 1.96 (q, *J* = 6.5 Hz, 2H), 1.90 – 1.85 (m, 1H), 1.81 (td, *J* = 13.4, 4.3 Hz, 2H), 1.65 – 1.60 (m, 1H), 1.54 – 1.48 (m, 2H), 1.45 (d, *J* = 16.5 Hz, 1H), 1.38 – 1.36 (m, 1H), 1.36 – 1.33 (m, 2H), 1.33 – 1.30 (m, 1H), 1.13 (s, 3H), 1.09 (s, 3H), 1.06 (s, 3H), 1.04 (d, *J* = 7.0 Hz, 3H), 0.91 (s, 9H), 0.89 (s, 9H), 0.87 – 0.84 (m, 1H), 0.84 – 0.80 (m, 1H), 0.11 (s, 3H), 0.11 (s, 3H), 0.10 (s, 3H), 0.09 (s, 3H).

Preparation of α-bromo ketone 103:



A 25 mL round-bottom flask with a stir bar was charged with ketone **102** (27.6 mg, 0.040 mmol, 1.0 equiv) and THF (2.0 mL, 0.01 M total). The resulting solution was cooled to 0 °C and stirred at this temperature for 5 min. $PhNMe_3Br_3$ (15.8 mg, 0.042 mmol, 1.05 equiv) in THF (2.0 mL, 0.01 M total) was also cooled to 0 °C and was added by cannula in one portion, and the reaction was stirred at 0 °C for 30 min. The reaction was diluted with Et_2O (2 mL) and was washed with 1:1 st. $NaHSO_3$ aq. : st. $NaHCO_3$ aq. (2.0 mL) and brine (2.0

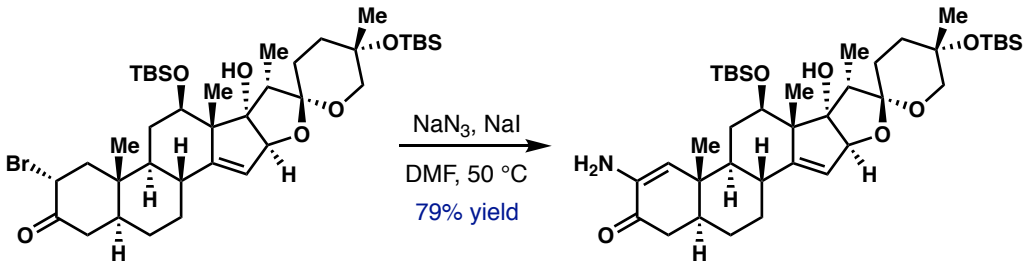
mL). The collected organics were dried over anhydrous Na_2SO_4 and concentrated in vacuo. The crude product was purified by flash column chromatography (10 to 15 to 20% EtOAc/hexanes) to afford α-bromo ketone **103** (22.7 mg, 0.030 mmol, 76% yield) as a colorless foam.[11i]

TLC (20% EtOAc/hexanes): R_f = 0.41 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.38 (t, *J* = 2.0 Hz, 1H), 4.73 (dd, *J* = 13.3, 6.2 Hz, 1H), 4.60 (dd, *J* = 2.3, 1.0 Hz, 1H), 4.04 (dd, *J* = 11.0, 5.0 Hz, 1H), 3.66 (d, *J* = 11.6 Hz, 1H), 3.45 (dd, *J* = 11.6, 2.8 Hz, 1H), 2.95 (s, 1H), 2.53 (dd, *J* = 12.9, 6.3 Hz, 1H), 2.48 – 2.37 (m, 2H), 2.21 – 2.10 (m, 1H), 2.03 – 1.98 (m, 1H), 1.94 (q, *J* = 6.9 Hz, 1H), 1.89 – 1.78 (m, 3H), 1.78 – 1.73 (m, 1H), 1.62 (p, *J* = 2.8, 2.4 Hz, 1H), 1.59 (q, *J* = 2.4, 2.0 Hz, 1H), 1.52 – 1.43 (m, 2H), 1.41 – 1.36 (m, 1H), 1.36 – 1.31 (m, 2H), 1.31 – 1.28 (m, 1H), 1.28 – 1.24 (m, 2H), 1.12 (d, *J* = 3.4 Hz, 6H), 1.07 (s, 3H), 1.03 (d, *J* = 7.0 Hz, 3H), 0.90 (s, 10H), 0.88 (s, 10H), 0.13 – 0.08 (m, 9H), 0.08 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 200.68, 155.21, 119.22, 107.37, 93.43, 91.37, 74.05, 69.83, 69.35, 56.41, 54.00, 51.22, 51.16, 48.23, 47.00, 43.78, 39.20, 33.99, 33.53, 31.39, 29.85, 28.81, 28.02, 27.66, 26.53, 26.16, 25.98, 18.33, 18.24, 15.17, 12.13, 8.04, -2.04, -2.19, -3.02, -4.10.

Preparation of steroid fragment α-enamino ketone 104:



In an N_2-filled glovebox, a dry 1 dr vial with a stir bar was charged with α-bromo ketone **103** (8.3 mg, 0.0108 mmol) and DMF (0.6 mL, 0.02 M). NaN_3 (7.1 mg, 0.11 mmol, 10.0 equiv) and NaI (1 mg, catalytic) were added. The reaction was removed from the glovebox and placed into a preheated oil bath at 50 °C, and the reaction was stirred at this temperature for 3 h. Upon completion, the reaction was brought to room temperature and poured into water (1 mL) and Et_2O (1 mL). The layers were separated and the aqueous extracted with Et2O (3 × 1 mL). Combined organics were washed with water (2 mL) and brine (2 mL), dried with Na_2SO_4, and were concentrated in vacuo. The crude product was purified by flash column chromatography (10 to 20 to 30% EtOAc/hexanes) to afford α-enamino ketone **104** (5.9 mg, 8.4 μmol, 79% yield) as a colorless foam.[27]

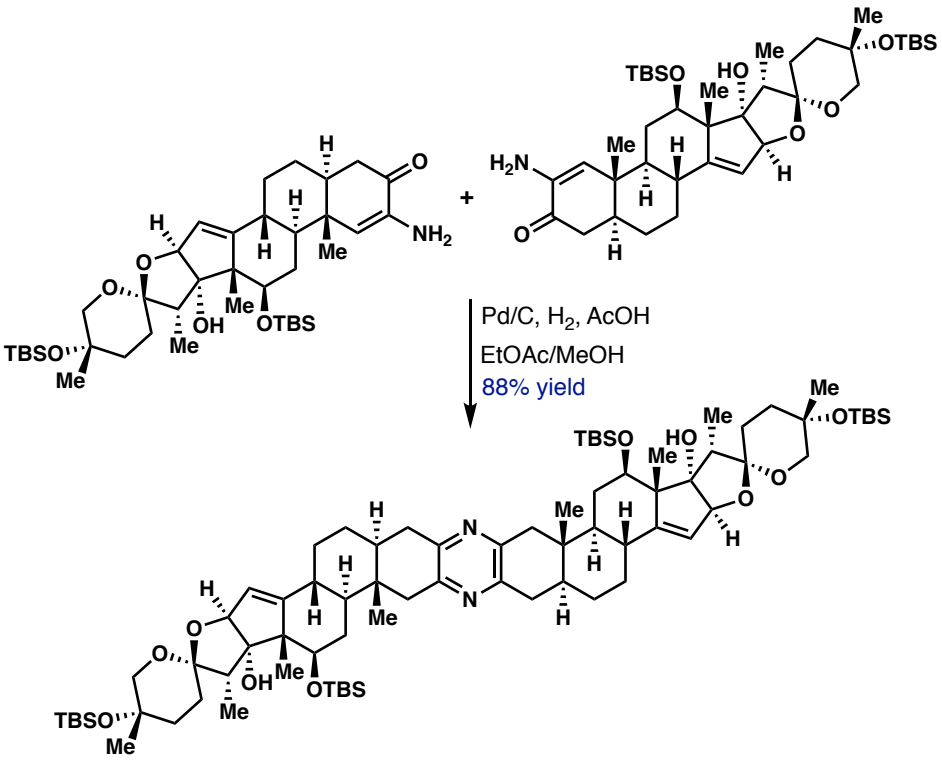
TLC (20% EtOAc/hexanes): R_f = 0.33 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.91 (s, 1H), 5.41 – 5.35 (m, 1H), 4.61 (dd, *J* = 2.2, 0.9 Hz, 1H), 4.06 (dd, *J* = 11.1, 4.8 Hz, 1H), 3.66 (d, *J* = 11.6 Hz, 1H), 3.49 (s, 2H), 3.44 (dd, *J* = 11.7, 2.8 Hz, 1H), 2.94 (s, 1H), 2.43 (dd, *J* = 17.9, 14.0 Hz, 1H), 2.30 (dd, *J* = 17.9, 4.2 Hz, 1H), 2.16 (td, *J* = 13.4, 4.6 Hz, 1H), 2.04 – 1.97 (m, 1H), 1.95 (d, *J* = 7.1 Hz, 1H), 1.93 – 1.89 (m, 1H), 1.89 – 1.85 (m, 1H), 1.85 – 1.82 (m, 1H), 1.79 (dd, *J* = 13.4, 4.2 Hz, 1H), 1.61 (dp, *J* = 10.6, 2.4 Hz, 1H), 1.54 – 1.49 (m, 1H), 1.47 (q, *J* = 3.6, 2.3 Hz, 1H), 1.44 –

1.41 (m, 1H), 1.41 – 1.36 (m, 1H), 1.36 – 1.26 (m, 2H), 1.12 (s, 3H), 1.08 (s, 3H), 1.03 (d, J = 7.4 Hz, 6H), 1.00 – 0.95 (m, 1H), 0.91 (s, 9H), 0.88 (s, 10H), 0.13 (d, J = 2.0 Hz, 6H), 0.10 (s, 3H), 0.08 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 195.55, 155.76, 138.22, 125.14, 118.85, 107.38, 93.53, 91.35, 74.24, 69.82, 69.37, 56.51, 48.86, 48.23, 44.57, 40.13, 38.27, 34.27, 33.99, 31.31, 28.80, 27.66, 27.08, 26.54, 26.18, 25.99, 18.34, 18.27, 15.29, 13.85, 8.05, -2.03, -2.18, -2.99, -4.03.

Preparation of pyrazine 105:



To a dry 1 dr vial with a stir bar was added α-enamino ketone **104** (3.6 mg, 0.005 mmol, 1.0 equiv) and Pd/C (10% loading, 1.2 mg, 30 wt %). The solids were taken up in EtOAc/MeOH (15:1, 0.40 mL, 0.04 M) and AcOH (1 drop, catalytic) was added. The

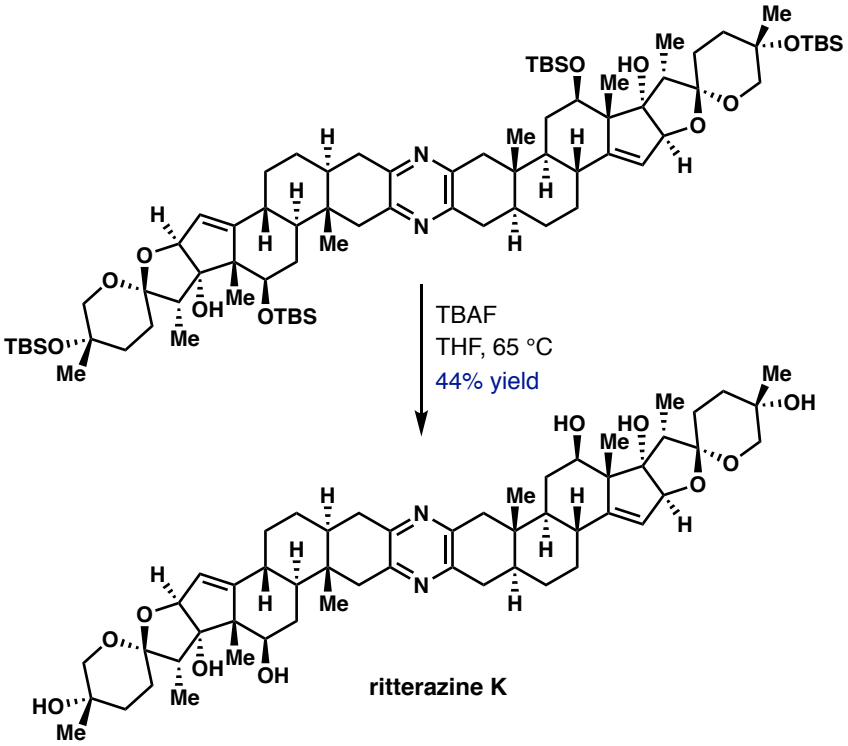
resulting mixture was sparged with H_2 (balloon) for 10 min and was stirred at room temperature under an atmosphere of H_2 for 4 h. Upon completion, the reaction was filtered through a shored plug of Celite, eluting with EtOAc, and was concentrated in vacuo. The crude product was purified by flash column chromatography (20 to 30 to 40% EtOAc/hexanes) to afford pyrazine **105** (3.0 mg, 2.19 μmol, 88% yield) as a colorless foam.[27]

TLC (20% EtOAc/hexanes): R_f = 0.29 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, C_5D_5N): δ 5.65 (t, *J* = 1.9 Hz, 2H), 5.04 (s, 3H), 4.38 (dd, *J* = 11.0, 4.9 Hz, 2H), 4.16 (s, 2H), 3.86 (d, *J* = 11.6 Hz, 2H), 3.44 (dd, *J* = 11.7, 2.7 Hz, 2H), 3.18 (d, *J* = 16.5 Hz, 2H), 3.00 (dd, *J* = 17.9, 5.3 Hz, 2H), 2.74 (dd, *J* = 18.4, 13.9 Hz, 4H), 2.34 (dd, *J* = 13.3, 4.5 Hz, 2H), 2.29 (d, *J* = 6.8 Hz, 2H), 2.11 – 2.06 (m, 2H), 2.04 (d, *J* = 3.0 Hz, 3H), 2.03 – 1.99 (m, 2H), 1.85 (dt, *J* = 11.0, 2.7 Hz, 2H), 1.76 (d, *J* = 12.5 Hz, 2H), 1.72 – 1.69 (m, 2H), 1.69 – 1.65 (m, 3H), 1.60 – 1.55 (m, 3H), 1.54 – 1.49 (m, 2H), 1.37 (d, *J* = 6.9 Hz, 7H), 1.31 (s, 7H), 1.09 (s, 6H), 1.02 (s, 18H), 1.01 (s, 19H), 0.93 – 0.90 (m, 2H), 0.87 (q, *J* = 3.4, 2.5 Hz, 2H), 0.84 (s, 6H), 0.28 (s, 6H), 0.23 (s, 6H), 0.17 (s, 6H), 0.15 (s, 6H), 0.14 – 0.12 (m, 2H).

^{13}C NMR (101 MHz, C_5D_5N): δ 156.96, 149.18, 149.05, 119.74, 108.10, 94.71, 91.81, 74.25, 70.14, 69.96, 57.02, 51.58, 48.68, 41.91, 36.42, 36.08, 34.64, 34.53, 32.32, 30.39, 29.68, 28.57, 28.47, 26.89, 26.69, 26.49, 18.84, 18.81, 16.20, 12.14, 9.18, 0.36, -1.67, -1.80, -3.14, -3.44.

Preparation of ritterazine K (98):



A 1 dr vial with a stir bar was charged with pyrazine **105** (2.75 mg, 2.0 μmol, 1.0 equiv) and THF (0.5 mL 0.004 M). TBAF (1.0 M in THF, 40 μL, 0.04 mmol, 20 equiv) was added. The reaction was placed in an oil bath pre-heated to 65 °C and was stirred at for 12 h. After cooling to room temperature, the reaction was concentrated, and the resulting residue taken up in EtOAc (1 mL) and washed with H_2O (8 × 1 mL). The organic layer was dried over anhydrous Na_2SO_4, concentrated in vacuo, and the crude product was purified by flash column chromatography (4 to 8 to 12% MeOH/DCM) to afford ritterazine K (**98**) (0.8 mg, 0.9 μmol, 44% yield) as a colorless glassy solid.[11i]

TLC (10% MeOH/DCM): R_f = 0.52 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, C_5D_5N): δ 5.74 (s, 2H), 5.63 (t, *J* = 1.8 Hz, 2H), 5.21 (d, *J* = 2.0 Hz, 2H), 5.12 (s, 2H), 4.75 (d, *J* = 2.2 Hz, 2H), 4.22 (ddd, *J* = 11.5, 4.9, 2.0 Hz, 2H), 4.04 (d, *J* = 11.4 Hz, 2H), 3.65 (td, *J* = 5.2, 2.2 Hz, 4H), 3.49 – 3.41 (m, 2H), 3.15 (d, *J* = 16.7 Hz, 2H), 2.96 (dd, *J* = 17.9, 5.4 Hz, 2H), 2.69 (dt, *J* = 17.6, 6.9 Hz, 4H), 2.64 – 2.49 (m, 3H), 2.24 (d, *J* = 6.9 Hz, 2H), 2.21 (dd, *J* = 8.9, 4.3 Hz, 2H), 2.15 (h, *J* = 4.7 Hz, 4H), 2.10 (d, *J* = 10.9 Hz, 2H), 1.95 – 1.85 (m, 4H), 1.84 – 1.78 (m, 4H), 1.76 (d, *J* = 3.3 Hz, 2H), 1.66 (q, *J* = 5.7, 4.2 Hz, 2H), 1.64 – 1.57 (m, 4H), 1.56 (d, *J* = 3.7 Hz, 2H), 1.36 (s, 6H), 1.33 (s, 4H), 1.31 (s, 3H), 1.26 (s, 7H), 0.91 (s, 1H), 0.89 (s, 2H), 0.88 (s, 1H), 0.80 (s, 6H).

^{13}C NMR (101 MHz, $(CD_3)_2CO$): δ 174.67, 149.23, 148.96, 120.14, 108.17, 94.01, 93.79, 76.07, 70.20, 66.24, 56.01, 53.68, 48.34, 46.24, 42.31, 36.76, 36.01, 34.43, 34.20, 33.05, 32.63, 29.05, 28.75, 27.77, 27.29, 26.51, 25.67, 23.32, 14.35, 12.71, 11.97, 7.85.

Note: NMR chemical shift references for C_5D_5N vary by source. There is no information on the reference shift used in the isolation report from Fusetani et al.[26] Herein, the reference shifts provided by Cambridge Isotope Laboratories, Inc., from whom our C_5D_5N was purchased, are used. For ^{1}H NMR: 8.74, 7.58, 7.22. These values lead to a consistent shift of 0.03 ppm in ^{1}H NMR between our chemical shifts (at 297K) and those from Fusetani (at 303K).[26] Raw Δδ and those normalized for this shift (norm) are presented below. We were unable to obtain satisfactory ^{13}C NMR data in C_5D_5N due to poor locking, and thus utilized $(CD_3)_2CO$. Despite this, our peaks in $(CD_3)_2CO$ each have matching counterparts with close chemical shifts in the ^{13}C NMR data in C_5D_5N reported by Fusetani.[26]

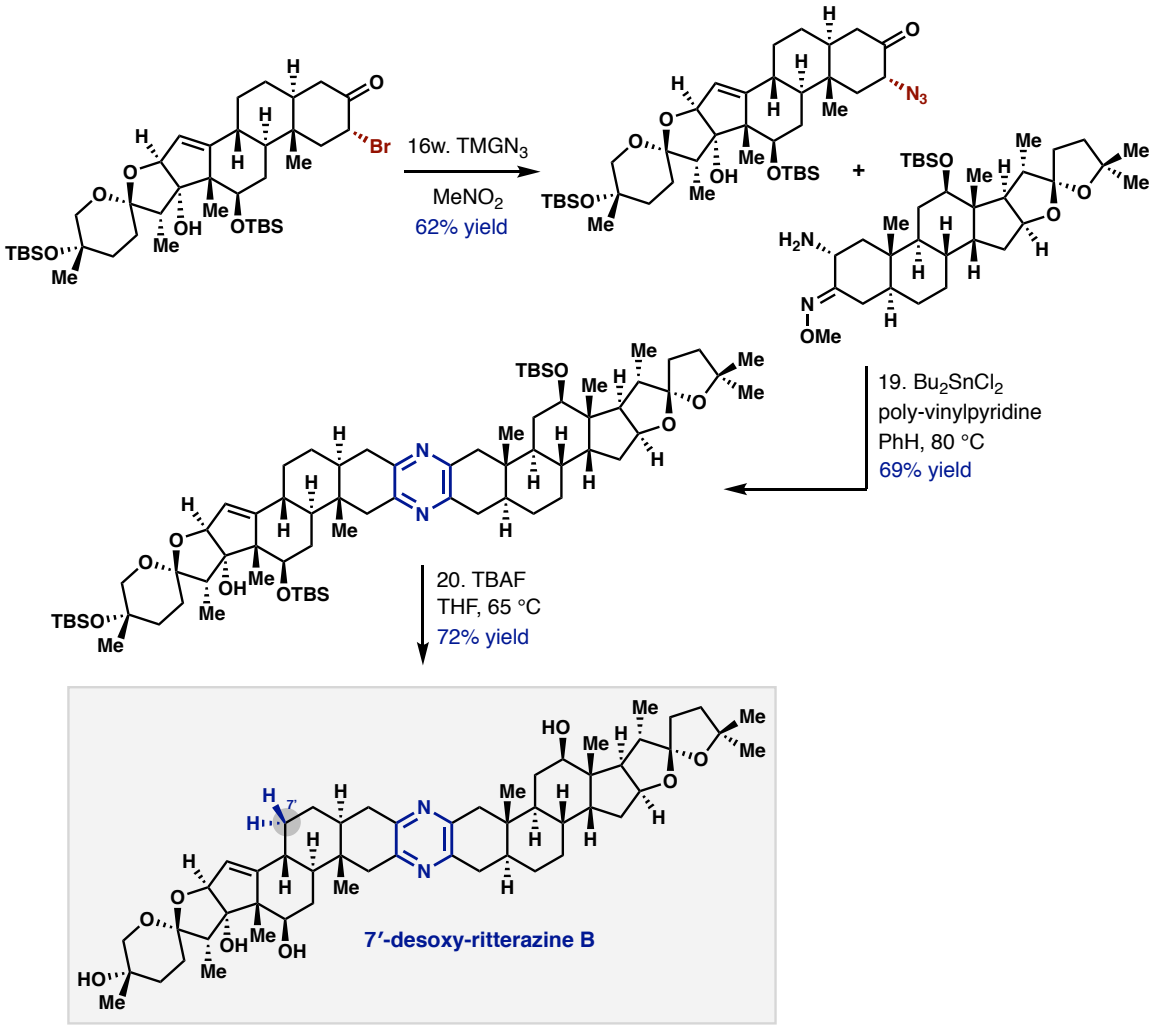
Table S4. Comparison of spectroscopic data for natural (Fusetani)[26] and synthetic (Reisman) ritterazine K (98)

^{1}H dimeric half				
proton	natural (Fusetani)	synthetic (Reisman)	Δδ	norm. Δδ (–0.03)
1α'	2.67	2.69	0.02	-0.01
1β'	3.13	3.16	0.03	0.00
4α'	2.92	2.96	0.04	0.01
4β'	2.67	2.70	0.03	0.00
5'	1.63	1.66	0.03	0.00
6α'	1.55	1.58	0.03	0.00
6β'	1.27	1.29	0.02	-0.01
7α'	1.35	1.39	0.04	0.01
7β'	1.75	1.78	0.03	0.00
8'	2.09	2.11	0.02	-0.01
9'	0.99	1.01	0.02	-0.01
11α'	2.13	2.15	0.02	-0.01
11β'	1.82	1.85	0.03	0.00
12'	4.19	4.22	0.03	0.00
15'	5.60	5.63	0.03	0.00
16'	5.18	5.21	0.03	0.00
18'	1.33	1.38	0.05	0.02
19'	0.78	0.80	0.02	-0.01
20'	2.21	2.24	0.03	0.00
21'	1.29	1.33	0.04	0.01
23α'	2.50	2.58	0.08	0.05
23β'	1.58	1.61	0.03	0.00
24α'	1.88	1.91	0.03	0.00
24β'	2.18	2.21	0.03	0.00
26α'	3.62	3.65	0.03	0.00
26β'	4.02	4.05	0.03	0.00
27'	1.22	1.26	0.04	0.01
12'-OH	4.72	4.75	0.03	0.00
17'-OH	5.08	5.12	0.04	0.01
25'-OH	–	5.75	–	–

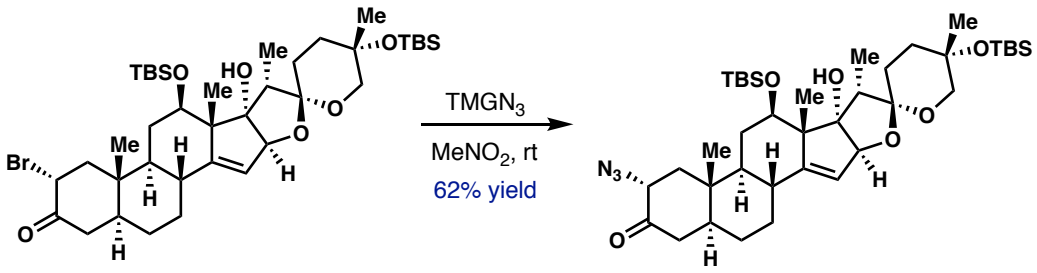
^{13}C dimeric half			
carbon	natural (Fusetani, C_5D_5N)	synthetic (Reisman, $(CD_3)_2CO$)	Δδ
1'	46.0	46.2	0.2
2'	148.6	149.0	0.4
3'	148.7	149.2	0.5
4'	35.8	36.0	0.2
5'	41.7	42.3	0.6
6'	28.2	27.8	-0.4
7'	28.9	28.8	-0.1
8'	34.0	34.4	0.4
9'	52.9	53.7	0.8
10'	36.3	36.8	0.5
11'	29.2	29.0	-0.2
12'	75.6	76.1	0.5
13'	56.0	56.0	0.0
14'	154.8	174.7	19.9
15'	120.0	120.1	0.1
16'	93.7	94.0	0.3
17'	93.3	93.8	0.5
18'	13.0	12.7	-0.3
19'	11.7	12.0	0.3
20'	48.4	48.3	-0.1
21'	8.2	7.9	-0.3
22'	107.9	108.2	0.3
23'	27.7	27.3	-0.4
24'	33.3	33.0	-0.3
25'	65.7	66.2	0.5
26'	70.2	70.2	0.0
27'	27.0	26.5	-0.5

2.7.2.3 7′-desoxy-ritterazine B

Synthesis of 7′-desoxy-ritterazine B (99)



Preparation of western fragment α-azido ketone 106:

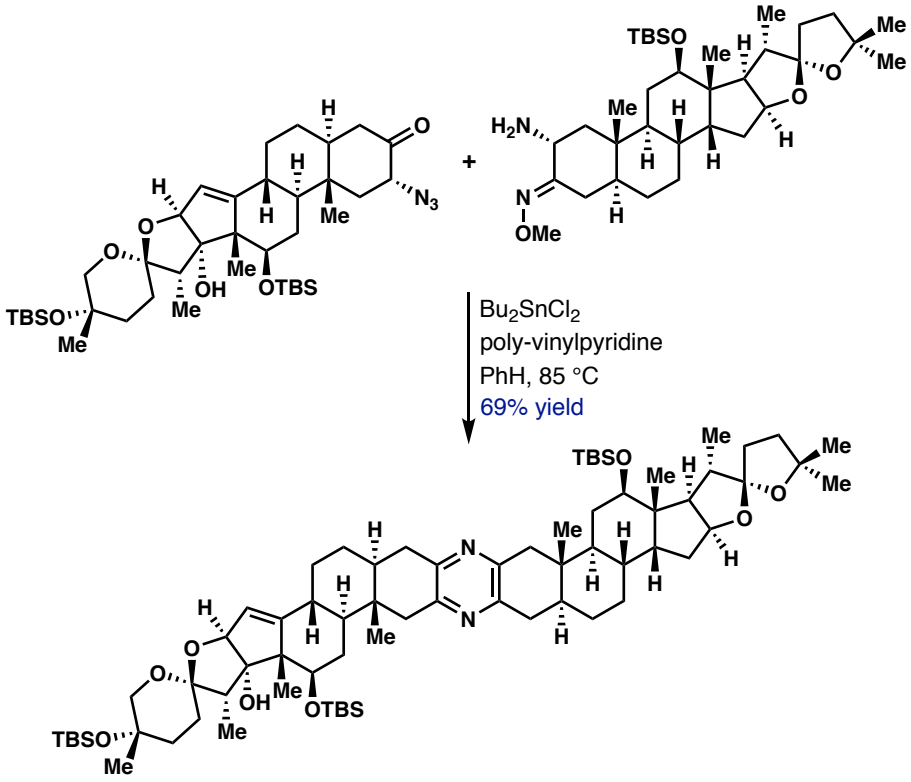


In an N_2-filled glovebox, a 10 mL round-bottom flask with a stir bar was charged with α-bromo ketone **103** (11.0 mg, 0.014 mmol) and $MeNO_2$ (1.43 mL, 0.01 M). $TMGN_3$ (9.1 mg, 0.057 mmol, 4.0 equiv) was added, and the reaction was stirred at room temperature for 4 h. Upon completion, the reaction was concentrated in vacuo. The crude product was purified by flash column chromatography (5 to 10 to 15 to 20 to 25% EtOAc/hexanes) to afford α-azido ketone **106** (6.5 mg, 8.9 μmol, 62% yield) as a colorless foam.

TLC (20% EtOAc/hexanes): R_f = 0.42 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, C_6D_6): δ 5.45 (t, *J* = 1.9 Hz, 1H), 4.93 (dd, *J* = 2.2, 1.0 Hz, 1H), 4.13 (dd, *J* = 10.8, 5.1 Hz, 1H), 3.75 (dd, *J* = 11.5, 4.1 Hz, 1H), 3.37 – 3.30 (m, 1H), 3.27 (dd, *J* = 13.1, 6.4 Hz, 1H), 2.99 (s, 1H), 2.26 (td, *J* = 13.2, 4.4 Hz, 1H), 2.08 (q, *J* = 6.9 Hz, 1H), 2.01 – 1.95 (m, 1H), 1.95 – 1.86 (m, 2H), 1.74 (dd, *J* = 11.5, 3.6 Hz, 1H), 1.67 (dd, *J* = 9.6, 5.6 Hz, 1H), 1.63 (t, *J* = 4.7 Hz, 1H), 1.61 – 1.58 (m, 1H), 1.58 – 1.55 (m, 1H), 1.49 – 1.44 (m, 1H), 1.42 – 1.39 (m, 1H), 1.38 – 1.35 (m, 2H), 1.32 (q, *J* = 4.3, 3.5 Hz, 3H), 1.30 (s, 2H), 1.28 – 1.23 (m, 1H), 1.19 (s, 3H), 1.06 (s, 1H), 1.04 (s, 8H), 1.03 (s, 9H), 0.92 (s, 3H), 0.46 (s, 3H), 0.24 (s, 3H), 0.16 (s, 3H), 0.13 (s, 3H), 0.07 (s, 3H).

Preparation of pyrazine 107:



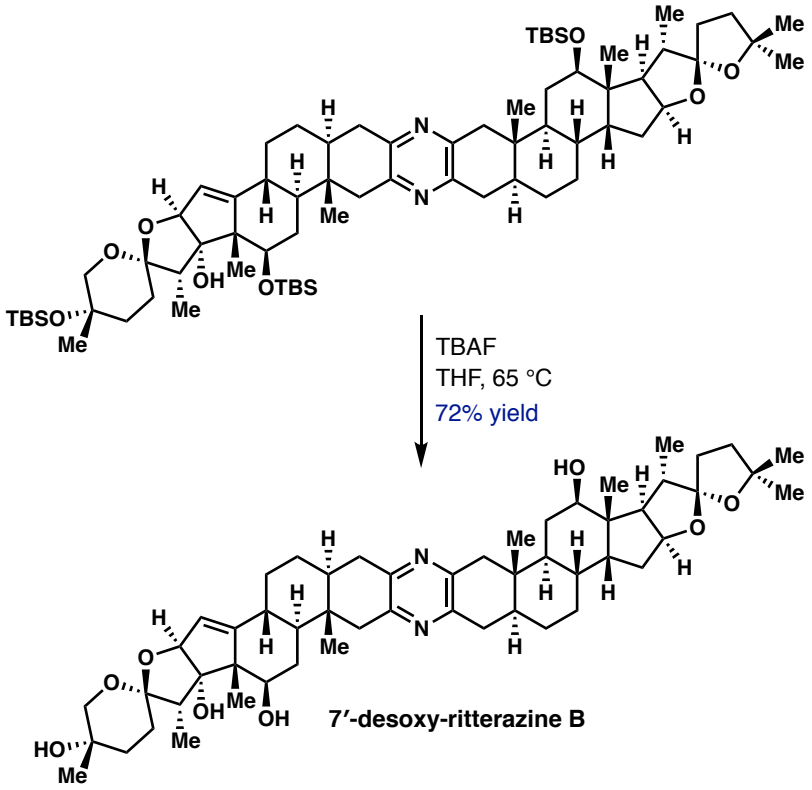
In an N_2-filled glovebox, a dry 1 dr vial with a stir bar was charged with Bu_2SnCl_2 (0.46 mg, 1.5 μmol, 20 mol %) and polyvinylpyridine (11.0 mg, 200 wt %). α-azido ketone **106** (5.48 mg, 7.5 μmol, 1.0 equiv) in PhH (0.75 mL) and α-amino methoxime **93** (4.42 mg, 7.5 μmol, 1.0 equiv) in PhH (0.75 mL) were then added. The vial was removed from the glovebox, placed in an oil bath preheated to 85 °C, and the resulting mixture was heated to reflux for 6 h. After cooling to room temperature, the reaction was filtered through a cotton plug and was concentrated in vacuo. The crude product was purified by flash column chromatography (20 to 30 to 40% EtOAc/hexanes) to afford pyrazine **107** (6.3 mg, 5.1 μmol, 69% yield) as a colorless foam.[11d]

TLC (40% Et_2O/hexanes): R_f = 0.38 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, C_5D_5N): δ 5.64 (s, 1H), 5.61 (d, *J* = 9.0 Hz, 1H), 4.83 (s, 2H), 4.40 – 4.34 (m, 1H), 4.15 (d, *J* = 10.6 Hz, 1H), 3.85 (t, *J* = 9.6 Hz, 2H), 3.57 (dd, *J* = 11.4, 4.3 Hz, 1H), 3.19 (t, *J* = 17.4 Hz, 2H), 3.00 (dt, *J* = 17.9, 5.3 Hz, 2H), 2.83 (d, *J* = 16.7 Hz, 1H), 2.78 – 2.67 (m, 4H), 2.56 (dt, *J* = 23.9, 7.4 Hz, 2H), 2.30 (qd, *J* = 13.6, 13.2, 5.3 Hz, 5H), 2.19 (s, 1H), 2.15 – 1.96 (m, 12H), 1.85 (dd, *J* = 13.3, 7.6 Hz, 6H), 1.68 (d, *J* = 16.4 Hz, 9H), 1.60 (s, 3H), 1.41 (s, 4H), 1.36 (d, *J* = 7.2 Hz, 7H), 1.29 (d, *J* = 3.3 Hz, 5H), 1.25 (d, *J* = 6.9 Hz, 7H), 1.20 – 1.17 (m, 6H), 1.12 (s, 3H), 1.08 (s, 4H), 1.01 (s, 7H), 1.00 (d, *J* = 3.1 Hz, 10H), 0.98 (s, 6H), 0.90 – 0.87 (m, 4H), 0.82 (d, *J* = 2.7 Hz, 3H), 0.78 (s, 3H), 0.27 (s, 3H), 0.23 (s, 3H), 0.16 (d, *J* = 2.7 Hz, 3H), 0.14 (s, 6H), 0.10 (d, *J* = 1.6 Hz, 2H).

^{13}C NMR (101 MHz, C_5D_5N): δ 149.67, 149.42, 149.16, 148.97, 123.34, 117.31, 94.65, 91.78, 82.04, 70.11, 57.82, 57.00, 49.32, 48.65, 47.98, 47.92, 46.44, 46.39, 45.48, 41.89, 41.86, 37.51, 36.39, 36.15, 34.52, 30.62, 30.35, 30.33, 29.97, 29.05, 26.89, 26.86, 26.66, 26.50, 26.46, 18.80, 18.77, 18.61, 16.16, 14.08, 12.20, 12.12, 9.13, 8.85, 0.34, -1.70, -1.83, -2.92, -3.15, -3.48, -4.40.

Preparation of 7′-desoxy-ritterazine B (99):



A dry 1 dr vial with a stir bar was charged with pyrazine **107** (6.0 mg, 4.9 μmol, 1.0 equiv) and THF (0.98 mL 0.005 M). TBAF (1.0 M in THF, 24.5 μL, 24.5 μmol, 5 equiv) was added. The reaction was placed in an oil bath pre-heated to 65 °C and was stirred at for 12 h. After cooling to room temperature, the reaction was concentrated, and the resulting residue taken up in EtOAc (1 mL) and washed with H_2O (8 × 1 mL). The organic layer was dried over anhydrous Na_2SO_4, concentrated in vacuo, and the crude product was purified by flash column chromatography (2 to 6 to 10% MeOH/DCM) to afford 7′-desoxy-ritterazine B (**99**) (3.1 mg, 3.5 μmol, 72% yield) as a colorless glassy solid.[7]

TLC (4% MeOH/DCM): R_f = 0.22 (*p*-anisaldehyde)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.36 (t, *J* = 1.8 Hz, 1H), 4.77 (s, 1H), 4.55 (t, *J* = 7.0 Hz, 1H), 3.87 (dd, *J* = 11.3, 5.5 Hz, 2H), 3.78 (s, 1H), 3.44 (s, 1H), 3.38 (ddd, *J* = 17.6, 11.4, 3.4 Hz, 2H), 3.01 – 2.95 (m, 4H), 2.91 (dd, *J* = 16.7, 6.2 Hz, 2H), 2.80 (dt, *J* = 17.9, 5.9 Hz, 2H), 2.63 (d, *J* = 11.2 Hz, 1H), 2.59 – 2.55 (m, 1H), 2.54 – 2.47 (m, 3H), 2.09 – 2.05 (m, 1H), 2.04 (d, *J* = 7.4 Hz, 2H), 2.00 – 1.97 (m, 2H), 1.96 (d, *J* = 1.7 Hz, 1H), 1.93 (d, *J* = 5.8 Hz, 2H), 1.89 (dt, *J* = 9.0, 4.3 Hz, 2H), 1.86 – 1.79 (m, 3H), 1.78 – 1.71 (m, 10H), 1.66 (d, *J* = 8.1 Hz, 7H), 1.62 (d, *J* = 7.0 Hz, 5H), 1.60 – 1.57 (m, 3H), 1.50 – 1.45 (m, 2H), 1.39 (d, *J* = 7.2 Hz, 4H), 1.36 (s, 4H), 1.18 (s, 3H), 1.14 (s, 3H), 1.13 (d, *J* = 6.9 Hz, 7H), 1.02 (d, *J* = 6.7 Hz, 4H), 0.99 (d, *J* = 3.6 Hz, 3H), 0.98 (s, 3H), 0.96 (d, *J* = 2.3 Hz, 4H), 0.80 (s, 3H).

^{1}H NMR (400 MHz, $(CD_3)_2CO$): δ 5.36 (t, *J* = 1.8 Hz, 1H), 4.75 (dd, *J* = 2.0, 0.9 Hz, 1H), 4.43 (t, *J* = 7.0 Hz, 1H), 3.85 (s, 1H), 3.84 – 3.80 (m, 1H), 3.79 – 3.74 (m, 1H), 3.67 (d, *J* = 1.8 Hz, 1H), 3.53 (d, *J* = 5.1 Hz, 1H), 3.44 (s, 1H), 3.40 (t, *J* = 5.7 Hz, 1H), 3.36 (dd, *J* = 11.5, 2.6 Hz, 1H), 2.90 – 2.84 (m, 2H), 2.78 (s, 2H), 2.73 (t, *J* = 5.5 Hz, 1H), 2.59 (dd, *J* = 9.6, 7.5 Hz, 2H), 2.55 (t, *J* = 5.9 Hz, 2H), 2.52 – 2.48 (m, 1H), 1.98 – 1.94 (m, 2H), 1.92 – 1.87 (m, 4H), 1.86 – 1.82 (m, 2H), 1.82 – 1.79 (m, 2H), 1.75 – 1.70 (m, 3H), 1.68 (ddd, *J* = 10.6, 4.1, 1.8 Hz, 4H), 1.61 (q, *J* = 5.5, 4.6 Hz, 3H), 1.56 (dd, *J* = 12.9, 2.0 Hz, 3H), 1.48 – 1.44 (m, 2H), 1.43 – 1.39 (m, 2H), 1.36 (ddd, *J* = 12.2, 4.4, 2.4 Hz, 3H), 1.14 (s, 4H), 1.12 (s, 3H), 1.09 (d, *J* = 7.0 Hz, 4H), 1.06 (s, 3H), 0.97 (d, *J* = 6.7 Hz, 8H), 0.87 – 0.85 (m, 4H), 0.81 – 0.79 (m, 4H).

^{13}C NMR (101 MHz, $(CD_3)_2CO$): δ 155.08, 149.39, 149.14, 149.08, 148.83, 120.15, 117.32, 108.18, 94.03, 93.81, 81.78, 80.08, 76.06, 72.37, 70.23, 66.24, 57.78, 56.01, 53.70, 48.83, 48.35, 46.65, 46.28, 45.92, 42.33, 42.28, 42.15, 38.07, 36.78, 36.29, 36.17, 36.07,

34.44, 33.52, 33.24, 33.06, 32.33, 30.88, 29.06, 28.92, 28.77, 27.29, 26.52, 14.66, 13.52, 12.71, 12.08, 11.98, 7.86.

2.8 REFERENCES

(1) Fukuzawa, S.; Matsunaga, S.; Fusetani, N. Isolation and Structure Elucidation of Ritterazines B and C, Highly Cytotoxic Dimeric Steroidal Alkaloids, from the Tunicate Ritterella Tokioka. *J. Org. Chem.* **1995**, *60*, 608–614.

(2) Moser, B. R. Review of Cytotoxic Cephalostatins and Ritterazines: Isolation and Synthesis. *J. Nat. Prod.* **2008**, *71*, 487–491.

(3) Komiya, T.; Fusetani, N.; Matsunaga, S.; Kubo, A.; Kaye, F. J.; Kelley, M. J.; Tamura, K.; Yoshida, M.; Fukuoka, M.; Nakagawa, K. Ritterazine B, a New Cytotoxic Natural Compound, Induces Apoptosis in Cancer Cells. *Cancer Chemother. Pharmacol.* **2003**, *51*, 202–208.

(4) Burgett, A. W. G.; Poulsen, T. B.; Wangkanont, K.; Anderson, D. R.; Kikuchi, C.; Shimada, K.; Okubo, S.; Fortner, K. C.; Mimaki, Y.; Kuroda, M.; Murphy, J. P.; Schwalb, D. J.; Petrella, E. C.; Cornella-Taracido, I.; Schirle, M.; Tallarico, J. A.; Shair, M. D. Natural Products Reveal Cancer Cell Dependence on Oxysterol-Binding Proteins. *Nat. Chem. Biol.* **2011**, *7*, 639–647.

(5) The structure of the eastern steroid was reassigned by Shair and coworkers. Phillips, S. T.; Shair, M. D. Syntheses of the Eastern Halves of Ritterazines B, F, G, and H, Leading to Reassignment of the 5,5-Spiroketal Stereochemistry of Ritterazines B and F. *J. Am. Chem. Soc.* **2007**, *129*, 6589–6598.

(6) Fukuzawa, S.; Matsunaga, S.; Fusetani, N. Isolation of 13 New Ritterazines from the Tunicate Ritterella Tokioka and Chemical Transformation of Ritterazine B1. *J. Org. Chem.* **1997**, *62*, 4484–4491.

(7) Budzikiewicz, H.; Flessner, T.; Jautelat, R.; Scholz, U.; Winterfeldt, E. *Progress in the Chemistry of Organic Natural Products*; Springer Vienna: Vienna, **2004**, 1–80.

(8) Imperatore, C.; Aiello, A.; D'Aniello, F.; Senese, M.; Menna, M. Alkaloids from Marine Invertebrates as Important Leads for Anticancer Drugs Discovery and Development. *Molecules* **2014**, *19*, 20391–20423.

(9) López-Antón, N.; Rudy, A.; Barth, N.; Schmitz, L. M.; Pettit, G. R.; Schulze-Osthoff, K.; Dirsch, V. M.; Vollmar, A. M. The Marine Product Cephalostatin 1 Activates an Endoplasmic Reticulum Stress-Specific and Apoptosome-Independent Apoptotic Signaling Pathway. *J. Biol. Chem.* **2006**, *281*, 33078–33086.

(10) Ambrose, A. J.; Santos, E. A.; Jimenez, P. C.; Rocha, D. D.; Wilke, D. V.; Beuzer, P.; Axelrod, J.; Kumar Kanduluru, A.; Fuchs, P. L.; Cang, H.; Costa-Lotufo, L. V.; Chapman, E.; La Clair, J. J. Ritterostatin GN1N, a Cephalostatin–Ritterazine Bis-Steroidal Pyrazine Hybrid, Selectively Targets GRP78. *ChemBioChem* **2017**, *18*, 506–510.

(11) (a) Jeong, J. U.; Fuchs, P. L. Synthesis of a 17-Deoxy, C-14,15-Dihydro Derivative of the North Spiroketal Moiety of the Cephalostatins. Conversion to a (+)-Trisdecacyclic C2 Symmetrical Pyrazine. *J. Am. Chem. Soc.* **1994**, *116*, 773–774. (b) Heathcock, C. H.; Smith, S. C. Synthesis and Biological Activity of Unsymmetrical Bis-Steroidal Pyrazines Related to the Cytotoxic Marine Natural Product Cephalostatin 1. *J. Org. Chem.* **1994**, *59*, 6828–6839. (c) Jeong, J. U.; Sutton, S. C.; Kim, S.; Fuchs, P. L. Biomimetic Total Syntheses of (+)-Cephalostatin 7, (+)-Cephalostatin 12, and (+)-Ritterazine K. *J. Am. Chem. Soc.* **1995**, *117*, 10157–10158. (d) Guo, C.; Bhandaru, S.; Fuchs, P. L.; Boyd, M. R. An Efficient Protocol for the Synthesis of Unsymmetrical Pyrazines. Total Synthesis of Dihydrocephalostatin 11. *J. Am. Chem. Soc.* **1996**, *118*, 10672–10673. (e) LaCour, T. G.; Guo, C.; Bhandaru, S.; Boyd, M. R.; Fuchs, P. L. Interphylal Product Splicing: The First Total Syntheses of Cephalostatin 1, the North Hemisphere of Ritterazine G, and the Highly Active Hybrid Analogue, Ritterostatin GN1N1. *J. Am. Chem. Soc.* **1998**, *120*, 692–707. (f) Jeong, J. U.; Guo, C.; Fuchs, P. L. Synthesis of the South Unit of Cephalostatin. 7. Total Syntheses of (+)-Cephalostatin 7, (+)-Cephalostatin 12, and (+)-Ritterazine K1. *J. Am. Chem. Soc.* **1999**, *121*, 2071–2084. (g) Kim, S.; Sutton, S. C.; Guo, C.; LaCour, T. G.; Fuchs, P. L. Synthesis of the North 1 Unit of the Cephalostatin Family from Hecogenin Acetate1. *J. Am. Chem. Soc.* **1999**, *121*, 2056–2070. (h) Lee, S.; Fuchs, P. L. The First Total Synthesis of (Corrected) Ritterazine M. *Org. Lett.* **2002**, *4*, 317–318. (i) Fortner, K. C.; Kato, D.; Tanaka, Y.; Shair, M. D. Enantioselective Synthesis of (+)-Cephalostatin 1. *J. Am. Chem. Soc.* **2010**, *132*, 275–280.

(12) (a) Liu, B.; De Brabander, J. K. Metal-Catalyzed Regioselective Oxy-Functionalization of Internal Alkynes: An Entry into Ketones, Acetals, and Spiroketals. *Organic Letters* **2006**, *8*, 4907–4910. (b) Tlais, S. F.; Dudley, G. B. A

Gold-Catalyzed Alkyne-Diol Cycloisomerization for the Synthesis of Oxygenated 5,5-Spiroketals. *Beilstein J. Org. Chem.* **2011**, *7*, 570–577. (c) Goodwin, J. A.; Aponick, A. Regioselectivity in the Au-Catalyzed Hydration and Hydroalkoxylation of Alkynes. *Chem. Commun.* **2015**, *51*, 8730–8741. (d) Pflästerer, D.; Rudolph, M.; Hashmi, A. S. K. Gold-Catalyzed Hydrofunctionalizations and Spiroketalizations of Alkynes as Key Steps in Total Synthesis. *Isr. J. Chem.* **2018**, *58*, 622–638.

(13) (a) Schönecker, B.; Lange, C.; Zheldakova, T.; Günther, W.; Görls, H.; Vaughan, G. Copper-Mediated Regio- and Stereoselective 12β-Hydroxylation of Steroids with Molecular Oxygen and an Unexpected 12β-Chlorination. *Tetrahedron* **2005**, *61*, 103–114. (b) Trammell, R.; See, Y. Y.; Herrmann, A. T.; Xie, N.; Díaz, D. E.; Siegler, Maxime. A.; Baran, P. S.; Garcia-Bosch, I. Decoding the Mechanism of Intramolecular Cu-Directed Hydroxylation of sp^3 C–H Bonds. *J. Org. Chem.* **2017**, *82*, 7887–7904.

(14) Kolb, H. C.; Van Nieuwenhze, M. S.; Sharpless, K. B. Catalytic Asymmetric Dihydroxylation. *Chem. Rev.* **1994**, *94*, 2483–2547.

(15) (a) Downie, I.; Holmes, J.; Lee, J. Preparation of Alkyl Chlorides Under Mild Conditions. *Chemistry and Industry* **1966**, *22*, 900. (b) Appel, R. Tertiary Phosphane/Tetrachloromethane, a Versatile Reagent for Chlorination, Dehydration, and P-N Linkage. *Angew. Chem. Int. Ed.* **1975**, *14*, 801–811.

(16) Nakagawa, T.; Kasatkin, A.; Sato, F. Highly Efficient Synthesis of Propargyl- and Allenyltitanium Reagents from Propargyl Halides or Propargyl Alcohol Derivatives. Practical Synthesis of Allenyl and Homopropargyl Alcohols. *Tetrahedron Lett.* **1995**, *36*, 3207–3210.

(17) (a) Smith, H.; Huges, G. A.; Douglas, G. H.; Wendt, G. R.; Buzby Jr., G. C.; Edgren, R. A.; Fisher, J.; Foell, T.; Gadsby, B.; Hartley, D.; Herbst, D.; Jansen, A. B. A.; Ledig, K.; McLoughlin, B. J.; McMenamin, J.; Pattison, T. W.; Phillips, P. C.; Rees, R.; Siddall, J.; Siuda, J.; Smith, L. L.; Tokolics, J.; Watson, D. H. P. Totally Synthetic Steroid Hormones. Part II. 13b-Alkylgona-1,3,5(10)-Trienes, 13b-Alkygon-4-En-3-Ones, and Related Compounds. *J. Chem. Soc.* **1964**, 4472–4492. (b) McKinney, A. R.; Suann, C. J.; Stenhouse, A. M. A Stereochemical Examination

of the Equine Metabolism of 17α-Methyltestosterone. *Anal. Chim. Acta* **2007**, *581*, 377–387.

(18) Noshita, T.; Sugiyama, T.; Kitazumi, Y.; Oritani, T. Phenolic Ferrier Reaction and Its Application to the Natural Product Synthesis. *Tetrahedron Lett.* **1994**, *35*, 8259–8262.

(19) See the Experimental Section 2.7.

(20) Calculated using the Entos software package at the BLYP/6-31G** theory level. See Manby, F.; Miller, T.; Bygrave, P.; Ding, F.; Dresselhaus, T.; Batista-Romero, F.; Buccheri, A.; Bungey, C.; Lee, S.; Meli, R.; Miyamoto, K.; Steinmann, C.; Tsuchiya, T.; Welborn, M.; Wiles, T.; Williams, Z. Entos: A Quantum Molecular Simulation Package. **2019**, *ChemRxiv*, No. 7762646.v2 (accessed 2019-09-17). It is possible that the ground state energy differences do not arise purely from the proposed steric interactions; there may be electronic contributions from the spiroketal to stabilization of the observed configuration.

(21) Chen, R.; Shen, Y.; Yang, S.; Zhang, Y. Conformational Design Principles in Total Synthesis. *Angew. Chem., Int. Ed.* **2020**, *59*, 14198–14210.

(22) (a) Krumpolc, M.; Rocek, J. Chromium(V) Oxidations of Organic Compounds. *Inorg. Chem.* **1985**, *24*, 617–621. (b) Wilde, N. C.; Isomura, M.; Mendoza, A.; Baran, P. S. Two-Phase Total Synthesis of (–)-Taxuyunnanine D. *J. Am. Chem. Soc.* **2014**, *136*, 4909–4912.

(23) (a) Turlik, A.; Chen, Y.; Scruse, A. C.; Newhouse, T. R. Convergent Total Synthesis of Principinol D, a Rearranged Kaurane Diterpenoid. *J. Am. Chem. Soc.* **2019**, *141*, 8088-8092. (b) Banwell, M. G.; Hockless, D. C. R.; McLeod, M. D. Chemoenzymatic Total Syntheses of the Sesquiterpene (−)-patchoulenone. *New J. Chem.* **2003**, *27*, 50−59. (c) For a review, see: (a) Szostak, M.; Spain, M.; Procter, D. J. Recent Advances in the Chemoselective Reduction of Functional Groups Mediated by Samarium(II) Iodide: A Single Electron Transfer Approach. *Chem. Soc. Rev.* **2013**, *42*, 9155–9183.

(24) Iwasaki, K.; Wan, K. K.; Oppedisano, A.; Crossley, S. W. M.; Shenvi, R. A. Simple, Chemoselective Hydrogenation with Thermodynamic Stereocontrol. *J. Am. Chem. Soc.* **2014**, *136*, 1300–1303.

(25) Staudinger, H.; Meyer, J. Über neue organische Phosphorverbindungen III. Phosphinmethylenderivate und Phosphinimine. *Helv. Chim. Acta* **1919**, *2*, 635–646.

(26) Fukizawa, S.; Matsunaga, S.; Fusetani, N. Ten More Ritterazines, Cytotoxic Steroidal Alkaloids from the Tunicate *Ritterella tokioka*. *Tetrahedron* **1995**, *51*, 6707–6716.

(27) Kramer, A.; Ullmann, U.; Winterfeldt, E. A Short Route to Cephalostatin Analogues. *J. Chem. Soc. Perkin. Trans. 1* **1993**, *23*, 2865–2867.

(28) (a) Wang, W.; Hu, S.; Lu, X.; Young, K. H.; Medeiros, L. J. Triple-hit B-cell Lymphoma with MYC, BCL2, and BCL6 Translocations/Rearrangements: Clinicopathologic Features of 11 Cases. *Am. J. Surg. Pathol.* **2015**, *39*, 1132–1139. (b) Zhang, J.; Weng, Z.; Huang, Y.; Li, M.; Wang, F.; Wang, Y.; Rao, H. High-grade B-Cell Lymphoma with MYC, BCL2, and/or BCL6 Translocations/Rearrangements: Clinicopathologic Features of 51 Cases in a Single Institution of South China. *Am. J. Surg. Pathol.* **2020**, *44*, 1602–1611.

(29) Nicolai, S.; Erard, S.; Fernández González, D.; Waser, J. Pd-Catalyzed Intramolecular Oxyalkynylation of Alkenes with Hypervalent Iodine. *Org. Lett.* **2010**, *12*, 384–387.

(30) Tietze, L. F. Preparation of Chiral Building Blocks for a Highly Convergent Vitamin E Synthesis. Systematic Investigations on the Enantioselectivity of the Sharpless Bishydroxylation. *Synthesis* **1998**, *6*, 873–878.

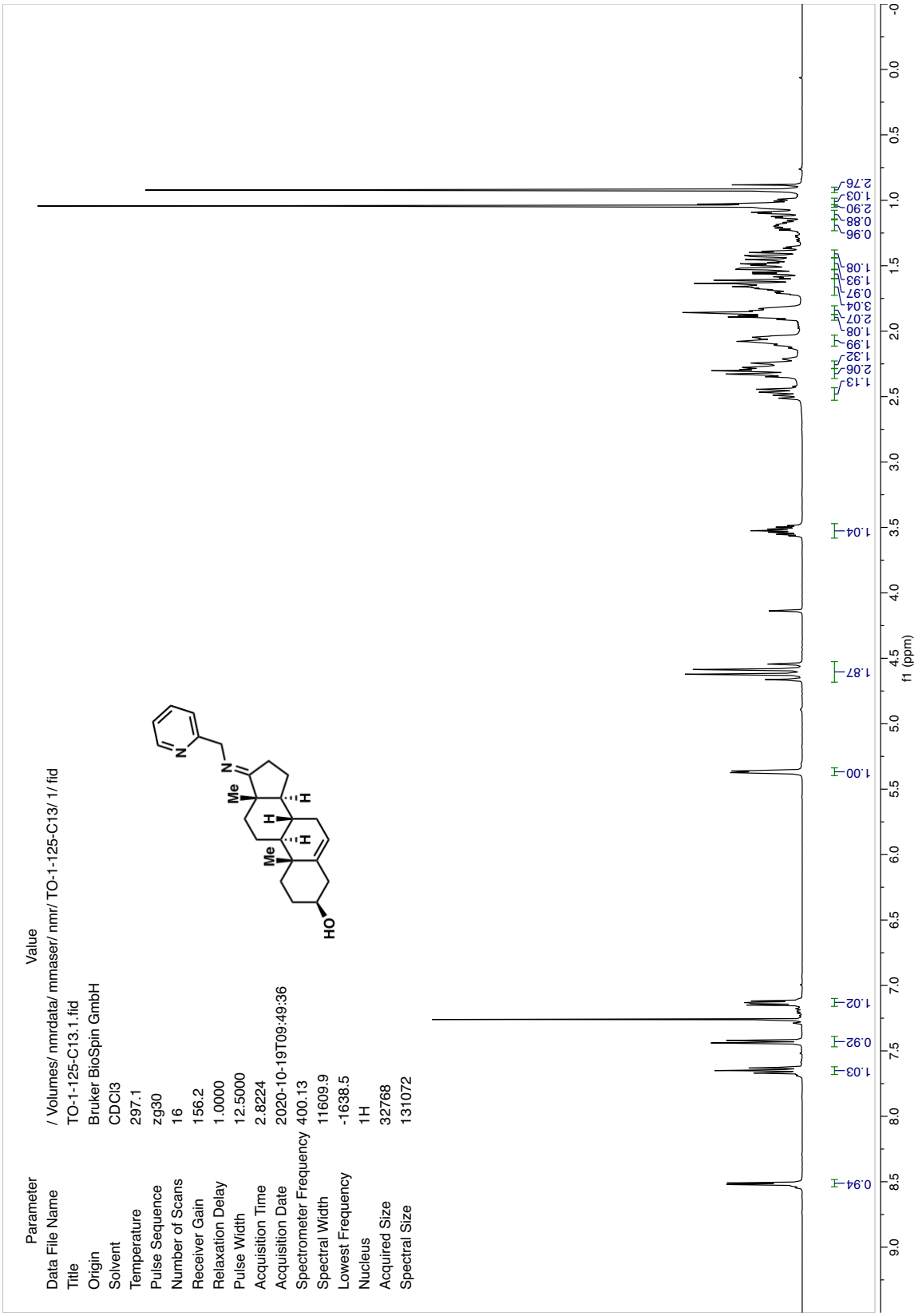
(31) Kloc, K.; Młochowski, J.; Syper, L. Synthesis of Novel Quinones with Silver (II) Dipicolinate as a New Selective Oxidant. *Chem. Lett.* **1980**, *9*, 725–728.

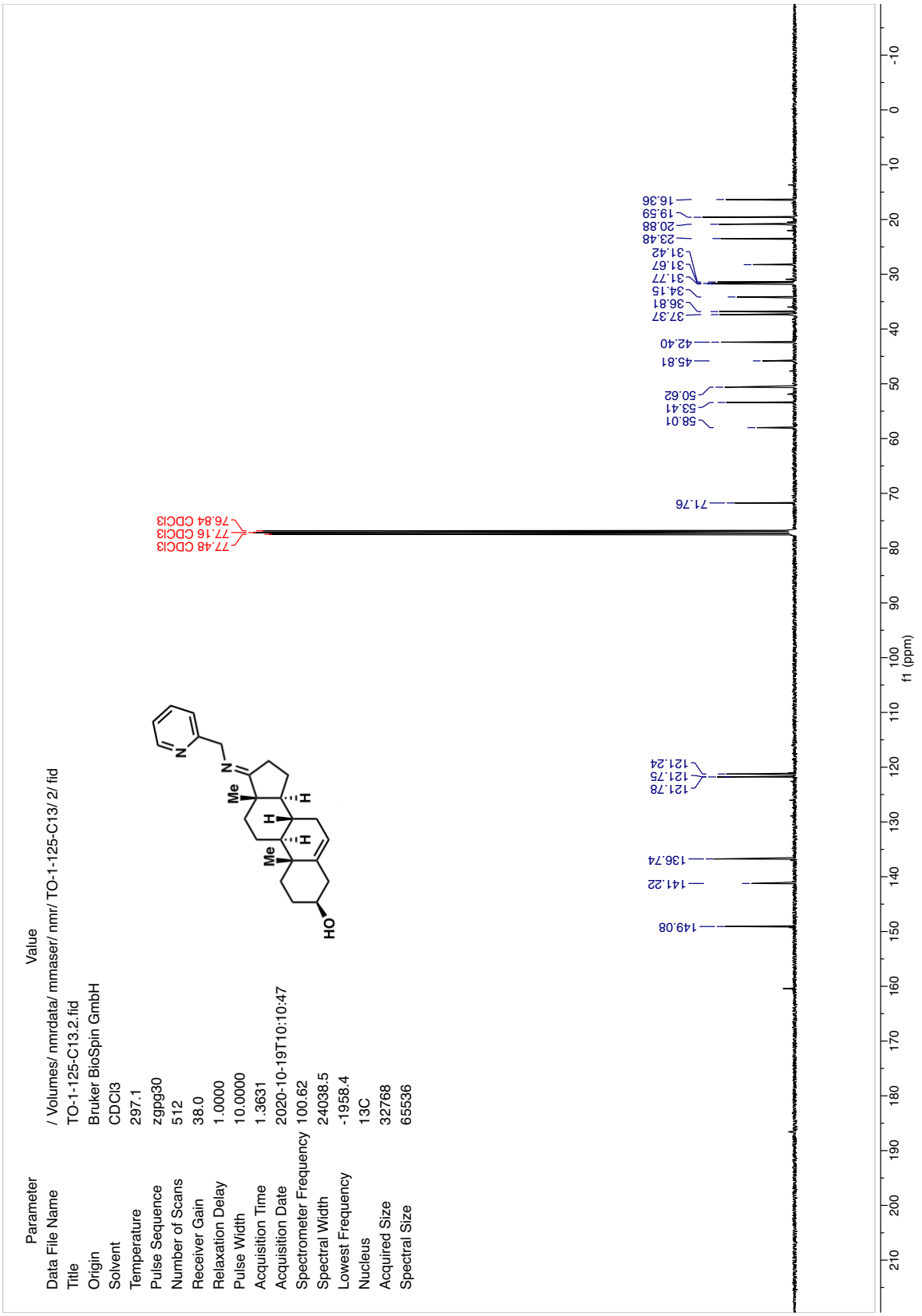
(32) (a) Krumpolc, M.; Rocek, J. Chromium(V) Oxidations of Organic Compounds. *Inorg. Chem.* **1985**, *24*, 617–621. (b) Kanda, Y.; Ishihara, Y.; Wilde, N. C.; Baran, P. S. Two-Phase Total Synthesis of Taxanes: Tactics and Strategies. *J. Org. Chem*. **2020**, *85*, 10293–10320.

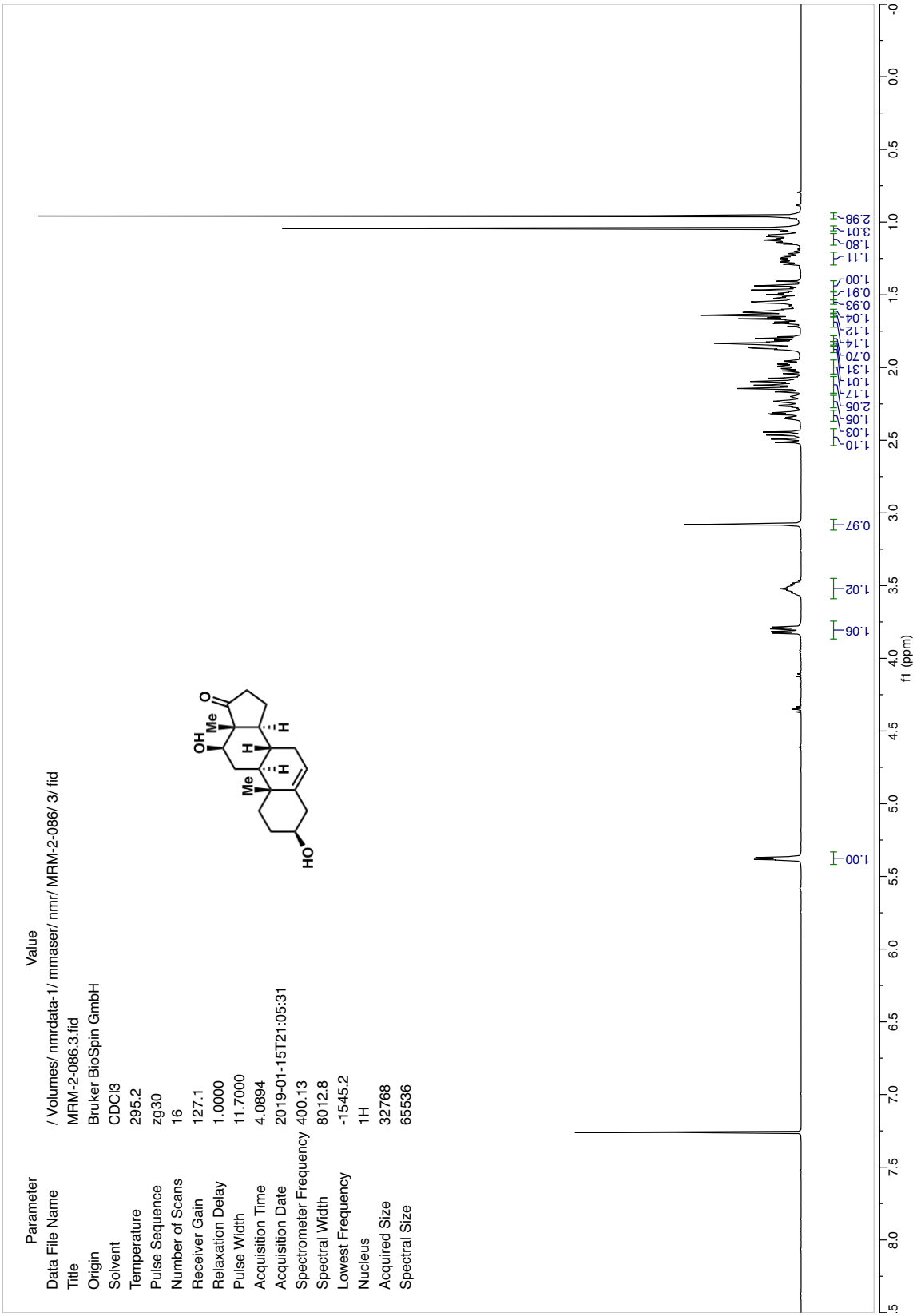
Appendix 2

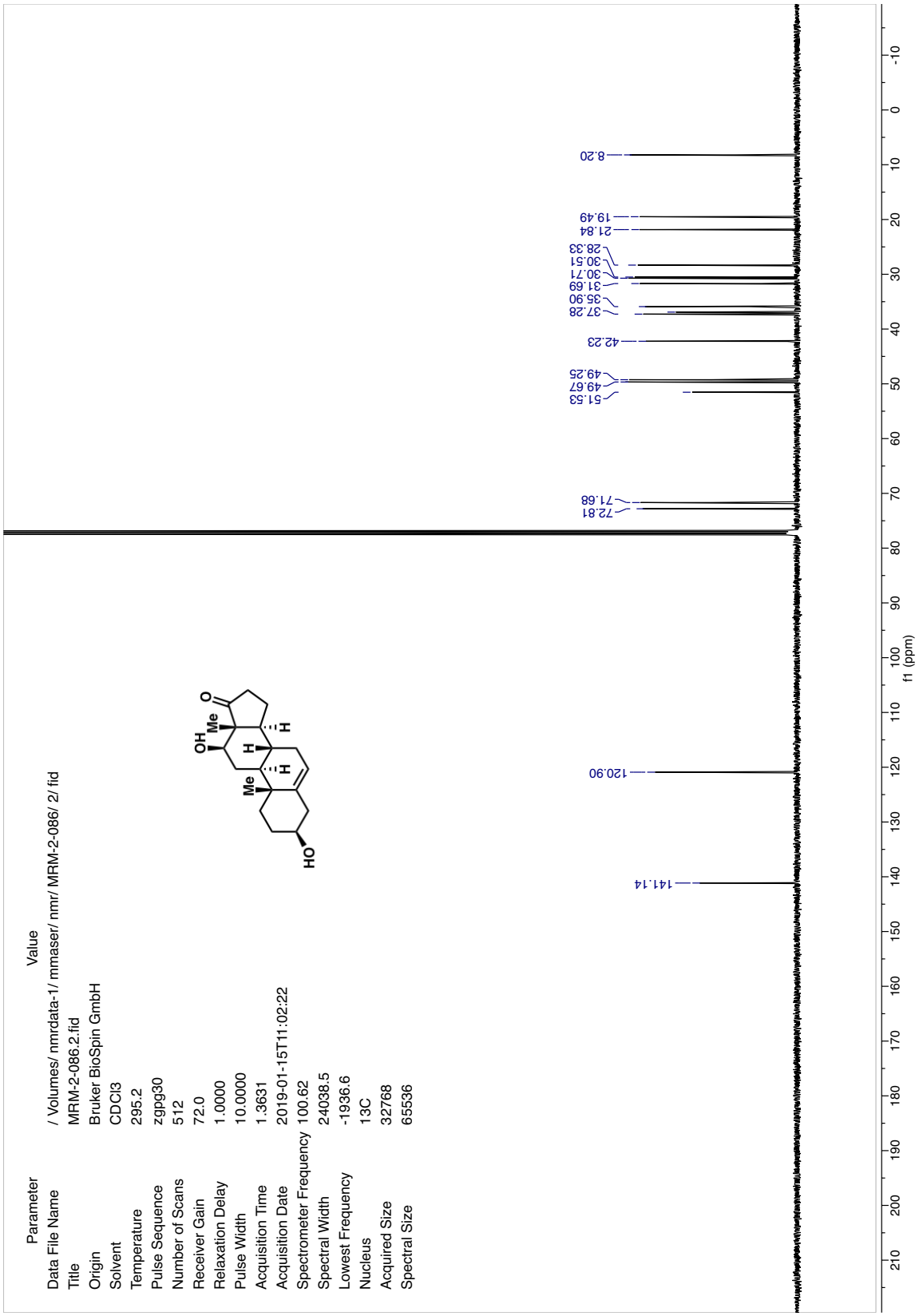
Spectra Relevant to Chapter 2:

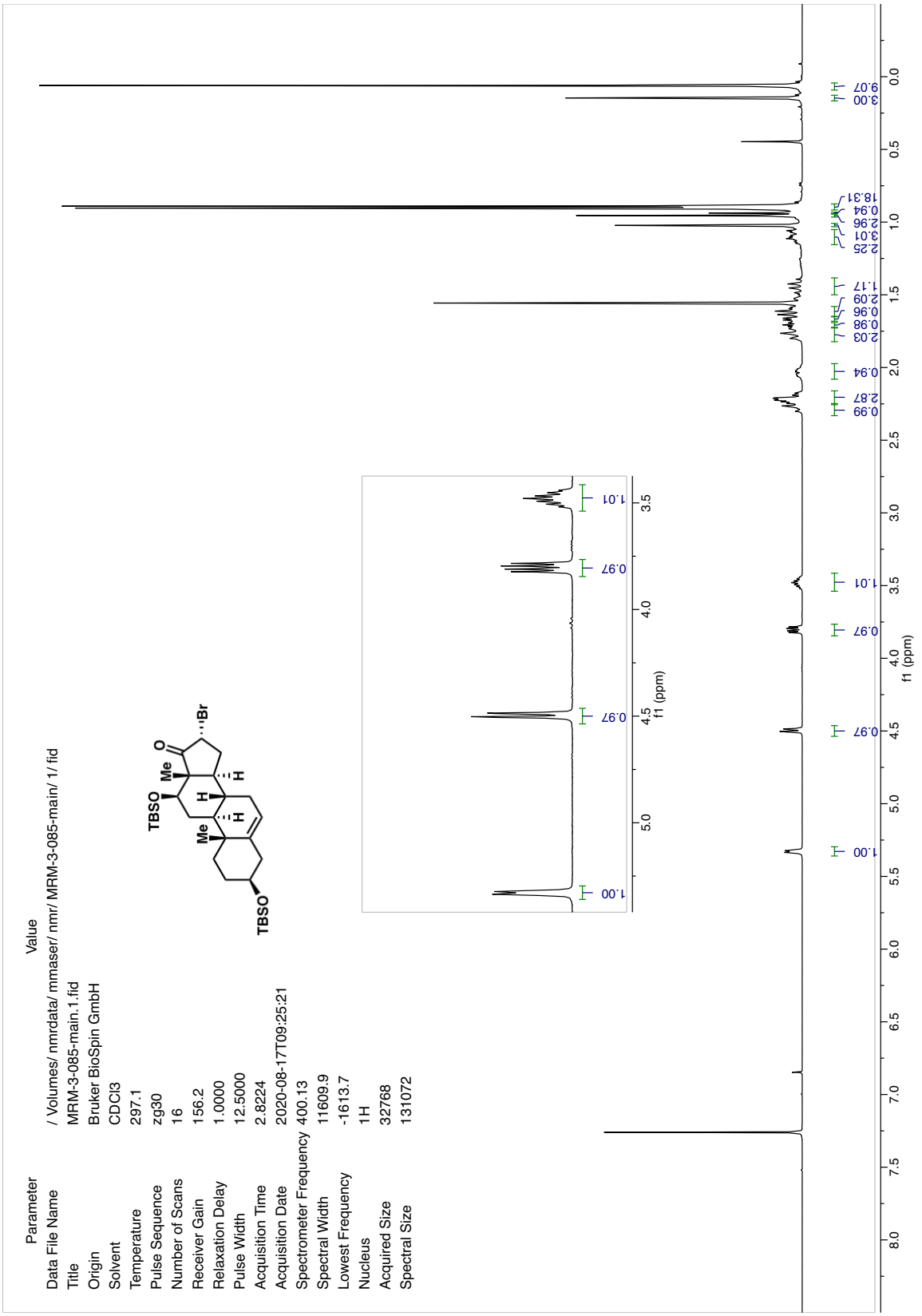
Total Synthesis of Ritterazine B and Analogs for the Treatment of Cancer

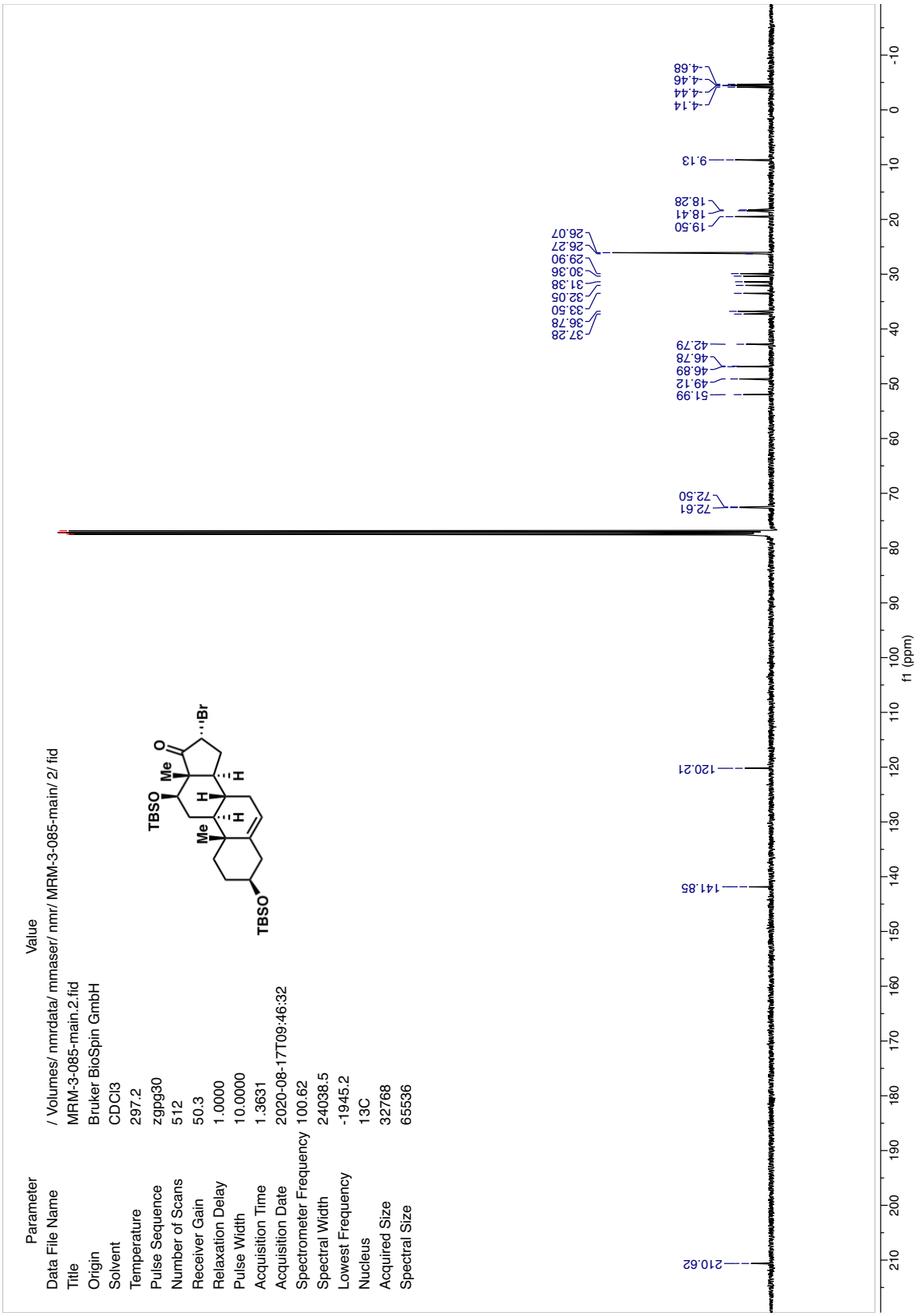

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ TO-1-125-C13/ 1/ fid
Title TO-1-125-C13.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-10-19T09:49:36
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1638.5
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
Me
Me
H
H
H
N
N
HO
0.94
1.03
0.92
1.02
1.00
1.87
1.04
1.13
2.06
1.32
1.99
1.08
2.07
3.04
0.97
1.93
1.08
0.96
0.88
2.90
1.03
2.76
f1 (ppm)

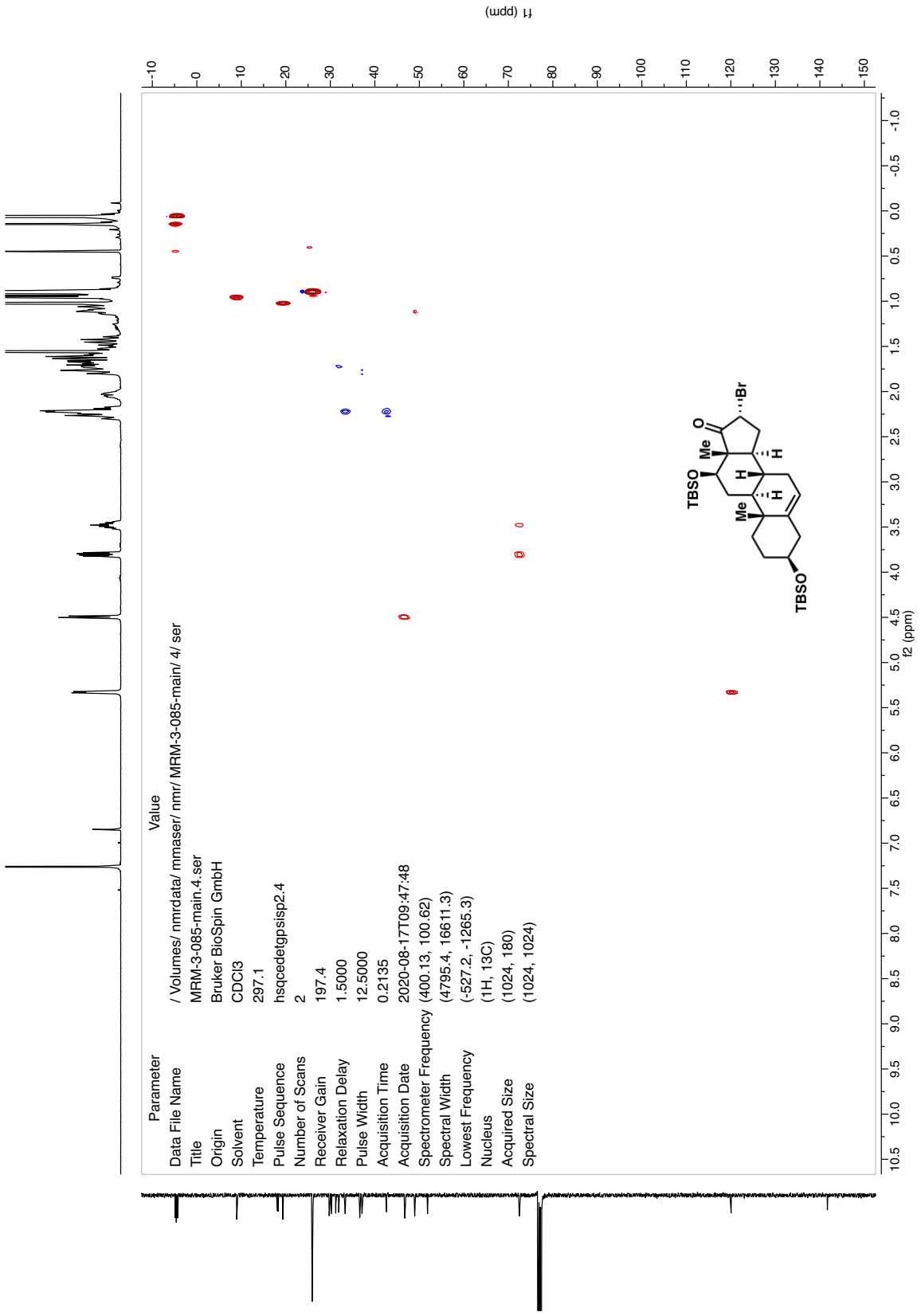

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ TO-1-125-C13/ 2/ fid
Title TO-1-125-C13.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 38.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-10-19T10:10:47
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1958.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
16.36
19.59
20.88
23.48
31.42
31.67
31.77
34.15
36.81
37.37
42.40
45.81
50.62
53.41
58.01
71.76
121.24
121.75
121.78
136.74
141.22
149.08
f1 (ppm)

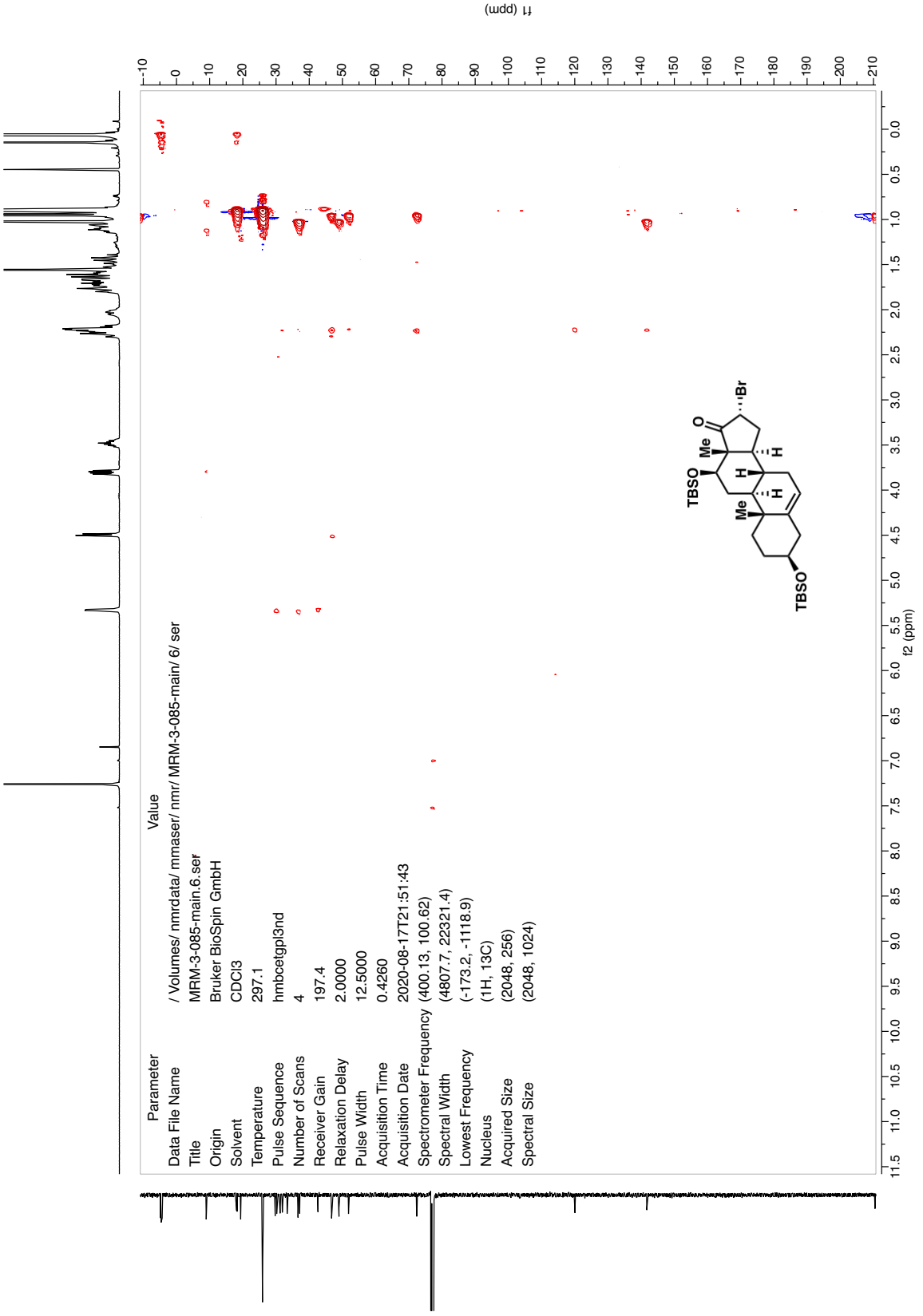

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-2-086/ 3/ fid
Title MRM-2-086.3.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 127.1
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-01-15T21:05:31
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OH
Me
O
H
H
Me
H
HO
1.00
1.06
1.02
0.97
1.10
1.03
1.05
2.05
1.17
1.01
1.31
0.70
1.14
1.12
1.04
0.93
0.91
1.00
1.11
1.80
3.01
2.98
f1 (ppm)

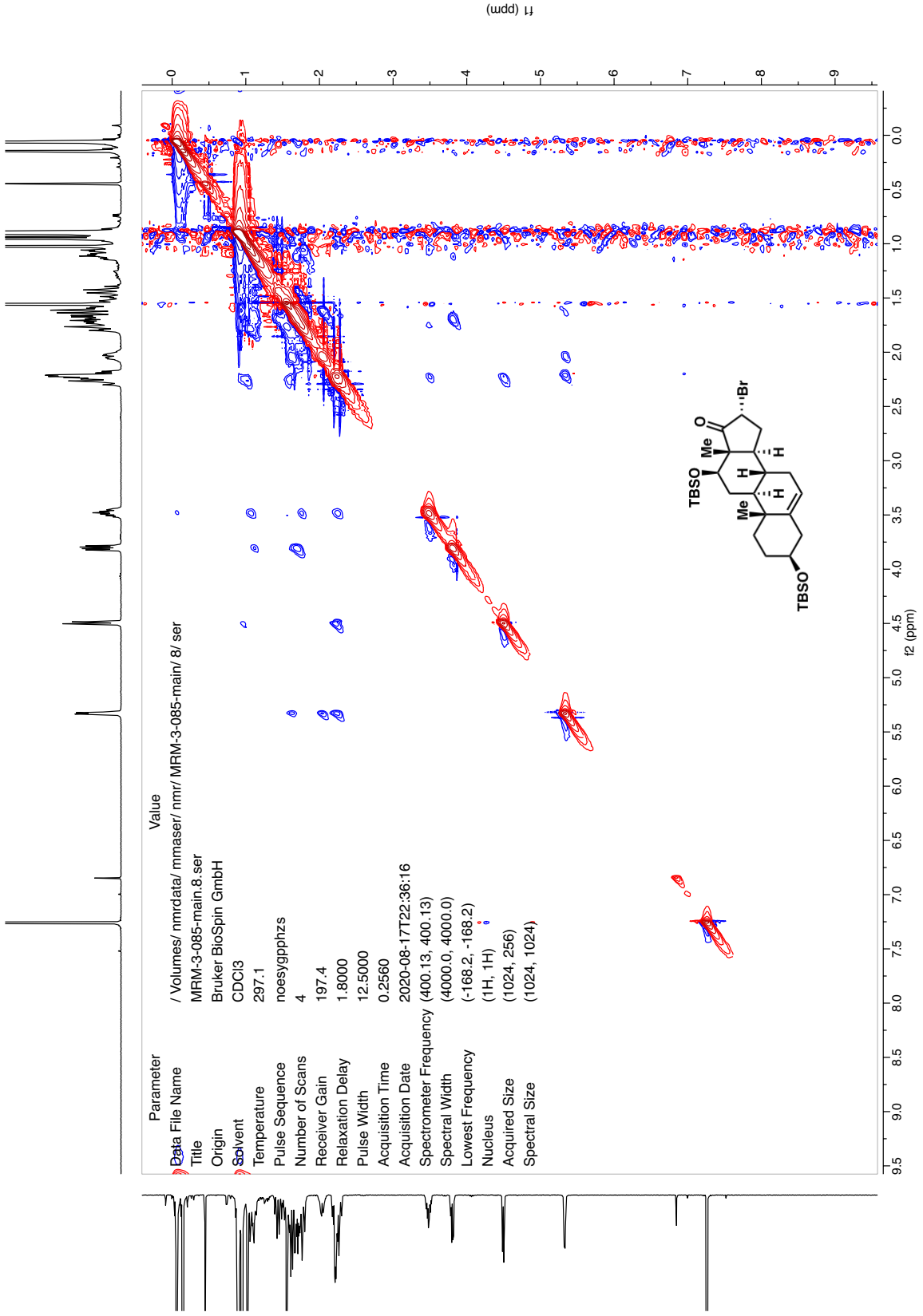

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-2-086/ 2/ fid
Title MRM-2-086.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-01-15T11:02:22
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1936.6
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OH
Me
O
H
Me
H
H
HO
8.20
19.49
21.84
28.33
30.51
30.71
31.69
35.90
37.28
42.23
49.25
49.67
51.53
71.68
72.81
120.90
141.14
f1 (ppm)

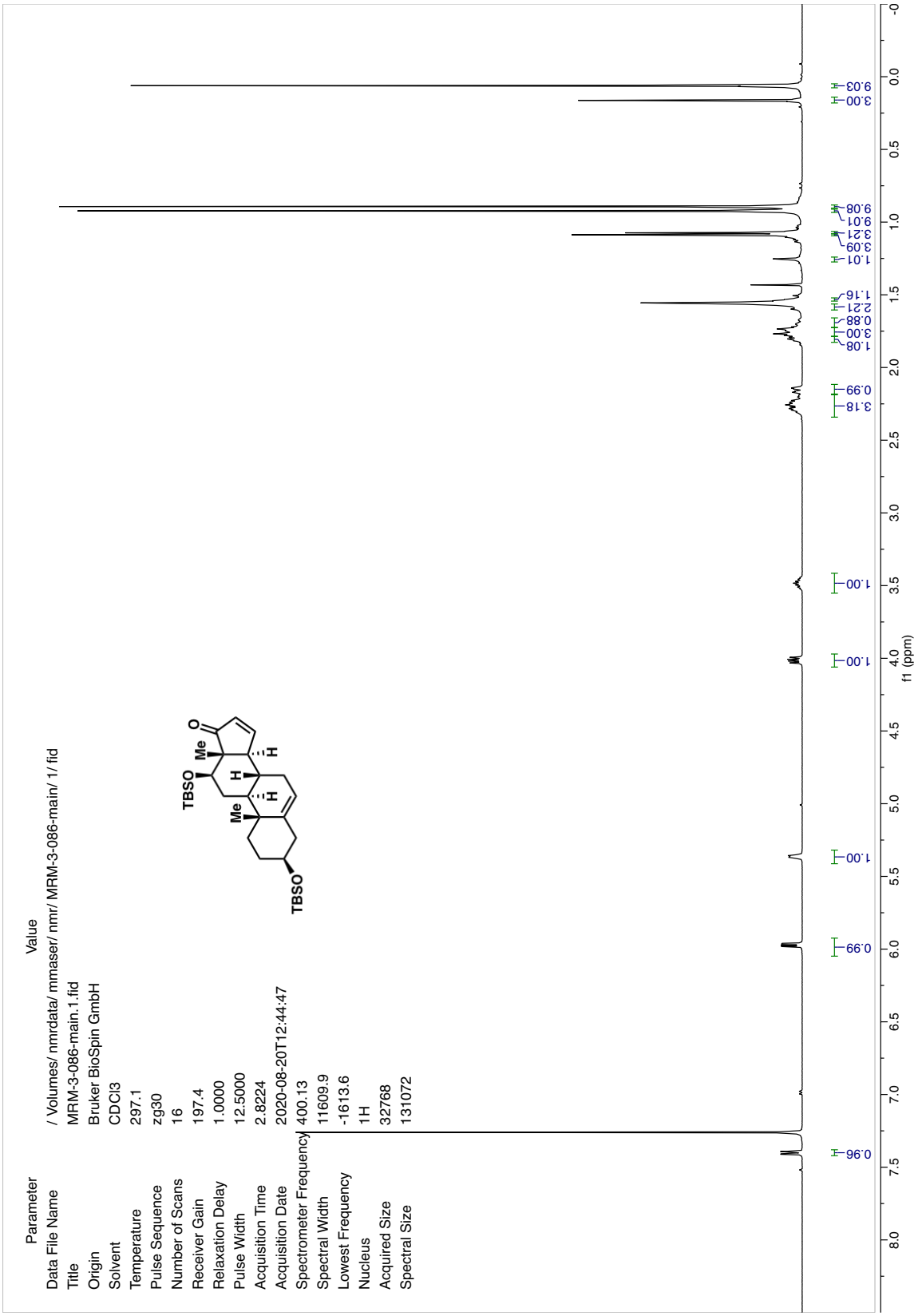

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-085-main/ 1/ fid
Title MRM-3-085-main.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-08-17T09:25:21
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.7
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
Me
O
Br
H
H
Me
H
H
TBSO
f1 (ppm)

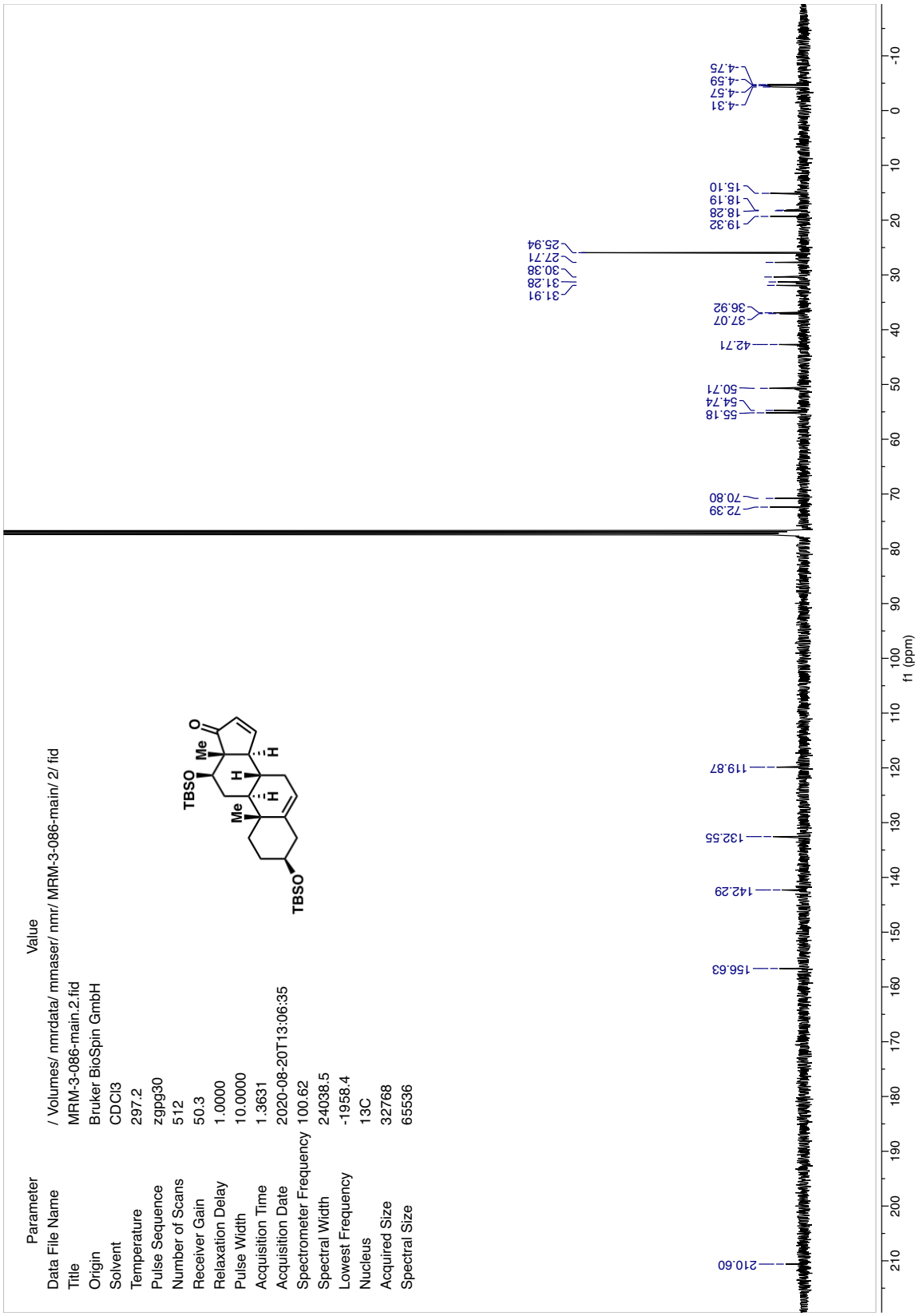

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-085-main/ 2/ fid
Title MRM-3-085-main.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-08-17T09:46:32
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.2
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
O
Br
H
H
H
Me
TBSO
210.62
141.85
120.21
77.28
77.00
76.72
72.61
72.50
51.99
49.12
46.89
46.78
42.79
37.28
36.78
33.50
32.05
31.38
30.36
29.90
26.27
26.07
19.50
18.41
18.28
9.13
4.68
4.46
4.44
4.14
f1 (ppm)

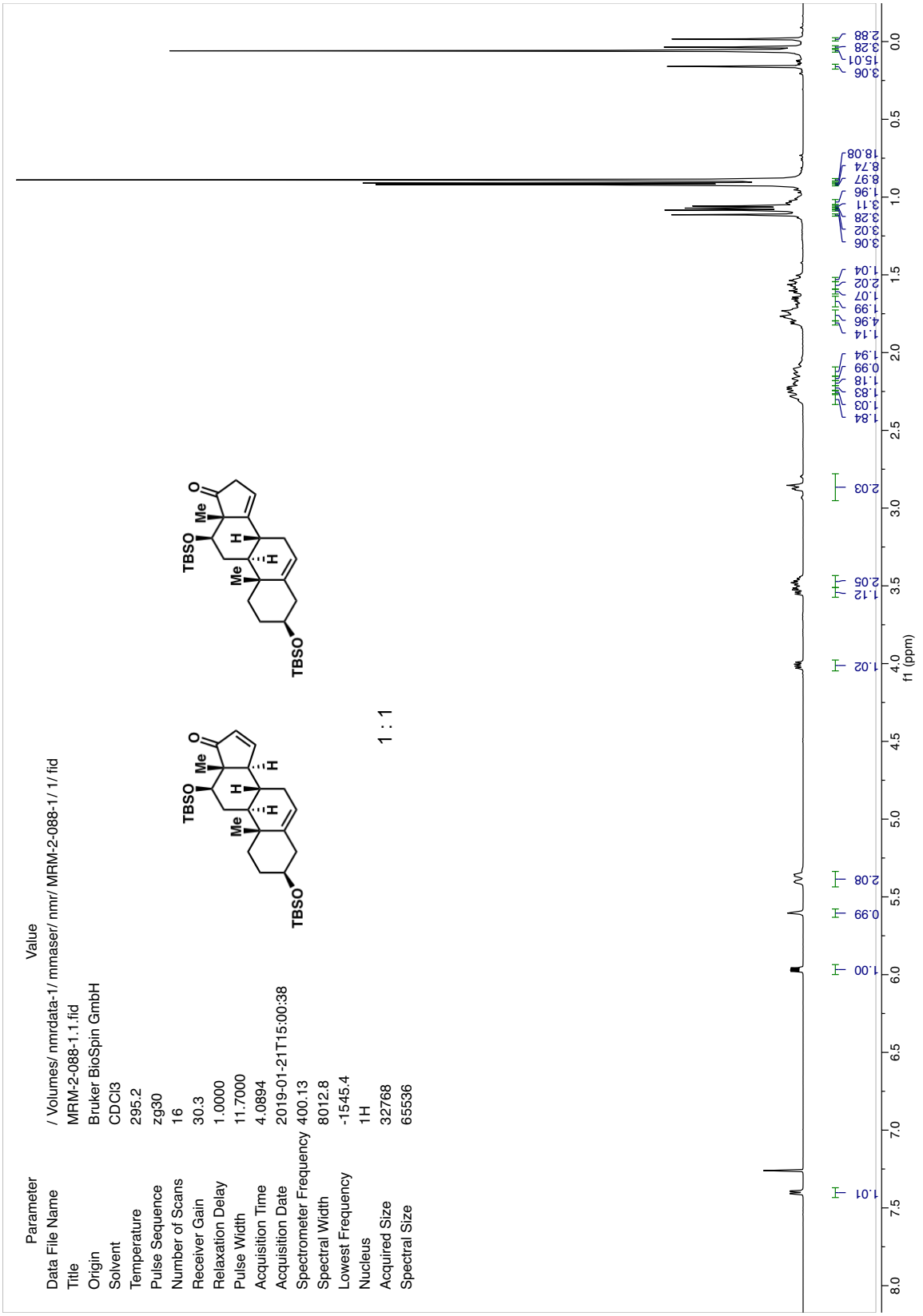

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-085-main/ 4/ ser
Title
MRM-3-085-main.4.ser
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
hsqcedetgpsisp2.4
Number of Scans
2
Receiver Gain
197.4
Relaxation Delay
1.5000
Pulse Width
12.5000
Acquisition Time
0.2135
Acquisition Date
2020-08-17T09:47:48
Spectrometer Frequency
(400.13, 100.62)
Spectral Width
(4795.4, 16611.3)
Lowest Frequency
(-527.2, -1265.3)
Nucleus
(1H, 13C)
Acquired Size
(1024, 180)
Spectral Size
(1024, 1024)
TBSO
Me
O
Br
H
H
H
Me
TBSO
f1 (ppm)
f2 (ppm)

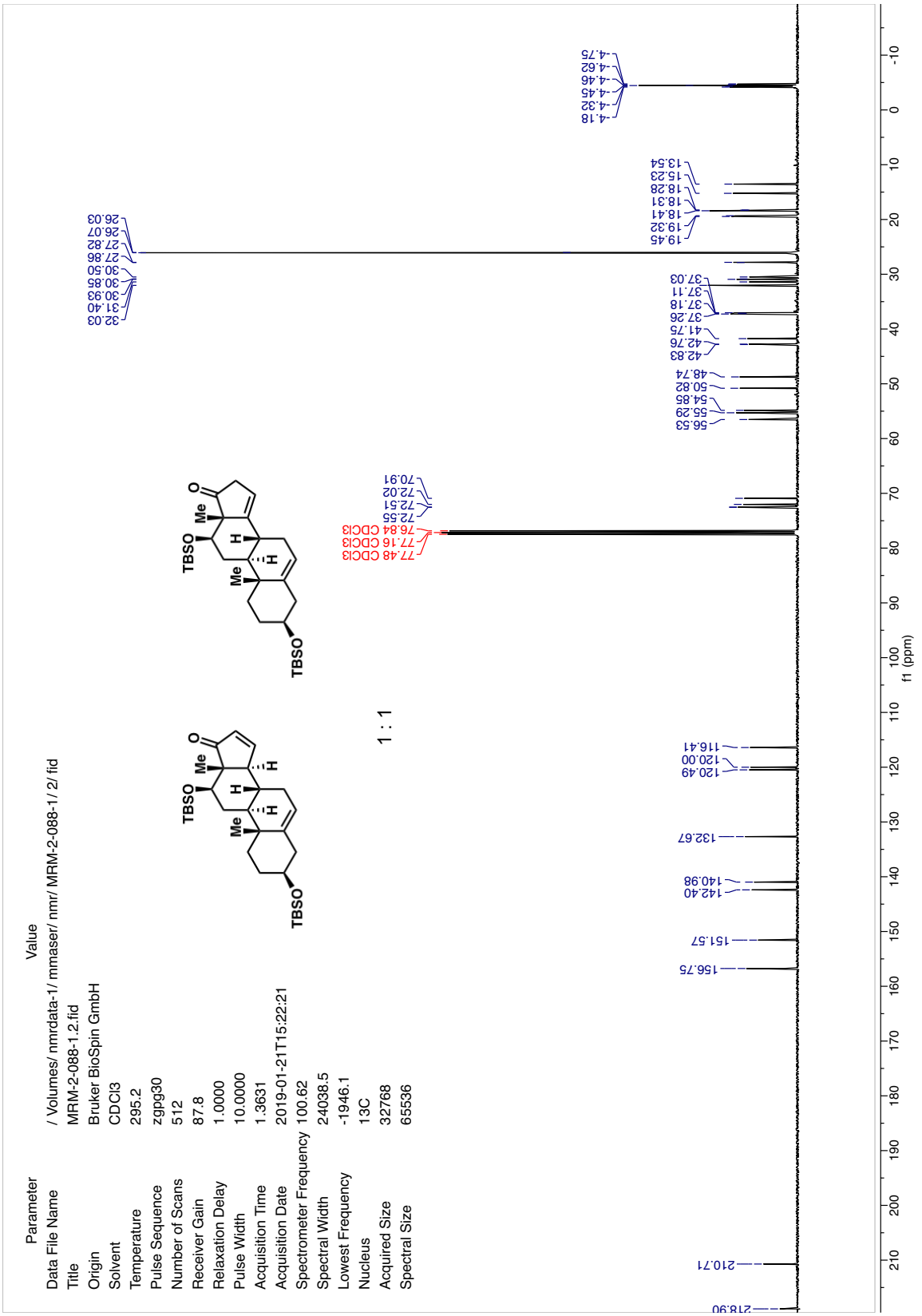

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-085-main/ 6/ ser
Title MRM-3-085-main.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-08-17T21:51:43
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-173.2, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
Me
O
Br
H
H
H
Me
TBSO

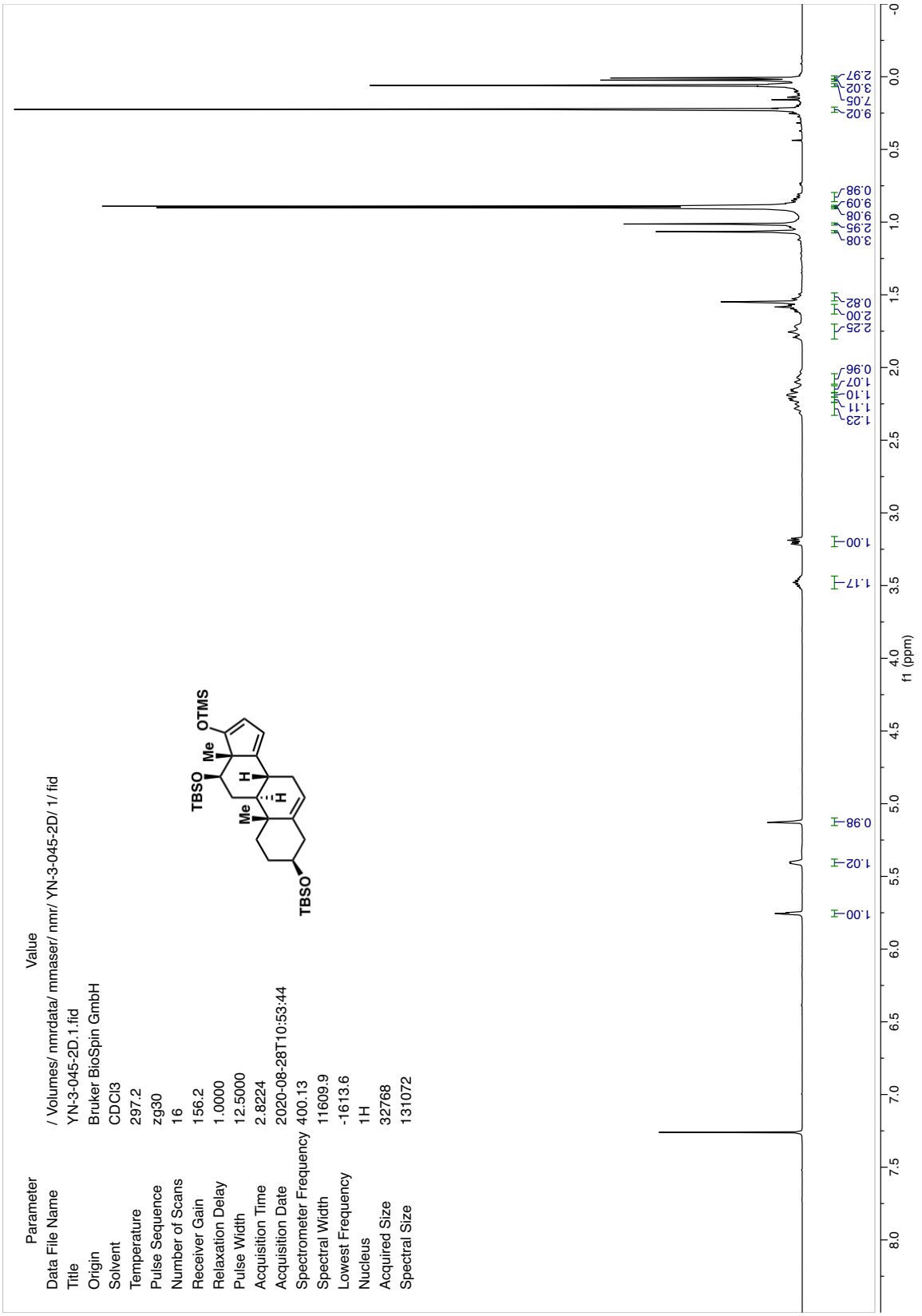

f1 (ppm)
f2 (ppm)
Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-085-main/ 8/ ser
Title
MRM-3-085-main.8.ser
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
noesygpphzs
Number of Scans
4
Receiver Gain
197.4
Relaxation Delay
1.8000
Pulse Width
12.5000
Acquisition Time
0.2560
Acquisition Date
2020-08-17T22:36:16
Spectrometer Frequency
(400.13, 400.13)
Spectral Width
(4000.0, 4000.0)
Lowest Frequency
(-168.2, -168.2)
Nucleus
(1H, 1H)
Acquired Size
(1024, 256)
Spectral Size
(1024, 1024)
TBSO
Me
O
Br
H
H
H
Me
TBSO

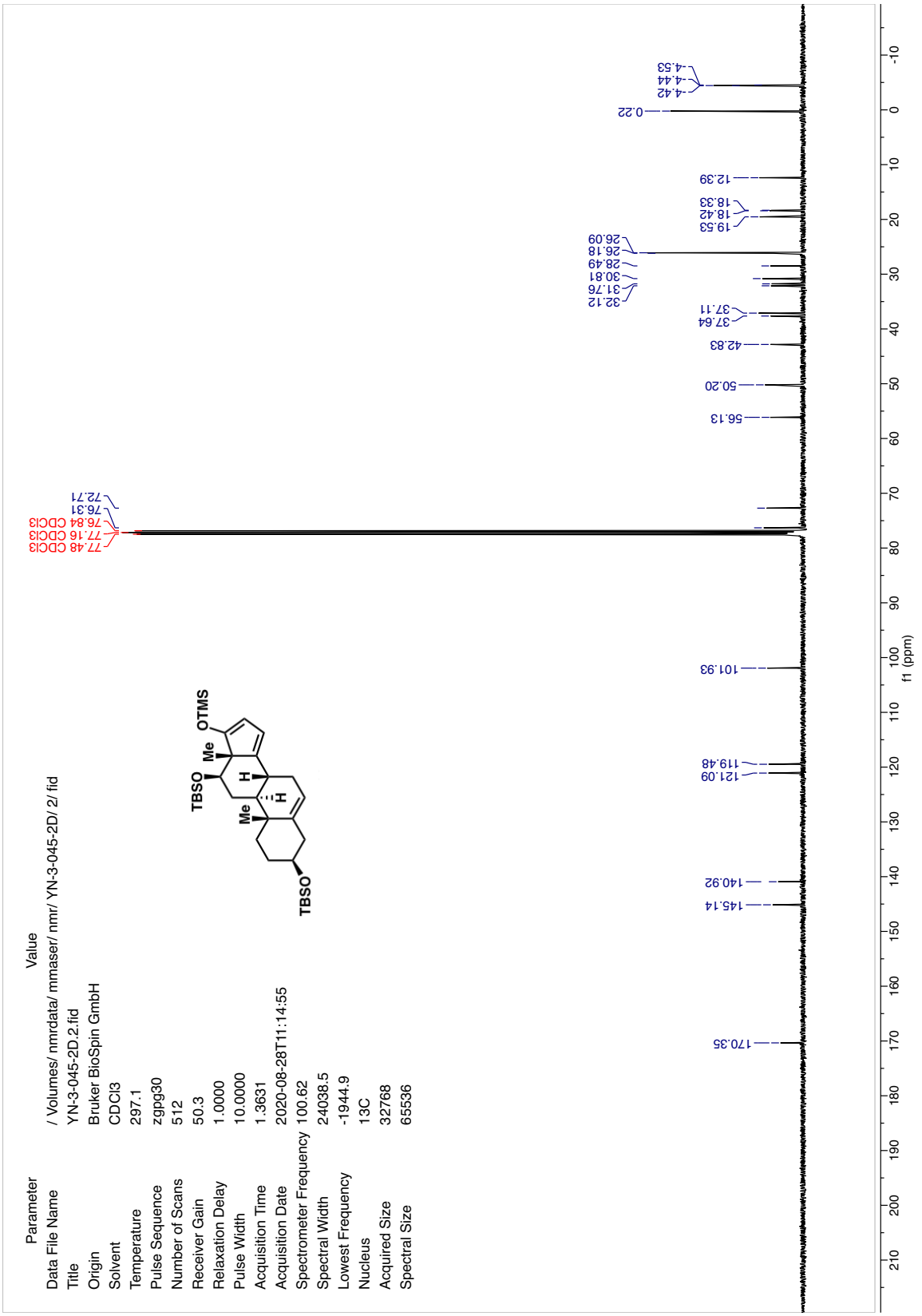

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-086-main/ 1/ fid
Title
MRM-3-086-main.1.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
zg30
Number of Scans
16
Receiver Gain
197.4
Relaxation Delay
1.0000
Pulse Width
12.5000
Acquisition Time
2.8224
Acquisition Date
2020-08-20T12:44:47
Spectrometer Frequency
400.13
Spectral Width
11609.9
Lowest Frequency
-1613.6
Nucleus
1H
Acquired Size
32768
Spectral Size
131072
TBSO
Me
O
H
H
Me
H
TBSO
9.03
3.00
9.08
9.01
3.21
3.09
1.01
1.16
2.21
0.88
3.00
1.08
0.99
3.18
1.00
1.00
1.00
0.99
0.96
f1 (ppm)

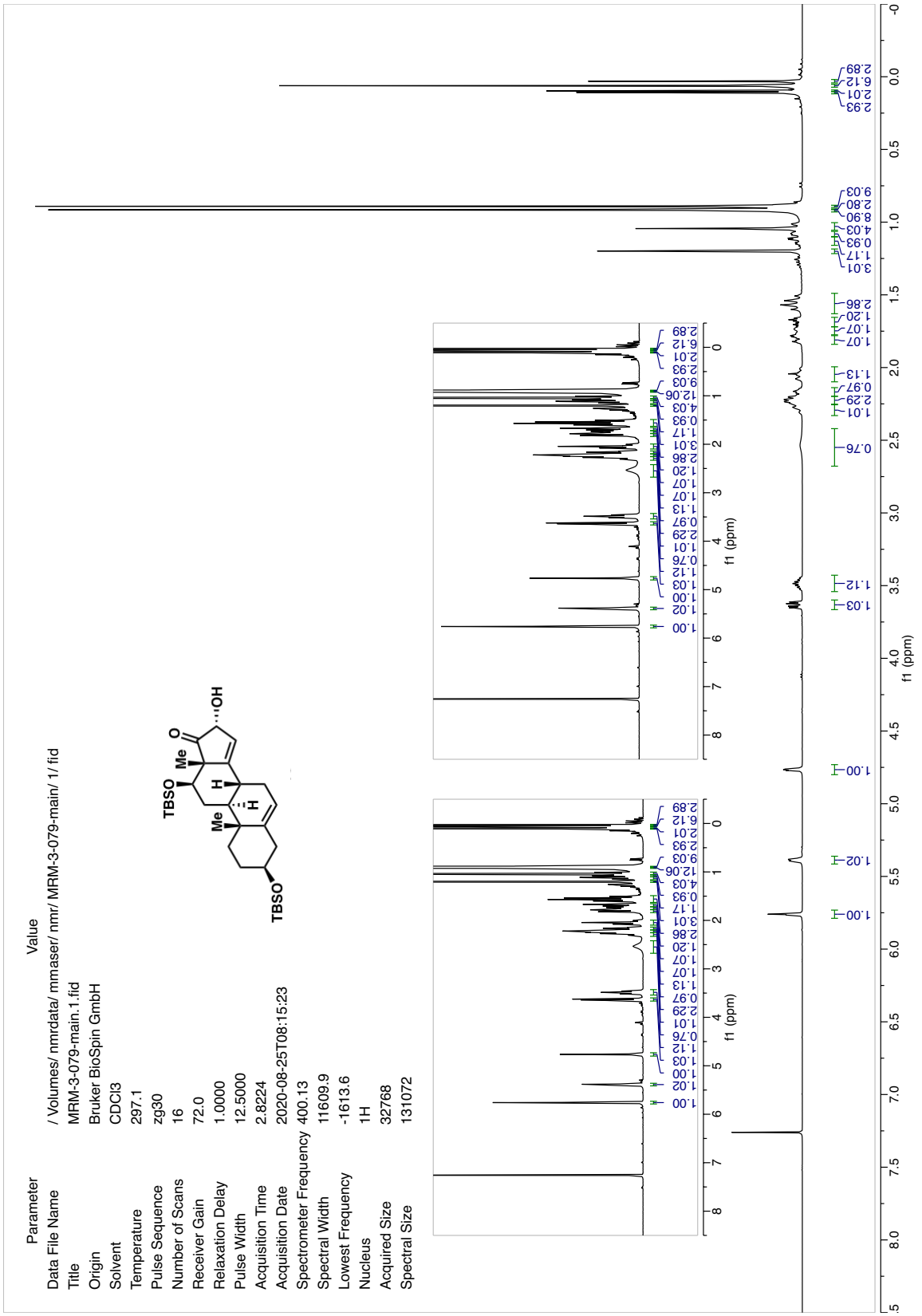

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-086-main/ 2/ fid
Title MRM-3-086-main.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-08-20T13:06:35
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1958.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
O
H
H
H
Me
TBSO
210.60
156.63
142.29
132.55
119.87
72.39
70.80
55.18
54.74
50.71
42.71
37.07
36.92
31.91
31.28
30.38
27.71
25.94
19.32
18.28
18.19
15.10
-4.31
-4.57
-4.59
-4.75
f1 (ppm)

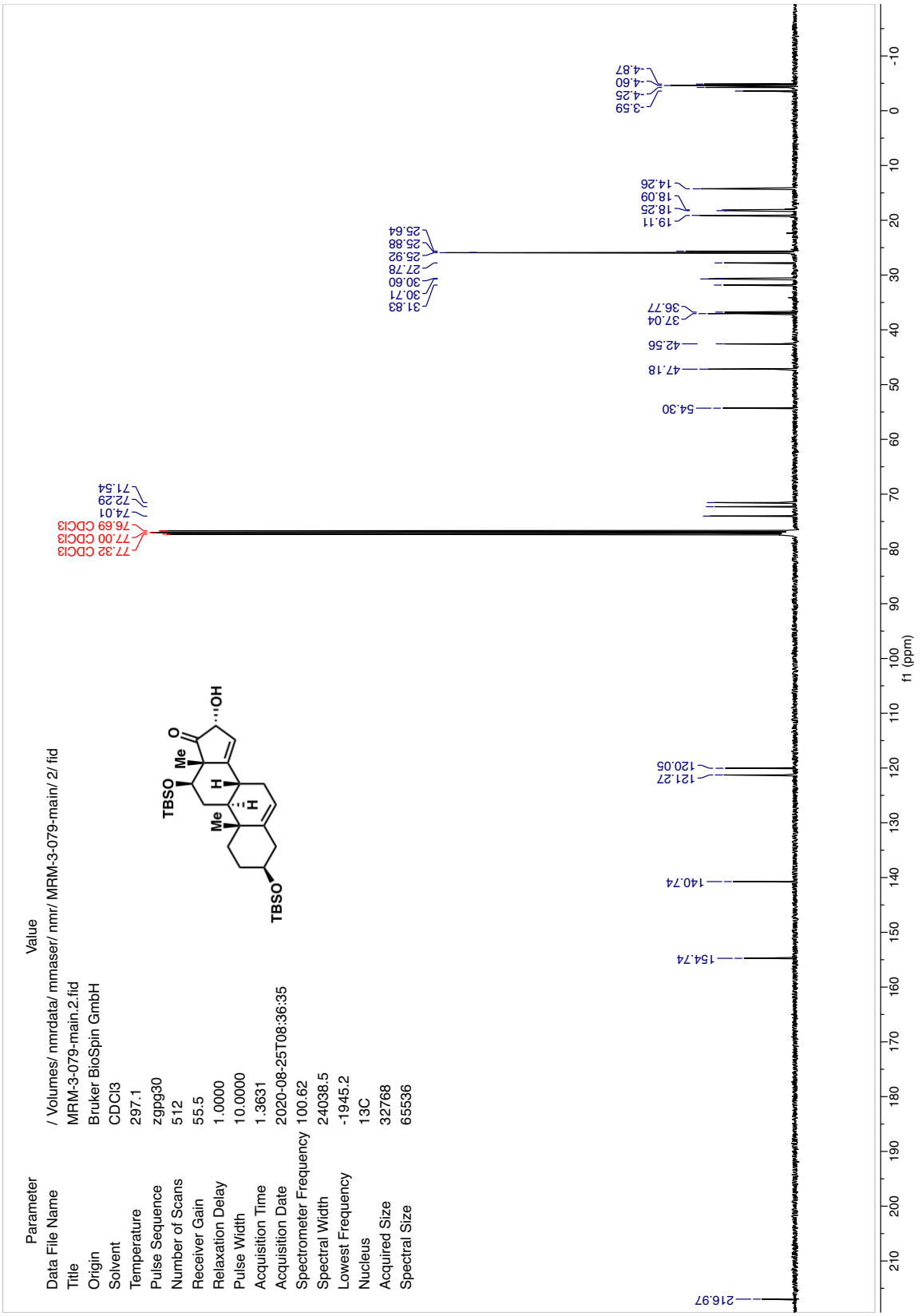

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-2-088-1/ 1/ fid
Title MRM-2-088-1.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 30.3
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-01-21T15:00:38
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
TBSO
Me
O
H
Me
H
TBSO
1 : 1
f1 (ppm)

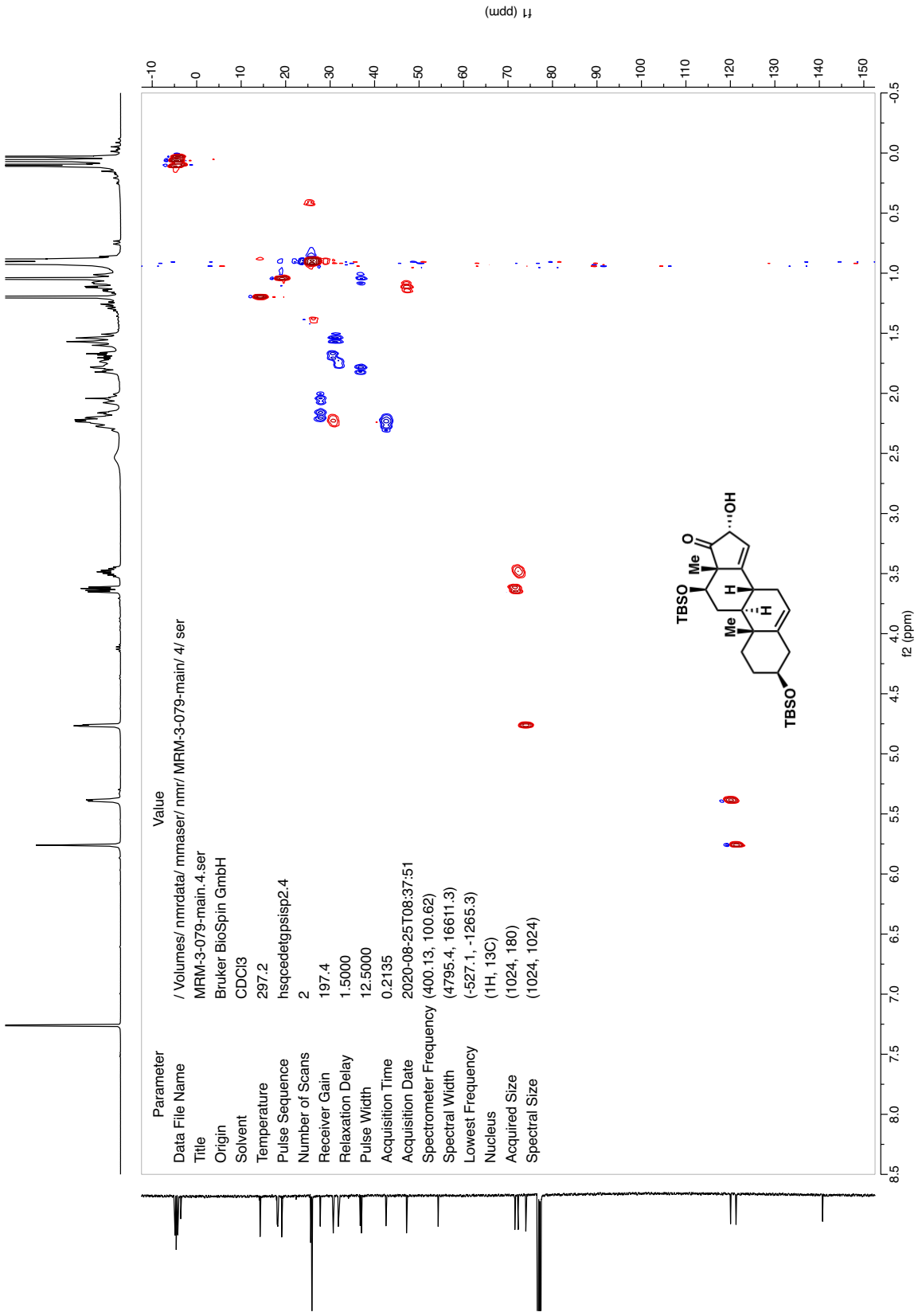

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-2-088-1/ 2/ fid
Title MRM-2-088-1.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 87.8
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-01-21T15:22:21
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1946.1
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
1 : 1
TBSO
Me
H
O
f1 (ppm)

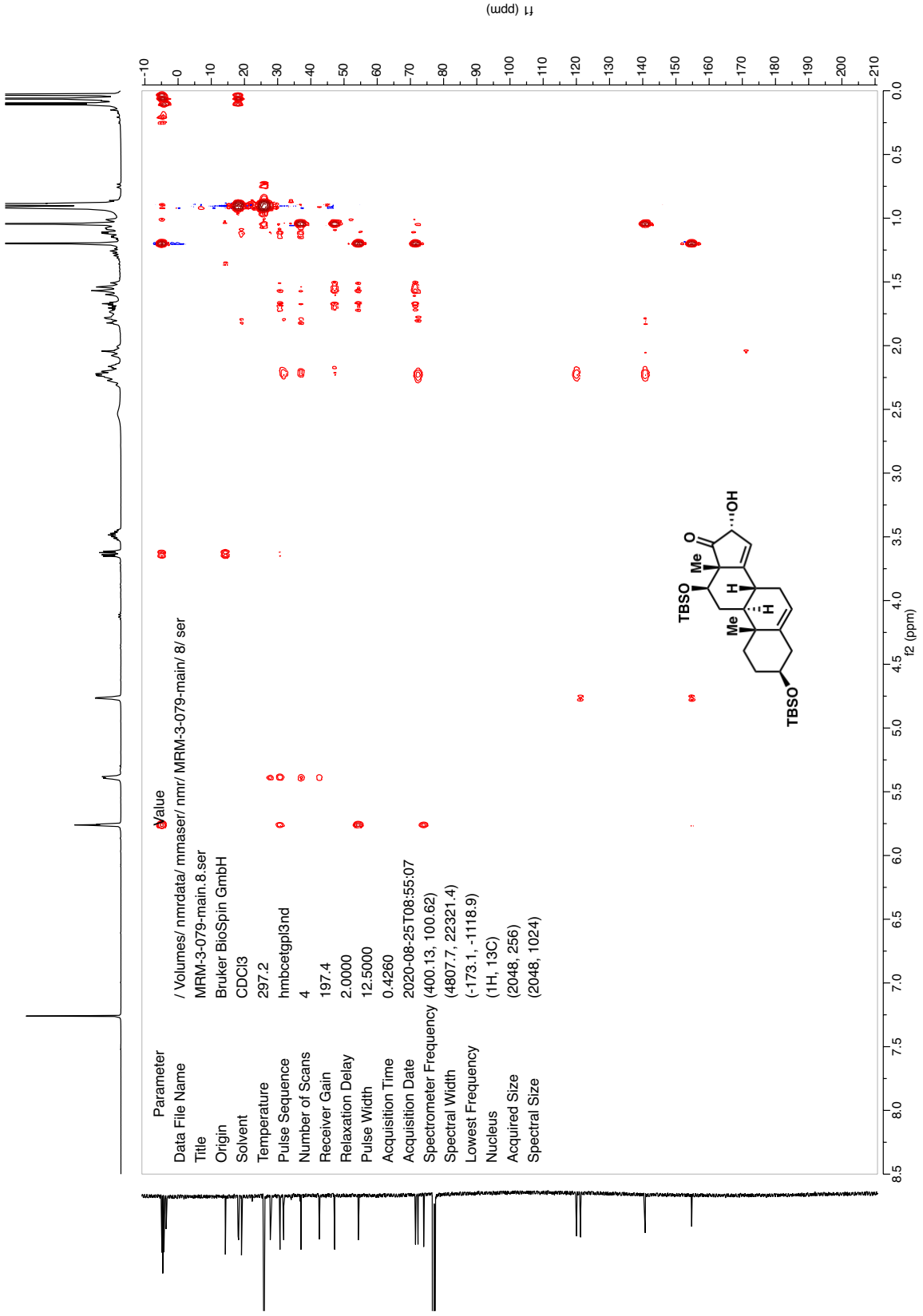

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-045-2D/ 1/ fid
Title YN-3-045-2D.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-08-28T10:53:44
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.6
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
Me
H
H
Me
TBSO
OTMS
f1 (ppm)

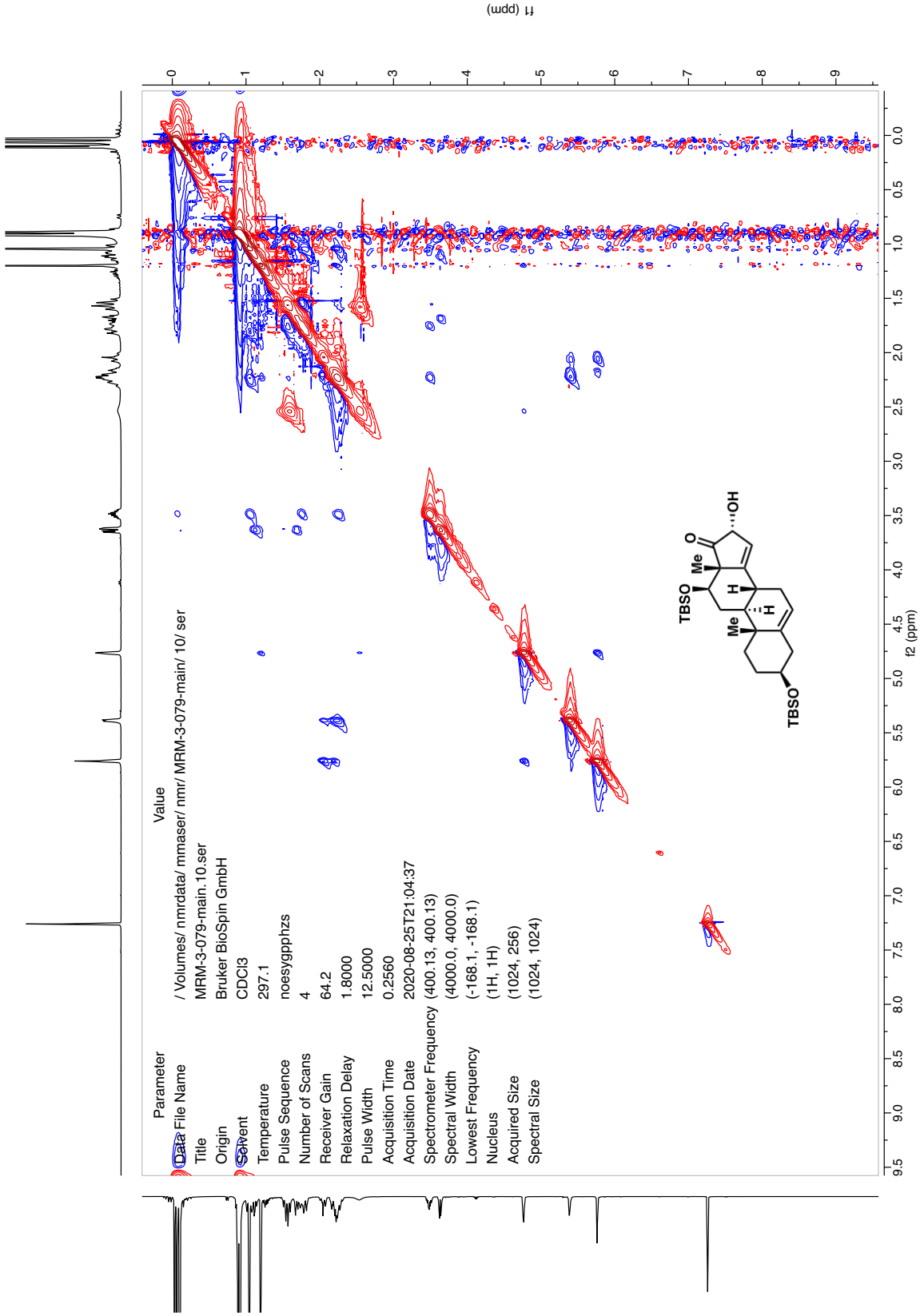

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-045-2D/ 2/ fid
Title YN-3-045-2D.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-08-28T11:14:55
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
H
H
Me
TBSO
OTMS
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
76.31
72.71
170.35
145.14
140.92
121.09
119.48
101.93
56.13
50.20
42.83
37.64
37.11
32.12
31.76
30.81
28.49
26.18
26.09
19.53
18.42
18.33
12.39
0.22
-4.42
-4.44
-4.53
f1 (ppm)

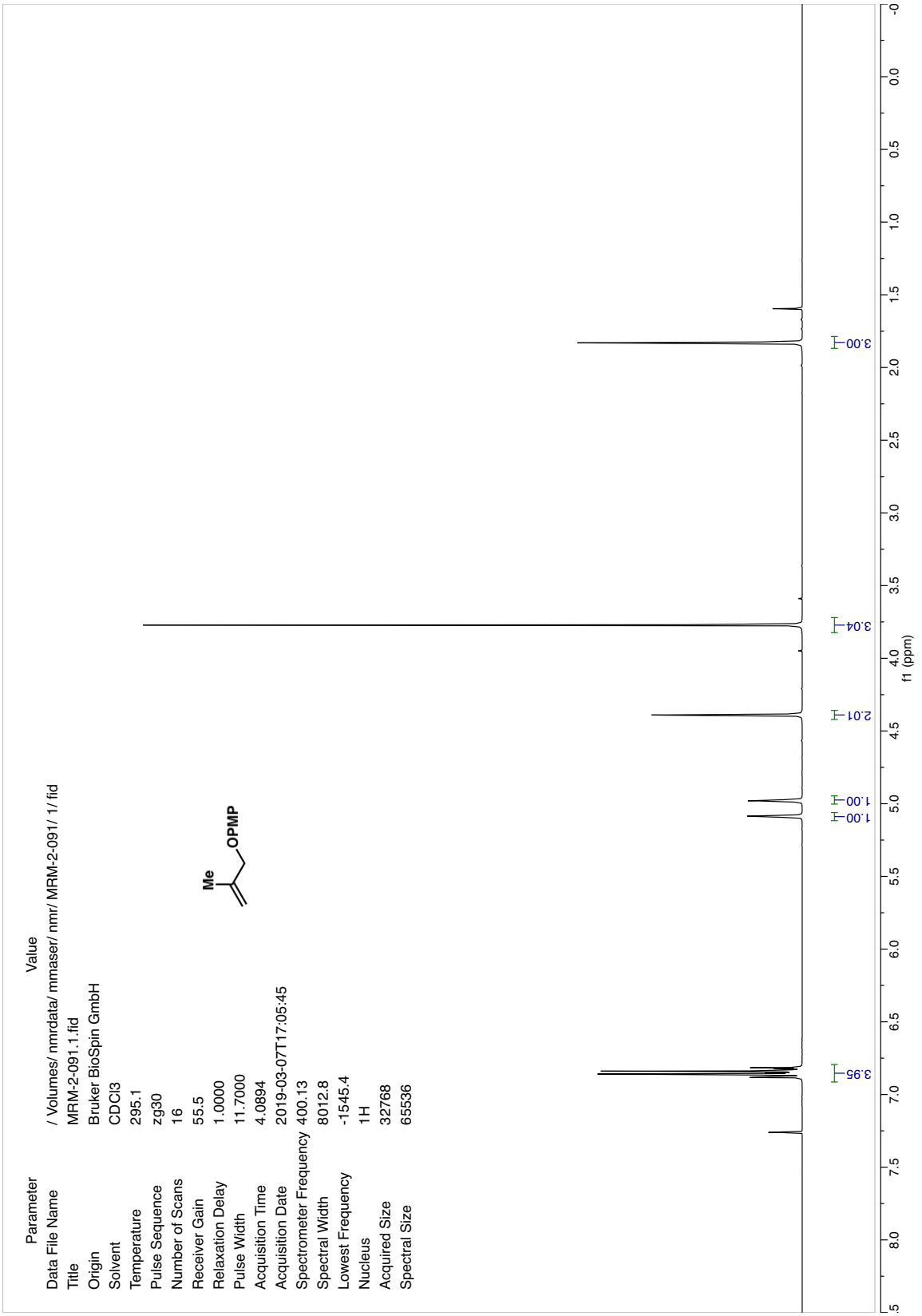

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-079-main/ 1/ fid
Title MRM-3-079-main.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-08-25T08:15:23
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.6
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
Me
O
OH
H
Me
H
H
TBSO
f1 (ppm)

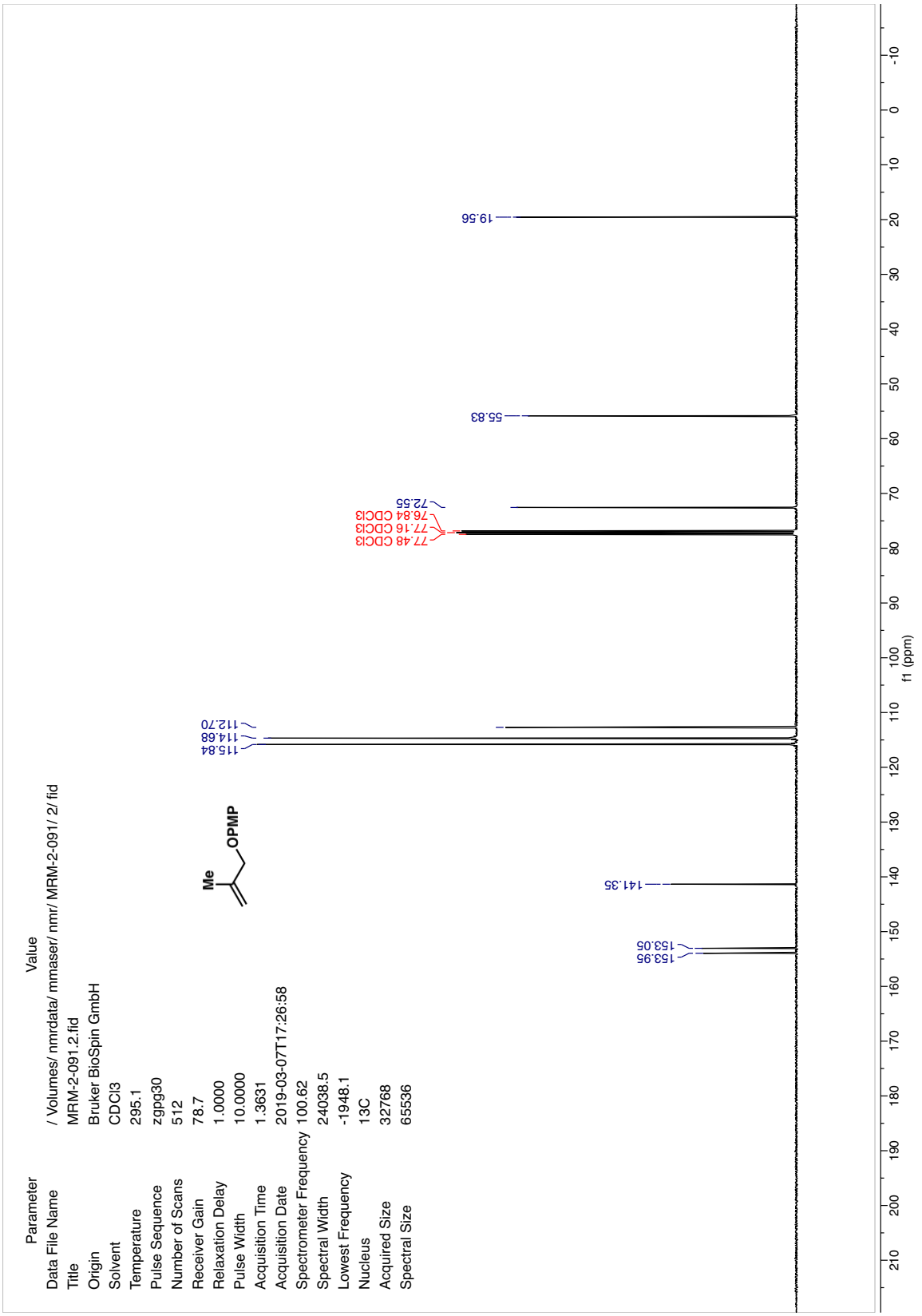

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-079-main/ 2/ fid
Title MRM-3-079-main.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-08-25T08:36:35
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.2
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
O
OH
H
Me
H
TBSO
77.32 CDCl3
77.00 CDCl3
76.69 CDCl3
74.01
72.29
71.54
54.30
47.18
42.56
37.04
36.77
31.83
30.71
30.60
27.78
25.92
25.88
25.64
19.11
18.25
18.09
14.26
-3.59
-4.25
-4.60
-4.87
216.97
154.74
140.74
121.27
120.05
f1 (ppm)

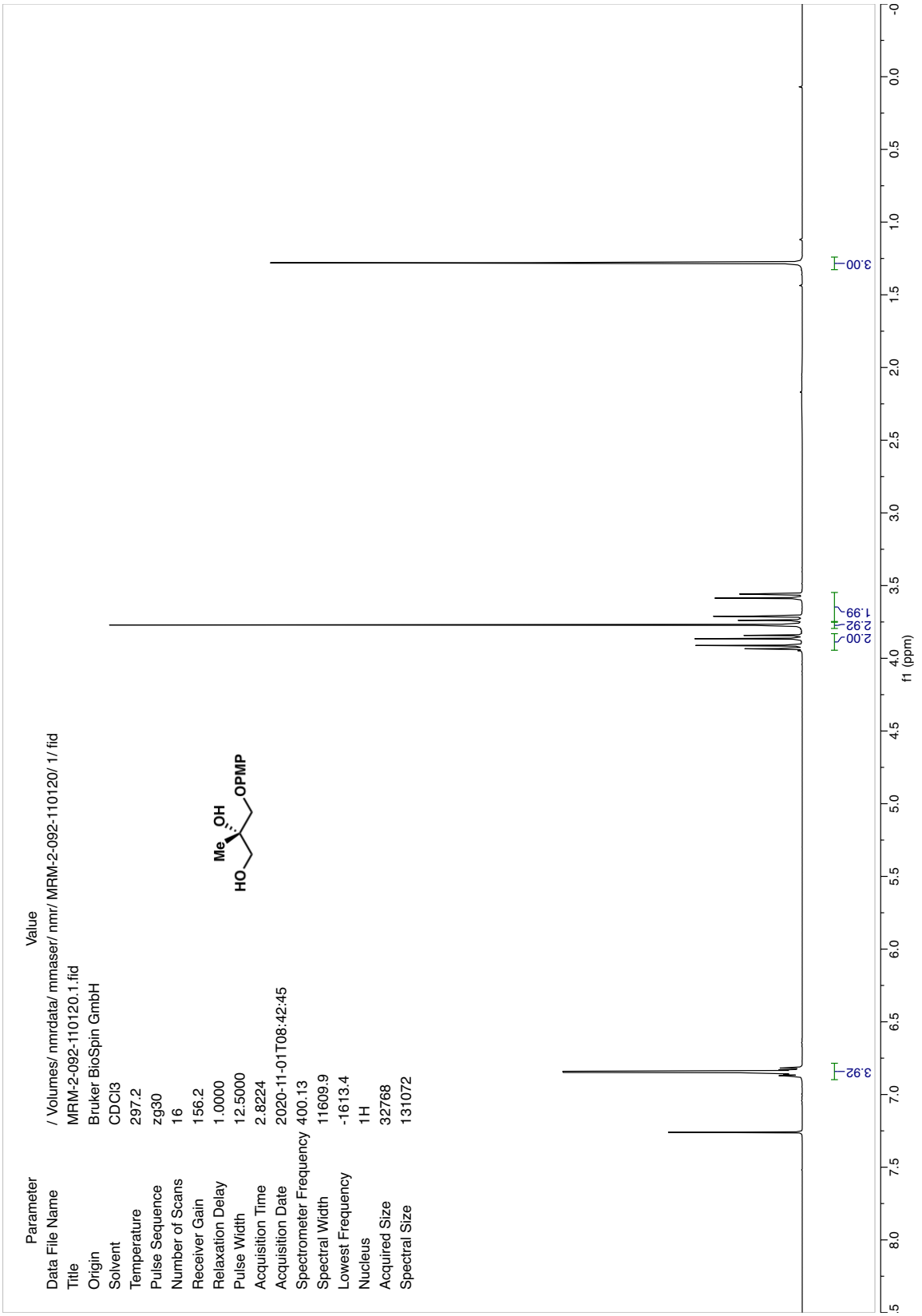

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-079-main/ 4/ ser
Title MRM-3-079-main.4.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-08-25T08:37:51
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-527.1, -1265.3)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
Me
O
OH
H
H
Me
TBSO

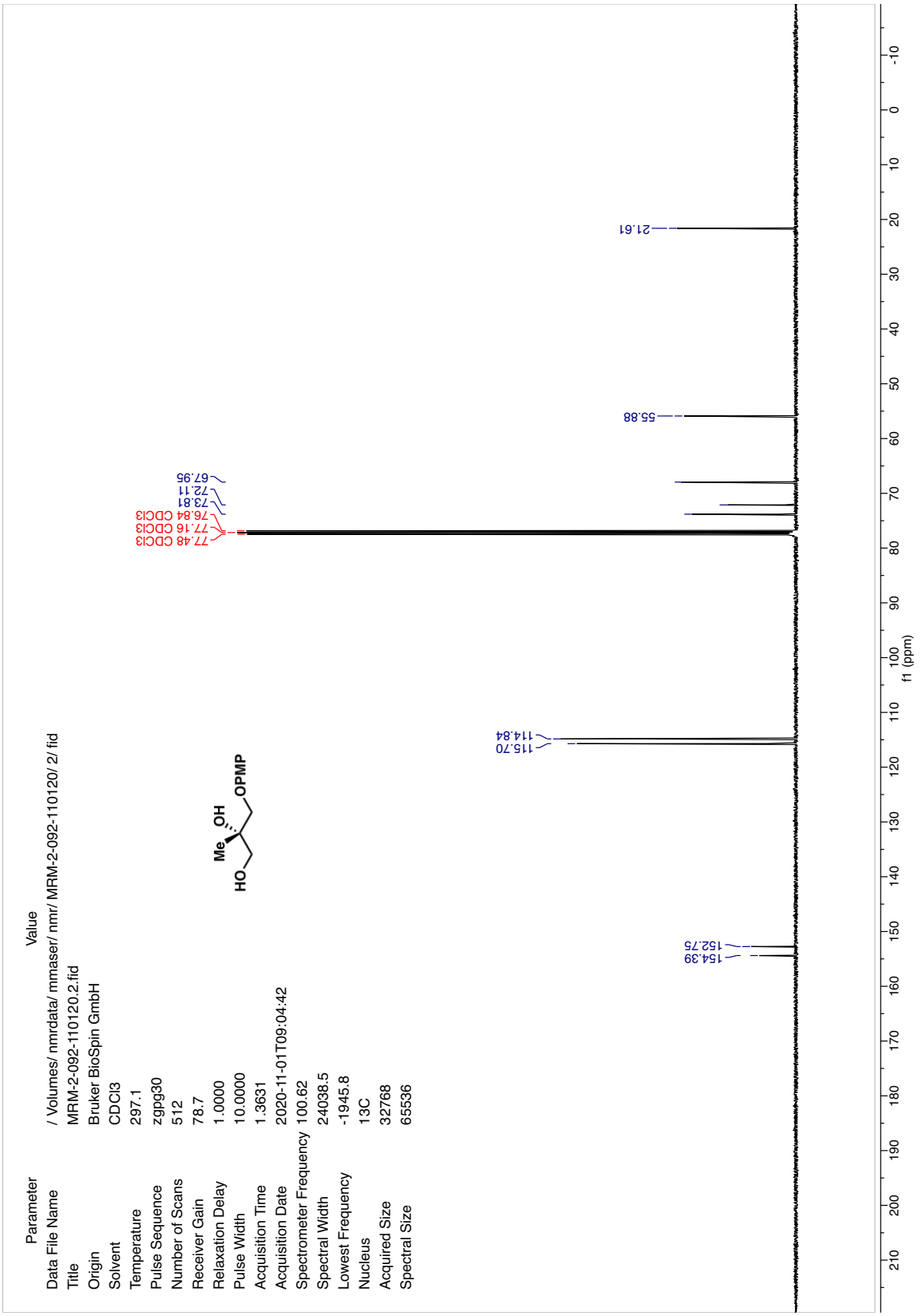

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-079-main/ 8/ ser
Title MRM-3-079-main.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-08-25T08:55:07
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-173.1, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
f1 (ppm)
f2 (ppm)
TBSO
Me
O
OH
H
Me
H
TBSO

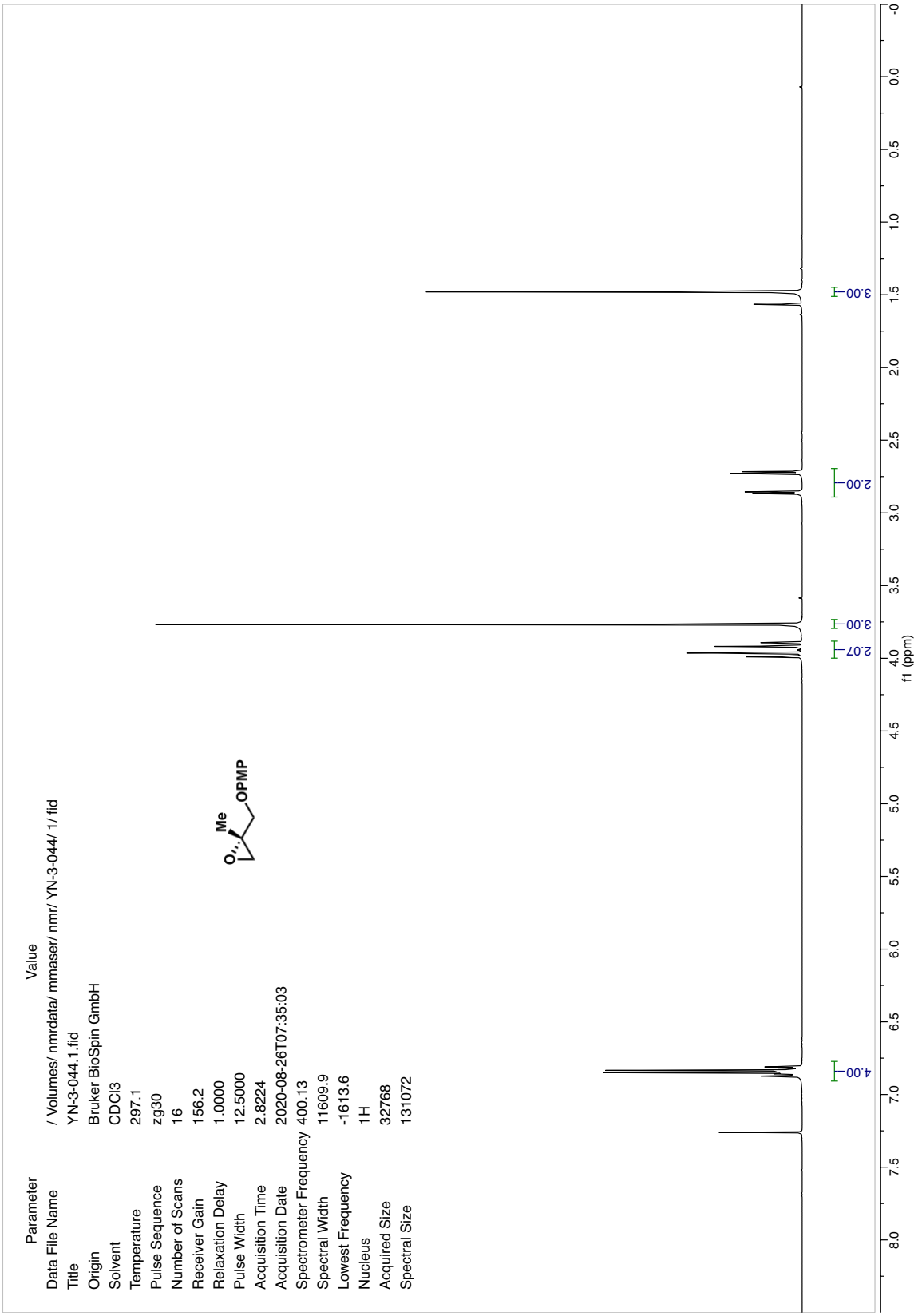

f1 (ppm)
f2 (ppm)
Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-079-main/ 10/ ser
Title
MRM-3-079-main.10.ser
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
noesygpphzs
Number of Scans
4
Receiver Gain
64.2
Relaxation Delay
1.8000
Pulse Width
12.5000
Acquisition Time
0.2560
Acquisition Date
2020-08-25T21:04:37
Spectrometer Frequency
(400.13, 400.13)
Spectral Width
(4000.0, 4000.0)
Lowest Frequency
(-168.1, -168.1)
Nucleus
(1H, 1H)
Acquired Size
(1024, 256)
Spectral Size
(1024, 1024)
TBSO
Me
O
OH
H
H
Me
TBSO

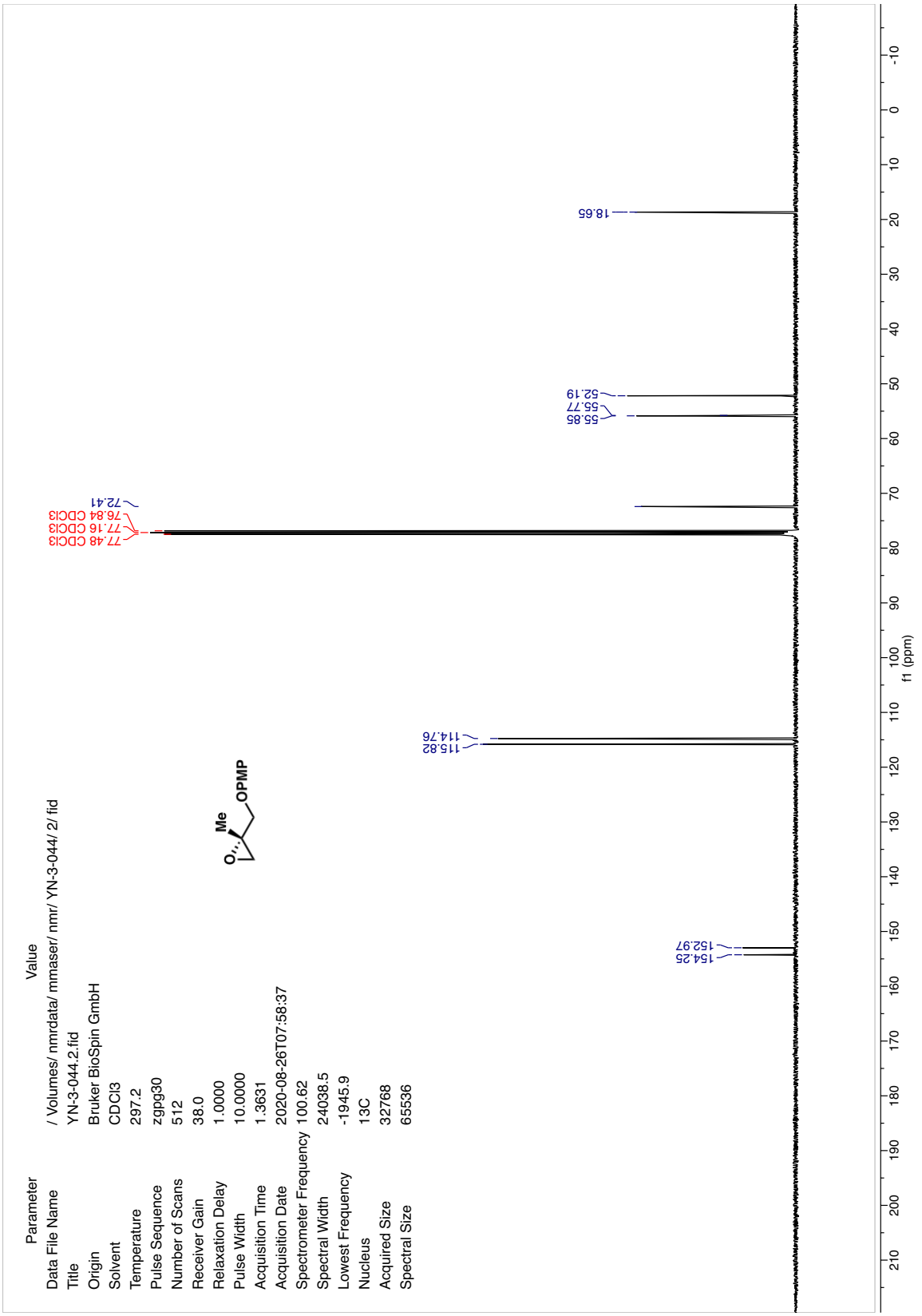

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-091/ 1/ fid
Title MRM-2-091.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-03-07T17:05:45
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
OPMP
3.95
1.00
1.00
2.01
3.04
3.00
f1 (ppm)

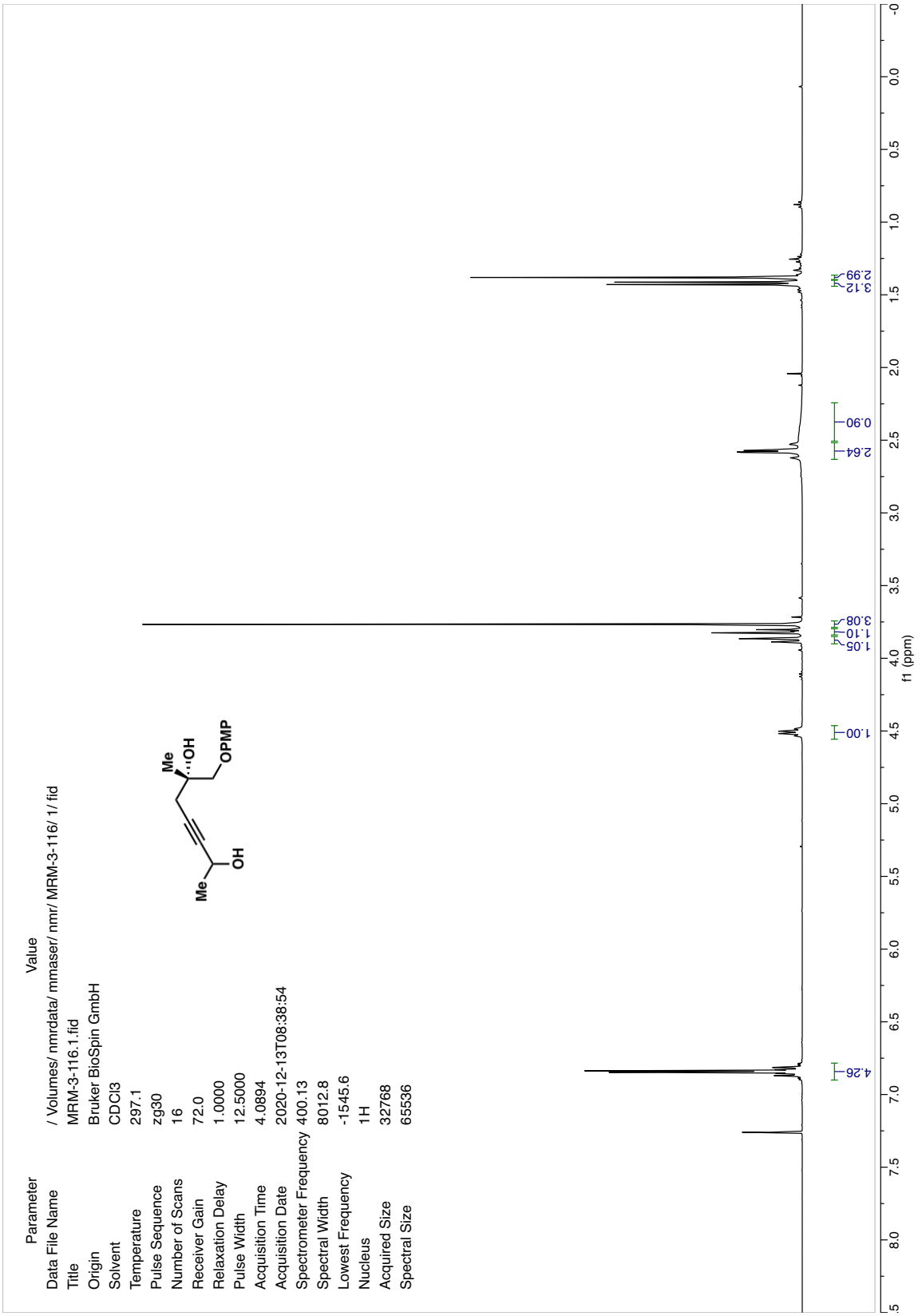

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-091/ 2/ fid
Title MRM-2-091.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-03-07T17:26:58
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1948.1
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OPMP
115.84
114.68
112.70
153.95
153.05
141.35
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
72.55
55.83
19.56
f1 (ppm)

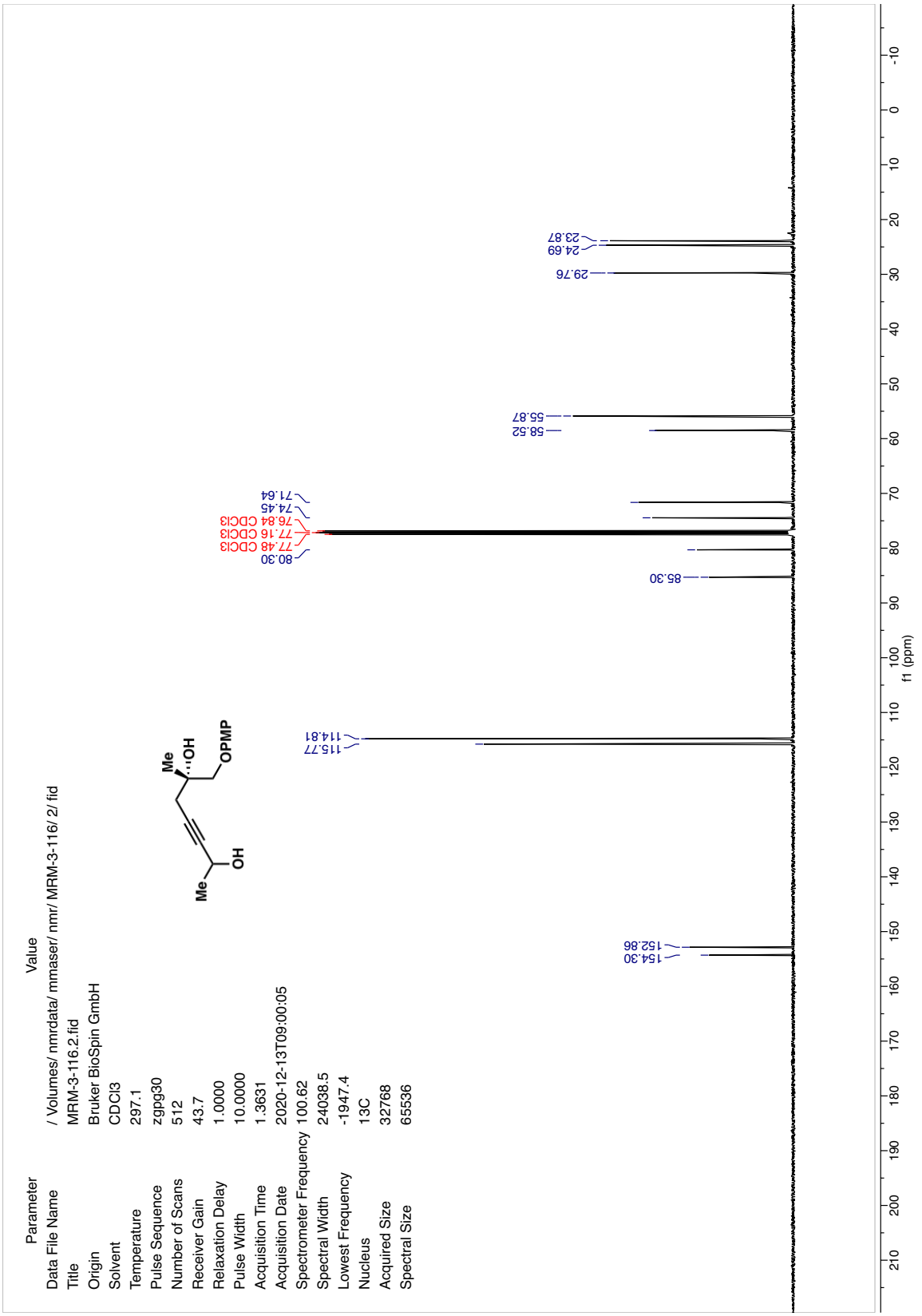

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-092-110120/ 1/ fid
Title MRM-2-092-110120.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-11-01T08:42:45
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.4
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
Me
OH
HO
OPMP
3.92
3.00
2.00
2.92
1.99
f1 (ppm)

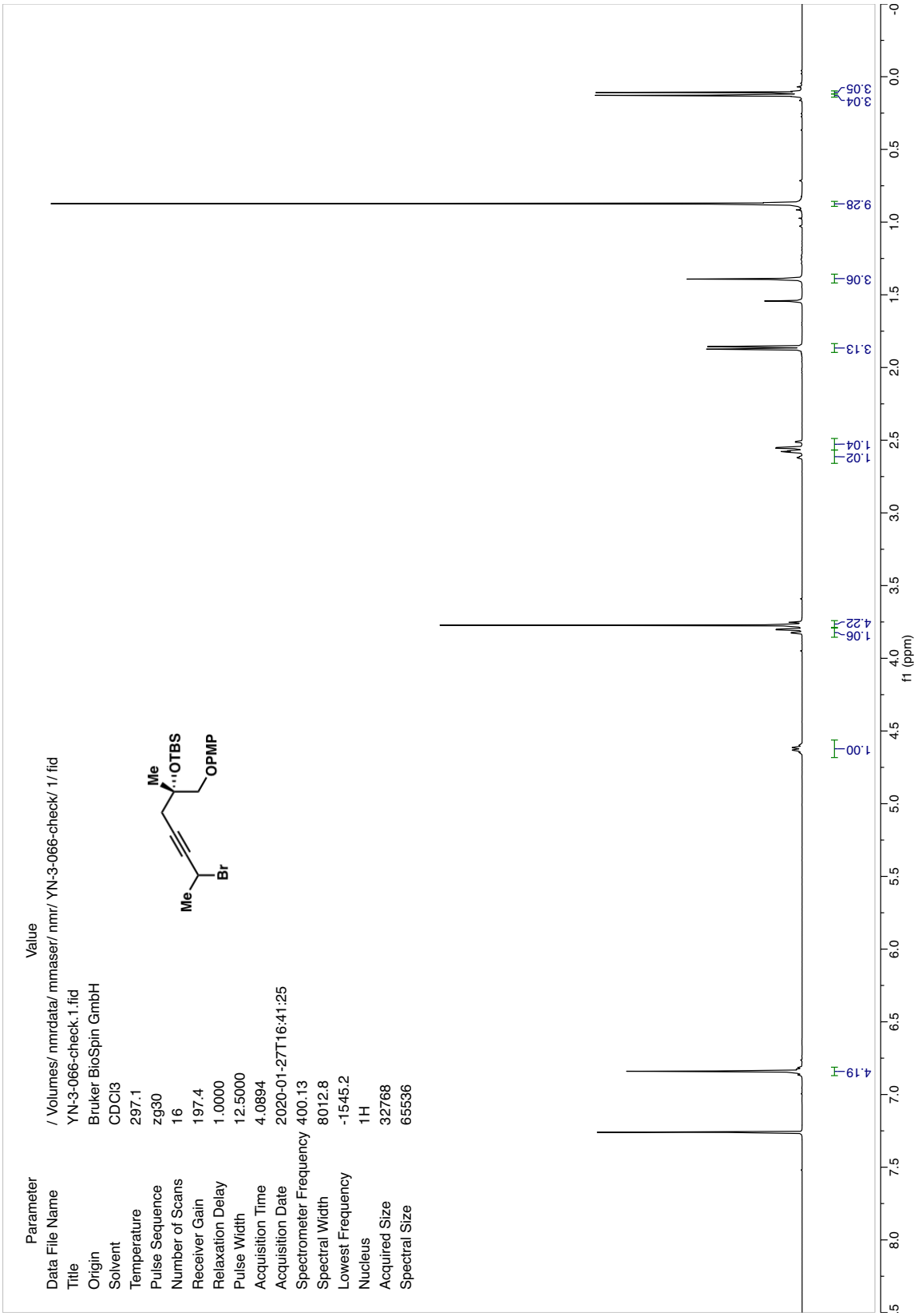

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-092-110120/ 2/ fid
Title MRM-2-092-110120.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-11-01T09:04:42
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.8
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OH
OPMP
HO
154.39
152.75
115.70
114.84
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
73.81
72.11
67.95
55.88
21.61
f1 (ppm)

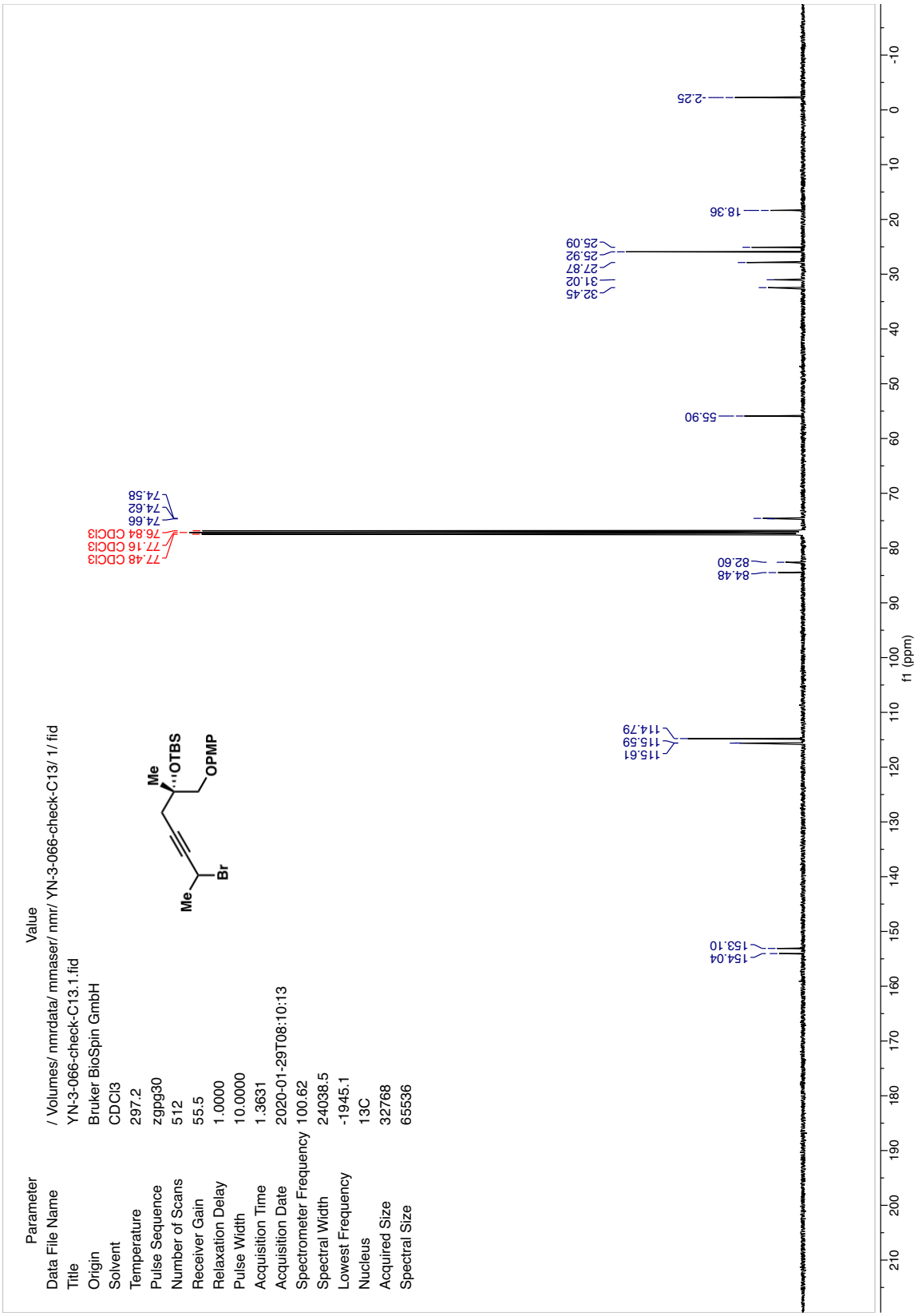

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-044/ 1/ fid
Title YN-3-044.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-08-26T07:35:03
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.6
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
Me
OPMP
4.00
2.07
3.00
2.00
3.00
f1 (ppm)

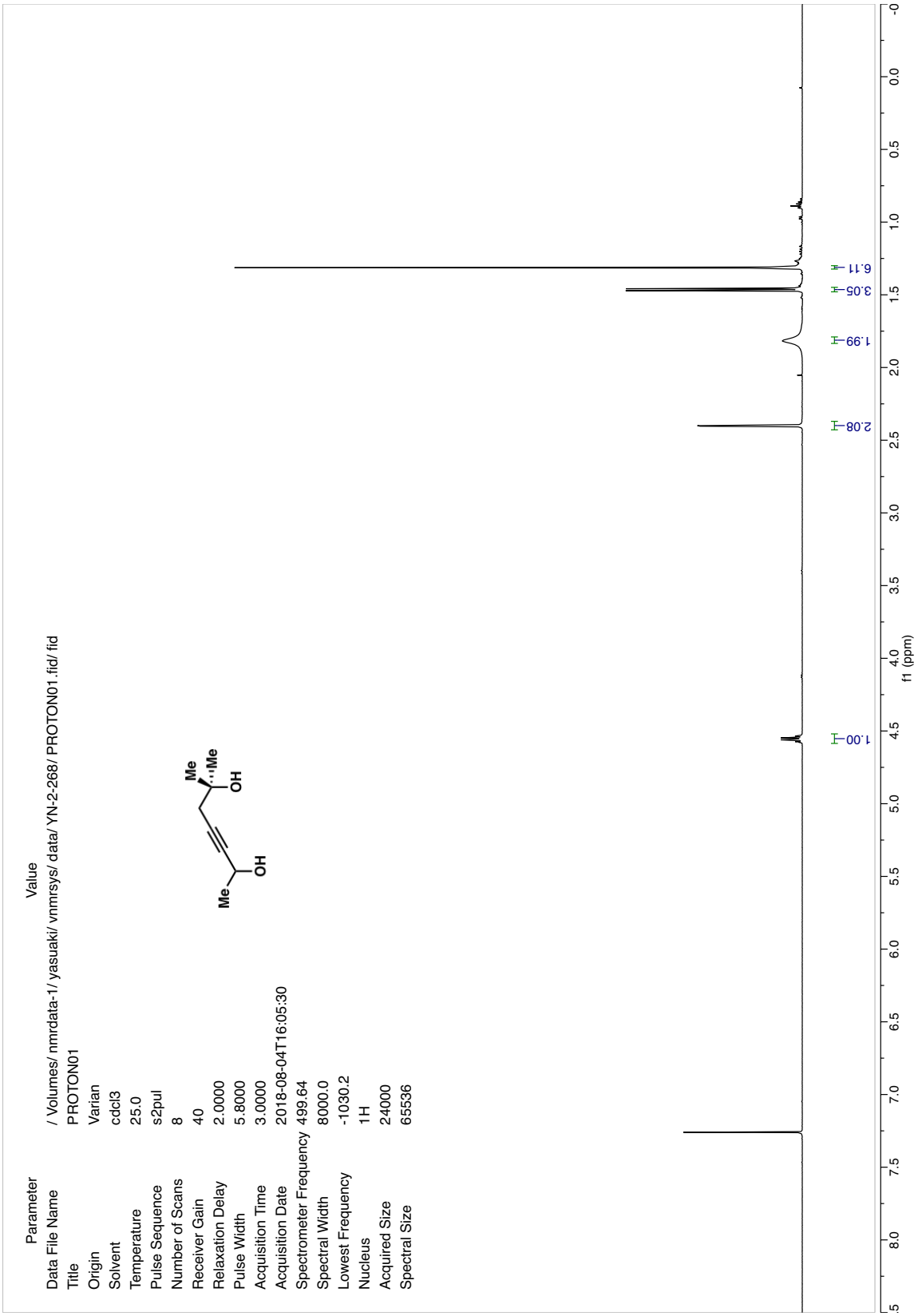

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-044/ 2/ fid
Title YN-3-044.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 38.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-08-26T07:58:37
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OPMP
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
72.41
154.25
152.97
115.82
114.76
55.85
55.77
52.19
18.65
f1 (ppm)

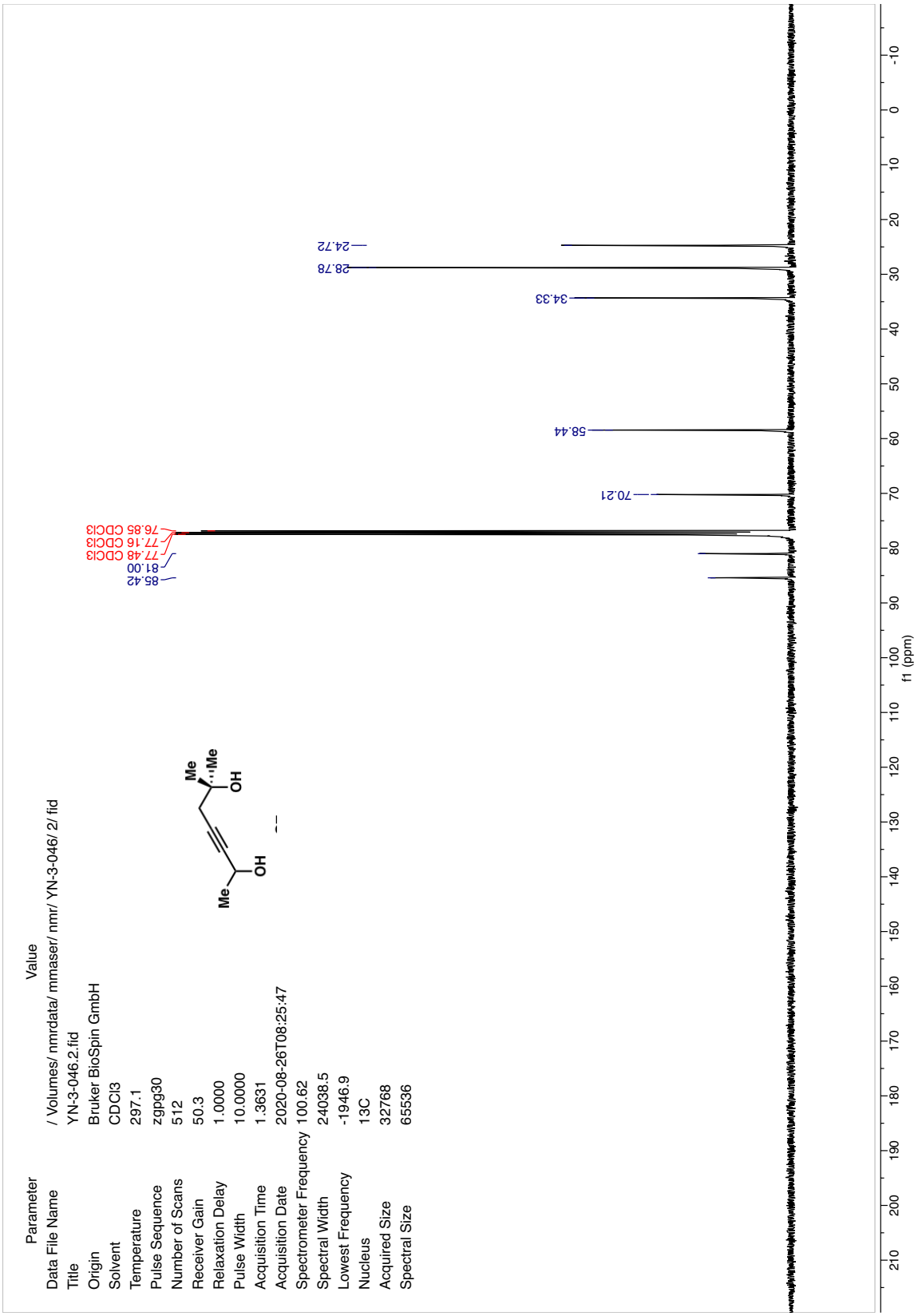

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-116/ 1/ fid
Title MRM-3-116.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-12-13T08:38:54
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.6
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
OH
OPMP
Me
OH
4.26
1.00
1.05
1.10
3.08
2.64
0.90
2.99
3.12
f1 (ppm)

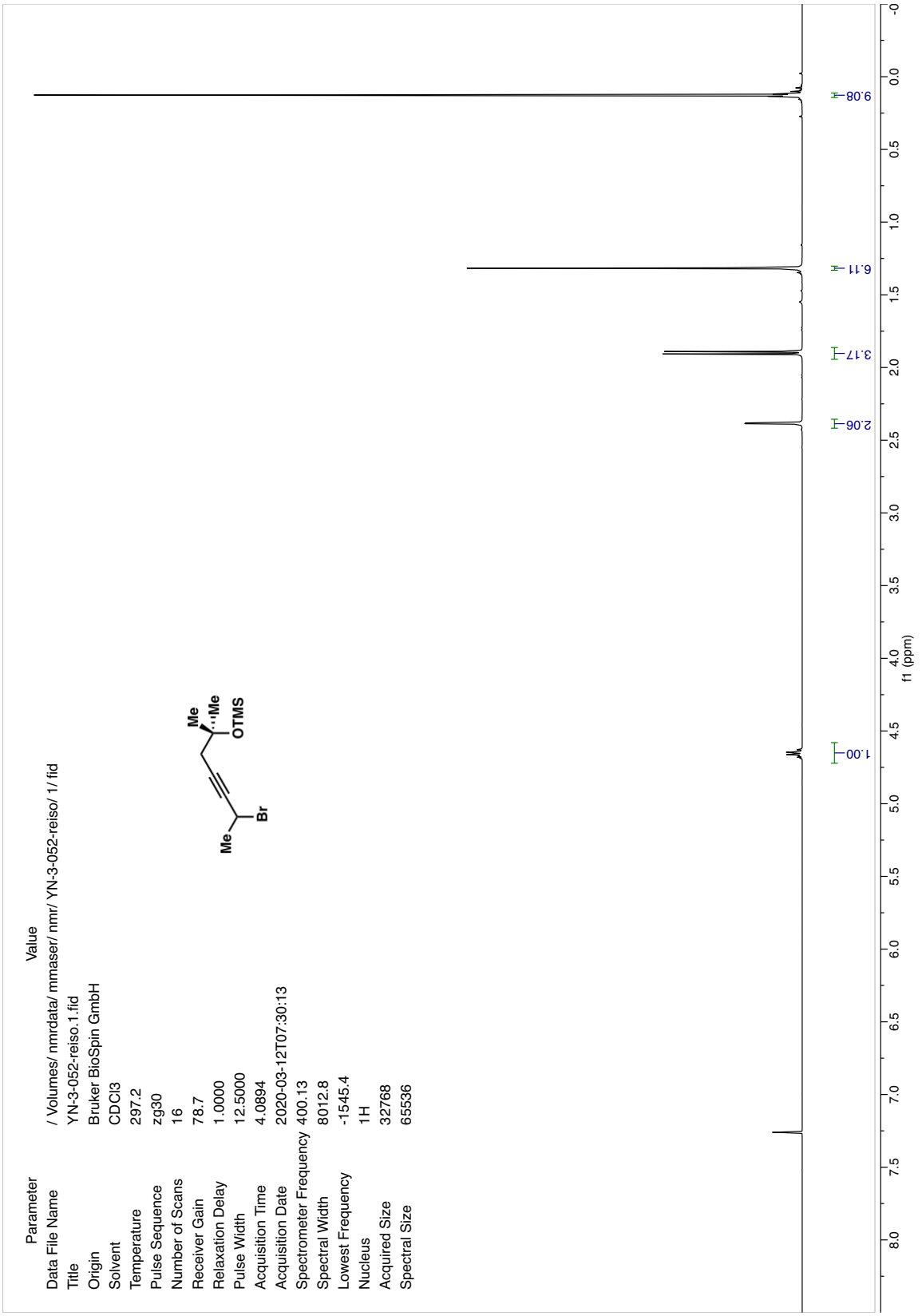

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-116/ 2/ fid
Title MRM-3-116.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 43.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-12-13T09:00:05
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1947.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OH
OPMP
Me
154.30
152.86
115.77
114.81
80.30
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
74.45
71.64
58.52
55.87
29.76
24.69
23.87
f1 (ppm)

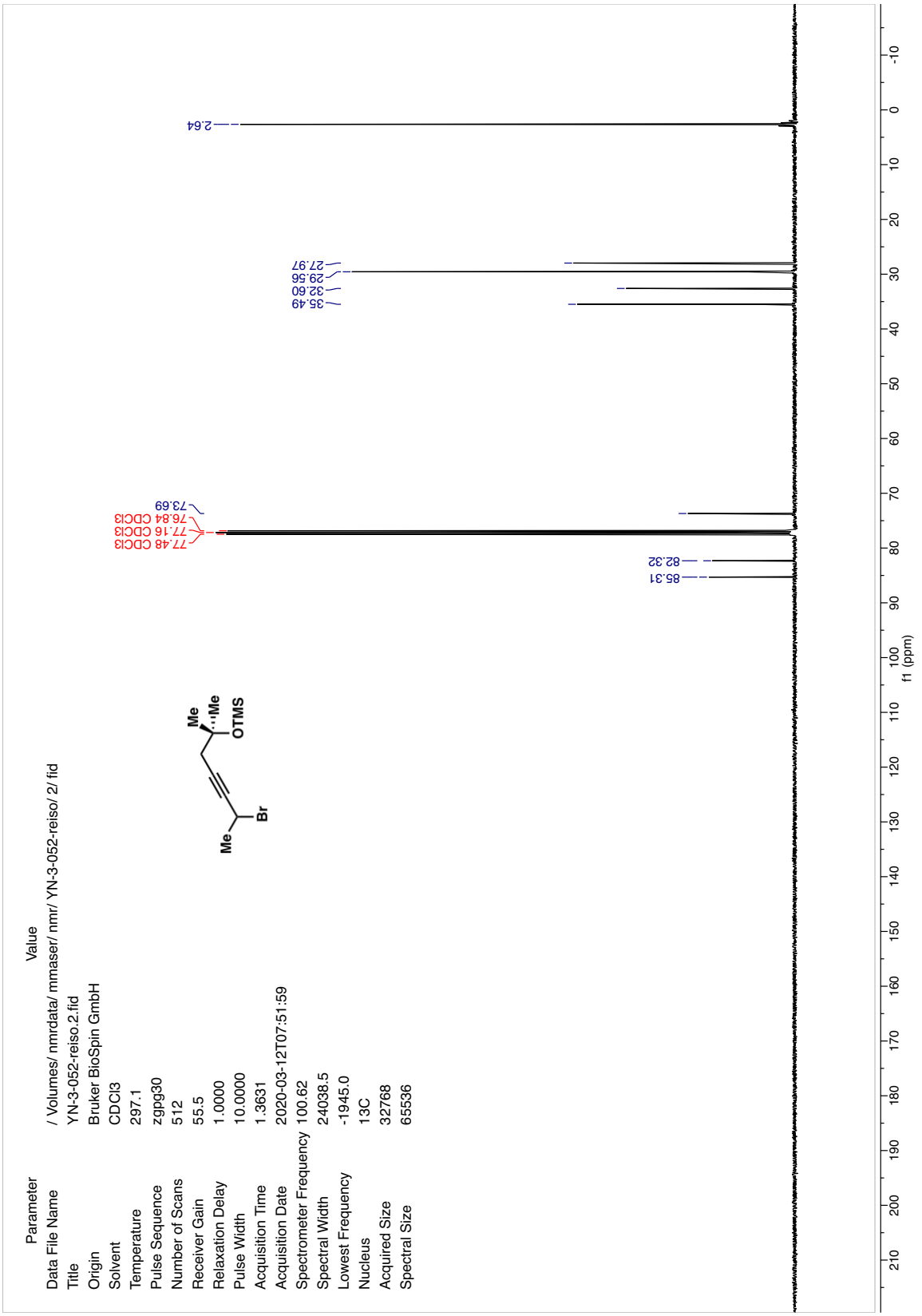

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-066-check/ 1/ fid
Title YN-3-066-check.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-01-27T16:41:25
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
OTBS
OPMP
Me
Br
4.19
1.00
1.06
4.22
1.02
1.04
3.13
3.06
9.28
3.04
3.05
f1 (ppm)

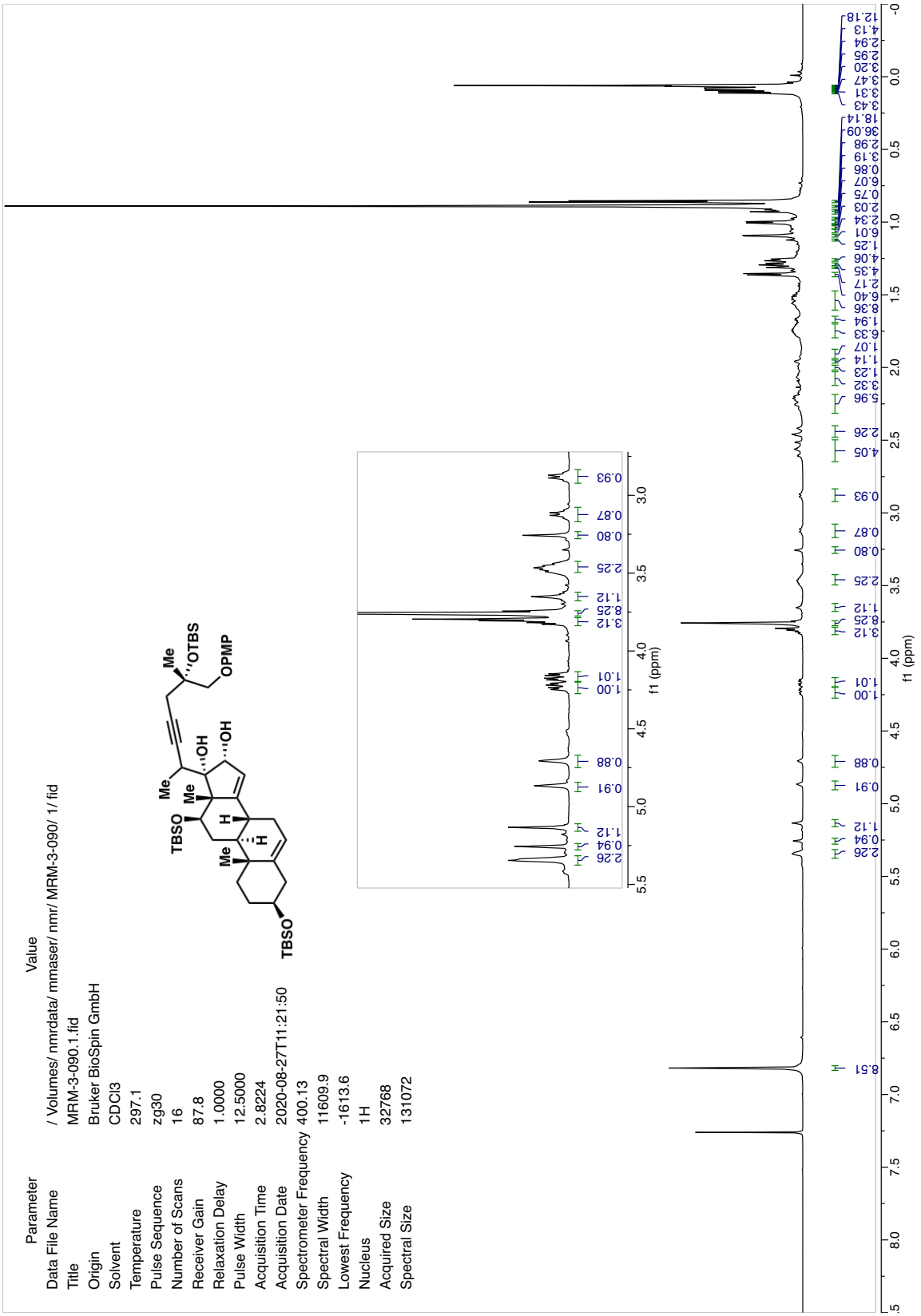

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-066-check-C13/ 1/ fid
Title YN-3-066-check-C13.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-01-29T08:10:13
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.1
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OTBS
OPMP
Br
Me
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
74.66
74.62
74.58
153.10
154.04
115.61
115.59
114.79
84.48
82.60
55.90
32.45
31.02
27.87
25.92
25.09
18.36
-2.25
f1 (ppm)

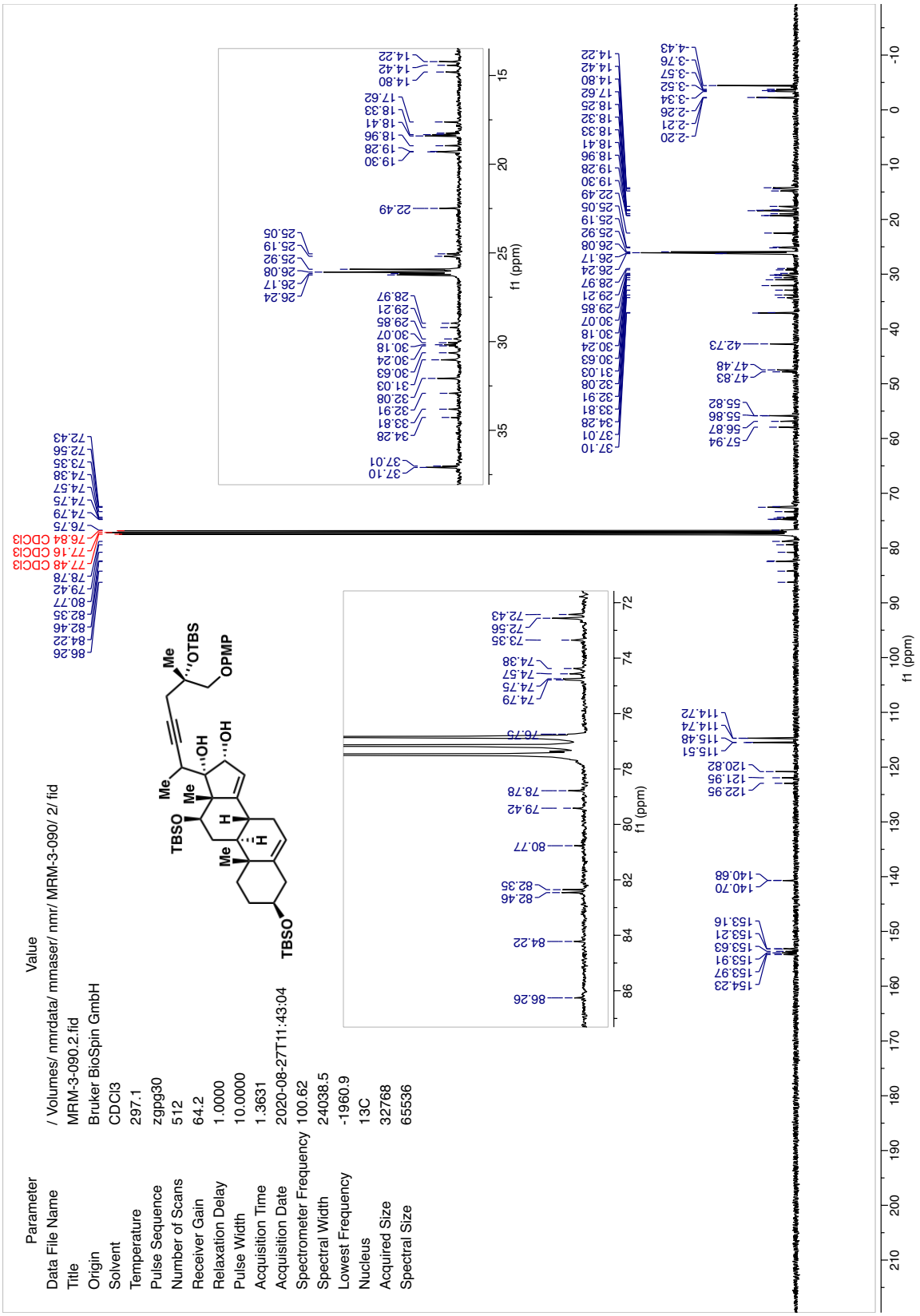

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ yasuaki/ vnmrsys/ data/ YN-2-268/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 8
Receiver Gain 40
Relaxation Delay 2.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2018-08-04T16:05:30
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1030.2
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
Me
Me
OH
OH
Me
6.11
3.05
1.99
2.08
1.00
f1 (ppm)

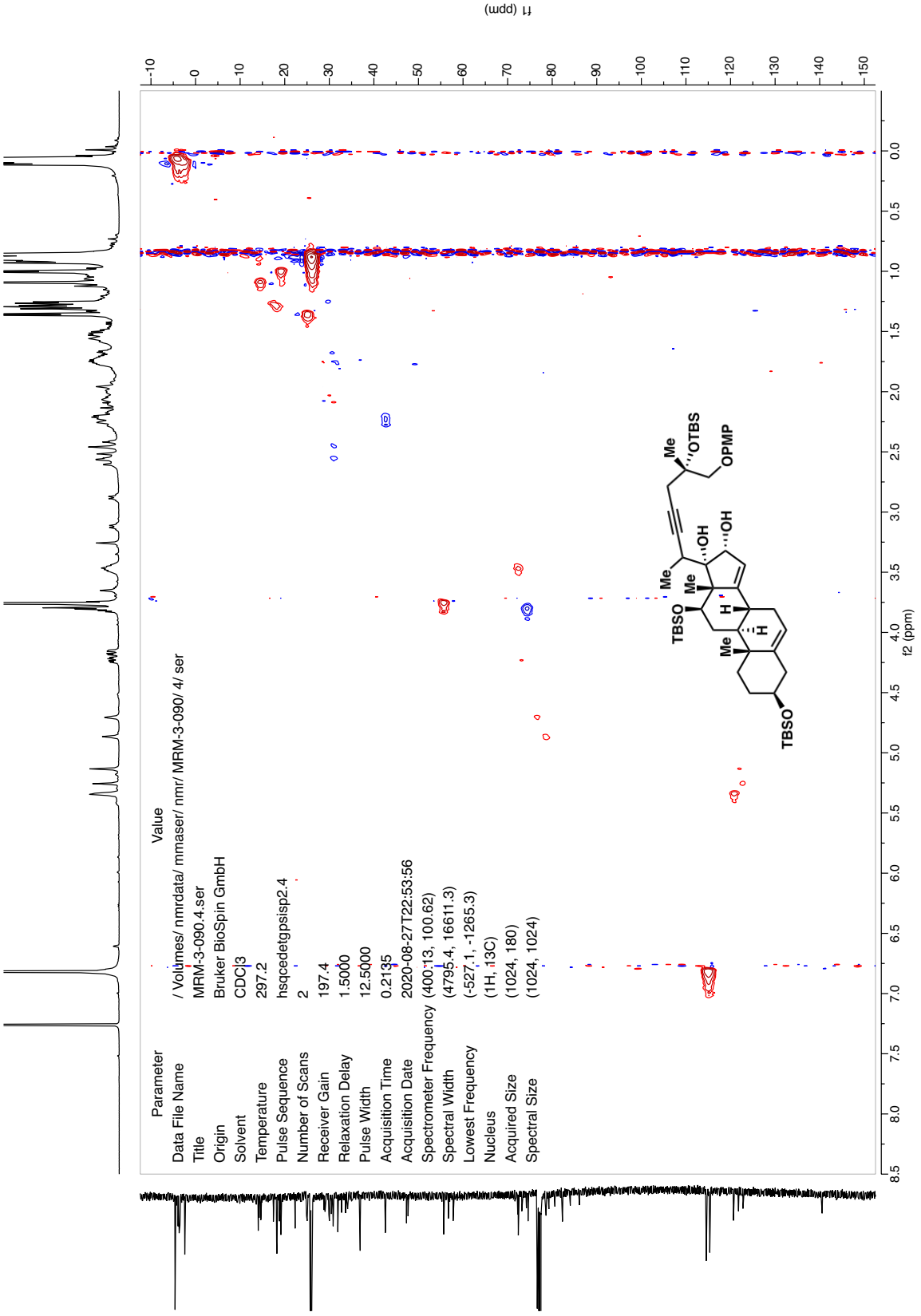

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-046/ 2/ fid
Title YN-3-046.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-08-26T08:25:47
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1946.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OH
Me
OH
Me
24.72
28.78
34.33
58.44
70.21
76.85 CDCl3
77.16 CDCl3
77.48 CDCl3
81.00
85.42
f1 (ppm)

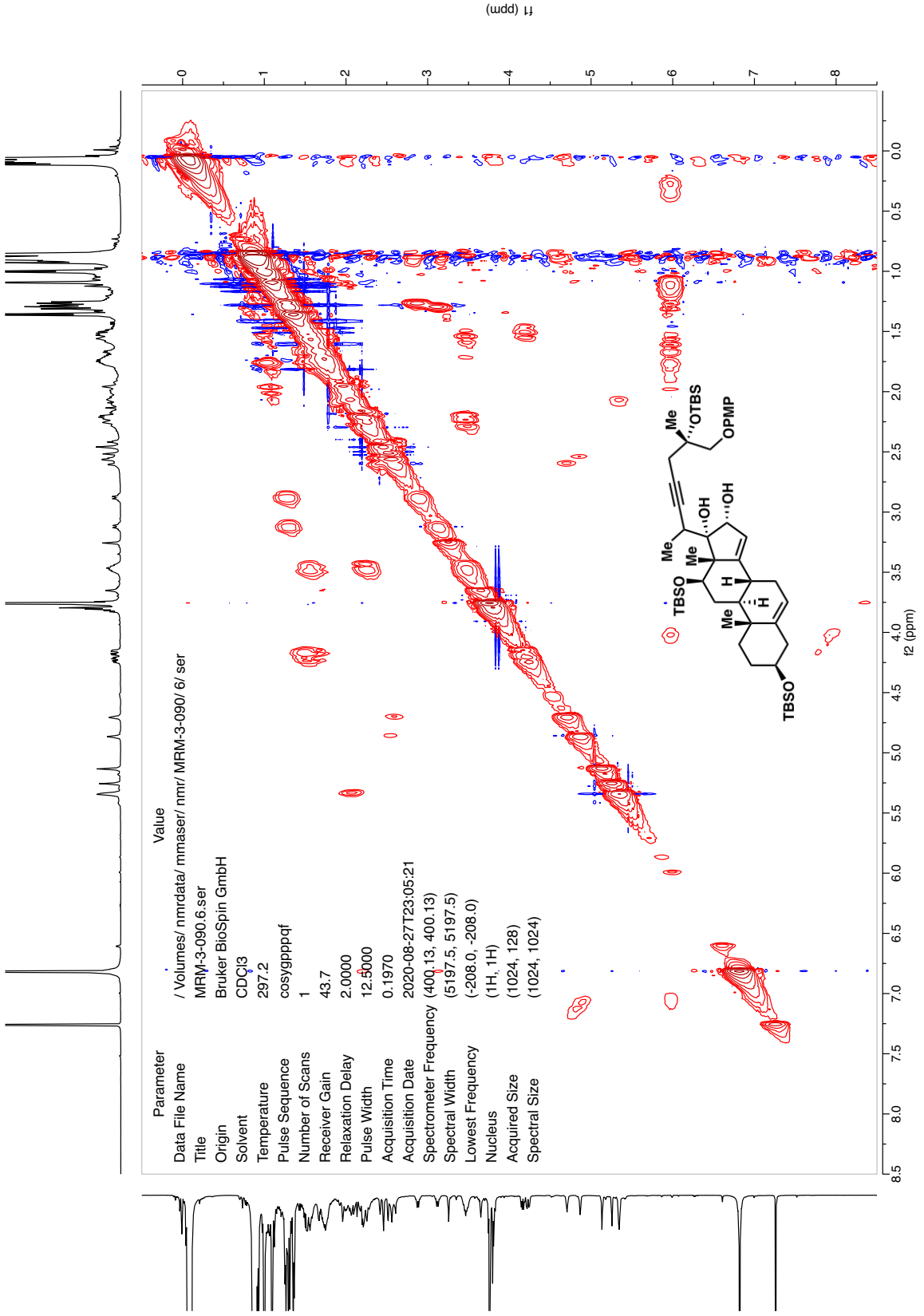

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-052-reiso/ 1/ fid
Title YN-3-052-reiso.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-03-12T07:30:13
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
Me
OTMS
Me
Br
9.08
6.11
3.17
2.06
1.00
f1 (ppm)

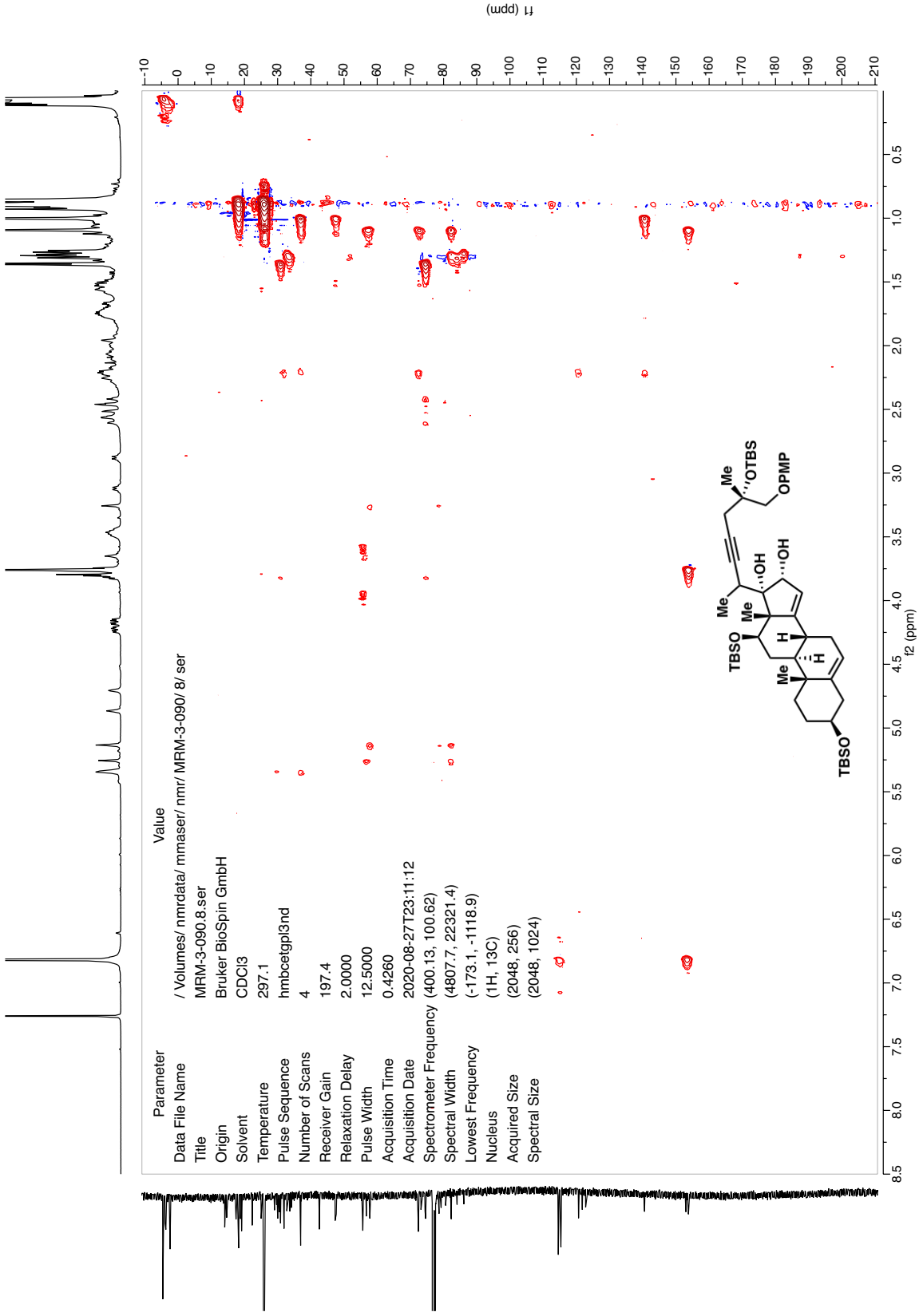

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-052-reiso/ 2/ fid
Title YN-3-052-reiso.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-03-12T07:51:59
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
Me
OTMS
Br
Me
2.64
27.97
29.56
32.60
35.49
73.69
76.84 CDCl3
77.16 CDCl3
77.48 CDCl3
82.32
85.31
f1 (ppm)

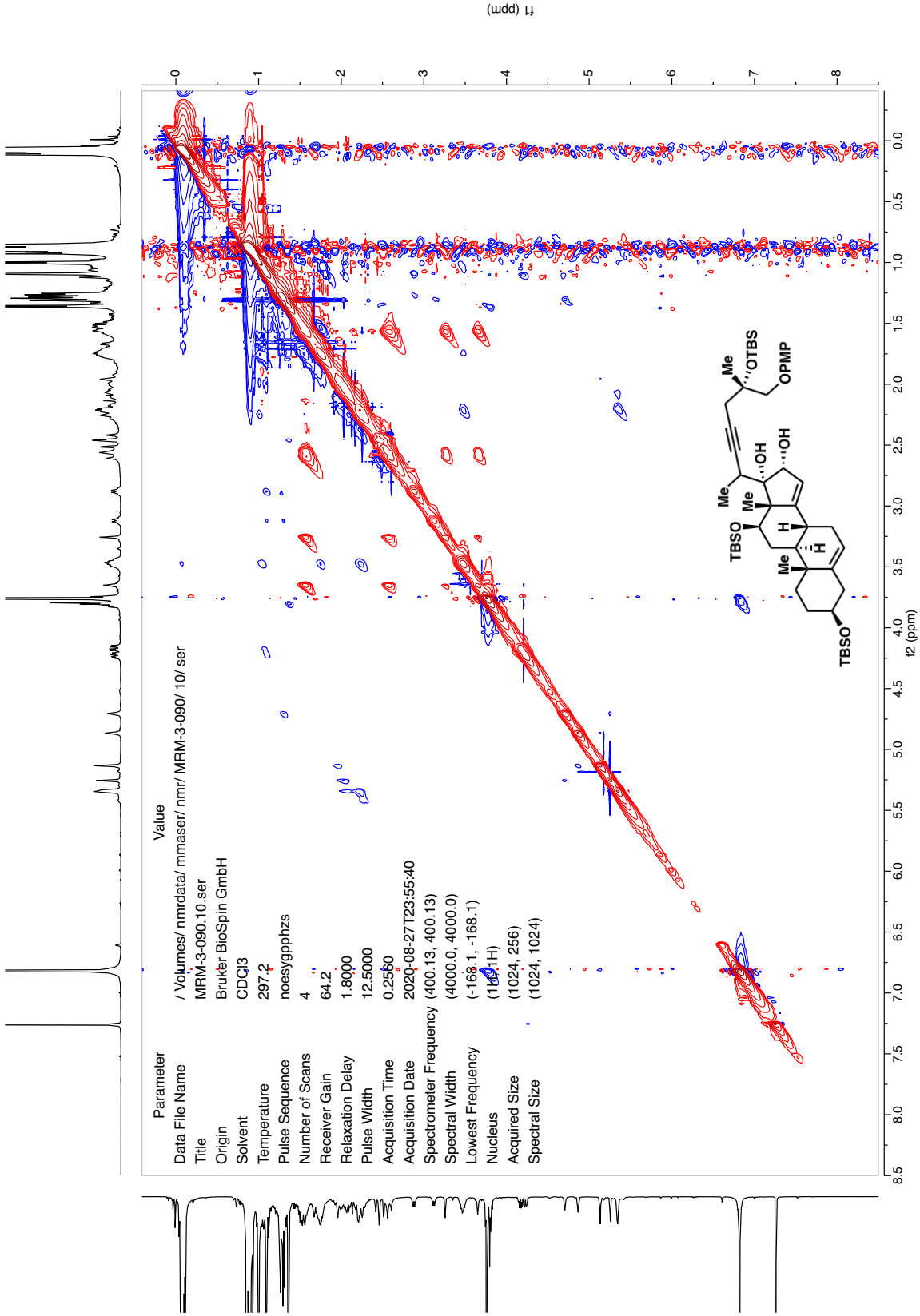

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-090/ 1/ fid
Title MRM-3-090.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 87.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-08-27T11:21:50
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.6
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
Me
Me
TBSO
H
H
Me
OH
OH
Me
OTBS
OPMP
f1 (ppm)

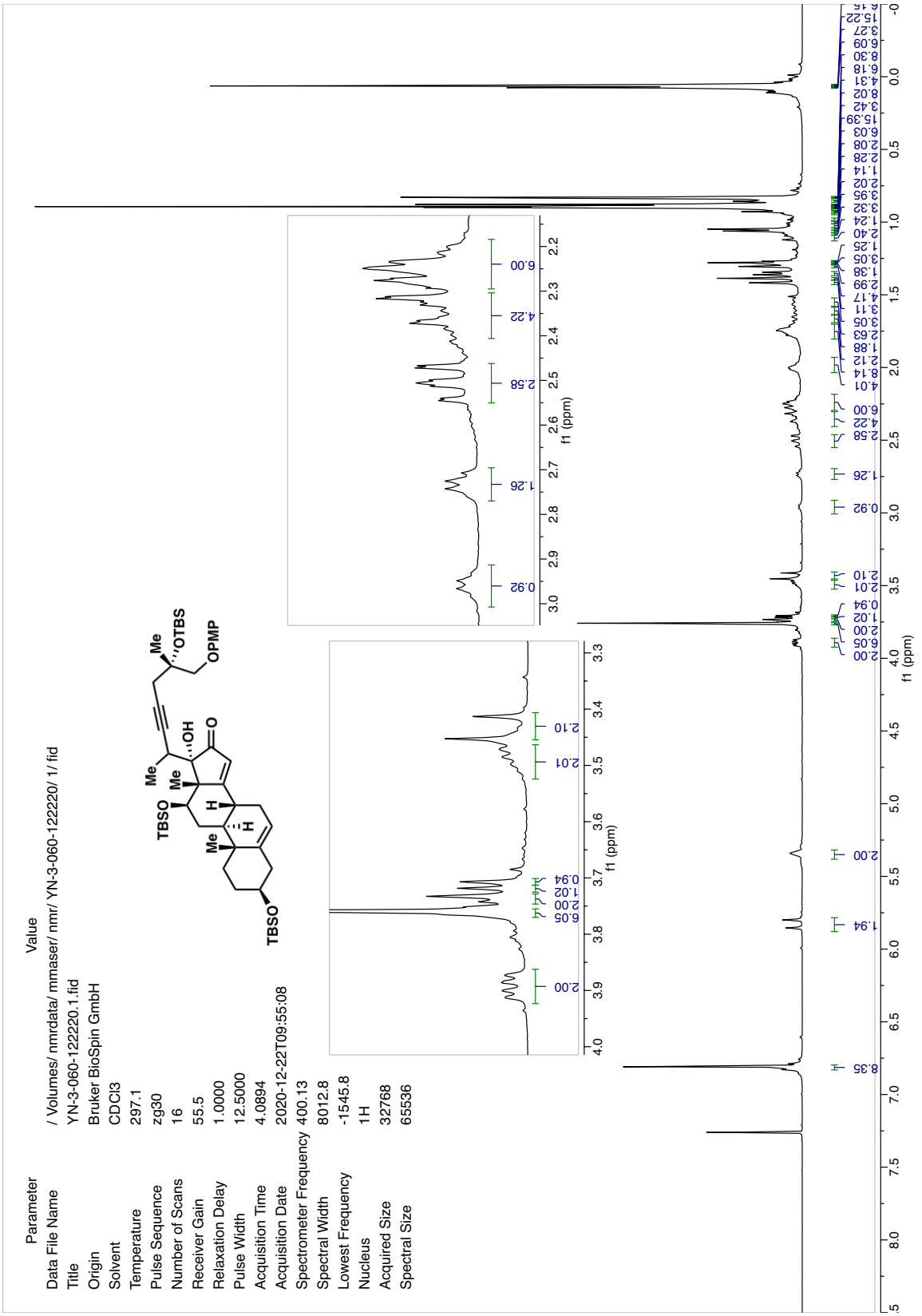

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-090/ 2/ fid
Title MRM-3-090.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-08-27T11:43:04
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1960.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
OTBS
OPMP
f1 (ppm)

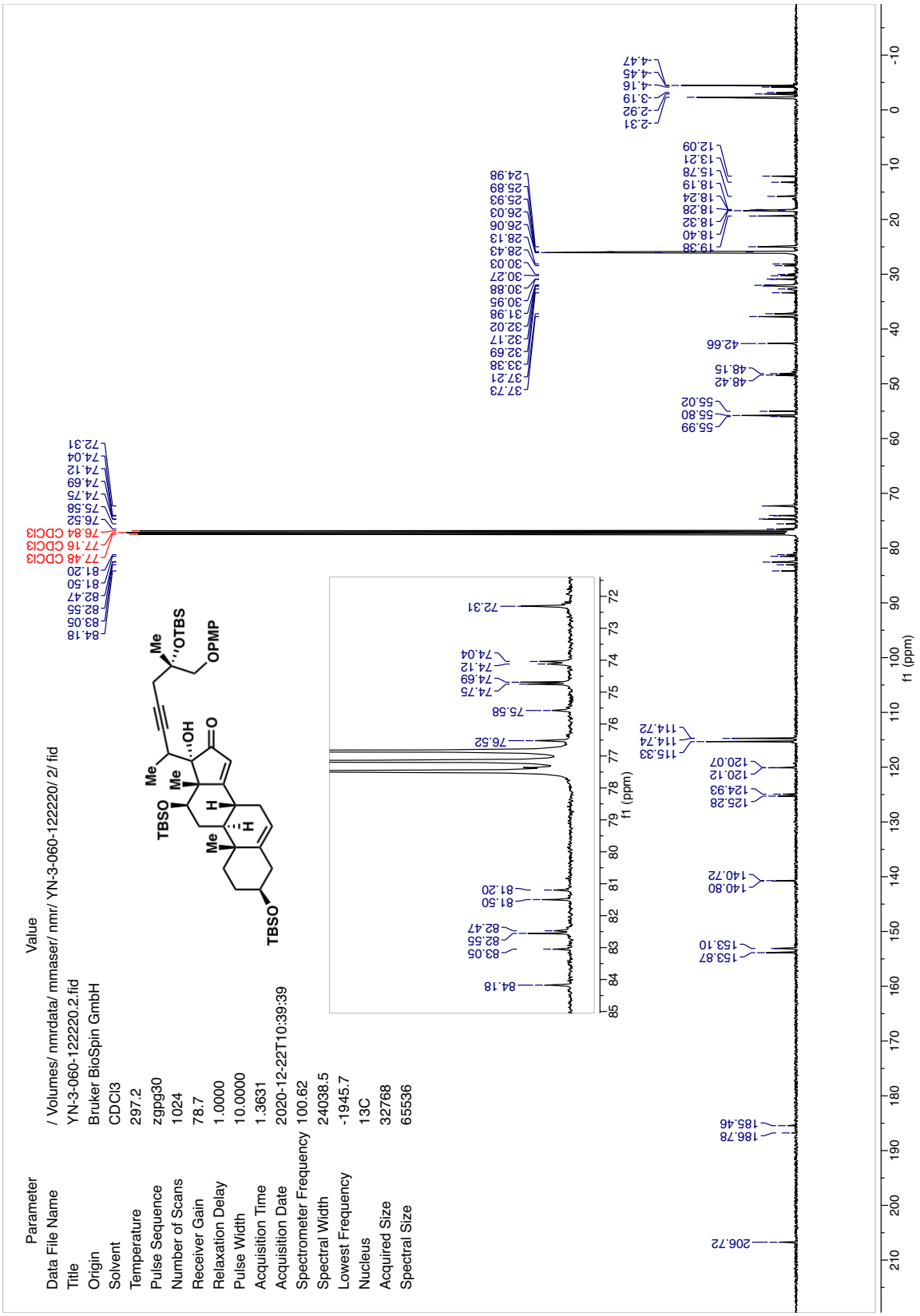

f1 (ppm)
f2 (ppm)
Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-090/ 4/ ser
Title
MRM-3-090.4.ser
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.2
Pulse Sequence
hsqcedetgpsisp2.4
Number of Scans
2
Receiver Gain
197.4
Relaxation Delay
1.5000
Pulse Width
12.5000
Acquisition Time
0.2135
Acquisition Date
2020-08-27T22:53:56
Spectrometer Frequency
(400.13, 100.62)
Spectral Width
(4795.4, 16611.3)
Lowest Frequency
(-527.1, -1265.3)
Nucleus
(1H, 13C)
Acquired Size
(1024, 180)
Spectral Size
(1024, 1024)
TBSO
Me
Me
H
H
Me
OH
OH
Me
OTBS
OPMP
TBSO

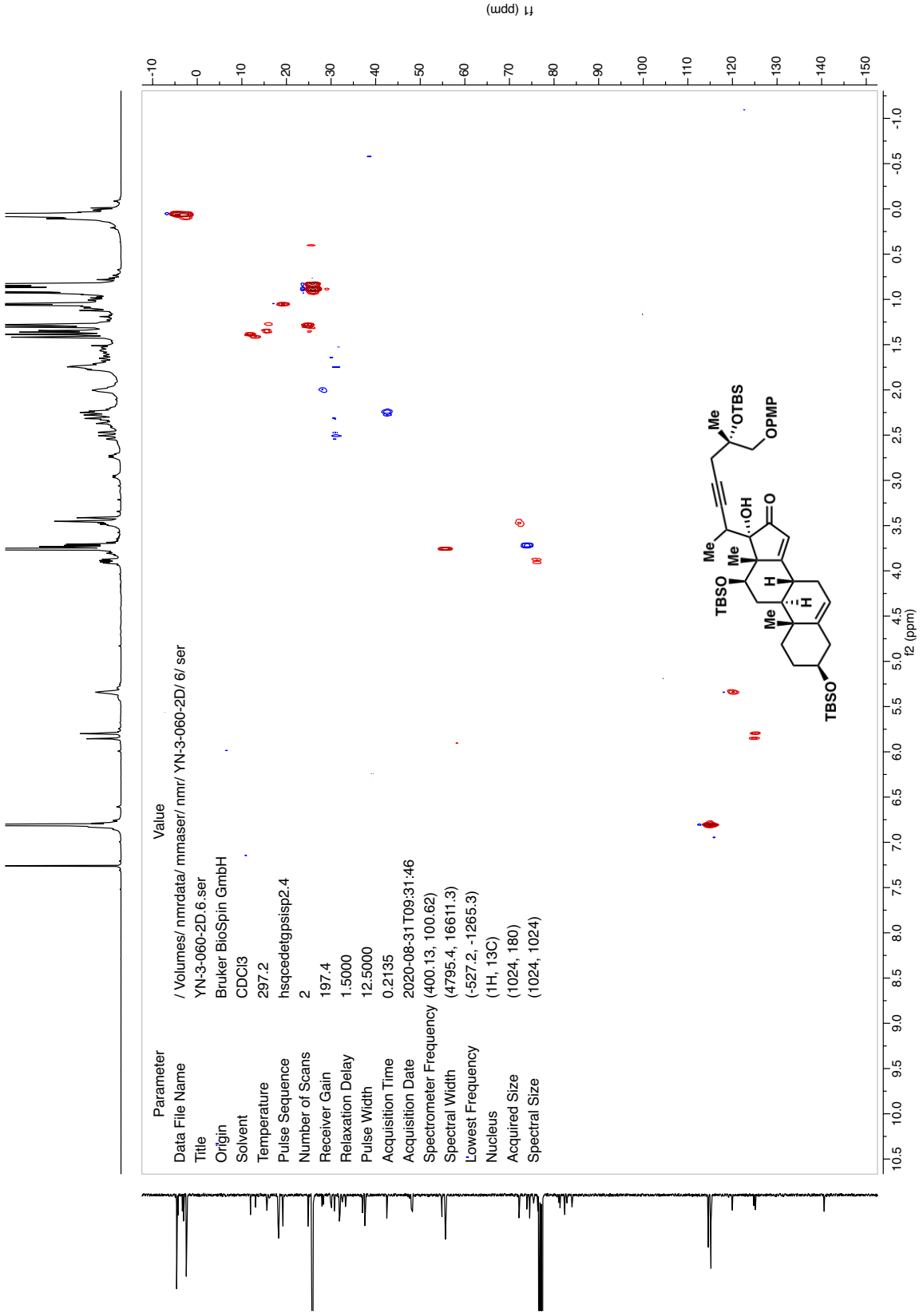

f1 (ppm)
f2 (ppm)
Parameter | Value
Data File Name | / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-090/ 6/ ser
Title | MRM-3-090.6.ser
Origin | Bruker BioSpin GmbH
Solvent | CDCl3
Temperature | 297.2
Pulse Sequence | cosygpppqf
Number of Scans | 1
Receiver Gain | 43.7
Relaxation Delay | 2.0000
Pulse Width | 12.5000
Acquisition Time | 0.1970
Acquisition Date | 2020-08-27T23:05:21
Spectrometer Frequency | (400.13, 400.13)
Spectral Width | (5197.5, 5197.5)
Lowest Frequency | (-208.0, -208.0)
Nucleus | (1H, 1H)
Acquired Size | (1024, 128)
Spectral Size | (1024, 1024)
TBSO
Me
OTBS
OPMP
OH
OH
Me
Me
TBSO
Me
H
H

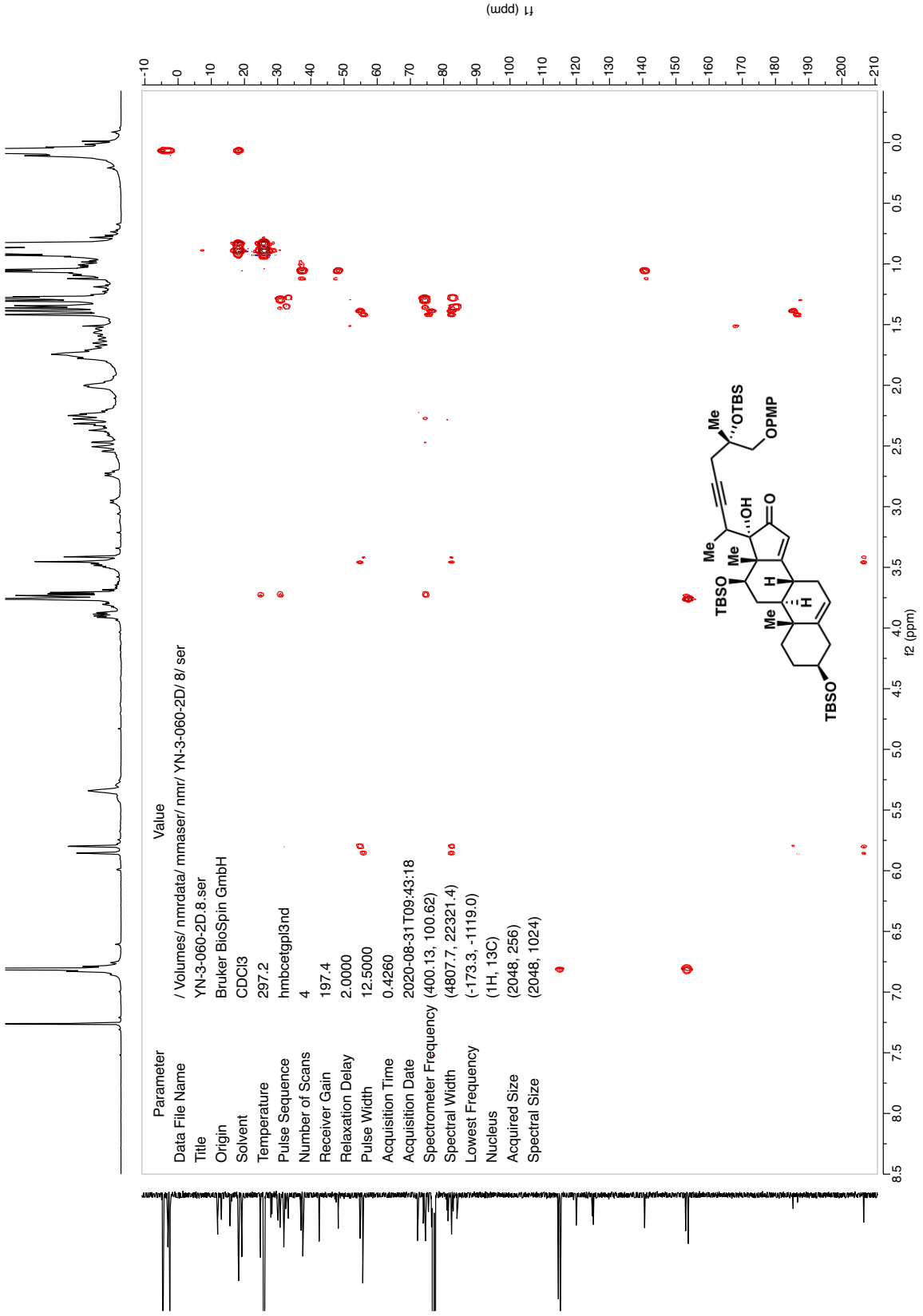

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-090/ 8/ ser
Title MRM-3-090.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-08-27T23:11:12
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-173.1, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
Me
Me
H
H
Me
OH
OH
Me
OTBS
OPMP
TBSO

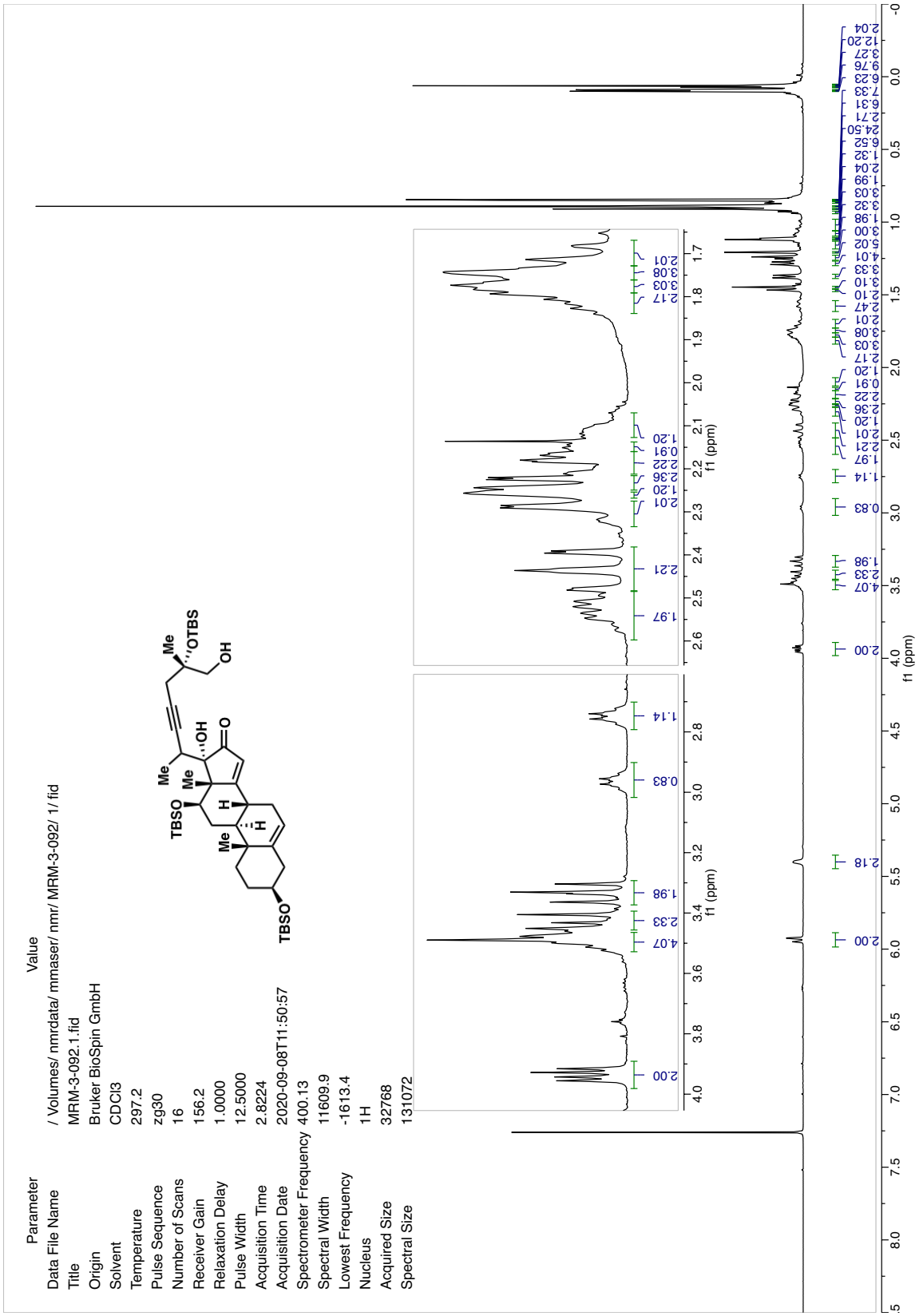

f1 (ppm)
f2 (ppm)
Parameter | Value
Data File Name | / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-090/ 10/ ser
Title | MRM-3-090.10.ser
Origin | Bruker BioSpin GmbH
Solvent | CDCl3
Temperature | 297.2
Pulse Sequence | noesygpphzs
Number of Scans | 4
Receiver Gain | 64.2
Relaxation Delay | 1.8000
Pulse Width | 12.5000
Acquisition Time | 0.2560
Acquisition Date | 2020-08-27T23:55:40
Spectrometer Frequency | (400.13, 400.13)
Spectral Width | (4000.0, 4000.0)
Lowest Frequency | (-168.1, -168.1)
Nucleus | (1H, 1H)
Acquired Size | (1024, 256)
Spectral Size | (1024, 1024)
Me
OTBS
OPMP
OH
OH
Me
Me
TBSO
H
H
Me
TBSO

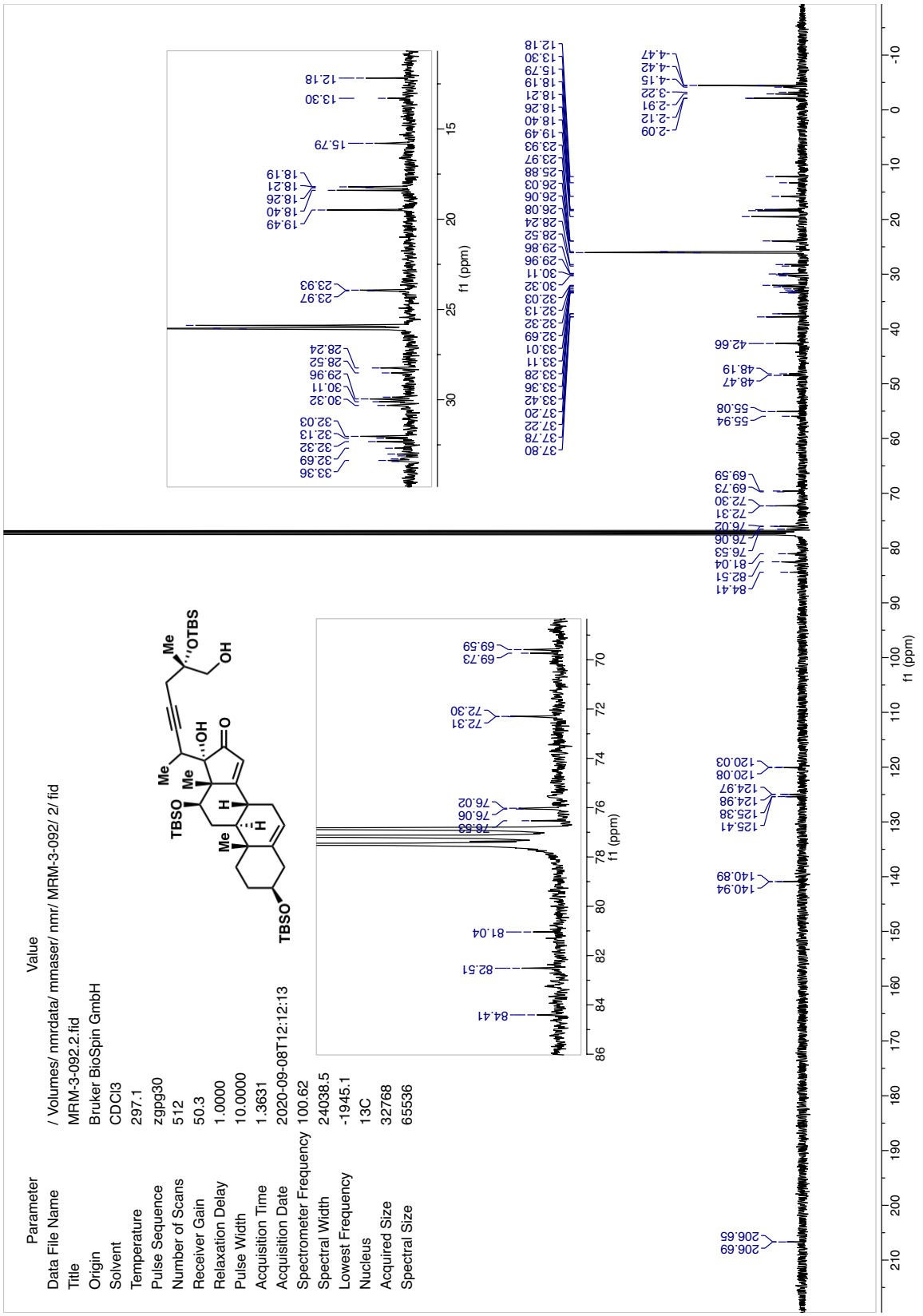

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-060-122220/ 1/ fid
Title YN-3-060-122220.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-12-22T09:55:08
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.8
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
TBSO
Me
Me
H
Me
H
TBSO
Me
OH
O
OTBS
OPMP
f1 (ppm)

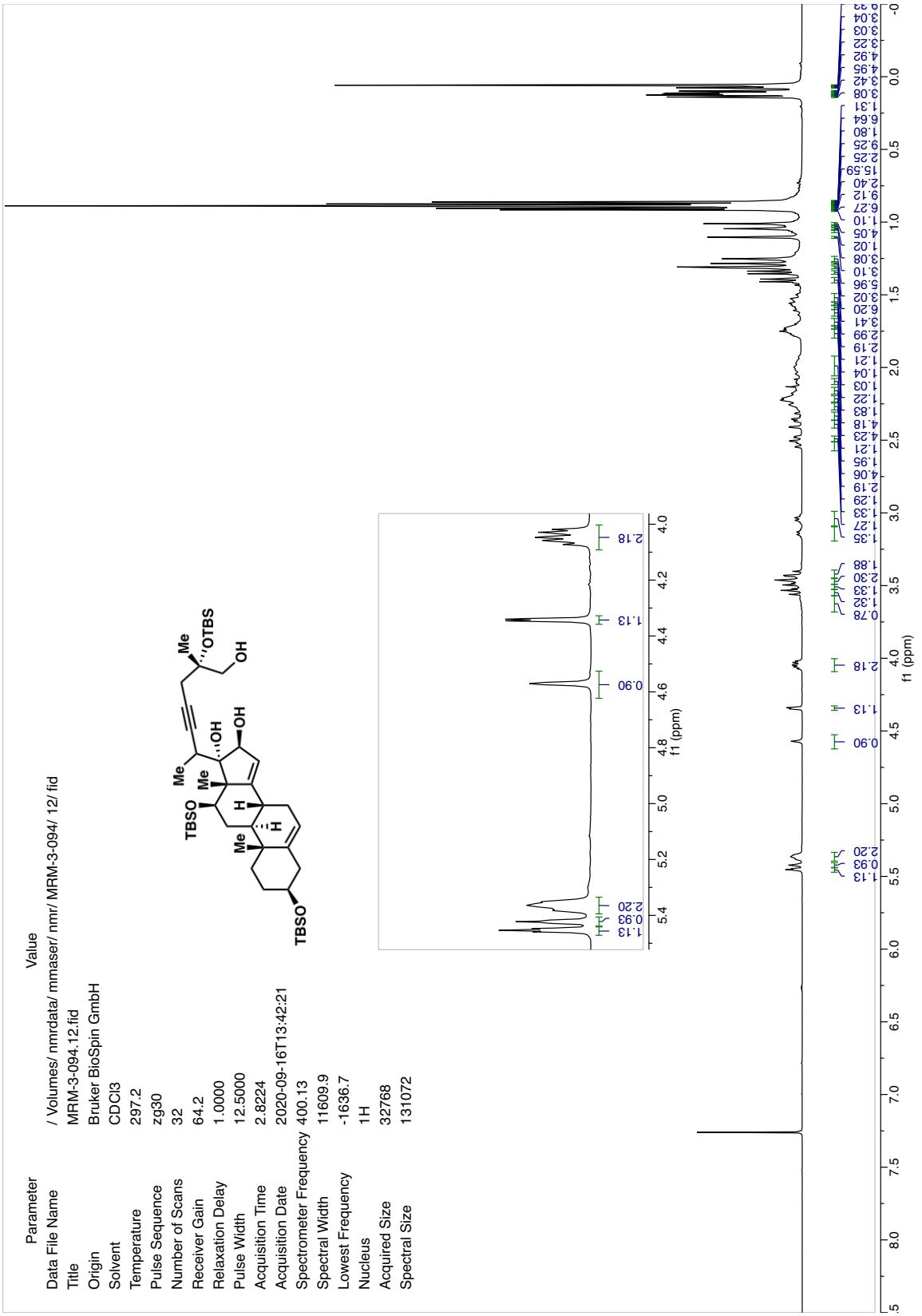

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-060-122220/ 2/ fid
Title YN-3-060-122220.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-12-22T10:39:39
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.7
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
OH
OTBS
OPMP
f1 (ppm)

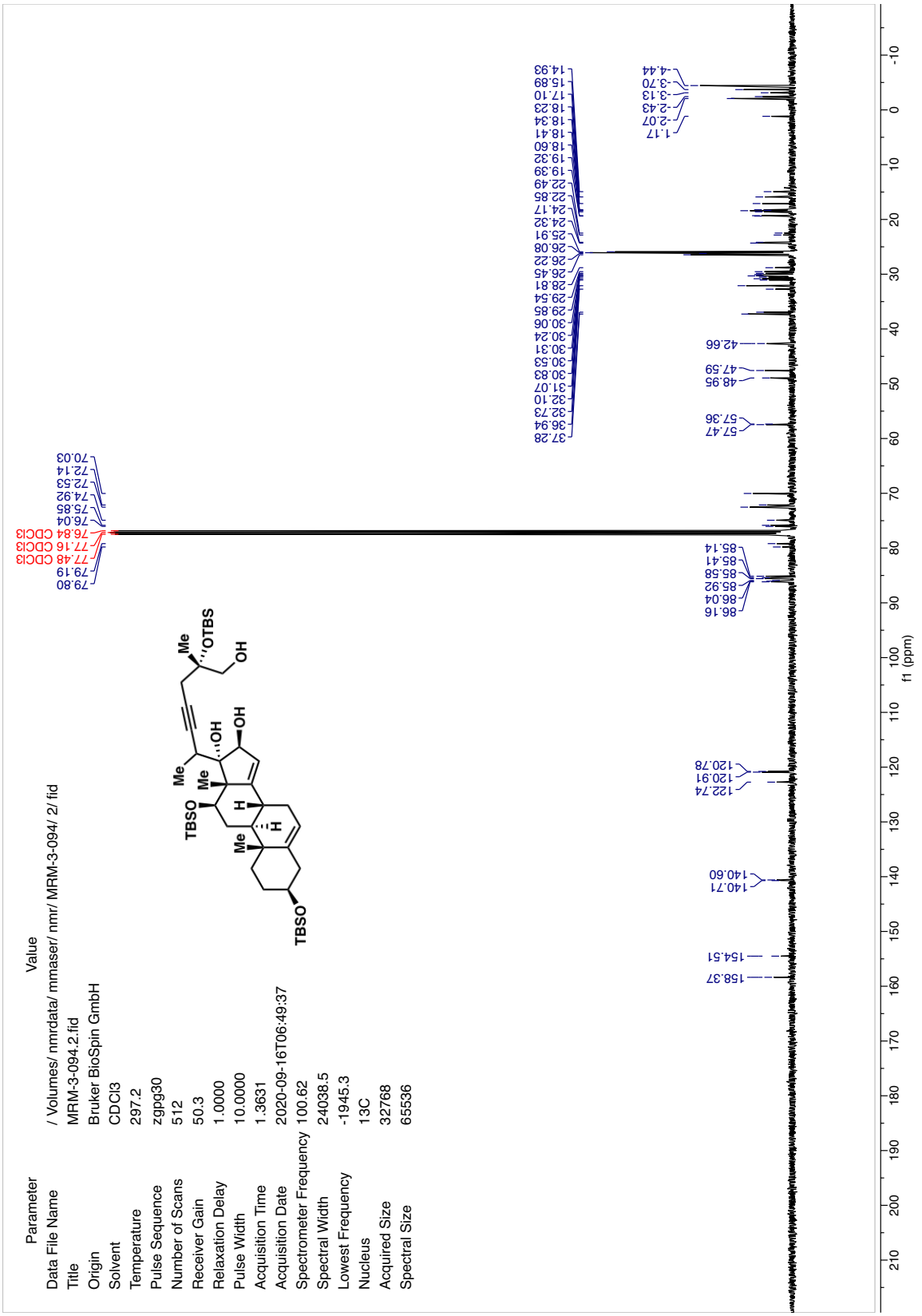

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-060-2D/ 6/ ser
Title YN-3-060-2D.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-08-31T09:31:46
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-527.2, -1265.3)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
Me
Me
Me
H
H
Me
OH
O
Me
OTBS
OPMP
TBSO

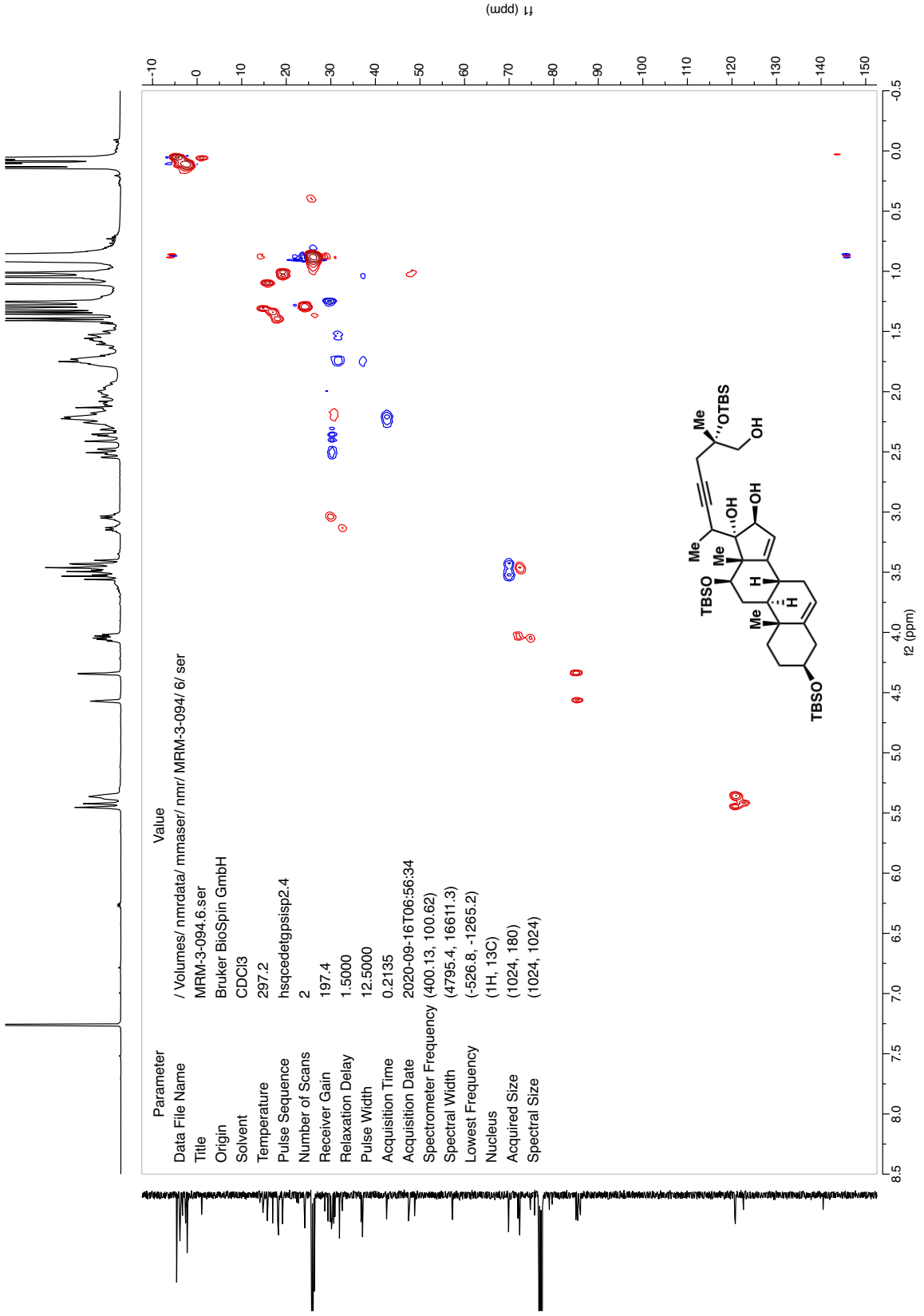

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-060-2D/ 8/ ser
Title YN-3-060-2D.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-08-31T09:43:18
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-173.3, -1119.0)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
Me
OTBS
OPMP
OH
O
H
H

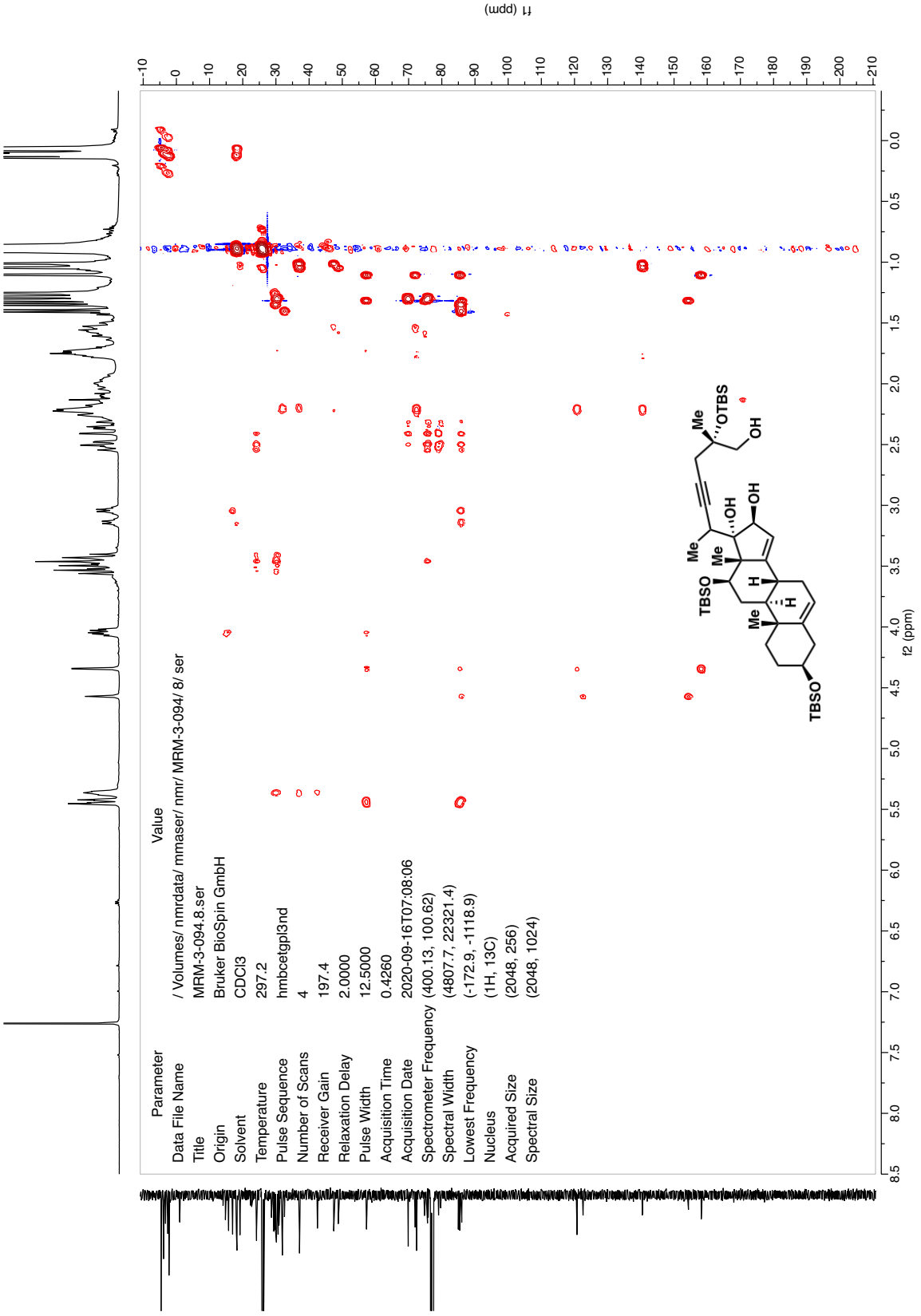

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-092/ 1/ fid
Title MRM-3-092.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-09-08T11:50:57
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.4
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
TBSO
OH
OTBS
Me
Me
Me
Me
H
H
O
f1 (ppm)

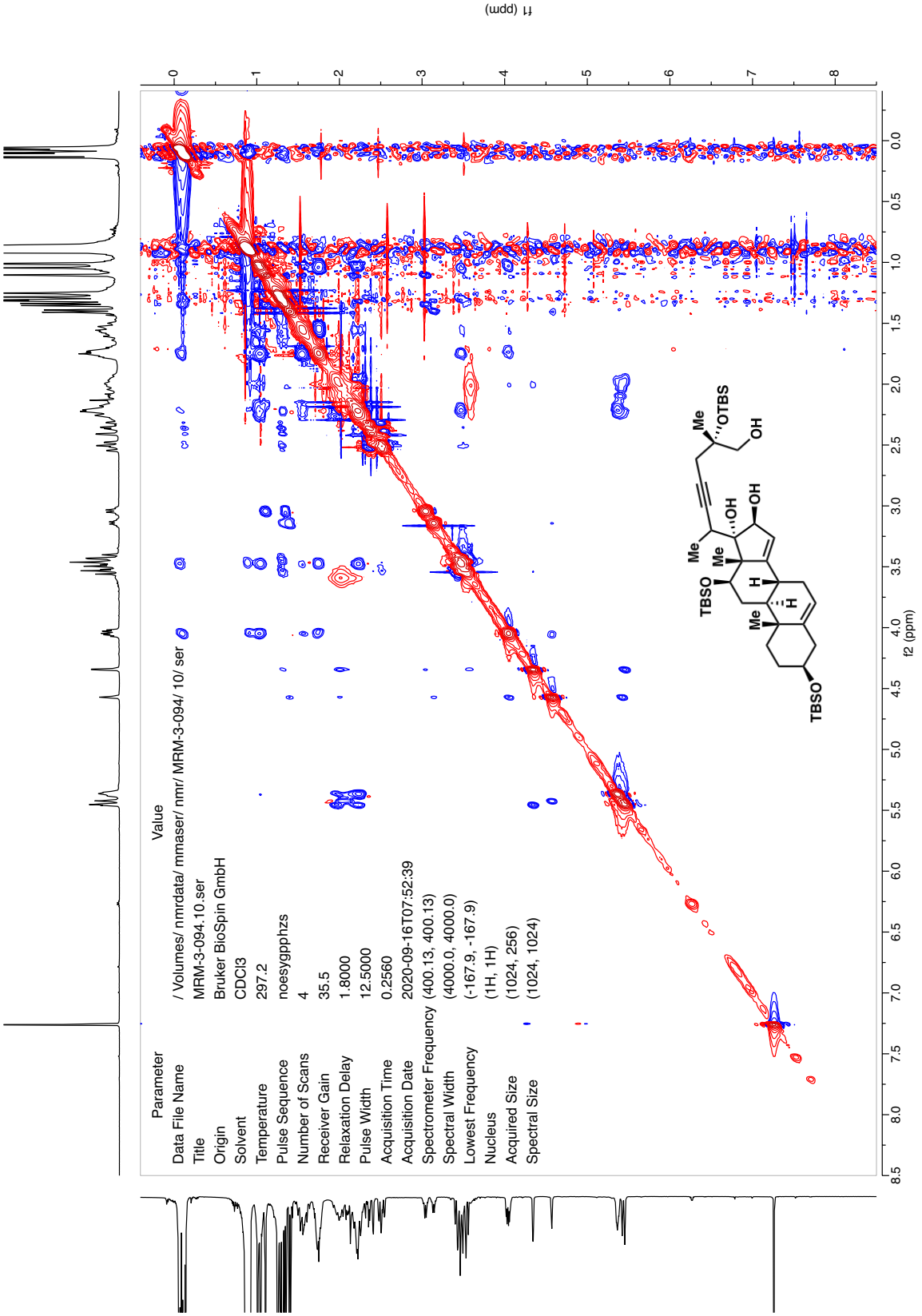

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-092/ 2/ fid
Title MRM-3-092.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-09-08T12:12:13
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.1
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
OTBS
OH
TBSO
OH
f1 (ppm)

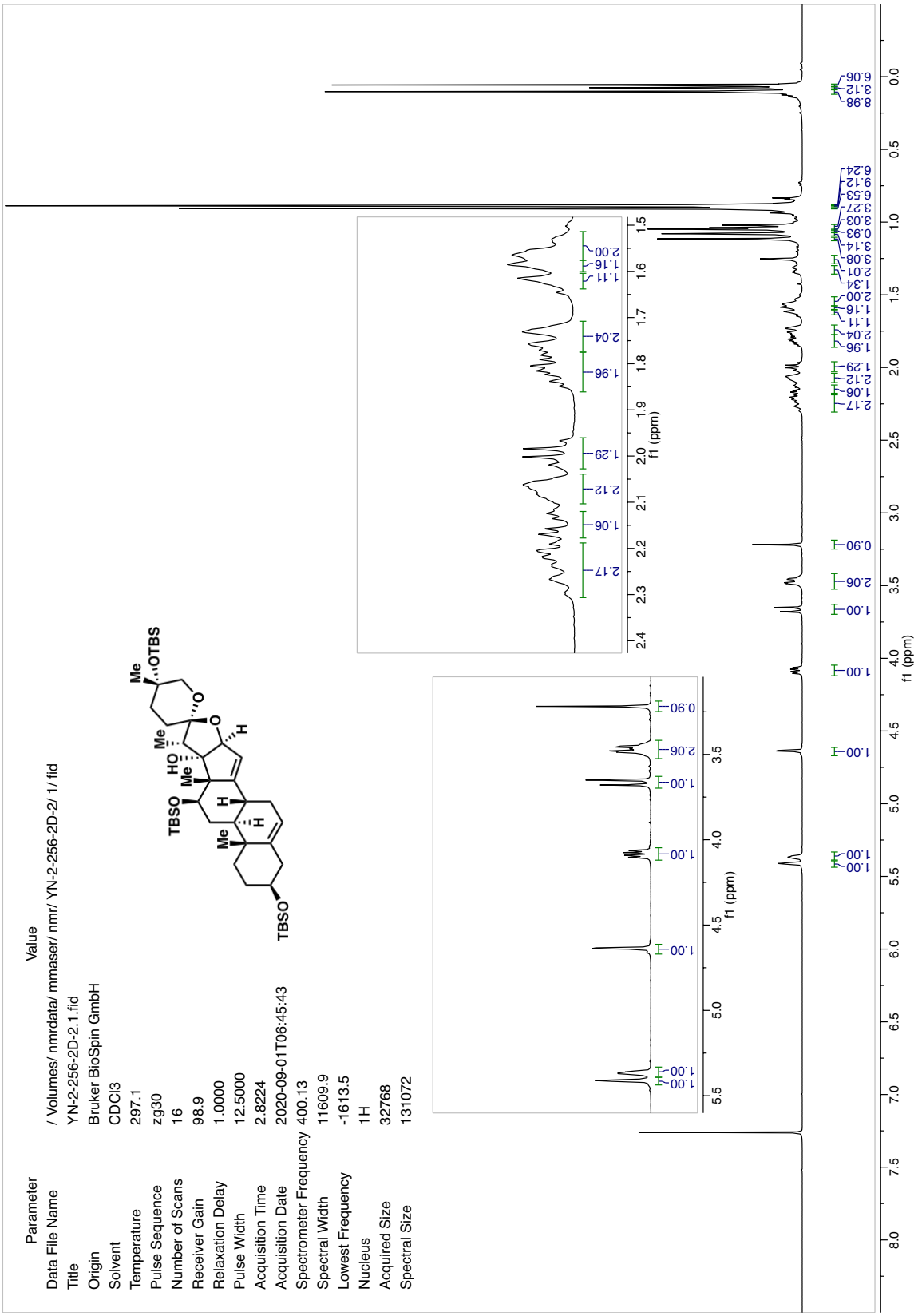

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-094/ 12/ fid
Title MRM-3-094.12.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-09-16T13:42:21
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1636.7
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
Me
Me
H
Me
H
TBSO
OH
OH
Me
OTBS
HO
f1 (ppm)

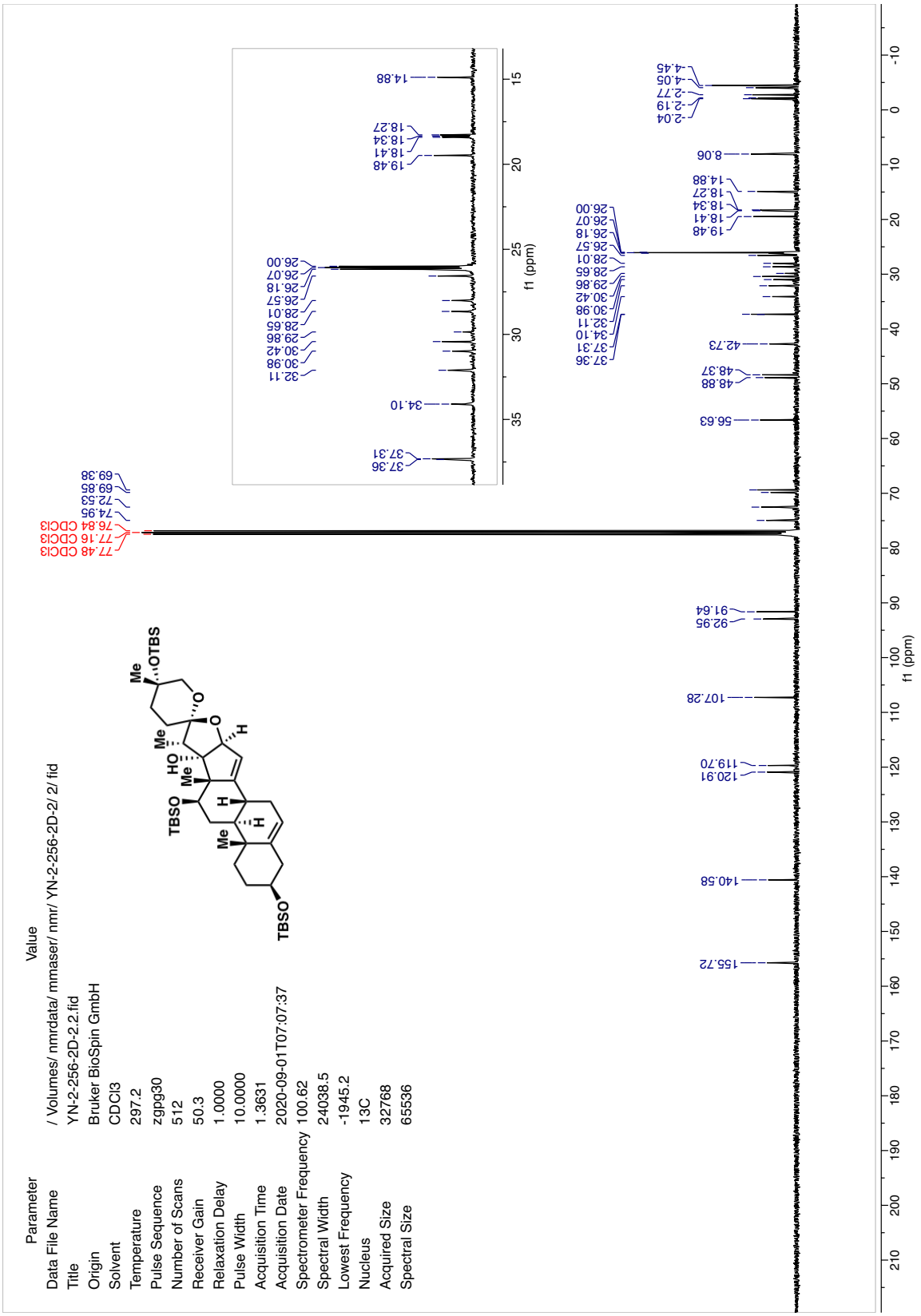

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-094/ 2/ fid
Title MRM-3-094.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-09-16T06:49:37
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
Me
Me
H
H
OH
OH
Me
OTBS
OH
TBSO
79.80
79.19
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
76.04
75.85
74.92
72.53
72.14
70.03
158.37
154.51
140.71
140.60
122.74
120.91
120.78
86.16
86.04
85.92
85.58
85.41
85.14
57.47
57.36
48.95
47.59
42.66
37.28
36.94
32.73
32.10
31.07
30.83
30.53
30.31
30.24
30.06
29.85
29.54
28.81
26.45
26.22
26.08
25.91
24.32
24.17
22.85
22.49
19.39
19.32
18.60
18.41
18.34
18.23
17.10
15.89
14.93
1.17
-2.07
-2.43
-3.13
-3.70
-4.44
f1 (ppm)

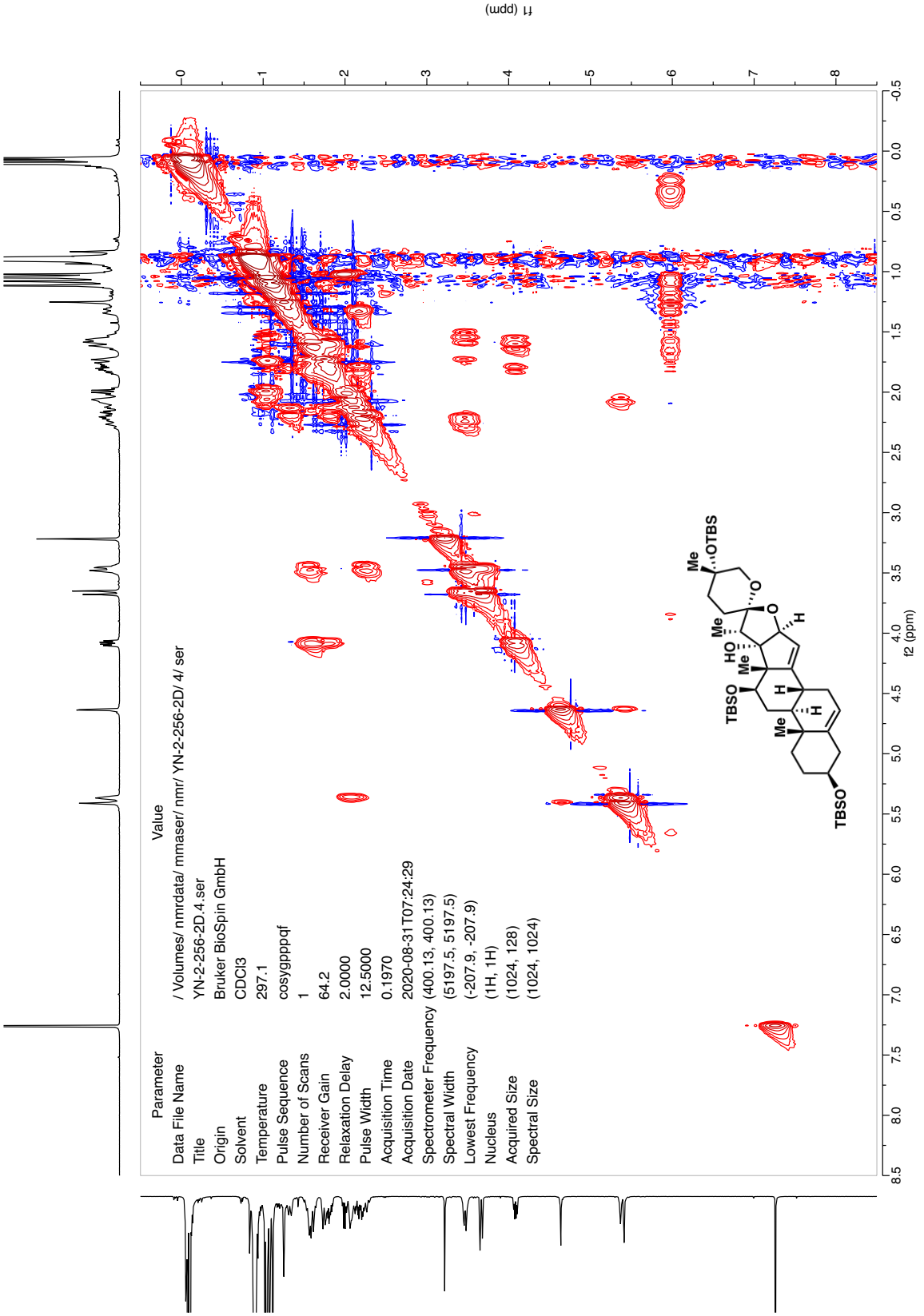

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-094/ 6/ ser
Title MRM-3-094.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-09-16T06:56:34
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-526.8, -1265.2)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
Me
Me
H
H
Me
TBSO
OH
OH
Me
OTBS
OH

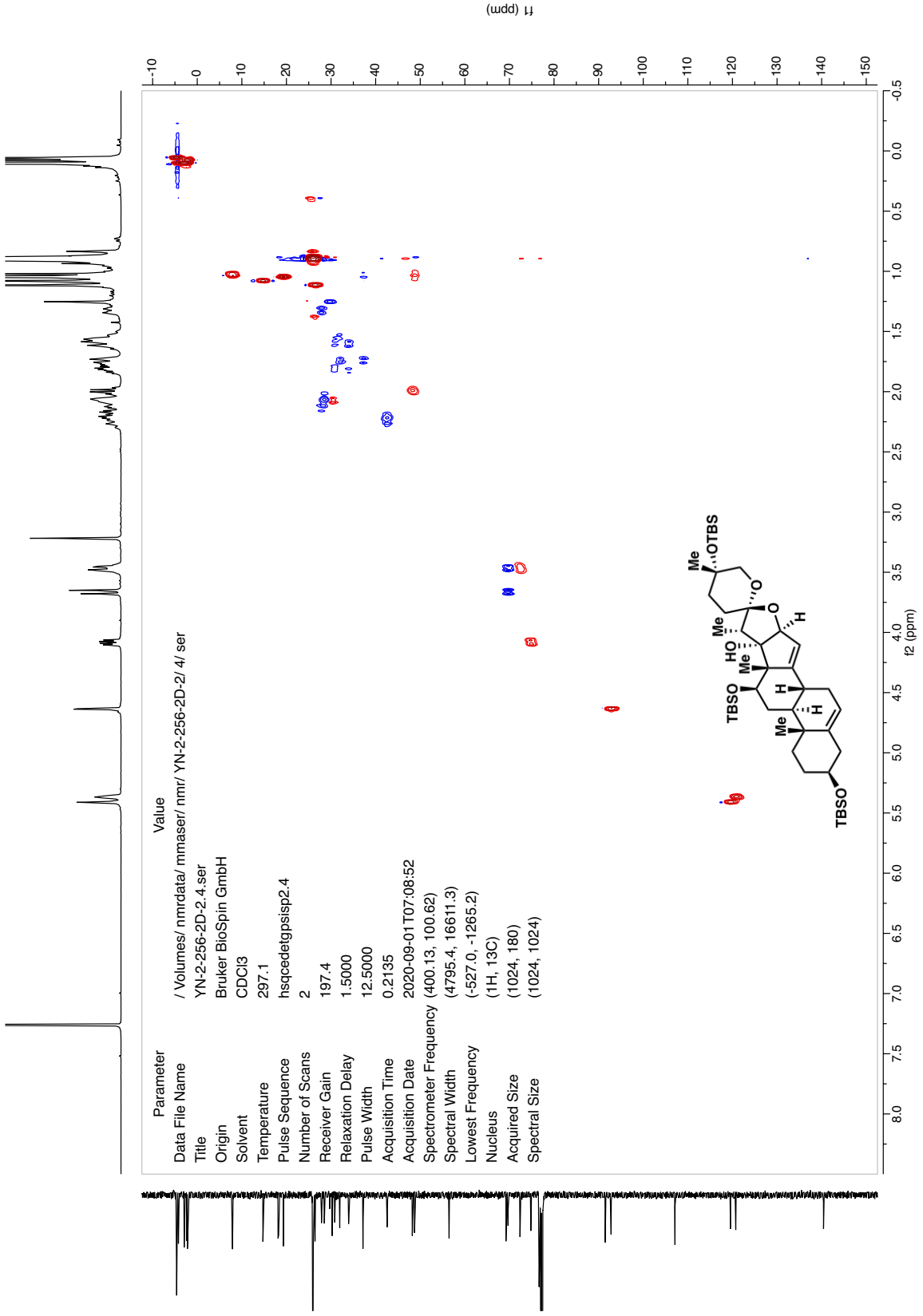

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-094/ 8/ ser
Title MRM-3-094.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-09-16T07:08:06
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-172.9, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
Me
Me
H
H
Me
OH
OH
Me
OTBS
OH
TBSO

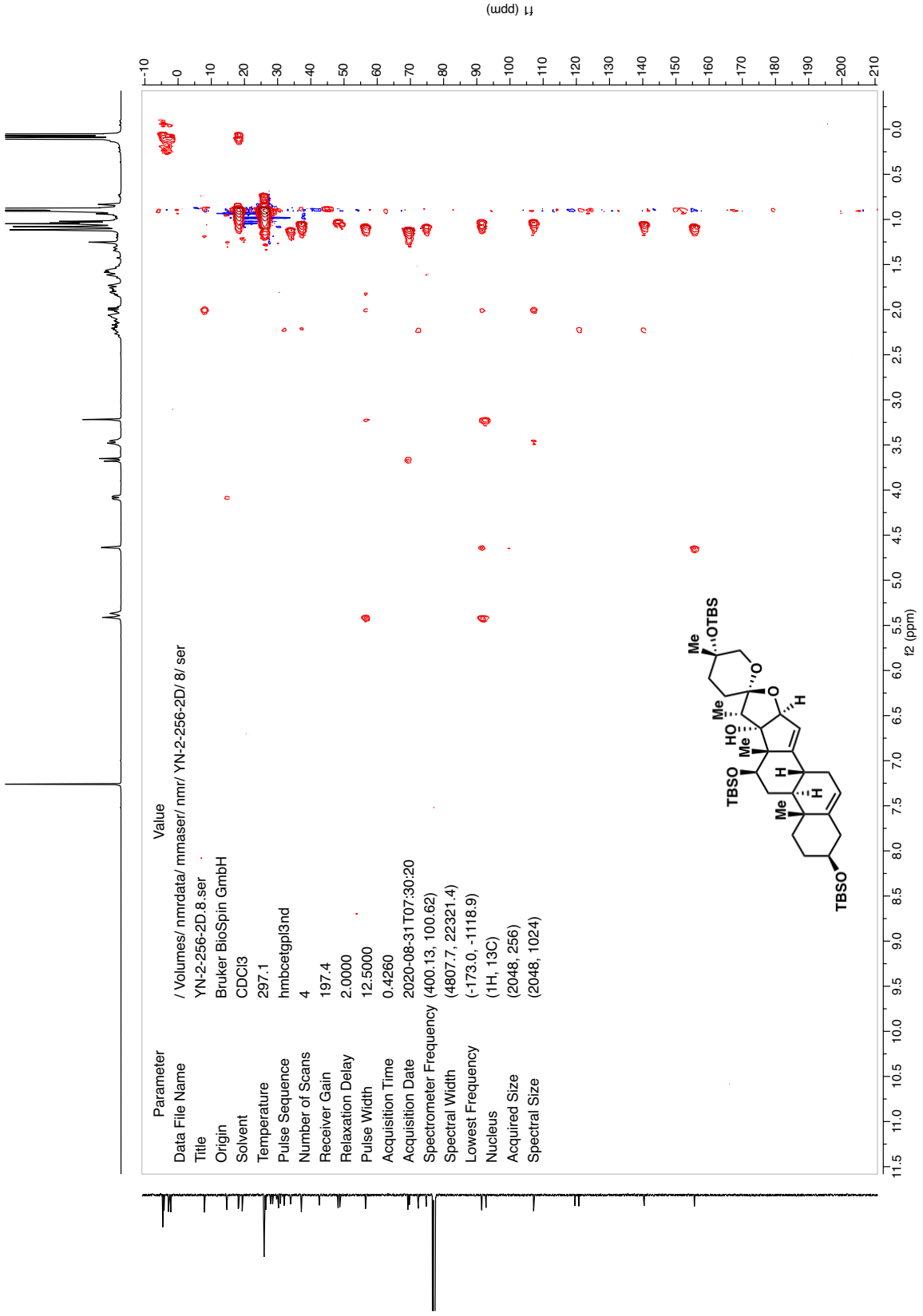

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-094/ 10/ ser
Title MRM-3-094.10.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence noesygpphzs
Number of Scans 4
Receiver Gain 35.5
Relaxation Delay 1.8000
Pulse Width 12.5000
Acquisition Time 0.2560
Acquisition Date 2020-09-16T07:52:39
Spectrometer Frequency (400.13, 400.13)
Spectral Width (4000.0, 4000.0)
Lowest Frequency (-167.9, -167.9)
Nucleus (1H, 1H)
Acquired Size (1024, 256)
Spectral Size (1024, 1024)
TBSO
Me
Me
H
H
Me
OH
OH
OH
Me
OTBS
TBSO

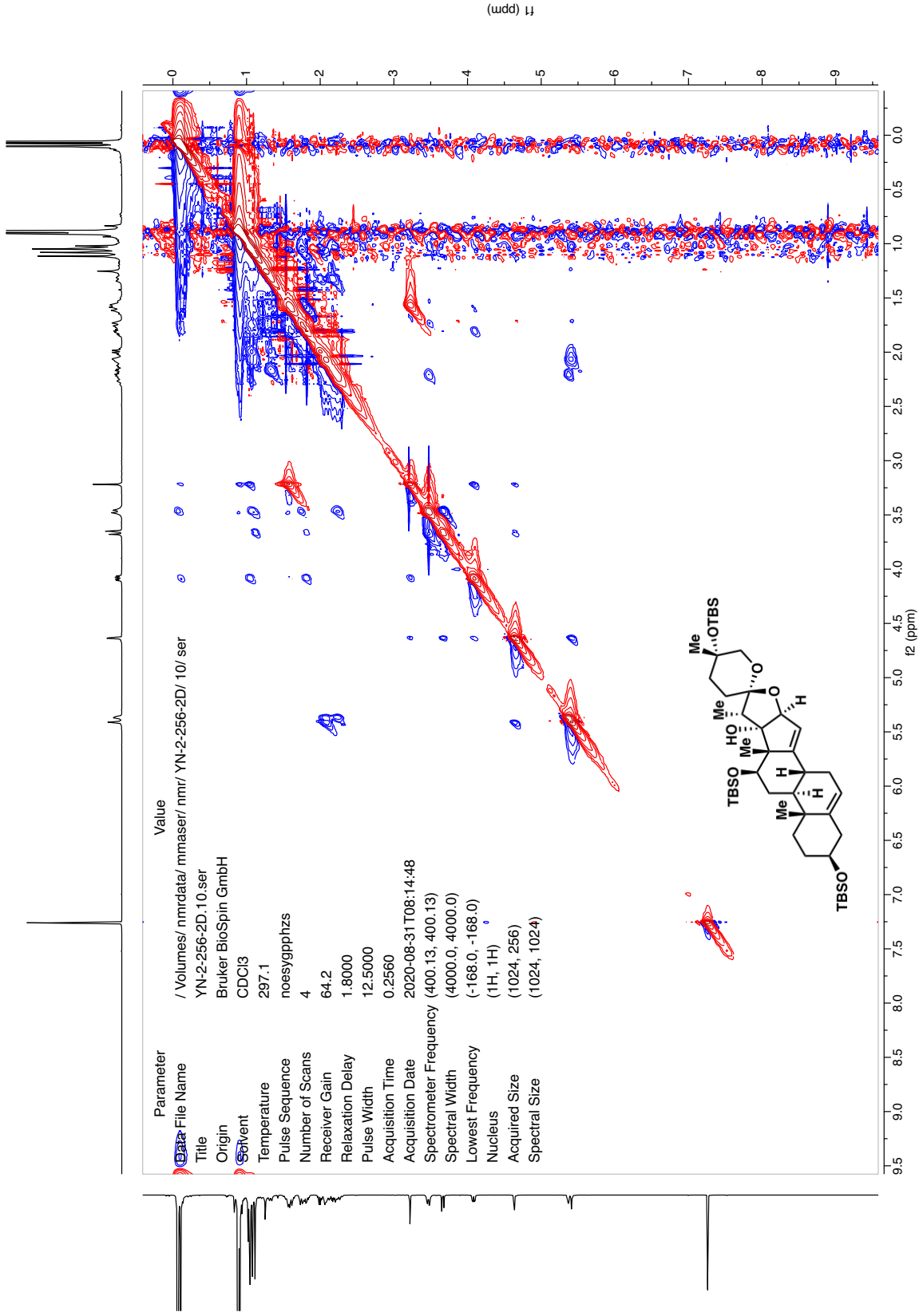

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-2-256-2D-2/ 1/ fid
Title YN-2-256-2D-2.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 98.9
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-09-01T06:45:43
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.5
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
Me
OTBS
HO
Me
TBSO
Me
H
H
H
O
O
TBSO
f1 (ppm)

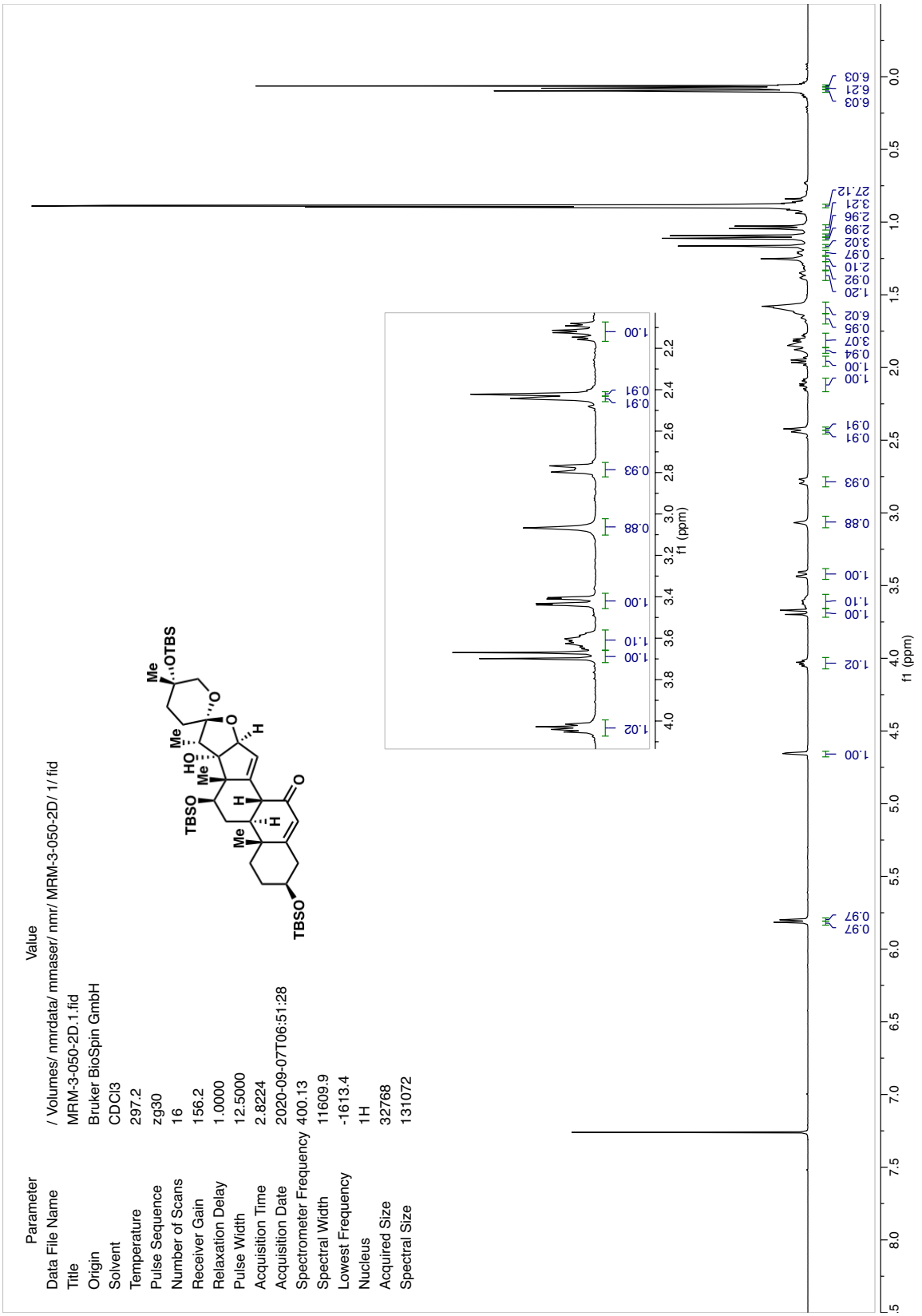

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-2-256-2D-2/ 2/ fid
Title YN-2-256-2D-2.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-09-01T07:07:37
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.2
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
OTBS
Me
HO
TBSO
f1 (ppm)

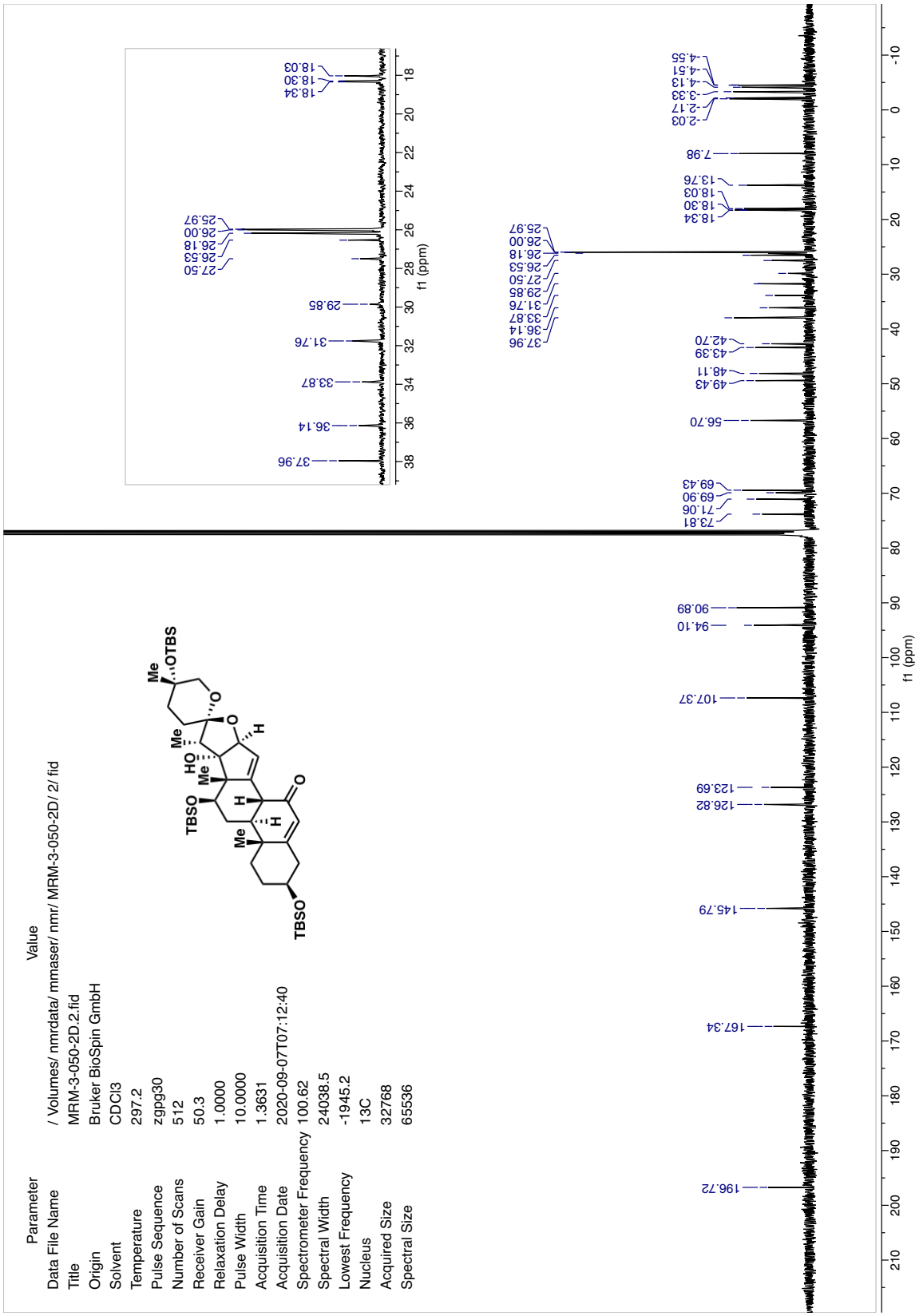

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-2-256-2D/ 4/ ser
Title YN-2-256-2D.4.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence cosygpppqf
Number of Scans 1
Receiver Gain 64.2
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.1970
Acquisition Date 2020-08-31T07:24:29
Spectrometer Frequency (400.13, 400.13)
Spectral Width (5197.5, 5197.5)
Lowest Frequency (-207.9, -207.9)
Nucleus (1H, 1H)
Acquired Size (1024, 128)
Spectral Size (1024, 1024)
Me
OTBS
O
O
H
Me
HO
Me
TBSO
H
H
Me
TBSO

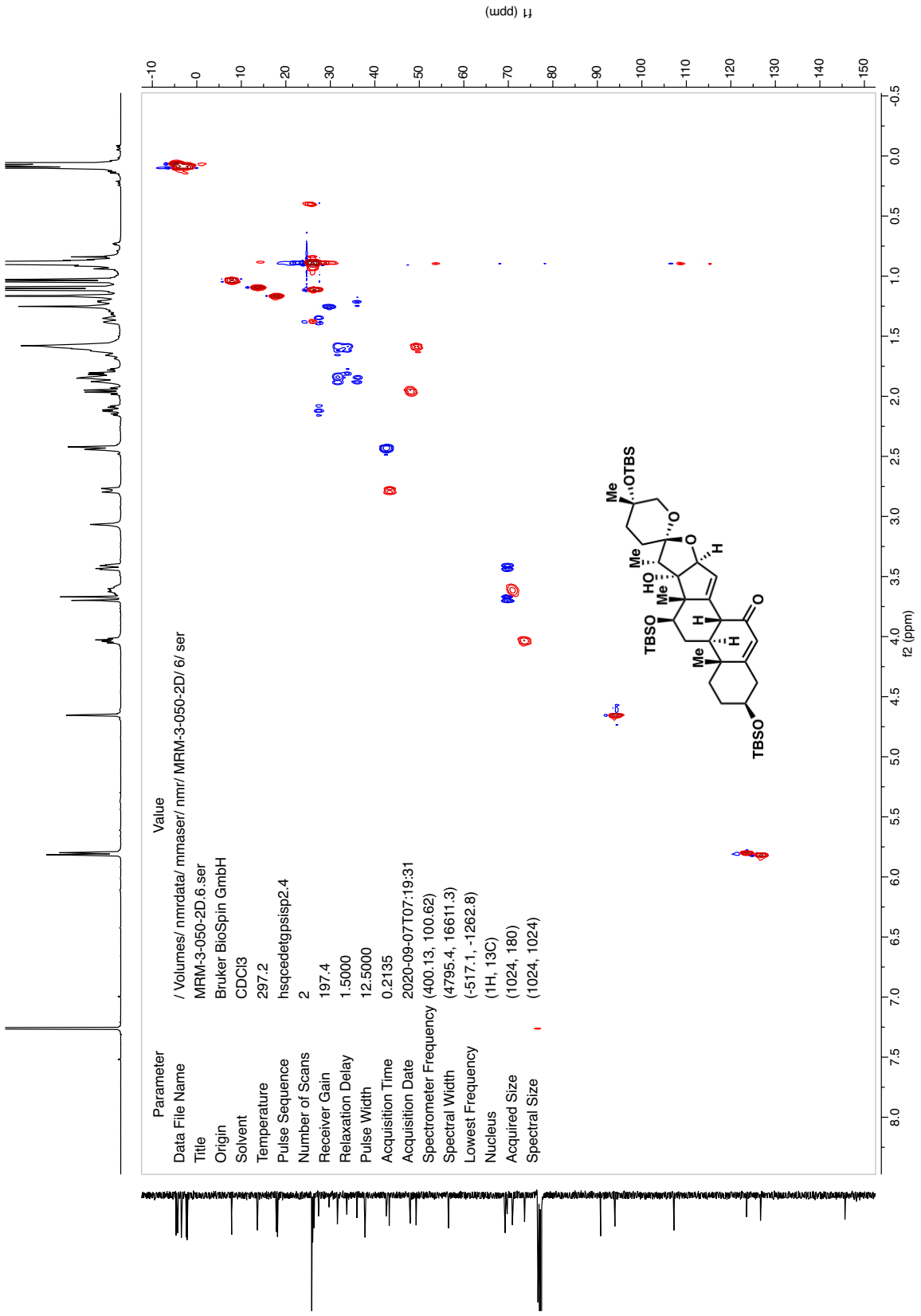

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-2-256-2D-2/ 4/ ser
Title YN-2-256-2D-2.4.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-09-01T07:08:52
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-527.0, -1265.2)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
HO
OTBS
Me
H

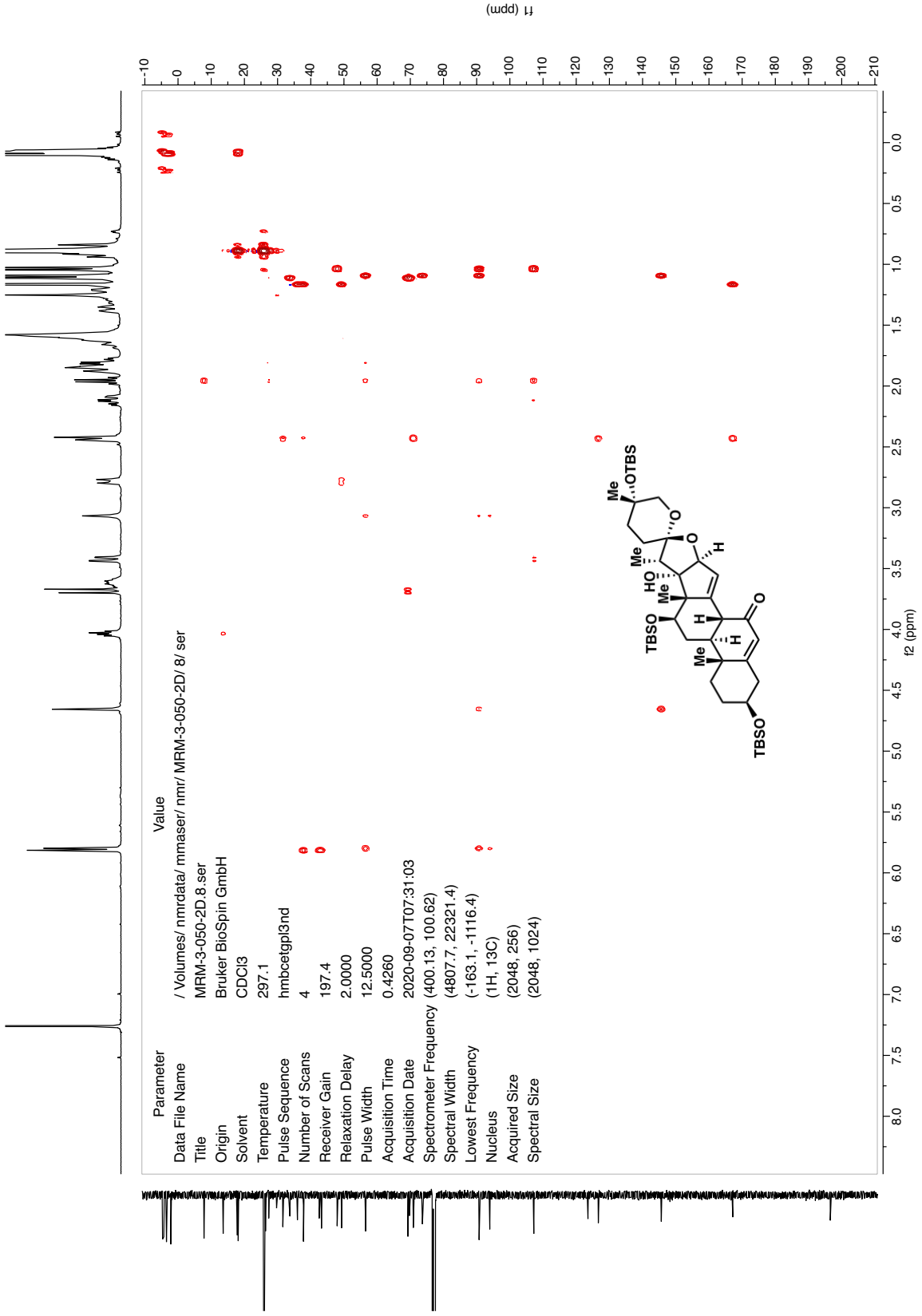

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-2-256-2D/ 8/ ser
Title YN-2-256-2D.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-08-31T07:30:20
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-173.0, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
Me
HO
Me
Me
H
H
H
Me
OTBS
TBSO

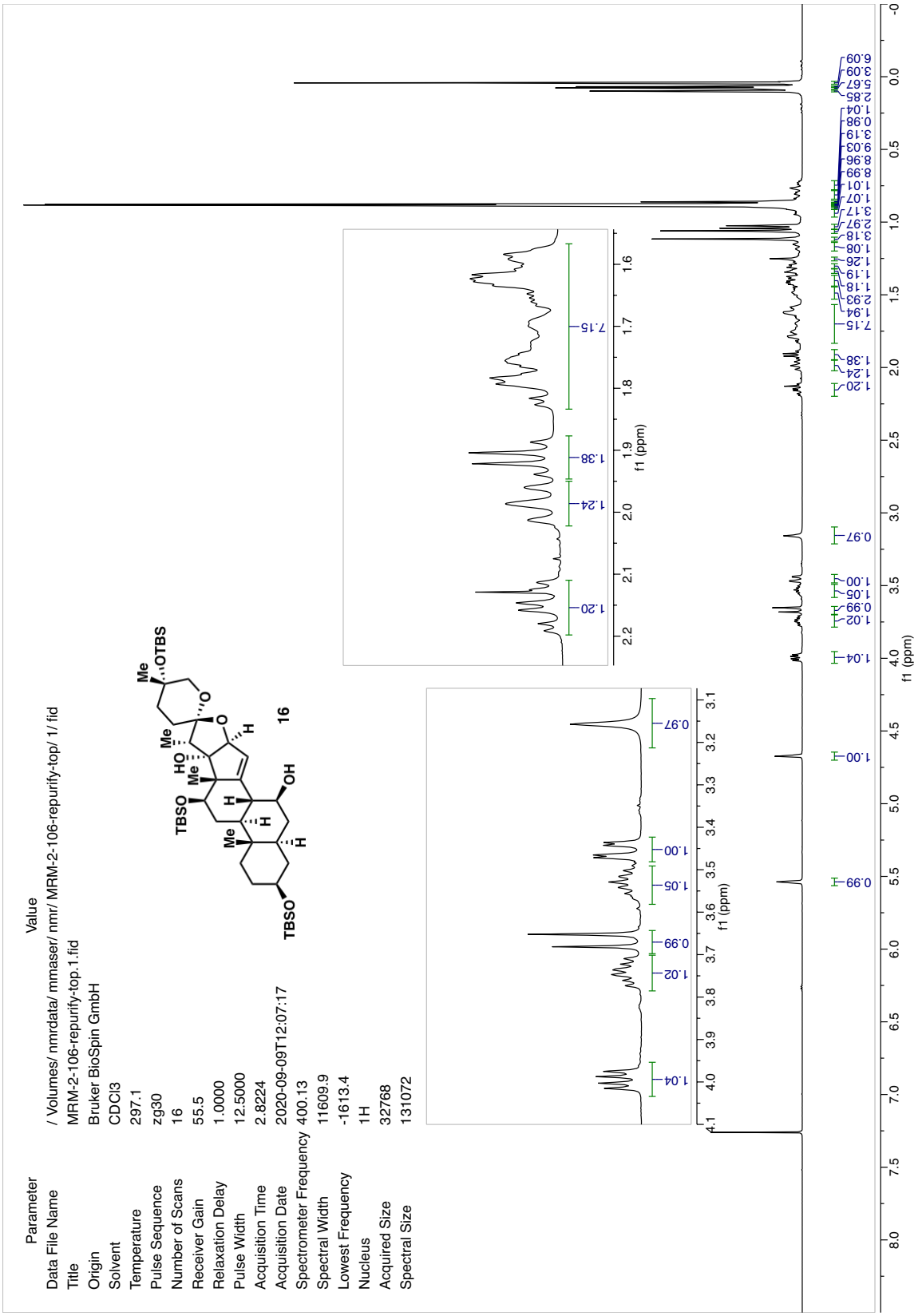

f1 (ppm)
f2 (ppm)
Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ YN-2-256-2D/ 10/ ser
Title
YN-2-256-2D.10.ser
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
noesygpphzs
Number of Scans
4
Receiver Gain
64.2
Relaxation Delay
1.8000
Pulse Width
12.5000
Acquisition Time
0.2560
Acquisition Date
2020-08-31T08:14:48
Spectrometer Frequency
(400.13, 400.13)
Spectral Width
(4000.0, 4000.0)
Lowest Frequency
(-168.0, -168.0)
Nucleus
(1H, 1H)
Acquired Size
(1024, 256)
Spectral Size
(1024, 1024)
Me
OTBS
O
O
H
Me
HO
Me
TBSO
H
H
Me
TBSO

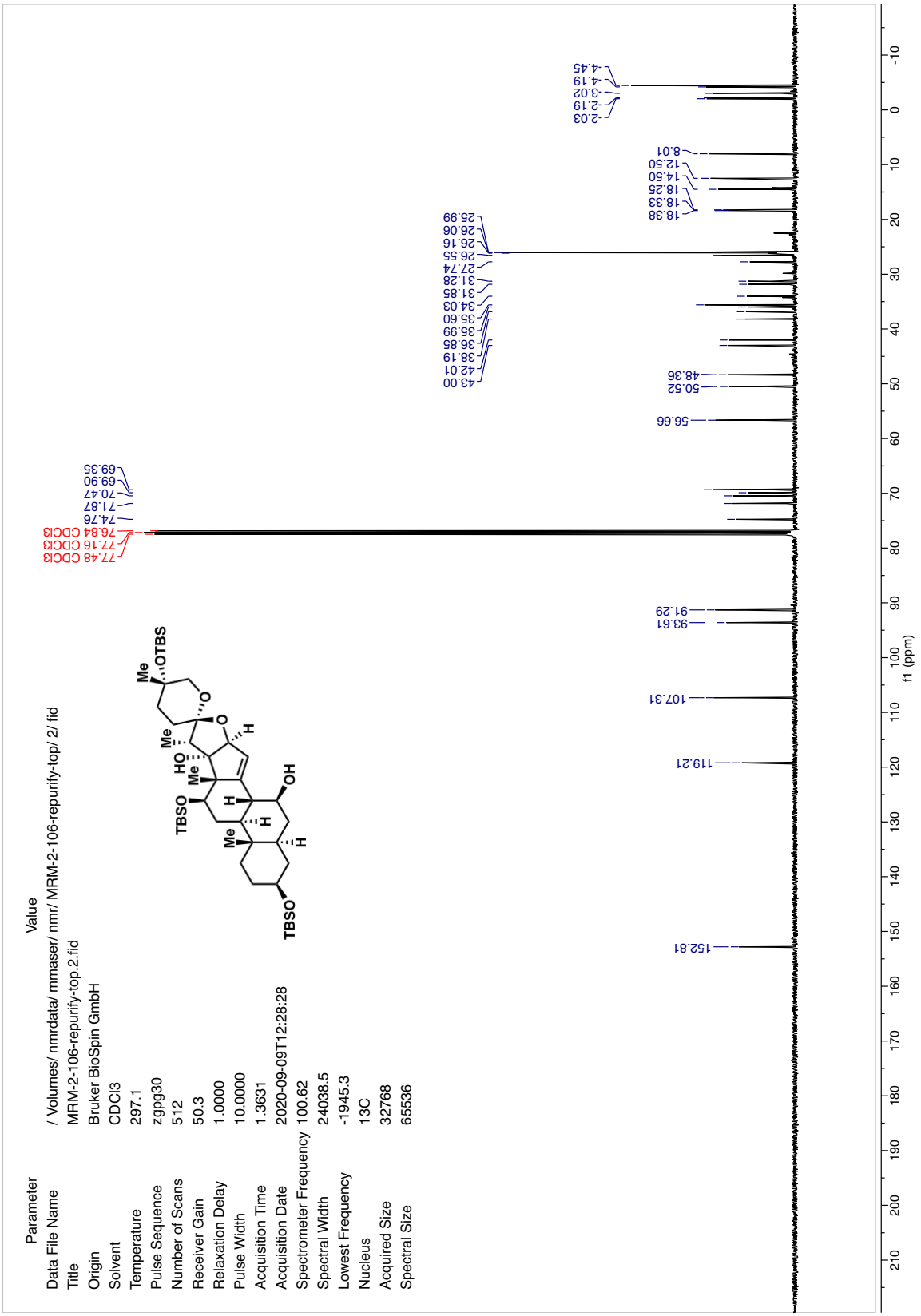

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-050-2D/ 1/ fid
Title MRM-3-050-2D.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-09-07T06:51:28
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.4
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
Me
OTBS
HO
Me
TBSO
Me
H
H
H
O
O
O
TBSO
f1 (ppm)

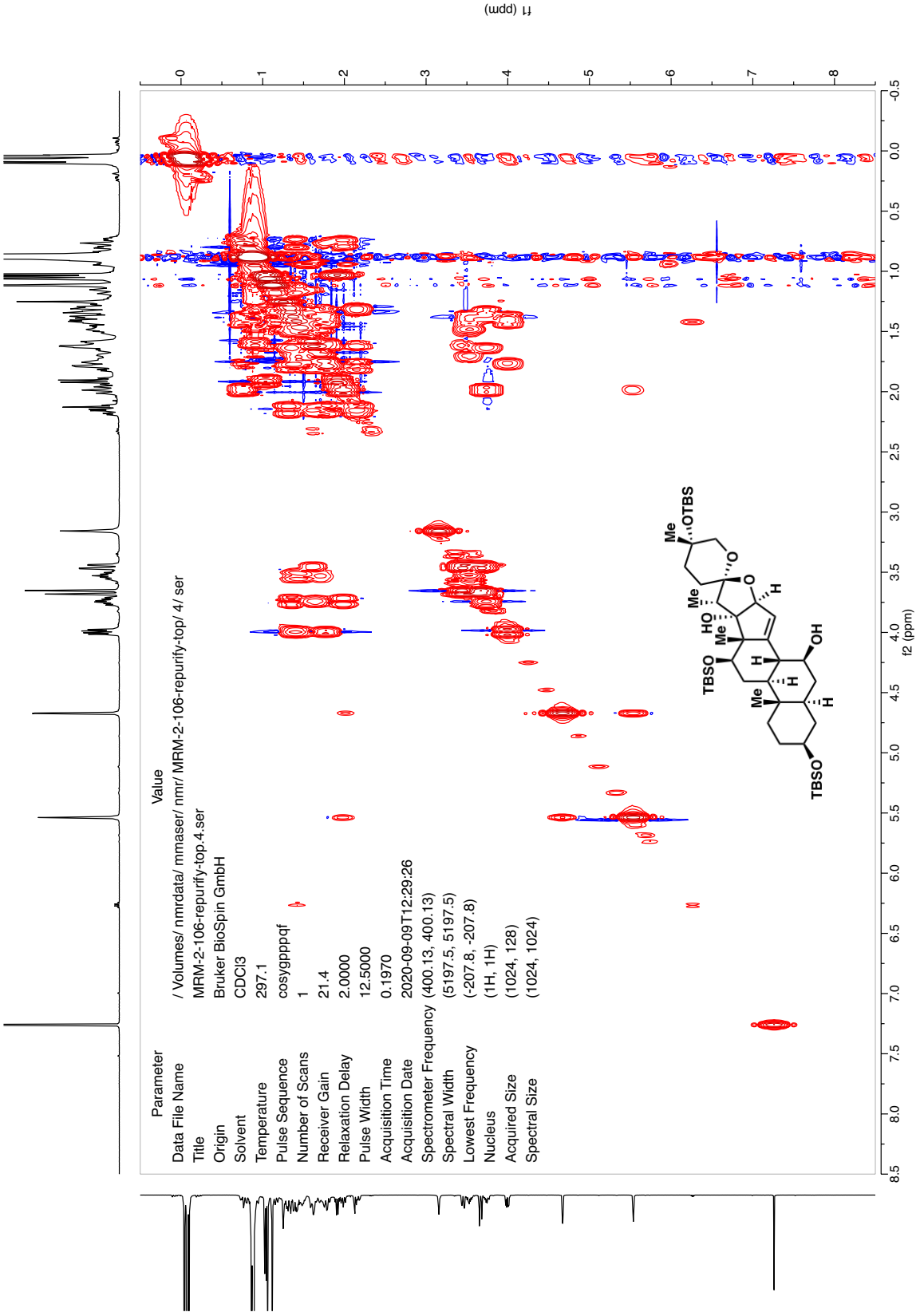

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-050-2D/ 2/ fid
Title MRM-3-050-2D.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-09-07T07:12:40
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.2
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
HO
Me
Me
H
H
H
O
O
O
OTBS
TBSO
f1 (ppm)

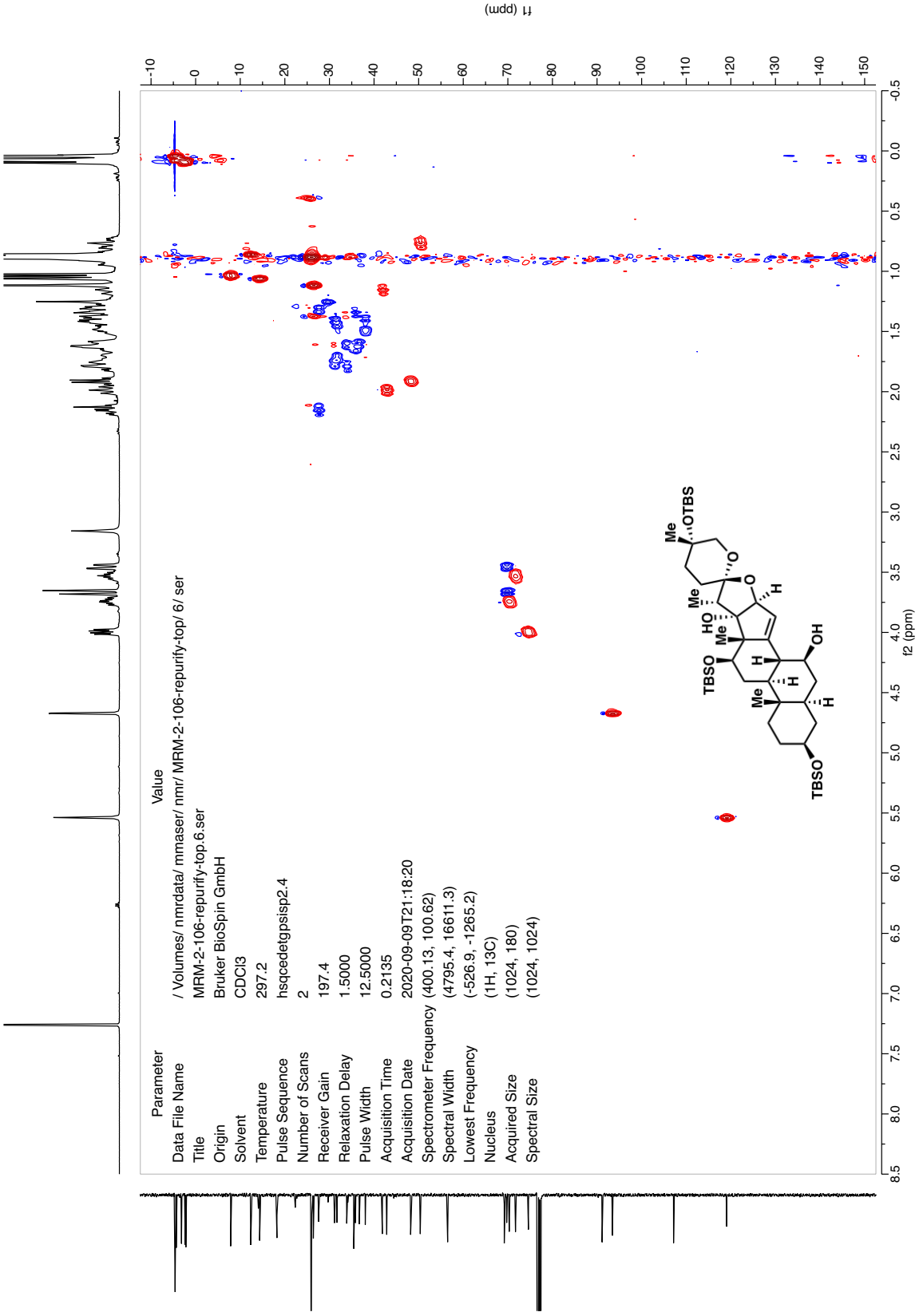

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-050-2D/ 6/ ser
Title MRM-3-050-2D.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-09-07T07:19:31
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-517.1, -1262.8)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
Me
OTBS
HO
Me
TBSO
Me
TBSO

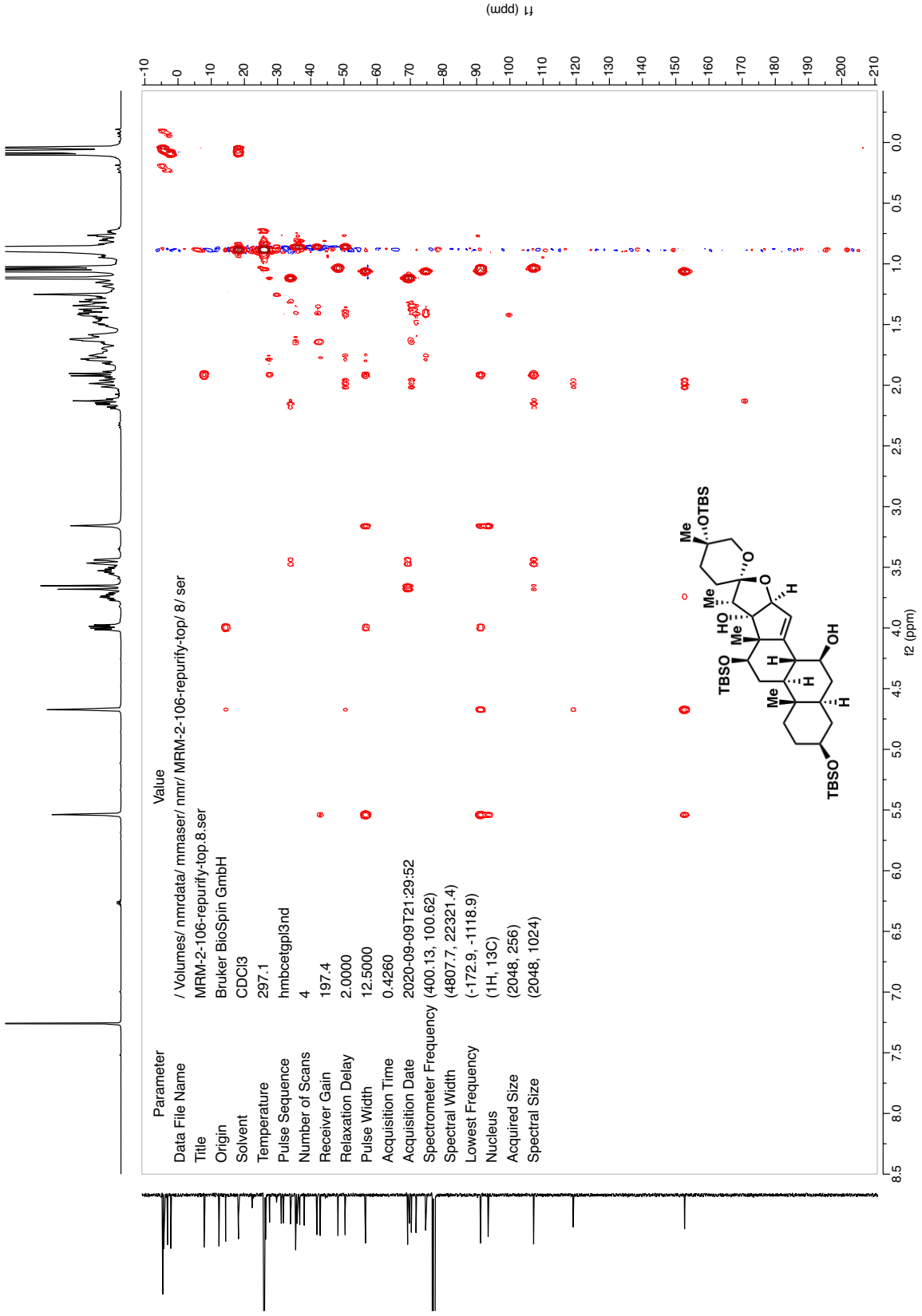

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-050-2D/ 8/ ser
Title
MRM-3-050-2D.8.ser
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
hmbcetgpl3nd
Number of Scans
4
Receiver Gain
197.4
Relaxation Delay
2.0000
Pulse Width
12.5000
Acquisition Time
0.4260
Acquisition Date
2020-09-07T07:31:03
Spectrometer Frequency
(400.13, 100.62)
Spectral Width
(4807.7, 22321.4)
Lowest Frequency
(-163.1, -1116.4)
Nucleus
(1H, 13C)
Acquired Size
(2048, 256)
Spectral Size
(2048, 1024)
f1 (ppm)
f2 (ppm)
TBSO
Me
HO
Me
Me
OTBS
TBSO

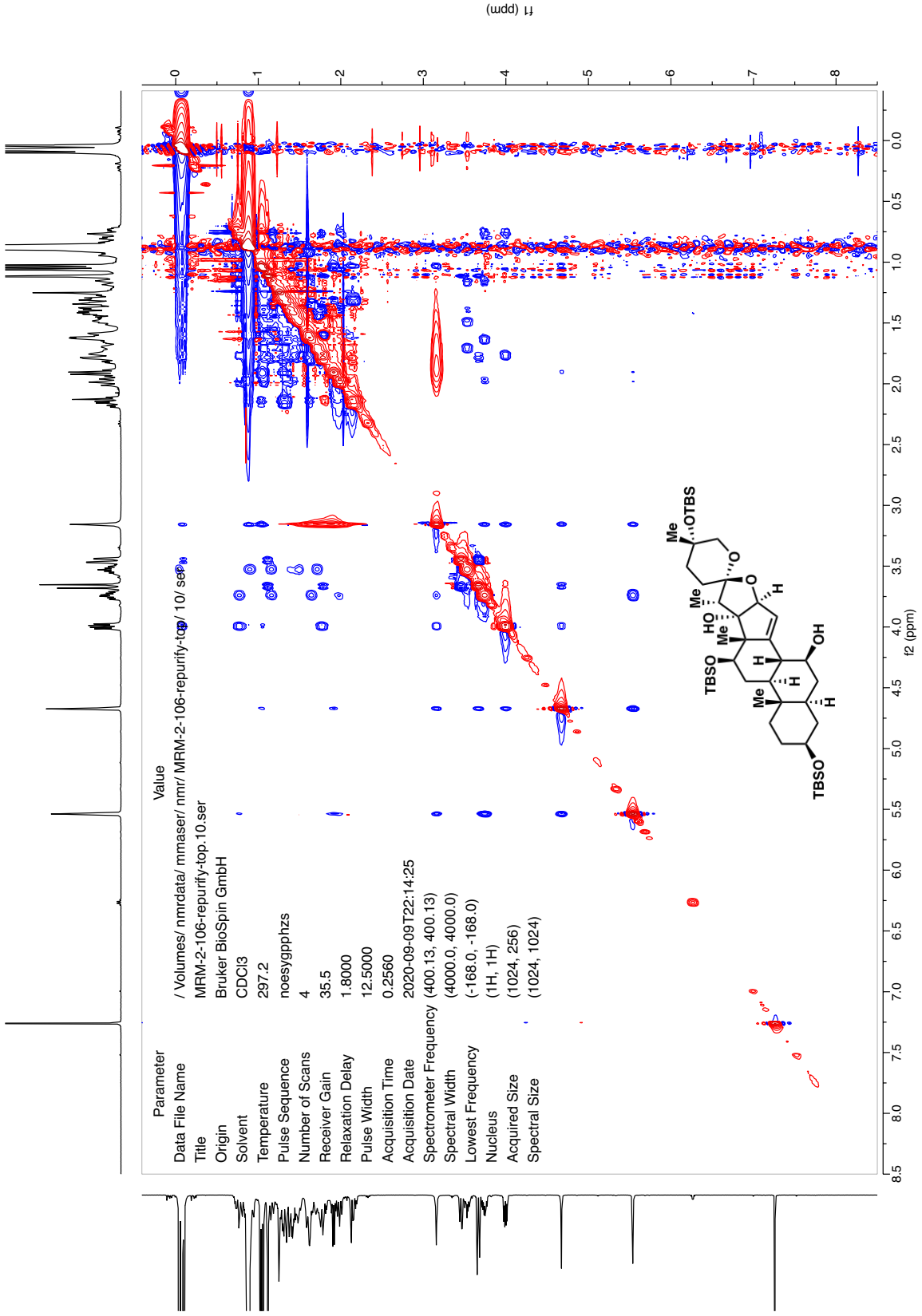

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-106-repurify-top/ 1/ fid
Title MRM-2-106-repurify-top.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-09-09T12:07:17
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.4
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
16
f1 (ppm)

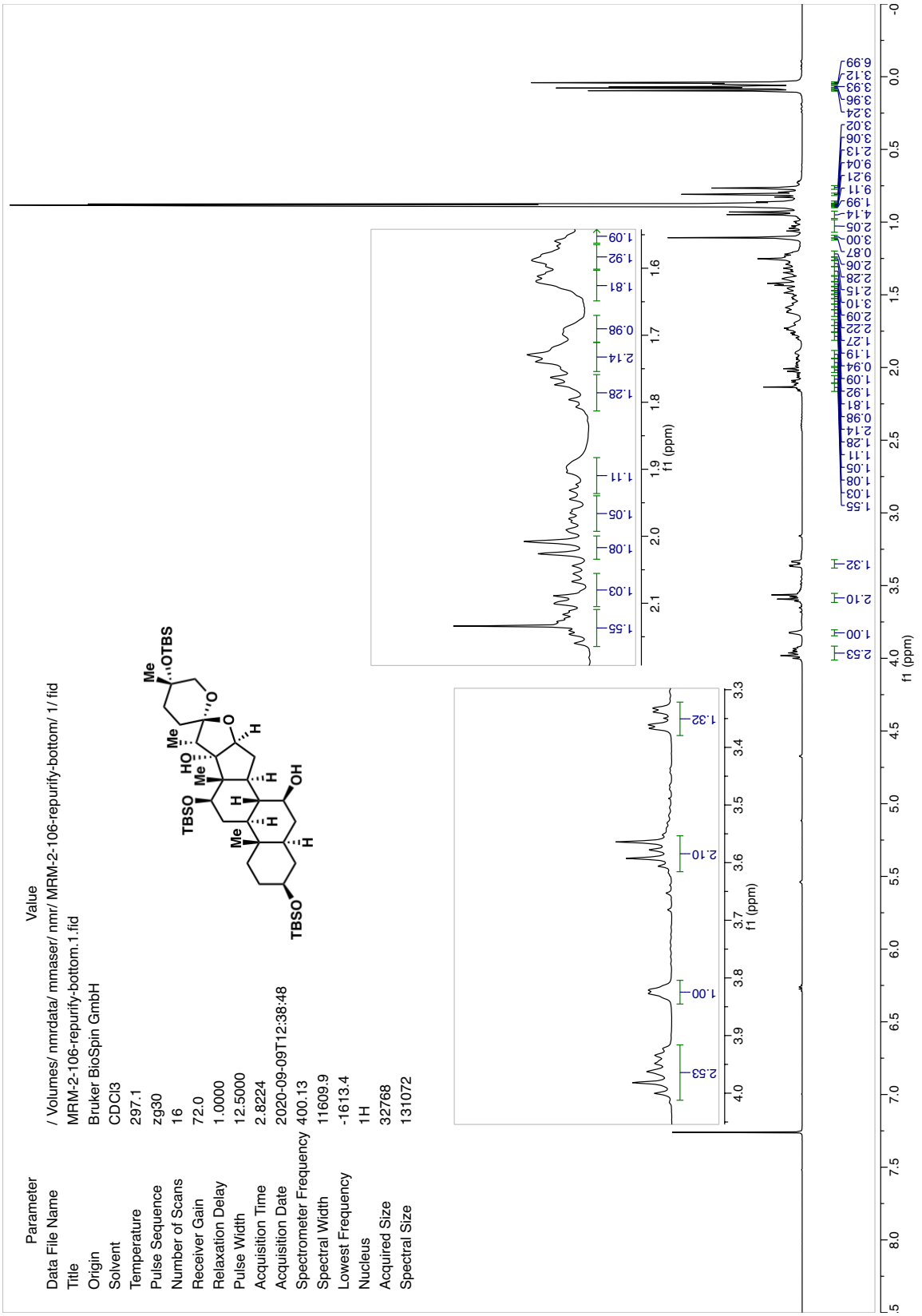

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-106-repurify-top/ 2/ fid
Title MRM-2-106-repurify-top.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-09-09T12:28:28
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.3
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OTBS
O
O
H
Me
HO
Me
TBSO
H
H
Me
H
OH
H
TBSO
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
74.76
71.87
70.47
69.90
69.35
43.00
42.01
38.19
36.85
35.99
35.60
34.03
31.85
31.28
27.74
26.55
26.16
26.06
25.99
18.38
18.33
18.25
14.50
12.50
8.01
-2.03
-2.19
-3.02
-4.19
-4.45
56.66
50.52
48.36
152.81
119.21
107.31
93.61
91.29
f1 (ppm)

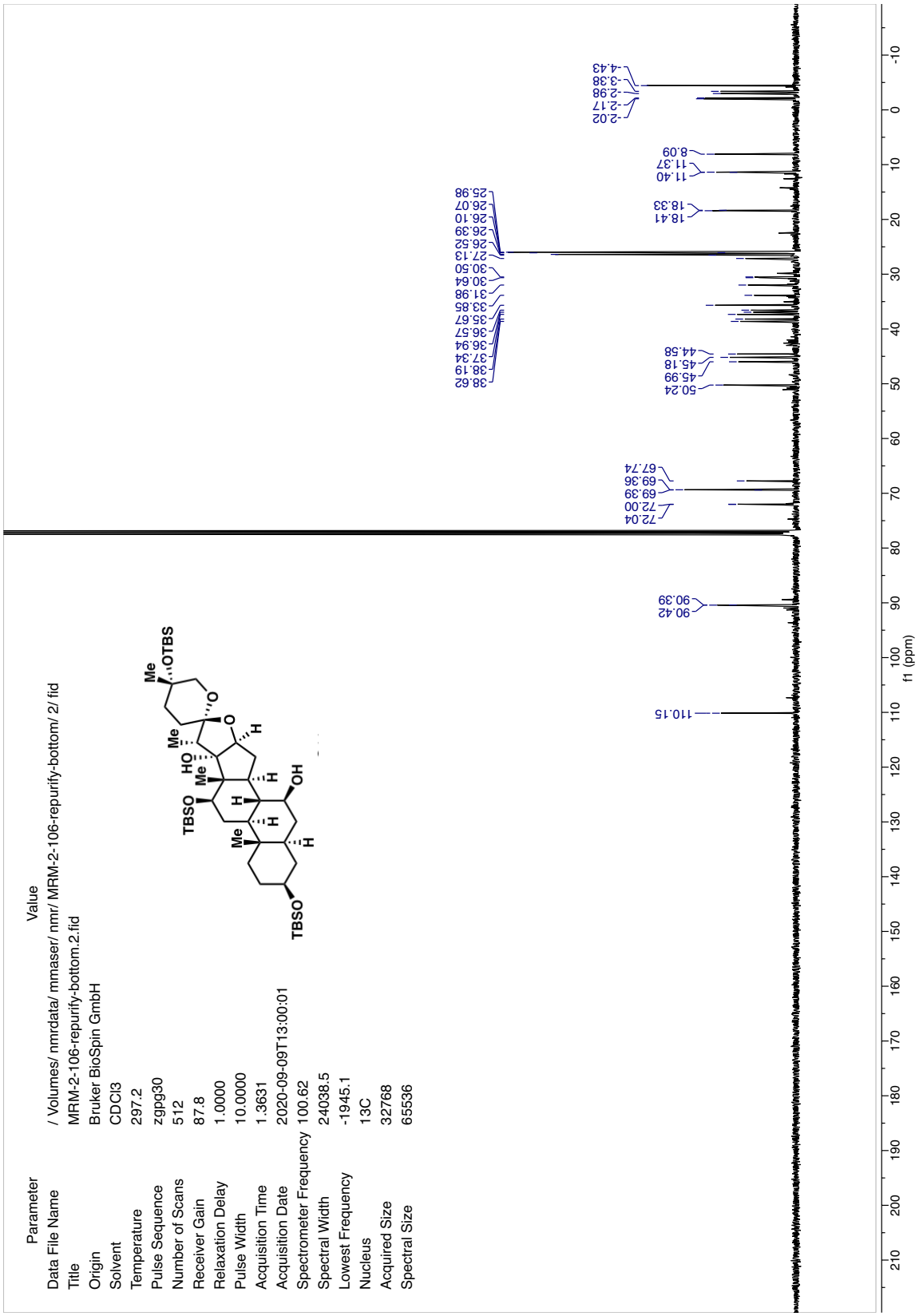

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-106-repurify-top/ 4/ ser
Title MRM-2-106-repurify-top.4.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence cosygpppqf
Number of Scans 1
Receiver Gain 21.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.1970
Acquisition Date 2020-09-09T12:29:26
Spectrometer Frequency (400.13, 400.13)
Spectral Width (5197.5, 5197.5)
Lowest Frequency (-207.8, -207.8)
Nucleus (1H, 1H)
Acquired Size (1024, 128)
Spectral Size (1024, 1024)
TBSO
OTBS
OH
HO
Me
H

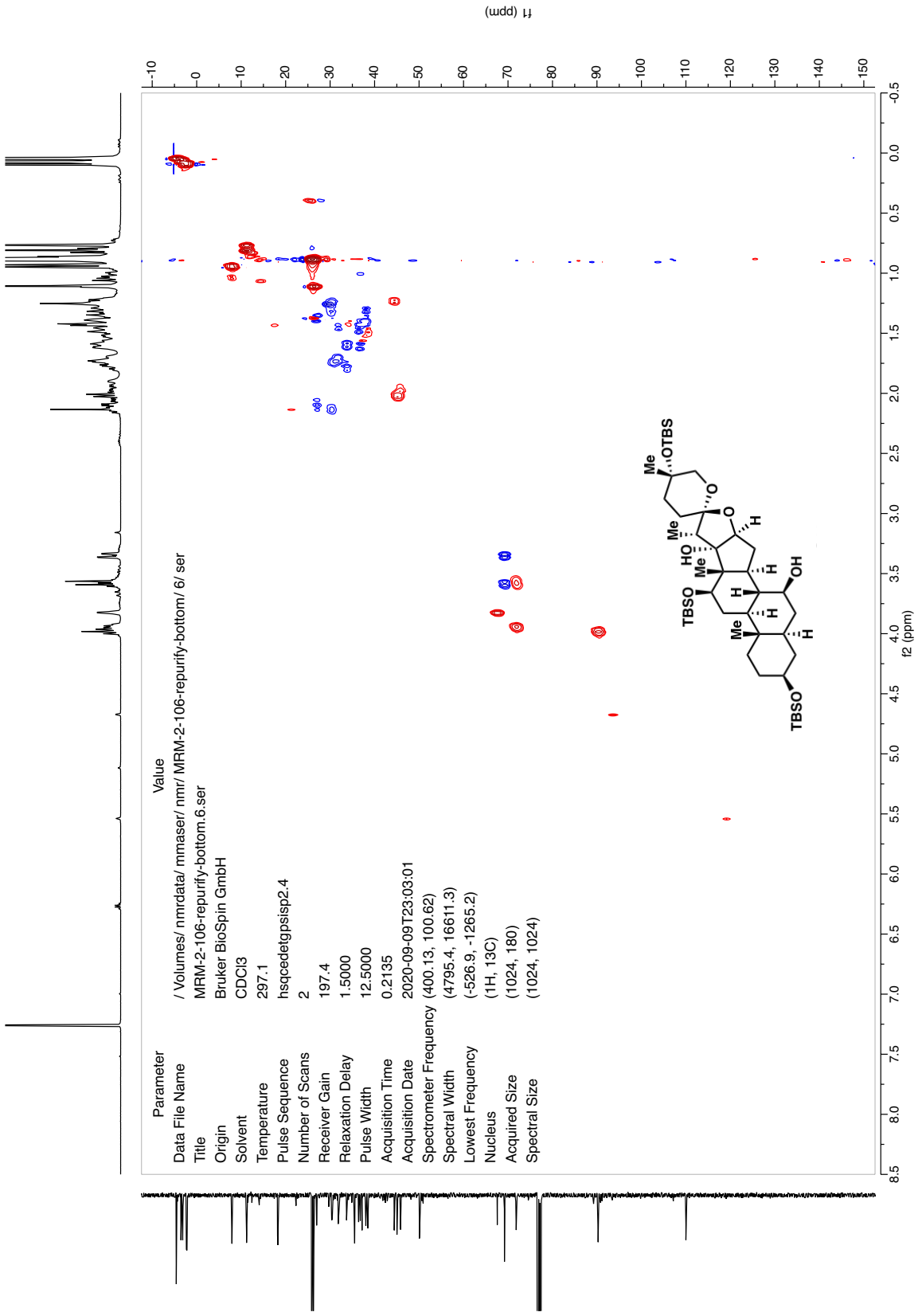

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-106-repurify-top/ 6/ ser
Title MRM-2-106-repurify-top.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-09-09T21:18:20
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-526.9, -1265.2)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
OH
Me
HO
OTBS
H

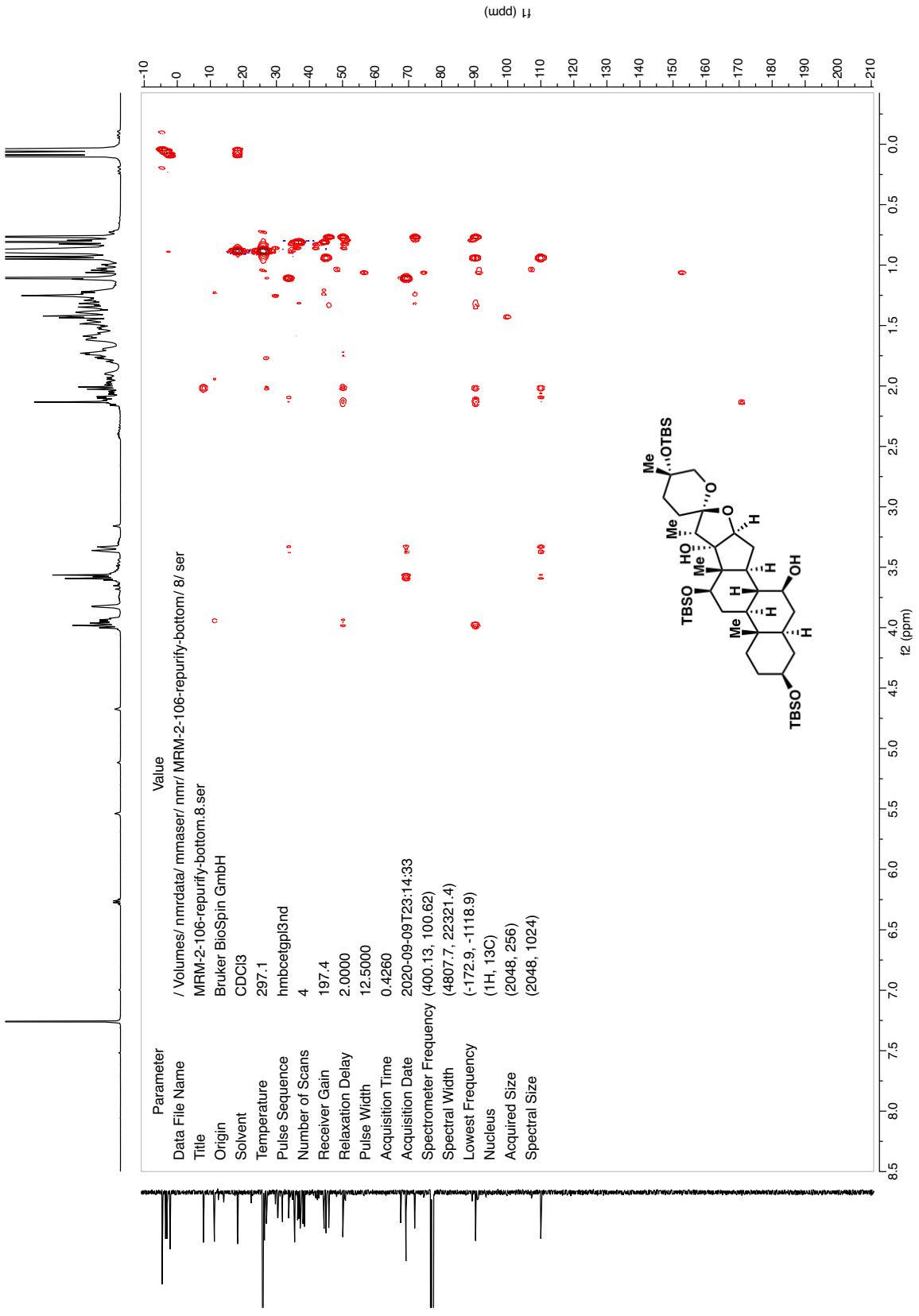

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-106-repurify-top/ 8/ ser
Title MRM-2-106-repurify-top.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-09-09T21:29:52
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-172.9, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
Me
OTBS
O
O
H
Me
HO
Me
TBSO
H
H
OH
Me
H
H
TBSO

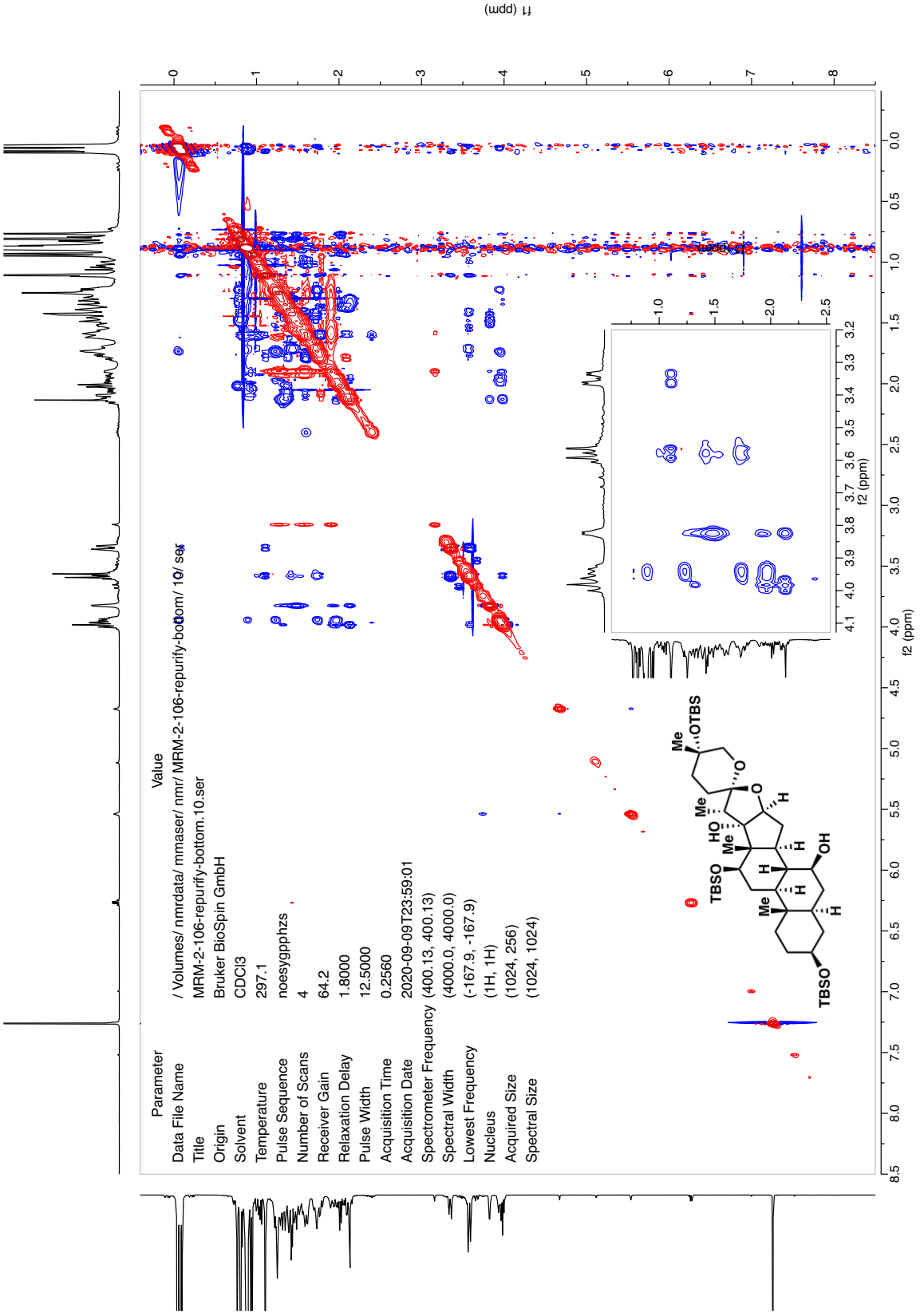

f1 (ppm)
f2 (ppm)
Parameter | Value
Data File Name | / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-106-repurify-top/ 10/ ser
Title | MRM-2-106-repurify-top.10.ser
Origin | Bruker BioSpin GmbH
Solvent | CDCl3
Temperature | 297.2
Pulse Sequence | noesygpphzs
Number of Scans | 4
Receiver Gain | 35.5
Relaxation Delay | 1.8000
Pulse Width | 12.5000
Acquisition Time | 0.2560
Acquisition Date | 2020-09-09T22:14:25
Spectrometer Frequency | (400.13, 400.13)
Spectral Width | (4000.0, 4000.0)
Lowest Frequency | (-168.0, -168.0)
Nucleus | (1H, 1H)
Acquired Size | (1024, 256)
Spectral Size | (1024, 1024)
TBSO
Me
HO
Me
TBSO
OTBS
Me
O
O
H
H
H
Me
OH
H

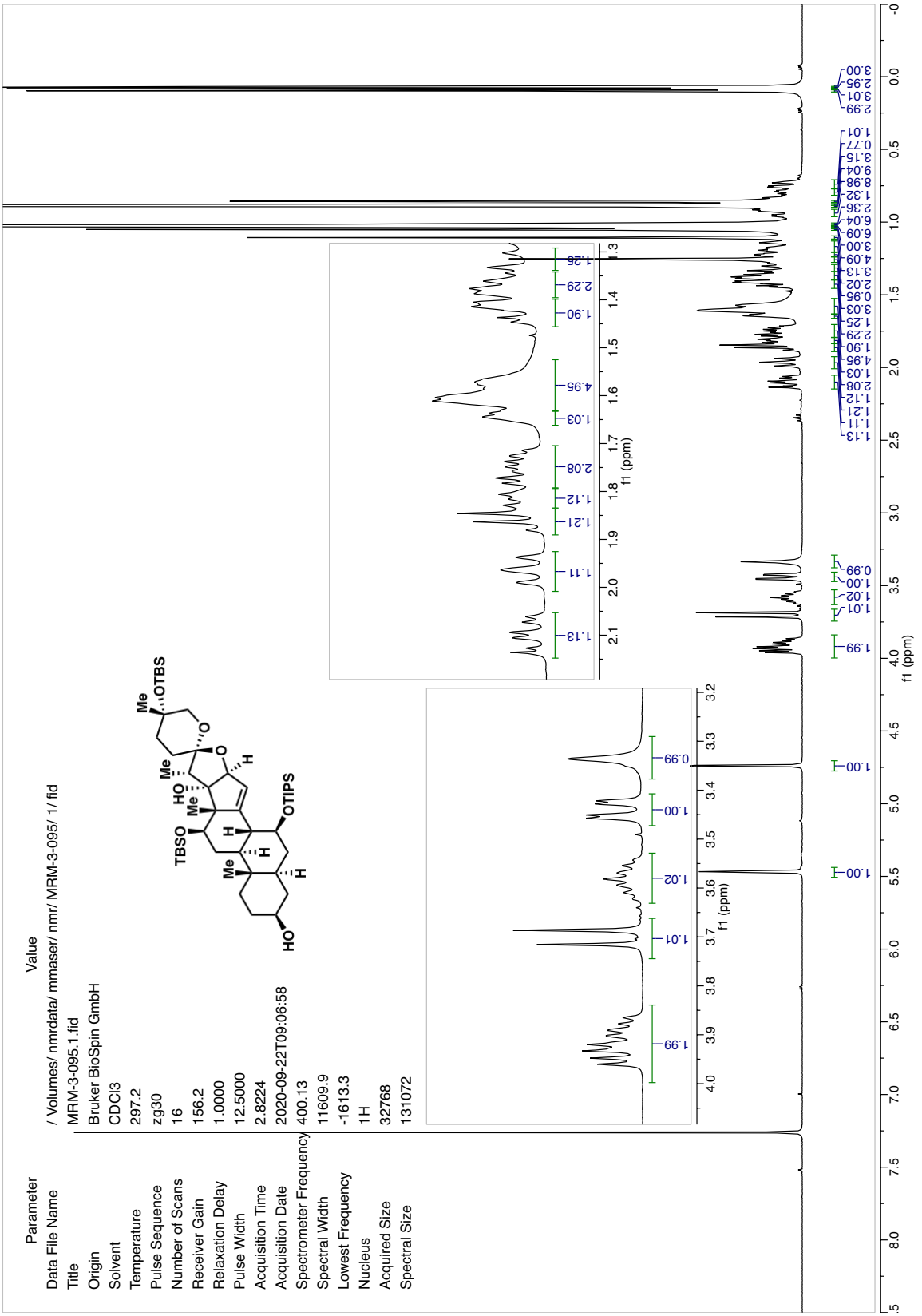

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-106-repurify-bottom/ 1/ fid
Title
MRM-2-106-repurify-bottom.1.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
zg30
Number of Scans
16
Receiver Gain
72.0
Relaxation Delay
1.0000
Pulse Width
12.5000
Acquisition Time
2.8224
Acquisition Date
2020-09-09T12:38:48
Spectrometer Frequency
400.13
Spectral Width
11609.9
Lowest Frequency
-1613.4
Nucleus
1H
Acquired Size
32768
Spectral Size
131072
Me
OTBS
Me
HO
Me
TBSO
H
H
H
Me
H
OH
H
TBSO
f1 (ppm)

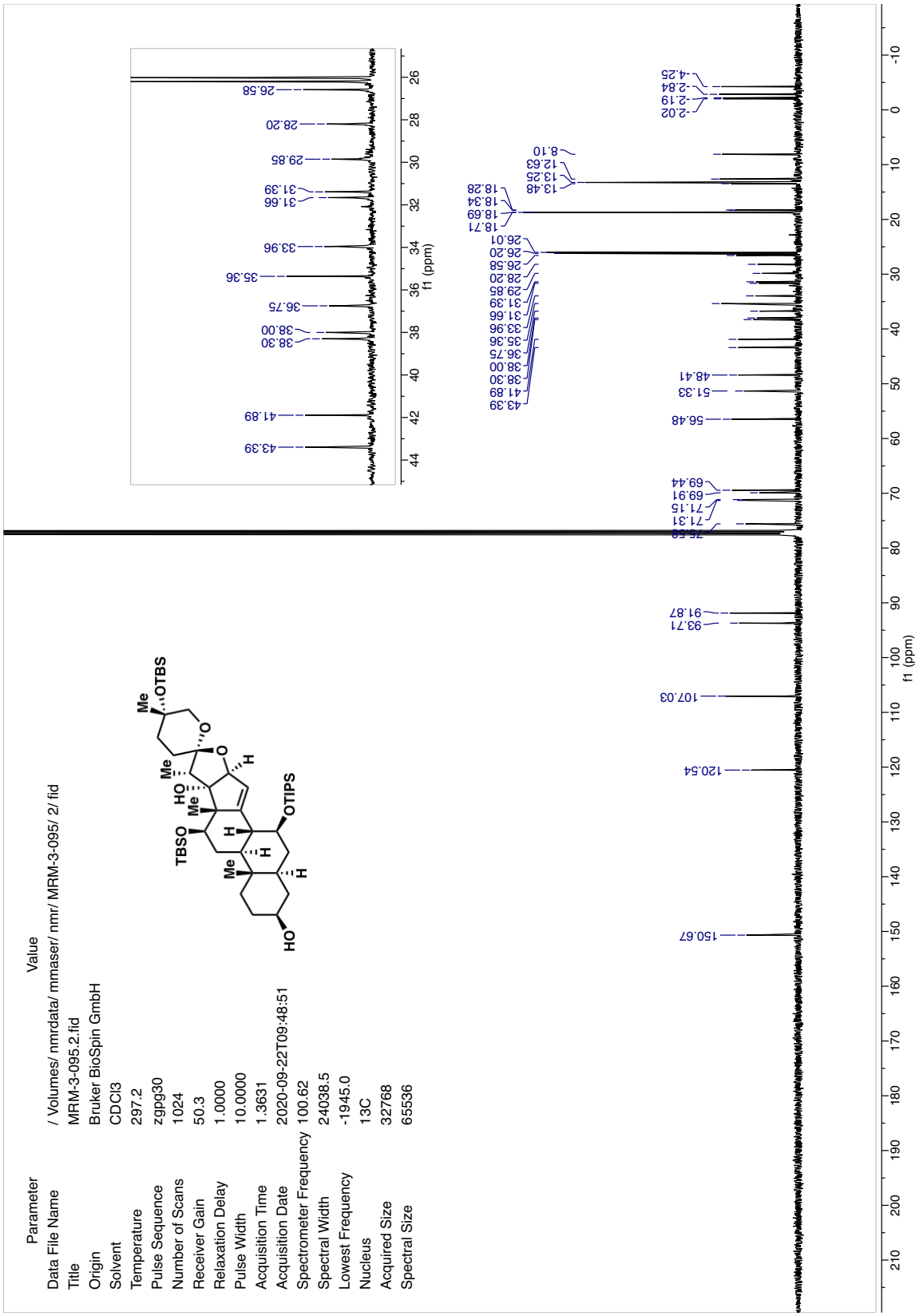

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-106-repurify-bottom/ 2/ fid
Title MRM-2-106-repurify-bottom.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 87.8
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-09-09T13:00:01
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.1
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OTBS
O
Me
HO
Me
TBSO
H
H
H
H
Me
H
OH
TBSO
-2.02
-2.17
-2.98
-3.38
-4.43
8.09
11.40
11.37
18.41
18.33
25.98
26.07
26.10
26.39
26.52
27.13
30.50
30.64
31.98
33.85
35.67
36.57
36.94
37.34
38.19
38.62
44.58
45.18
45.99
50.24
67.74
69.36
69.39
72.00
72.04
90.39
90.42
110.15
f1 (ppm)

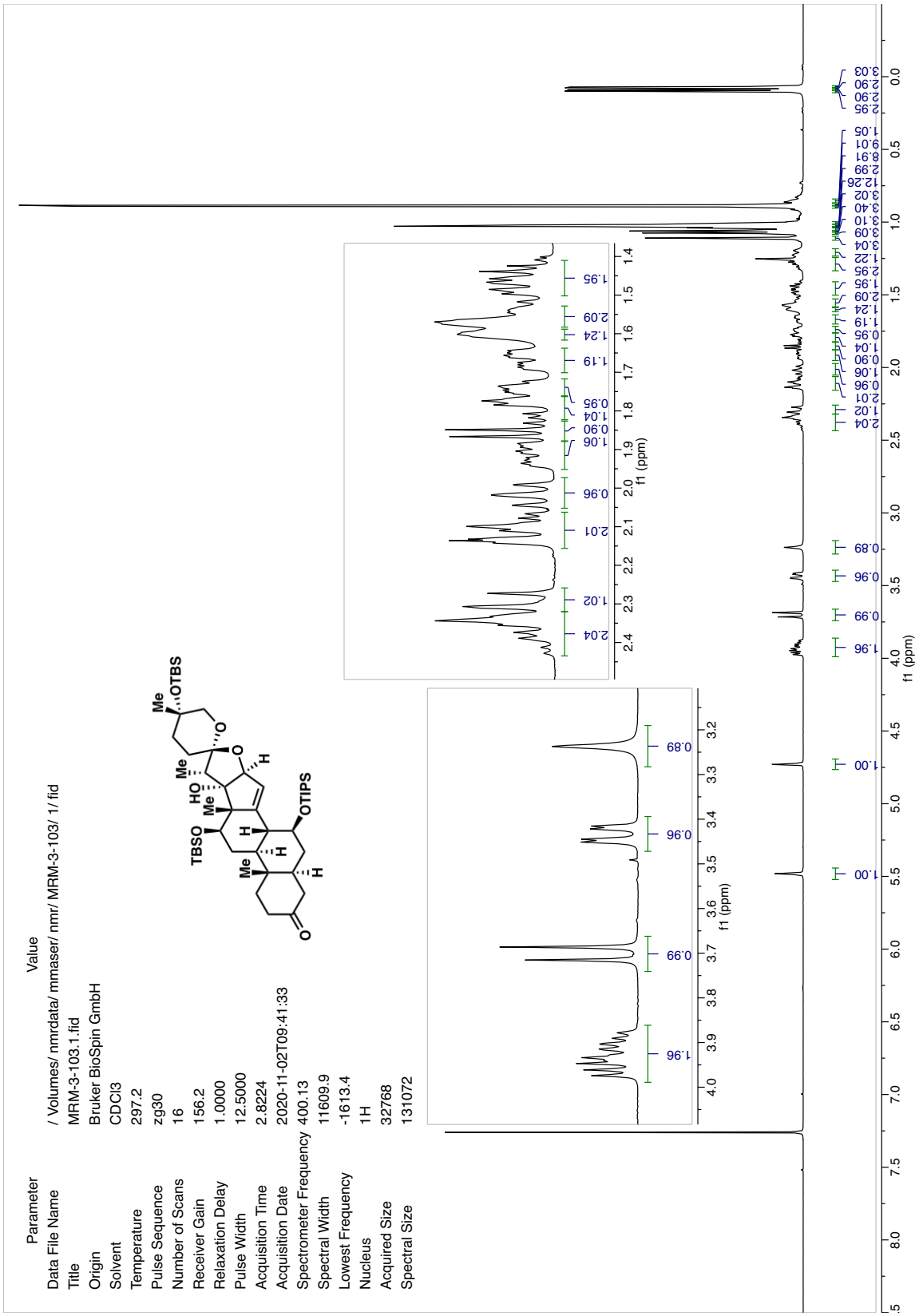

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-106-repurify-bottom/ 6/ ser
Title MRM-2-106-repurify-bottom.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-09-09T23:03:01
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-526.9, -1265.2)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
Me
OTBS
O
O
H
HO
Me
TBSO
H
H
H
OH
Me
H
TBSO

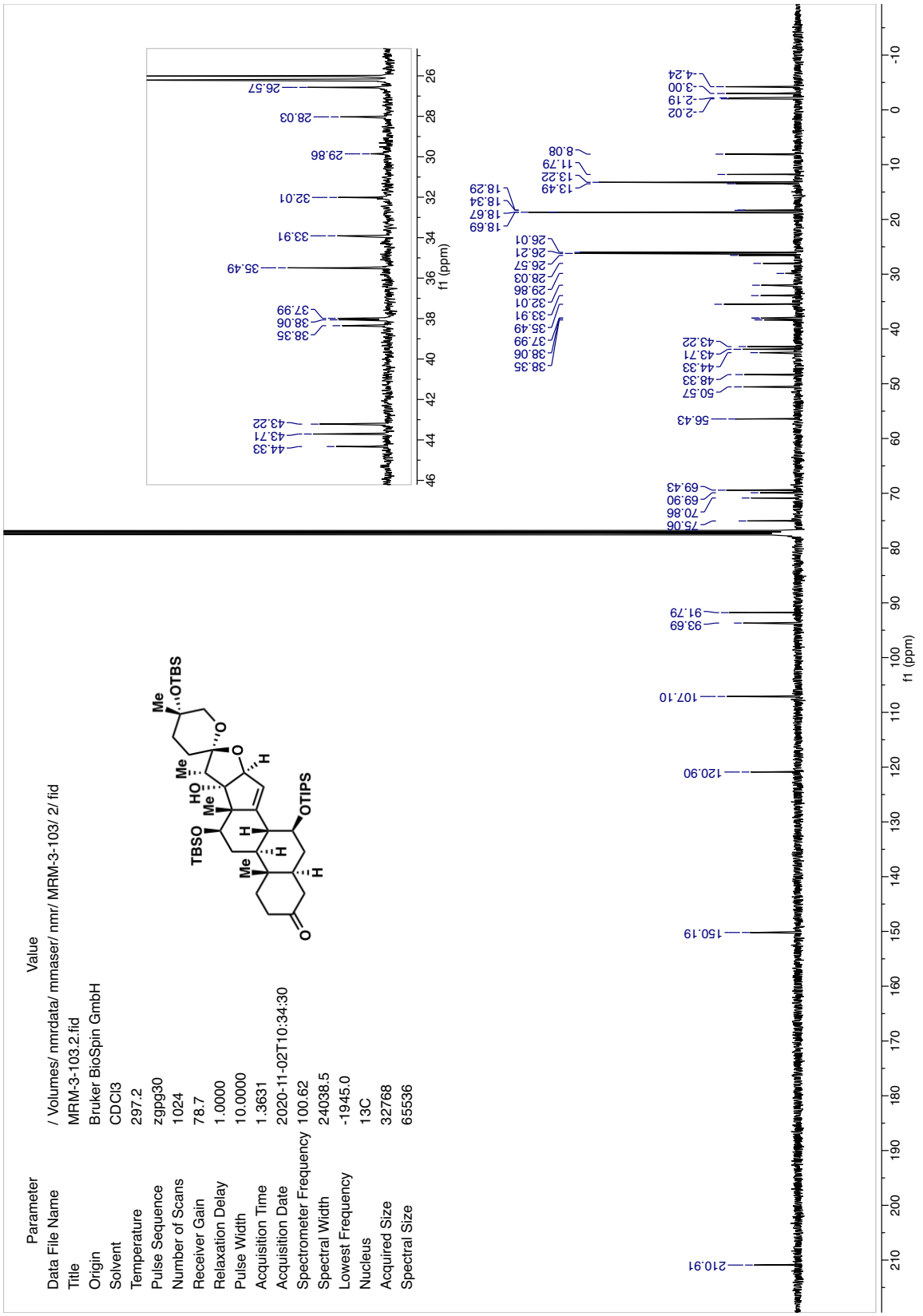

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-106-repurify-bottom/ 8/ ser
Title MRM-2-106-repurify-bottom.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-09-09T23:14:33
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-172.9, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
Me
OTBS
O
O
H
Me
HO
Me
TBSO
H
H
H
OH
Me
H
TBSO

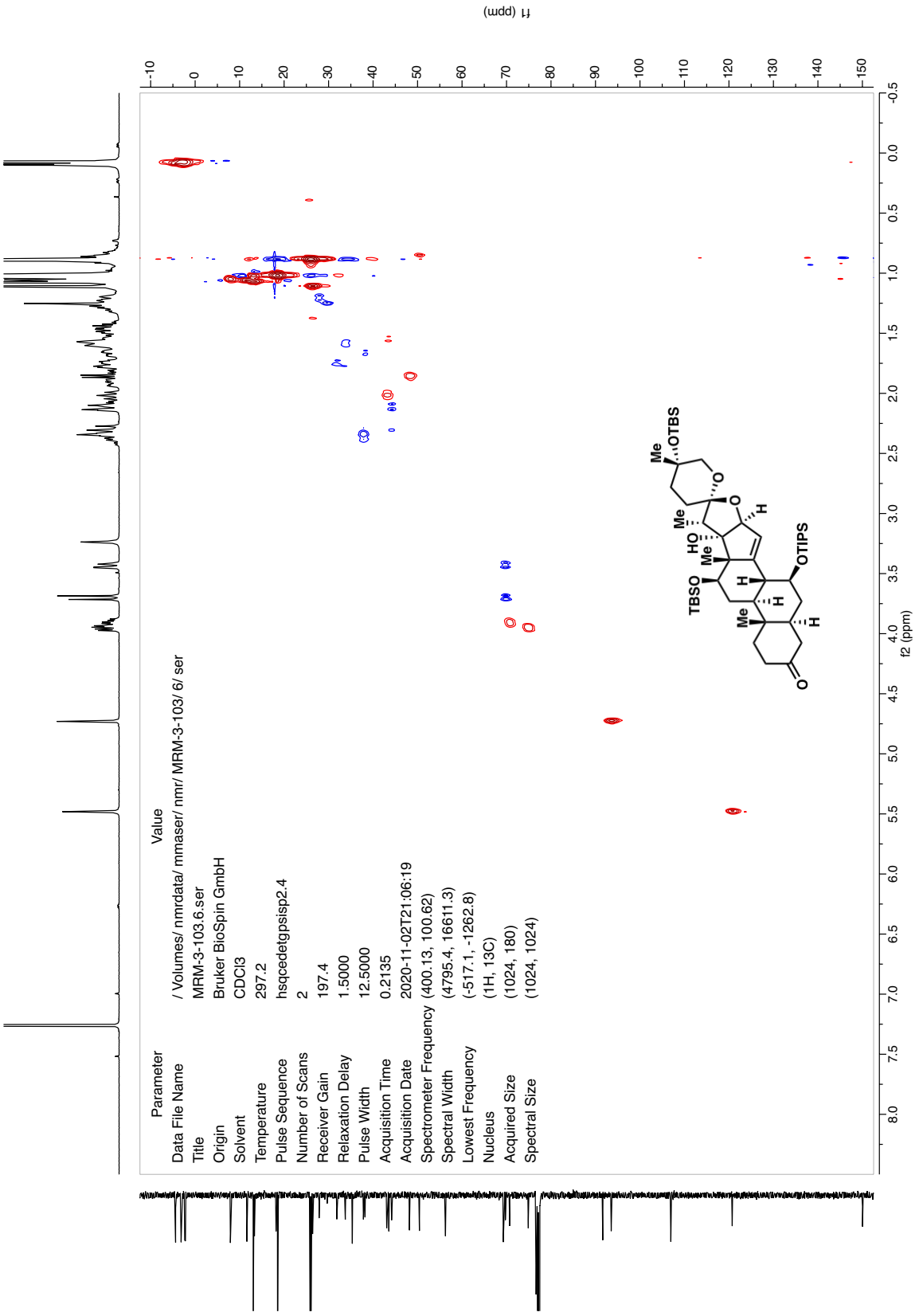

f1 (ppm)
f2 (ppm)
Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-106-repurify-bottom/ 10/ ser
Title
MRM-2-106-repurify-bottom.10.ser
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
noesygpphzs
Number of Scans
4
Receiver Gain
64.2
Relaxation Delay
1.8000
Pulse Width
12.5000
Acquisition Time
0.2560
Acquisition Date
2020-09-09T23:59:01
Spectrometer Frequency
(400.13, 400.13)
Spectral Width
(4000.0, 4000.0)
Lowest Frequency
(-167.9, -167.9)
Nucleus
(1H, 1H)
Acquired Size
(1024, 256)
Spectral Size
(1024, 1024)
TBSO
OTBS
Me
HO
OH
H

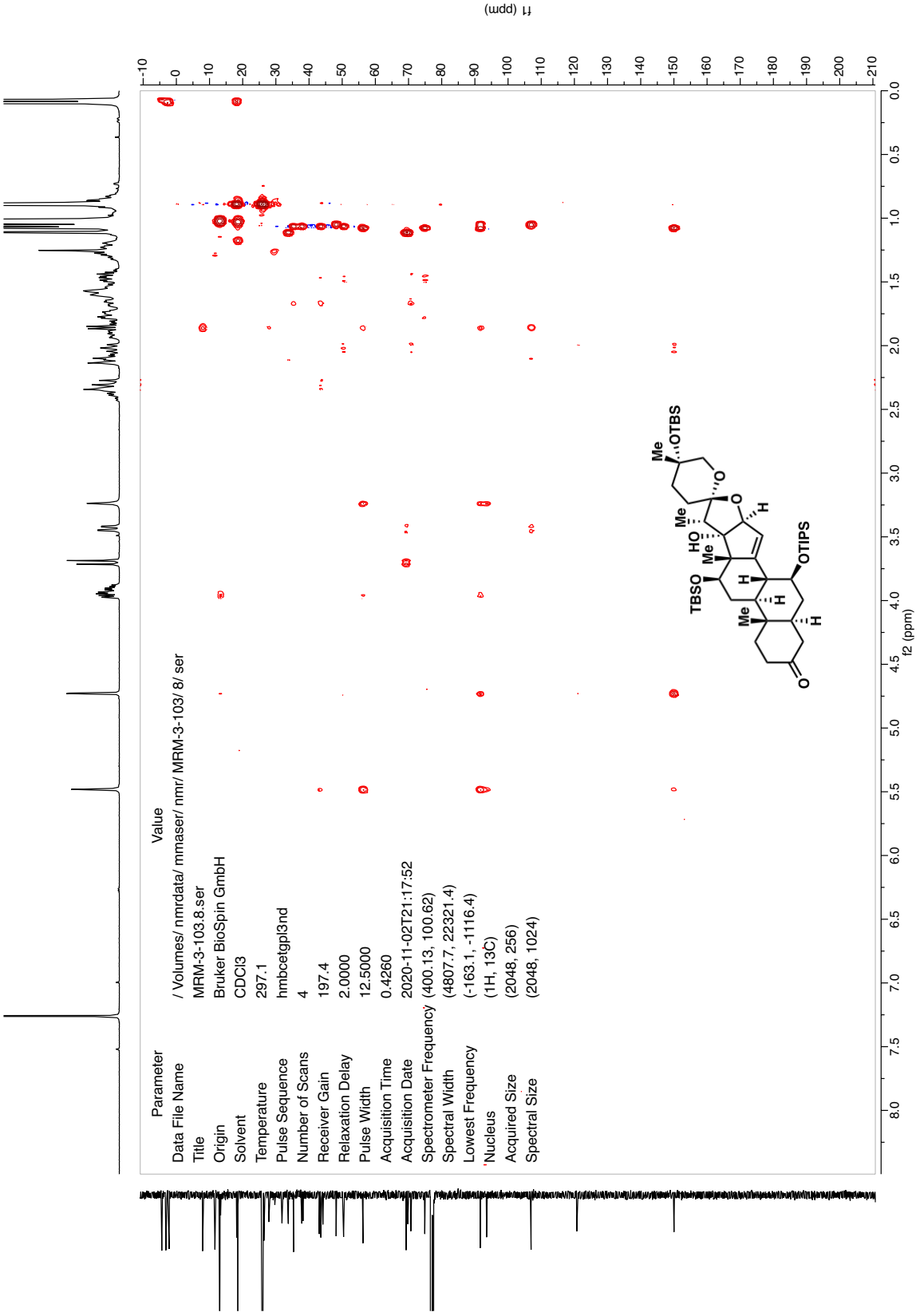

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-095/ 1/ fid
Title MRM-3-095.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-09-22T09:06:58
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.3
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
Me
OTBS
HO
Me
TBSO
Me
H
H
H
OTIPS
HO
f1 (ppm)

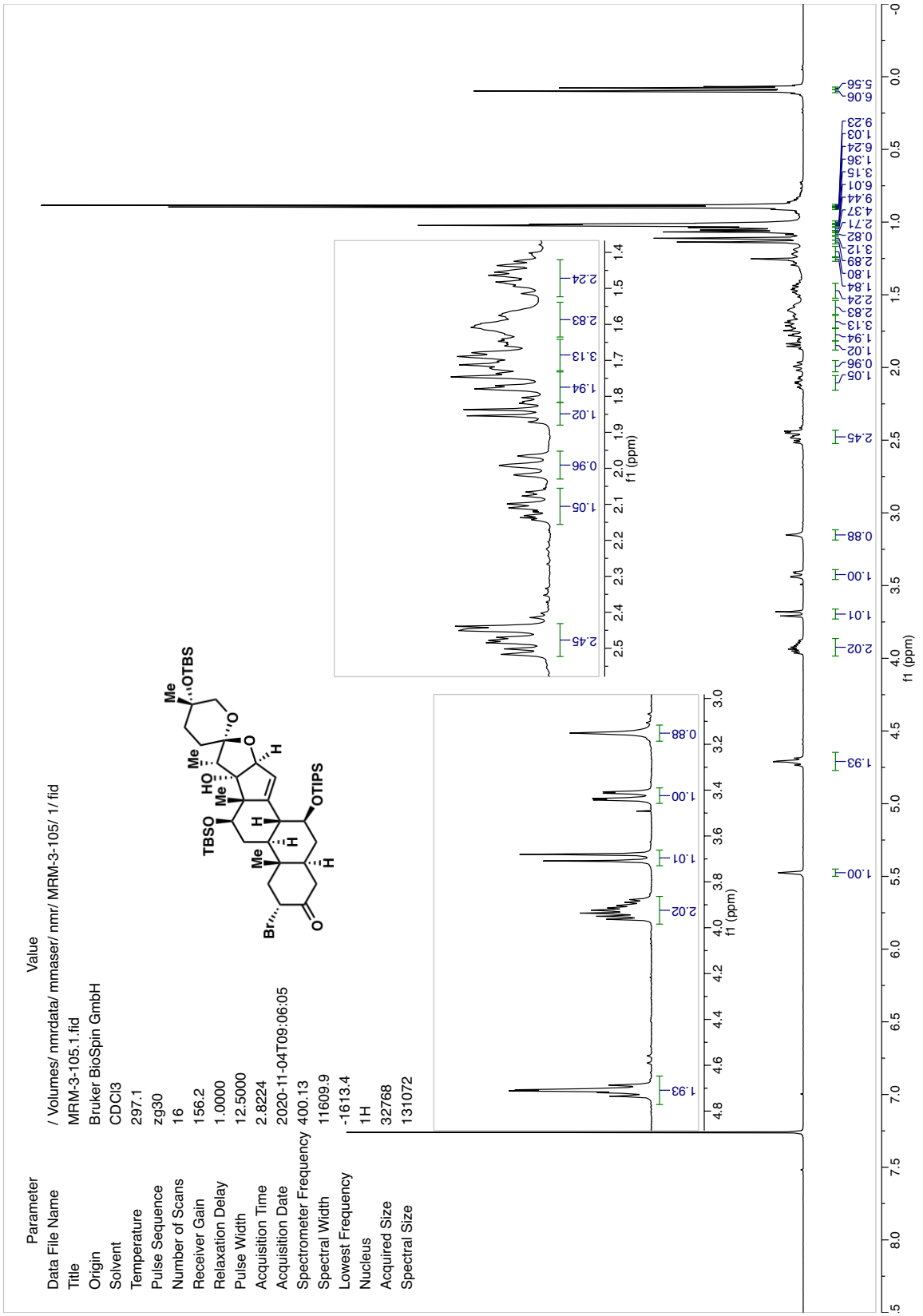

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-095/ 2/ fid
Title MRM-3-095.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-09-22T09:48:51
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OTBS
O
O
H
Me
HO
Me
TBSO
H
H
H
Me
OTIPS
H
HO
150.67
120.54
107.03
93.71
91.87
77.59
71.31
71.15
69.91
69.44
56.48
51.33
48.41
43.39
41.89
38.30
38.00
36.75
35.36
33.96
31.66
31.39
29.85
28.20
26.58
26.20
26.01
18.71
18.69
18.34
18.28
13.48
13.25
12.63
8.10
-2.02
-2.19
-2.84
-4.25
f1 (ppm)

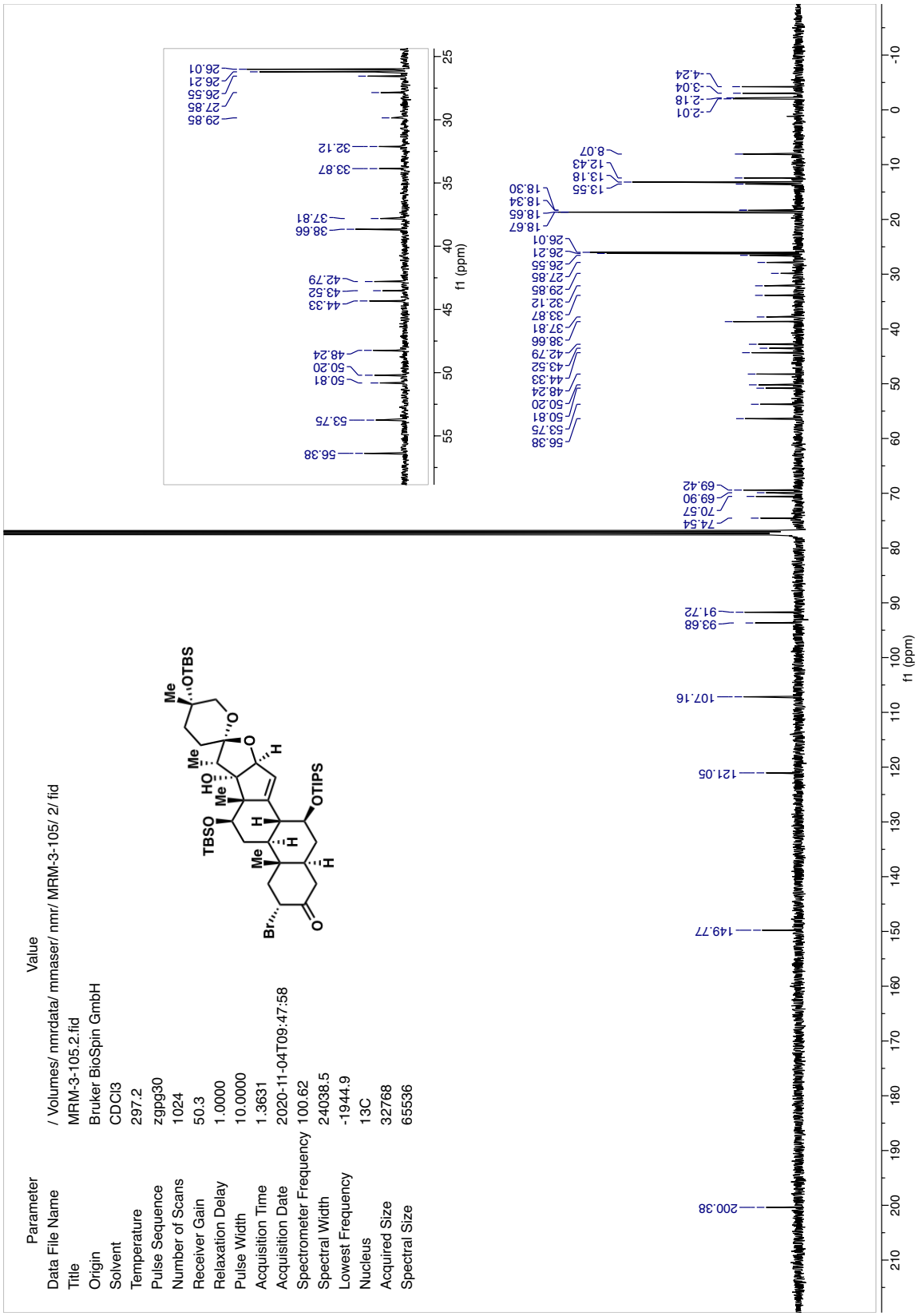

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-103/ 1/ fid
Title MRM-3-103.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-11-02T09:41:33
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.4
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
Me
OTBS
HO
Me
TBSO
Me
H
H
H
Me
OTIPS
H
O
O
O
f1 (ppm)

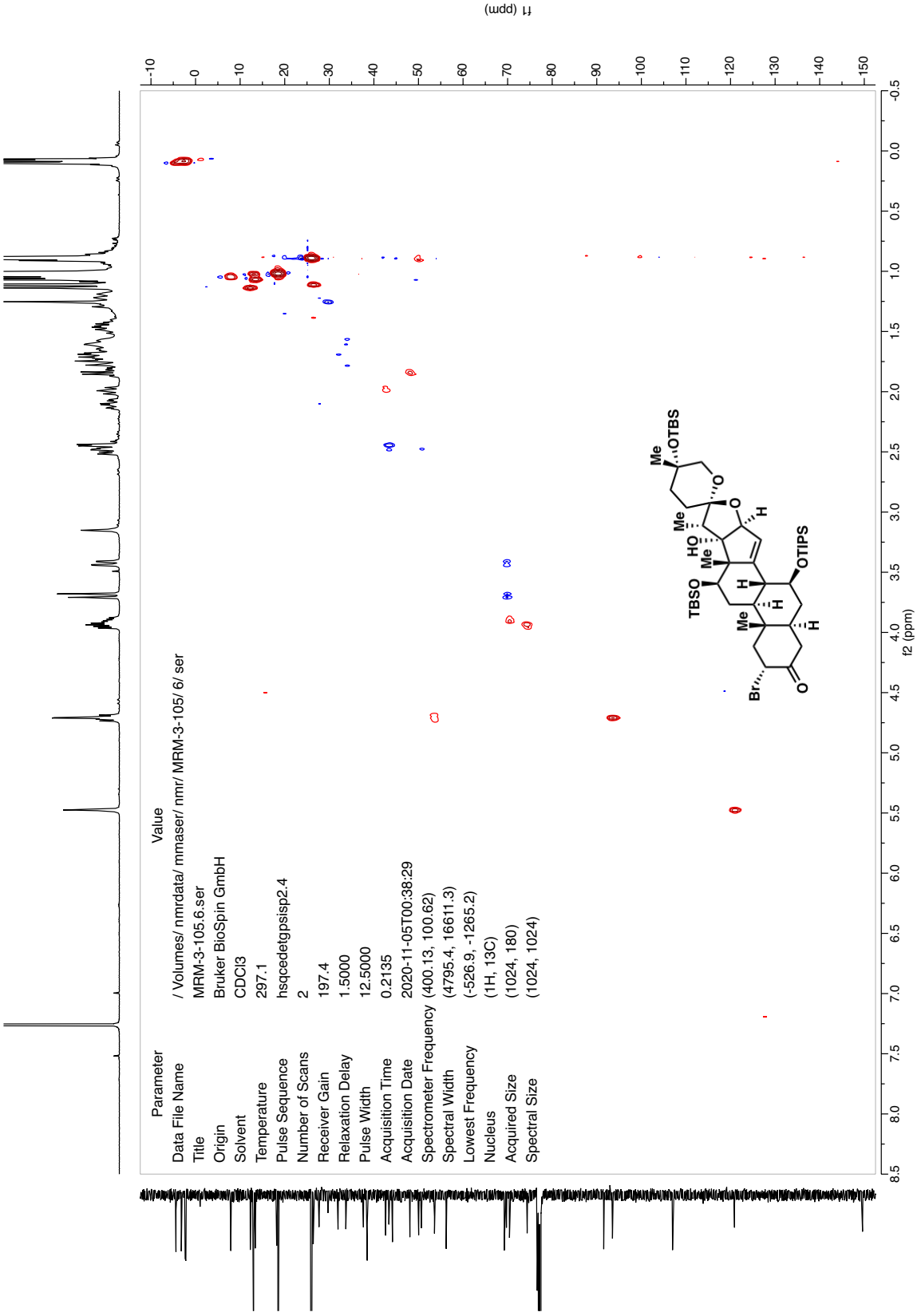

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-103/ 2/ fid
Title MRM-3-103.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-11-02T10:34:30
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OTBS
O
O
H
Me
HO
Me
TBSO
H
H
OTIPS
Me
H
O
210.91
150.19
120.90
107.10
93.69
91.79
75.06
69.90
70.86
69.43
56.43
50.57
48.33
44.33
43.71
43.22
38.35
38.06
37.99
35.49
33.91
32.01
29.86
28.03
26.57
26.21
26.01
18.69
18.67
18.34
18.29
13.49
13.22
11.79
8.08
-2.02
-2.19
-3.00
-4.24
f1 (ppm)

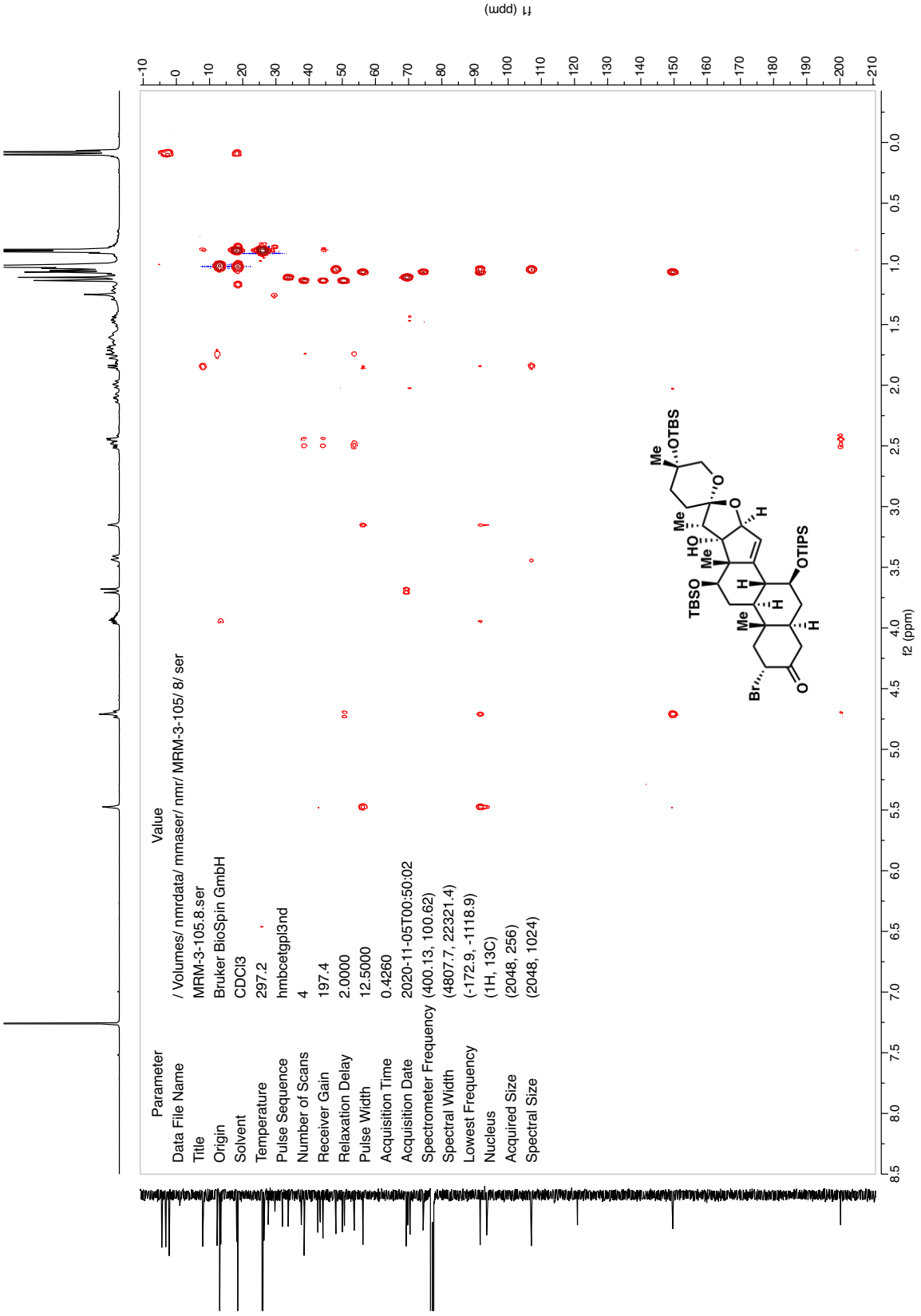

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-103/ 6/ ser
Title MRM-3-103.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-11-02T21:06:19
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-517.1, -1262.8)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
Me
OTBS
O
Me
HO
Me
H
TBSO
Me
H
H
OTIPS
H
O

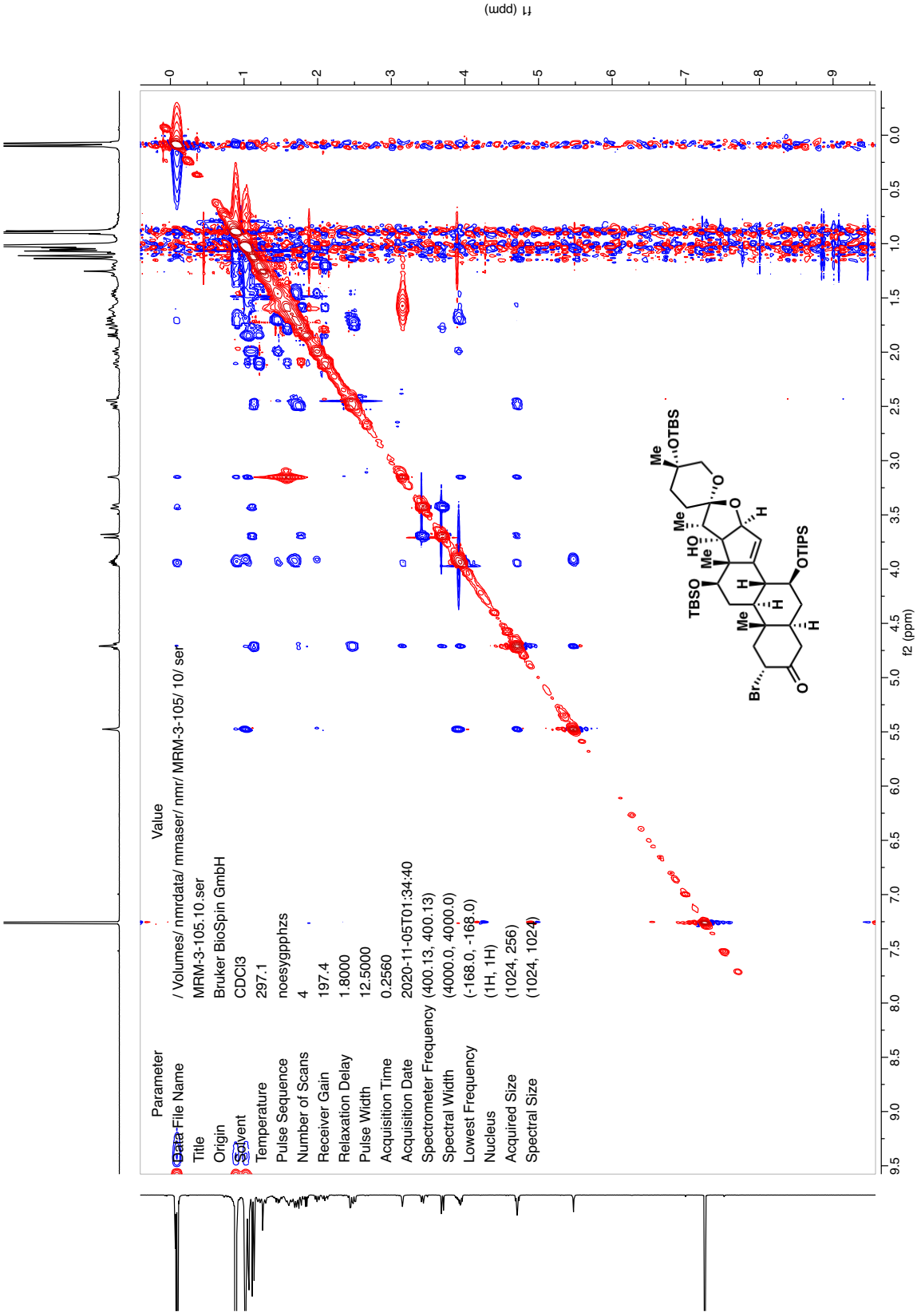

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-103/ 8/ ser
Title MRM-3-103.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-11-02T21:17:52
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-163.1, -1116.4)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
Me
OTBS
O
O
H
HO
Me
TBSO
Me
H
H
OTIPS
Me
H
O

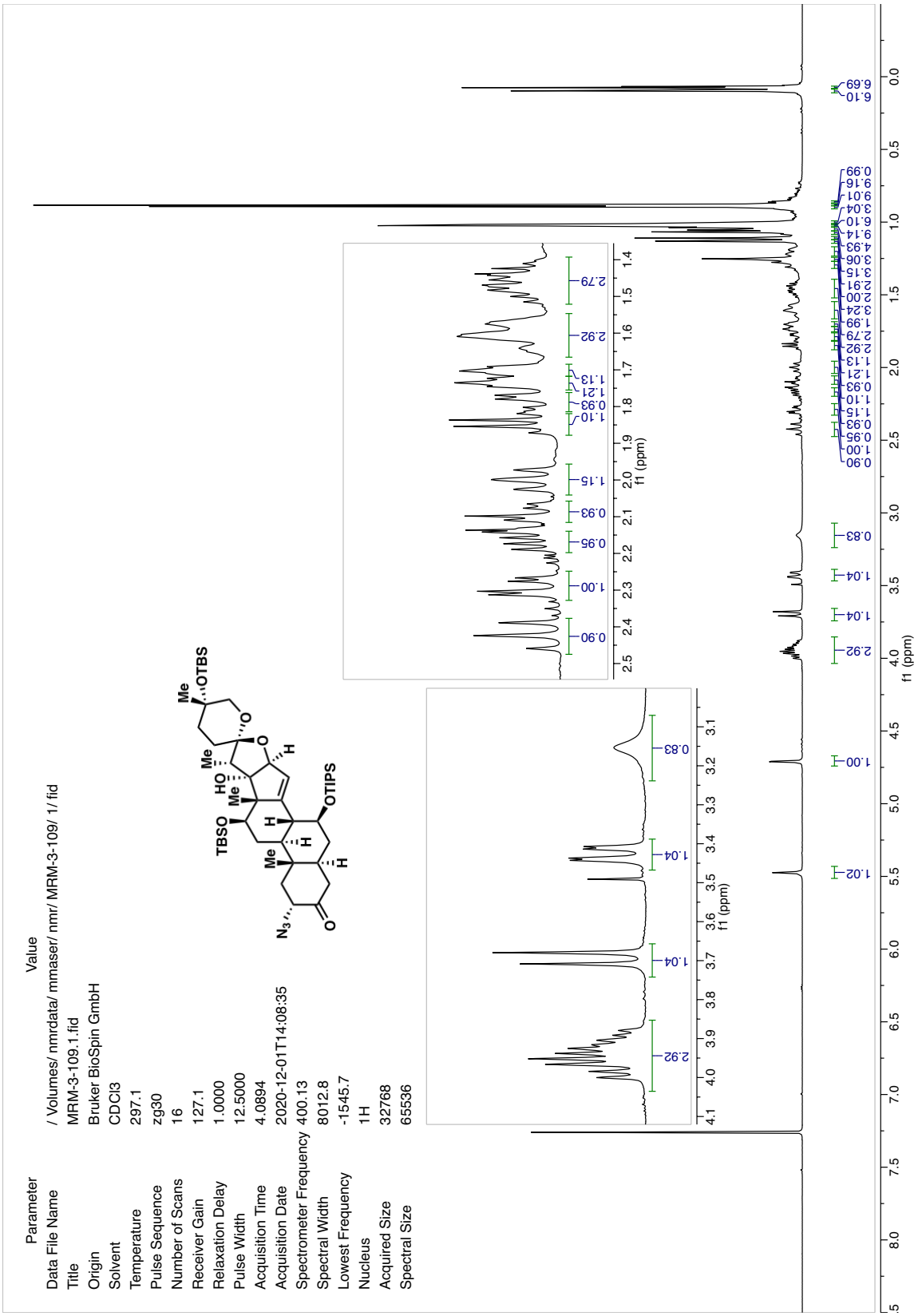

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-105/ 1/ fid
Title MRM-3-105.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-11-04T09:06:05
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.4
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
Me
OTBS
O
Me
HO
Me
H
TBSO
H
H
Me
H
OTIPS
Br
O
f1 (ppm)

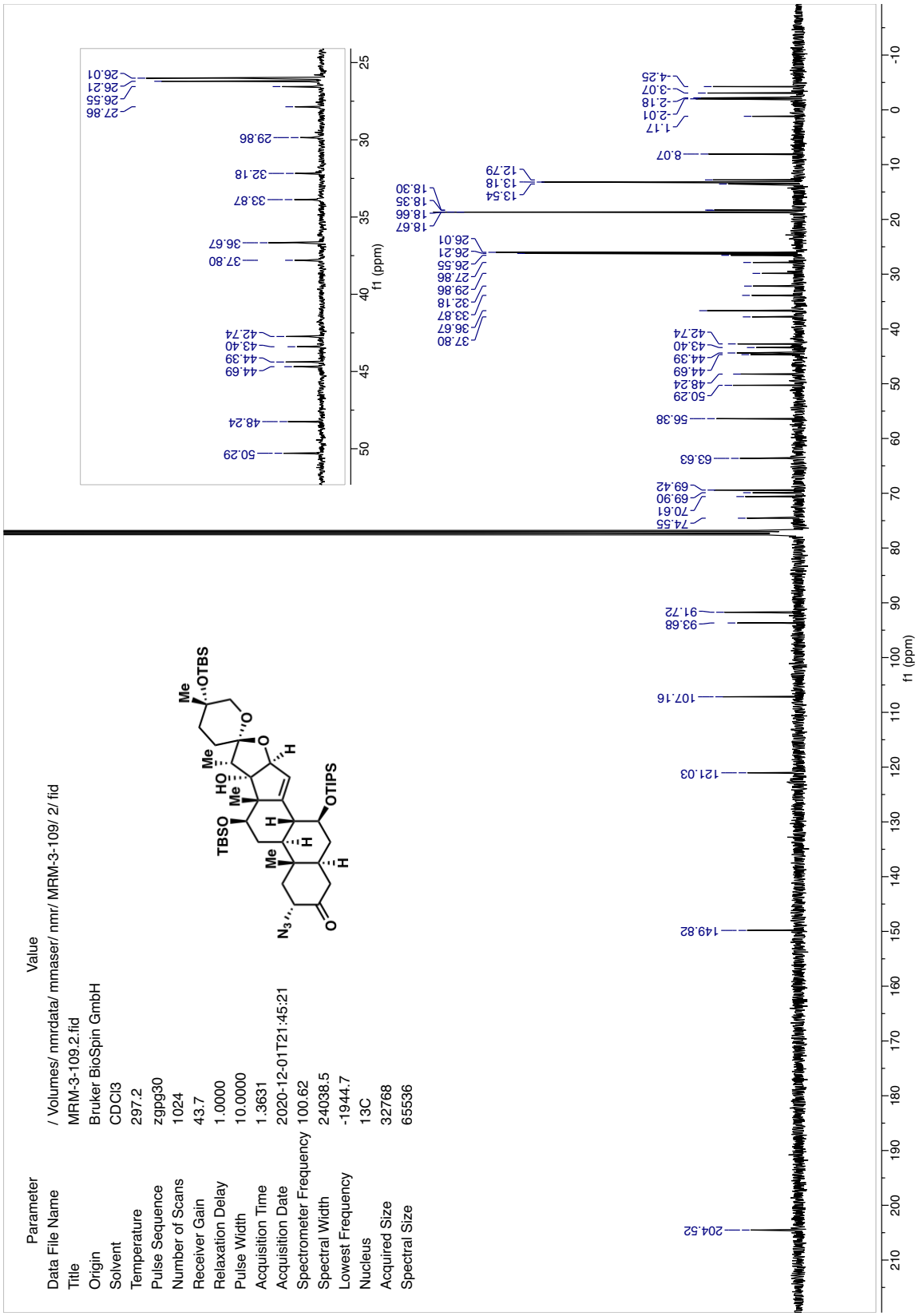

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-105/ 2/ fid
Title MRM-3-105.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-11-04T09:47:58
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OTBS
Me
O
Me
HO
Me
TBSO
H
H
H
Me
OTIPS
Br
O
200.38
149.77
121.05
107.16
93.68
91.72
74.54
70.57
69.90
69.42
56.38
53.75
50.81
50.20
48.24
44.33
43.52
42.79
38.66
37.81
33.87
32.12
29.85
27.85
26.55
26.21
26.01
18.67
18.65
18.34
18.30
13.55
13.18
12.43
8.07
-2.01
-2.18
-3.04
-4.24
f1 (ppm)

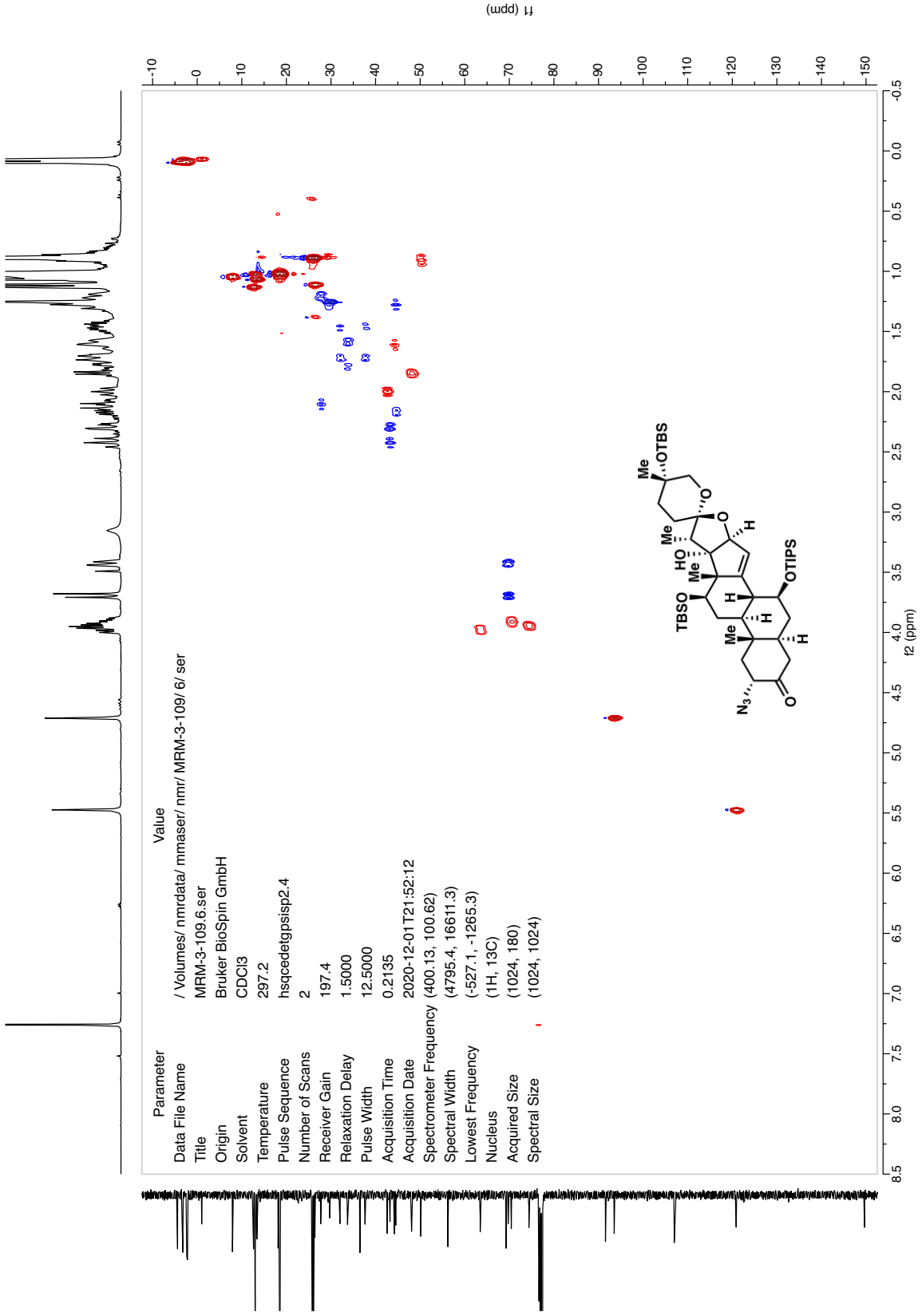

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-105/ 6/ ser
Title MRM-3-105.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-11-05T00:38:29
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-526.9, -1265.2)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
Me
OTBS
HO
Me
O
H
TBSO
H
OTIPS
Me
H
H
Br
O

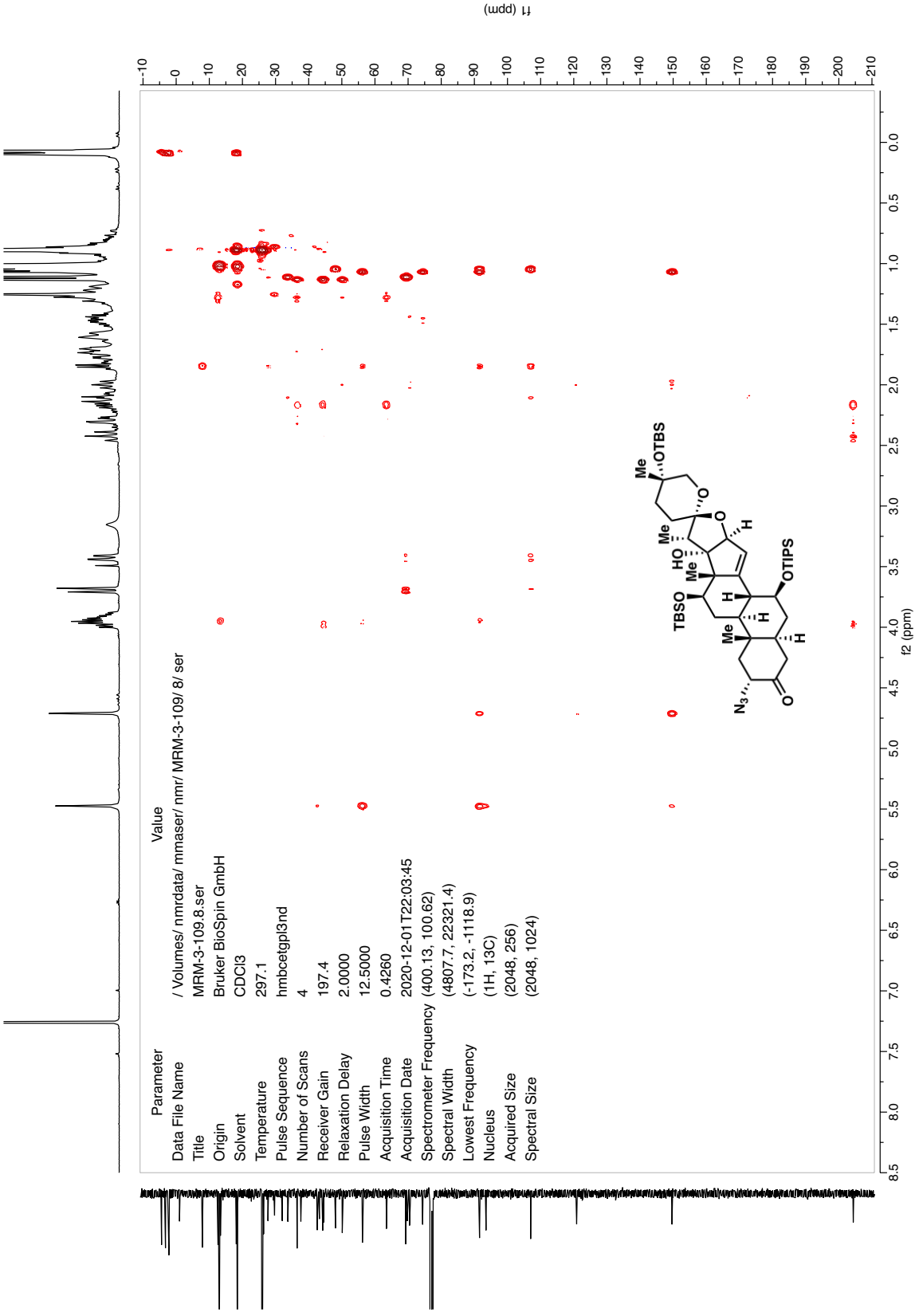

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-105/ 8/ ser
Title MRM-3-105.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-11-05T00:50:02
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-172.9, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
f1 (ppm)
f2 (ppm)
Me
OTBS
HO
Me
TBSO
Me
H
OTIPS
H
H
Me
Br
O
O
O
H
H

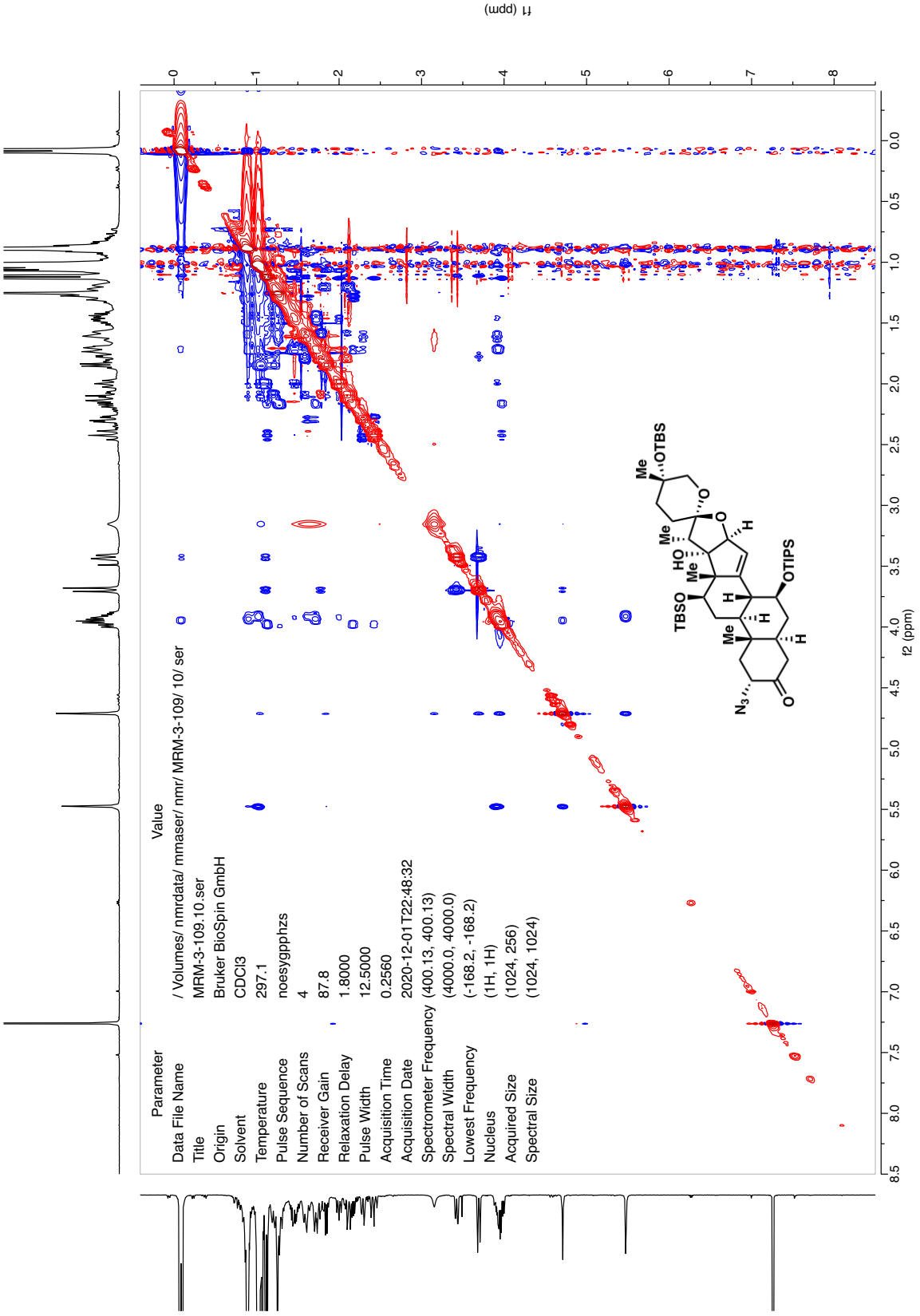

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-105/ 10/ ser
Title MRM-3-105.10.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence noesygpphzs
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 1.8000
Pulse Width 12.5000
Acquisition Time 0.2560
Acquisition Date 2020-11-05T01:34:40
Spectrometer Frequency (400.13, 400.13)
Spectral Width (4000.0, 4000.0)
Lowest Frequency (-168.0, -168.0)
Nucleus (1H, 1H)
Acquired Size (1024, 256)
Spectral Size (1024, 1024)
Me
OTBS
HO
Me
O
H
TBSO
Me
H
OTIPS
H
Me
H
Br
O

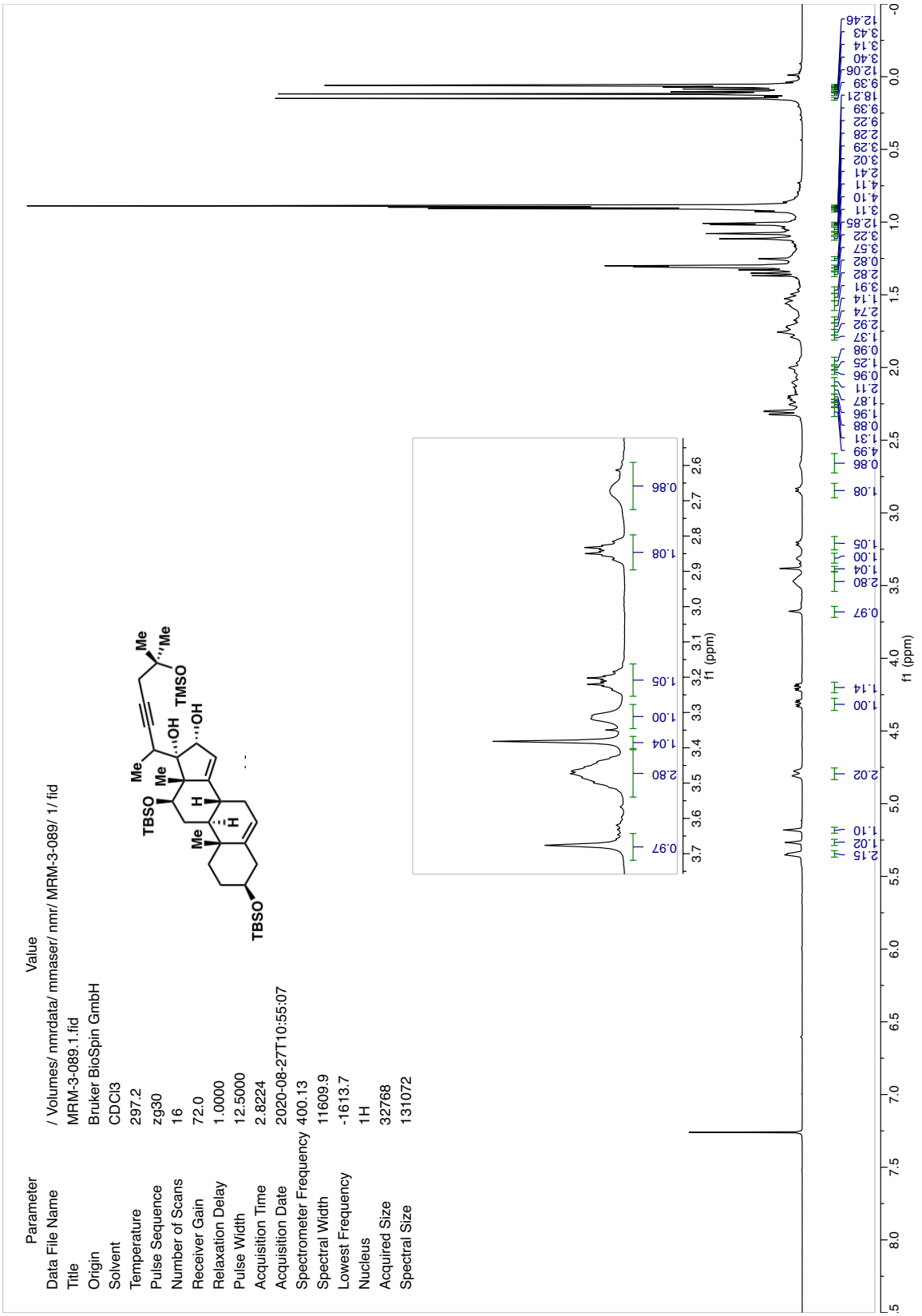

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-109/ 1/ fid
Title MRM-3-109.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 127.1
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-12-01T14:08:35
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.7
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
OTBS
O
O
H
HO
Me
TBSO
H
H
OTIPS
Me
H
O
N3
f1 (ppm)

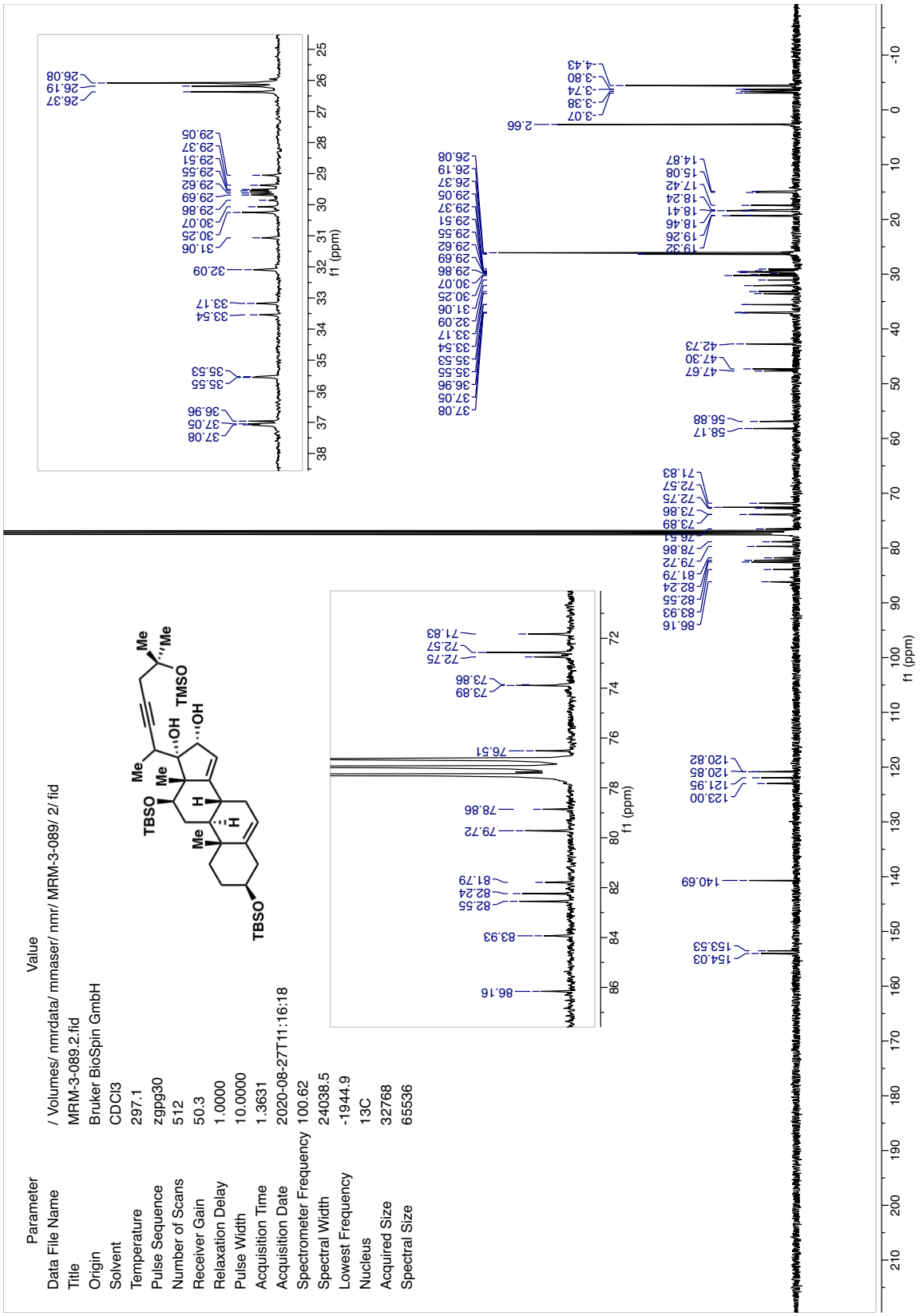

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-109/ 2/ fid
Title MRM-3-109.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 43.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-12-01T21:45:21
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.7
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
OTBS
O
O
H
Me
HO
Me
TBSO
H
OTIPS
H
H
Me
H
N3
O
204.52
149.82
121.03
107.16
93.68
91.72
74.55
70.61
69.90
69.42
63.63
56.38
50.29
48.24
44.69
44.39
43.40
42.74
37.80
36.67
33.87
32.18
29.86
27.86
26.55
26.21
26.01
18.67
18.66
18.35
18.30
13.54
13.18
12.79
8.07
1.17
-2.01
-2.18
-3.07
-4.25
f1 (ppm)

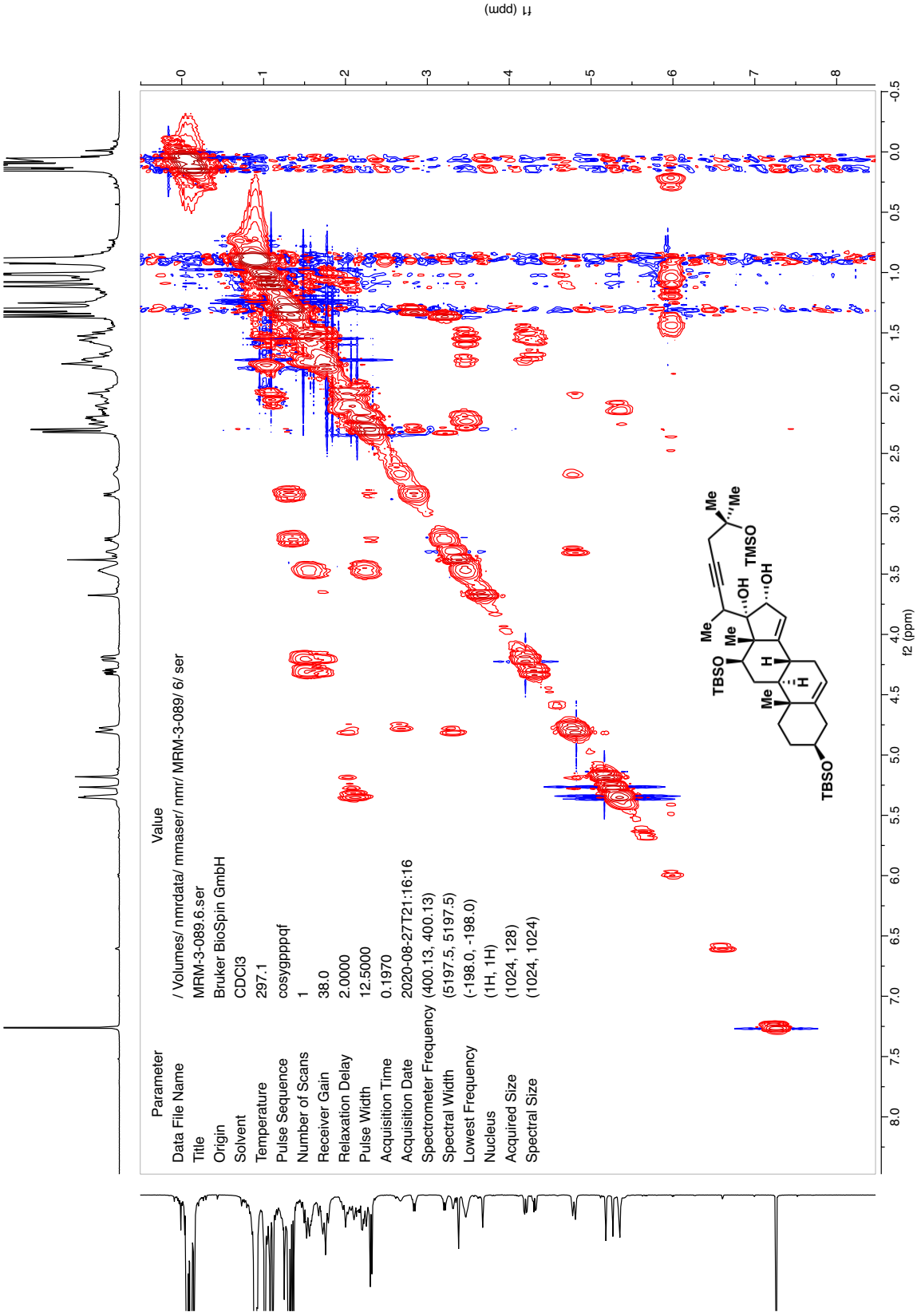

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-109/ 6/ ser
Title MRM-3-109.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-12-01T21:52:12
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-527.1, -1265.3)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
Me
OTBS
O
Me
HO
Me
TBSO
H
H
Me
H
OTIPS
H
N3
O

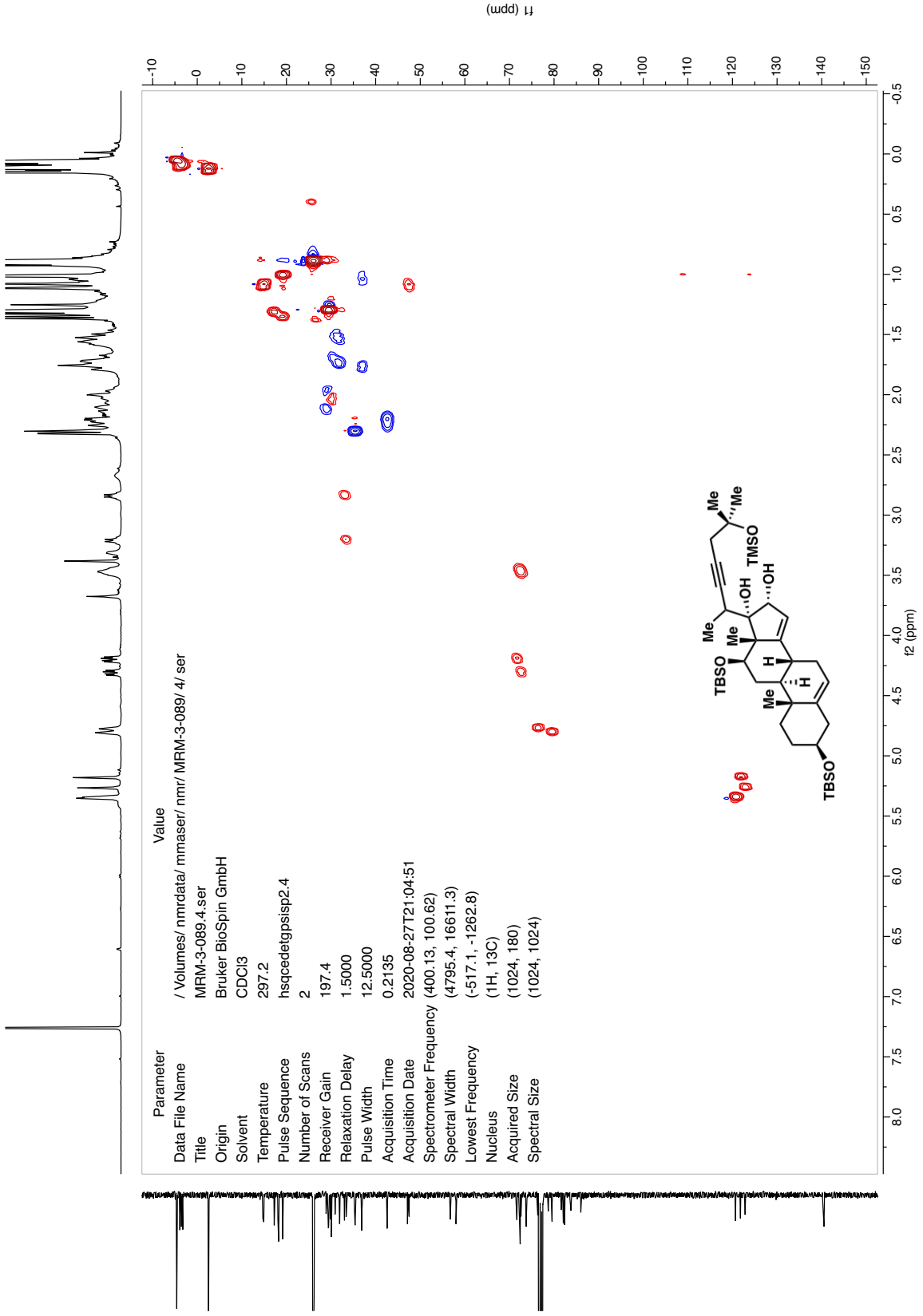

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-109/ 8/ ser
Title MRM-3-109.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-12-01T22:03:45
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-173.2, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
f1 (ppm)
f2 (ppm)
Me
OTBS
O
Me
HO
Me
H
TBSO
H
OTIPS
Me
H
H
N3
O

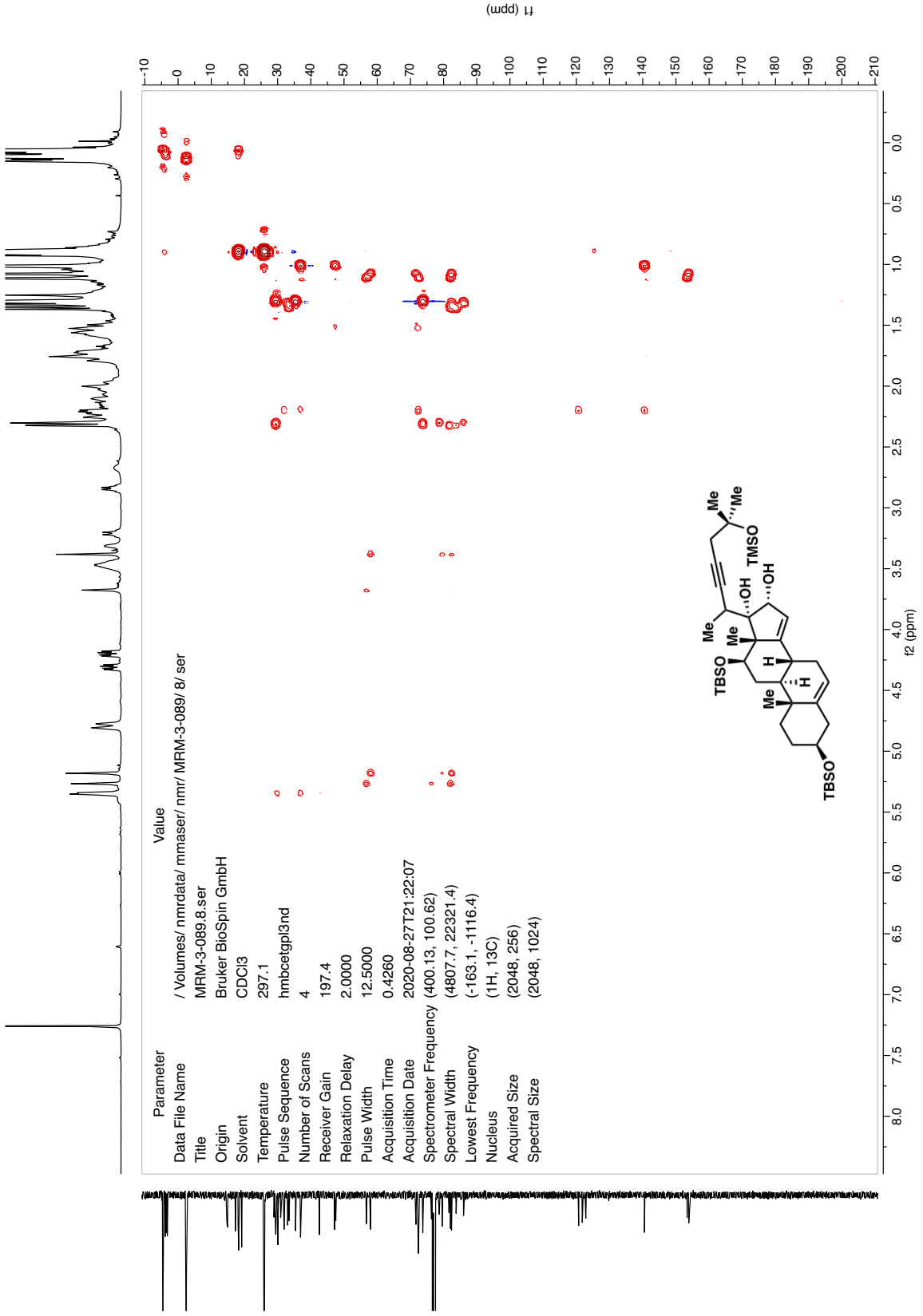

f1 (ppm)
f2 (ppm)
Parameter | Value
Data File Name | / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-109/ 10/ ser
Title | MRM-3-109.10.ser
Origin | Bruker BioSpin GmbH
Solvent | CDCl3
Temperature | 297.1
Pulse Sequence | noesygpphzs
Number of Scans | 4
Receiver Gain | 87.8
Relaxation Delay | 1.8000
Pulse Width | 12.5000
Acquisition Time | 0.2560
Acquisition Date | 2020-12-01T22:48:32
Spectrometer Frequency | (400.13, 400.13)
Spectral Width | (4000.0, 4000.0)
Lowest Frequency | (-168.2, -168.2)
Nucleus | (1H, 1H)
Acquired Size | (1024, 256)
Spectral Size | (1024, 1024)
Me
OTBS
O
Me
HO
Me
H
TBSO
H
H
OTIPS
Me
H
N3
O

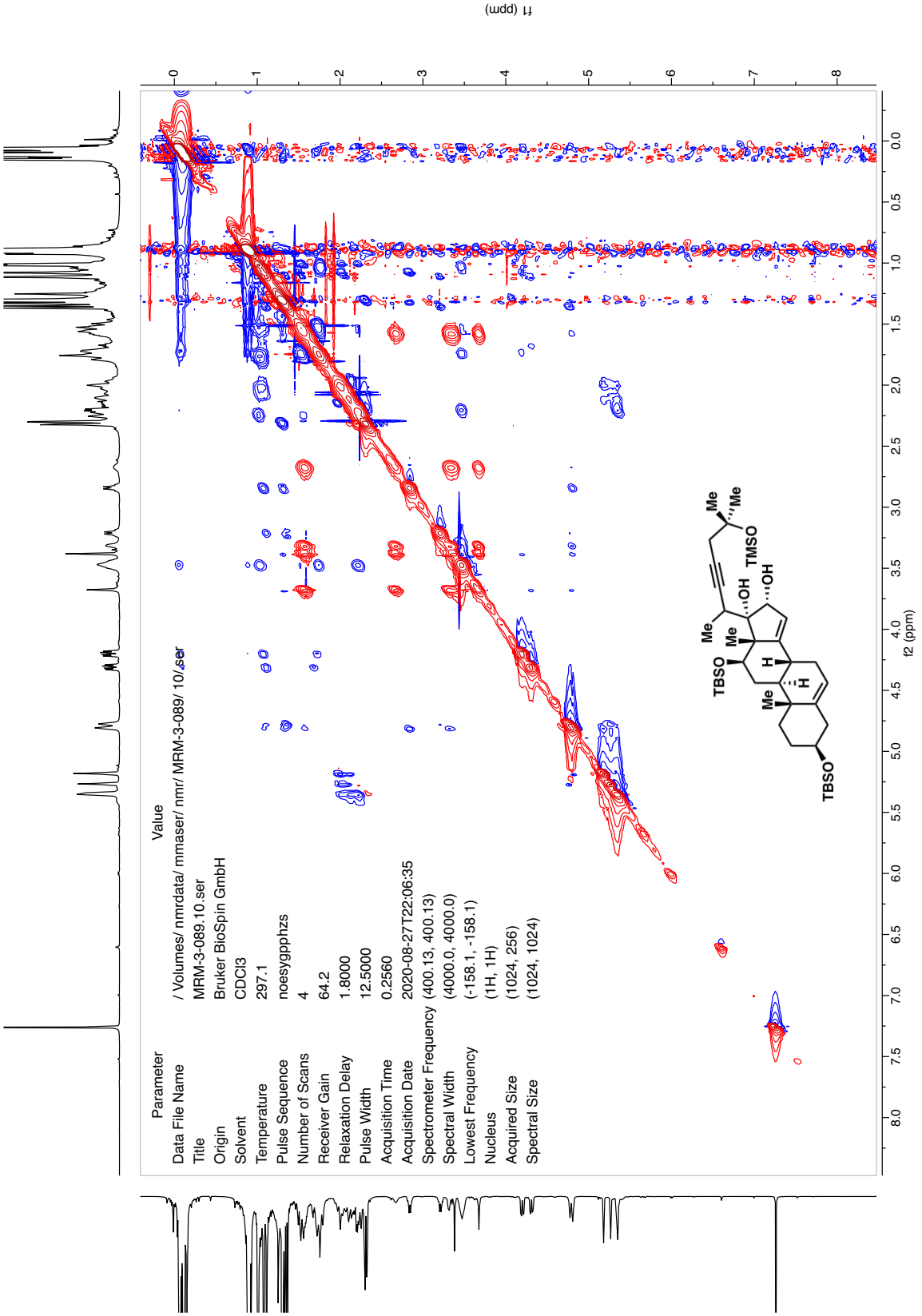

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-089/ 1/ fid
Title MRM-3-089.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-08-27T10:55:07
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.7
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
Me
Me
H
H
Me
OH
OH
TMSO
Me
Me
TBSO
f1 (ppm)

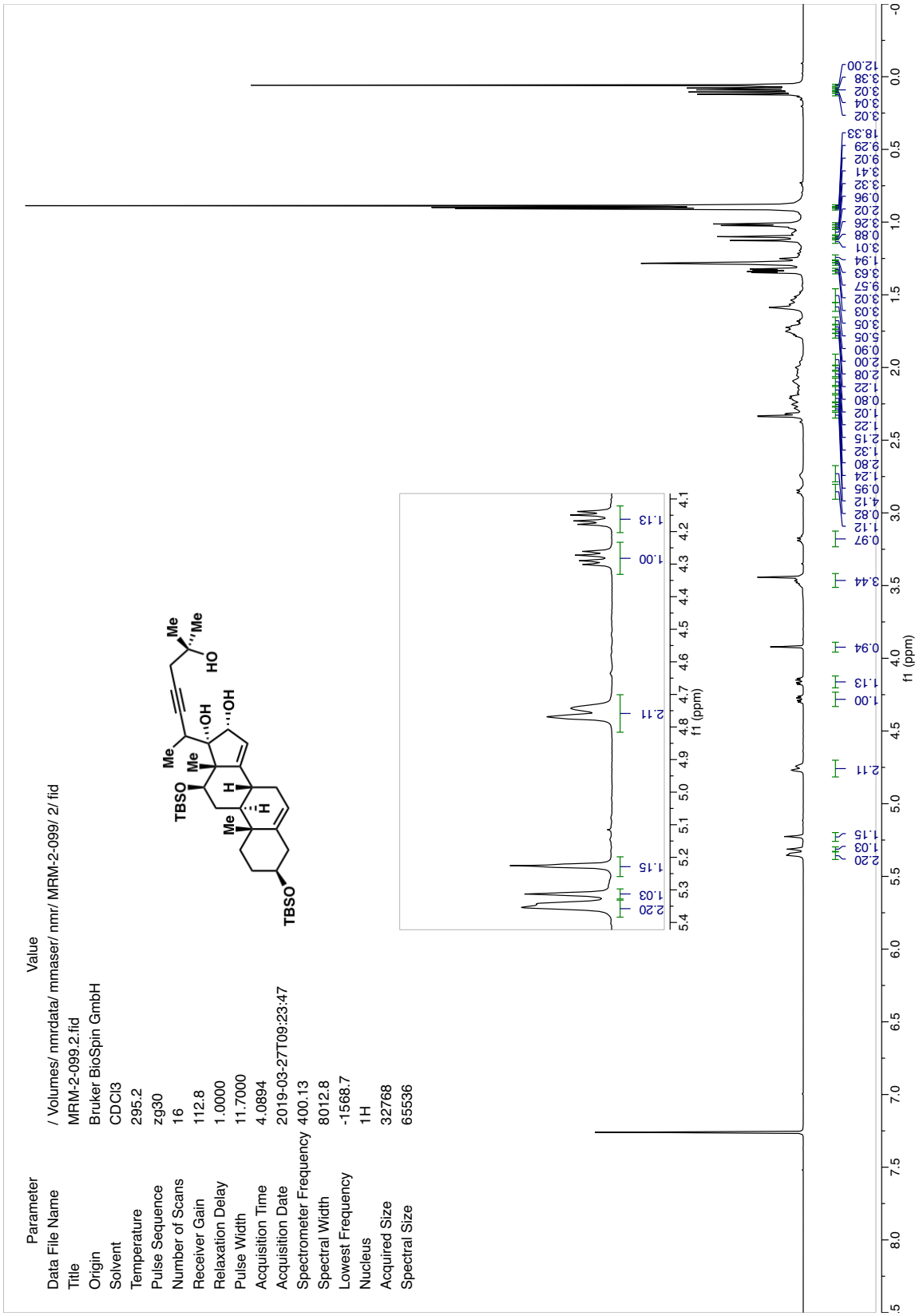

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-089/ 2/ fid
Title MRM-3-089.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-08-27T11:16:18
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
Me
H
H
Me
OH
OH
TMSO
Me
Me
TBSO
f1 (ppm)

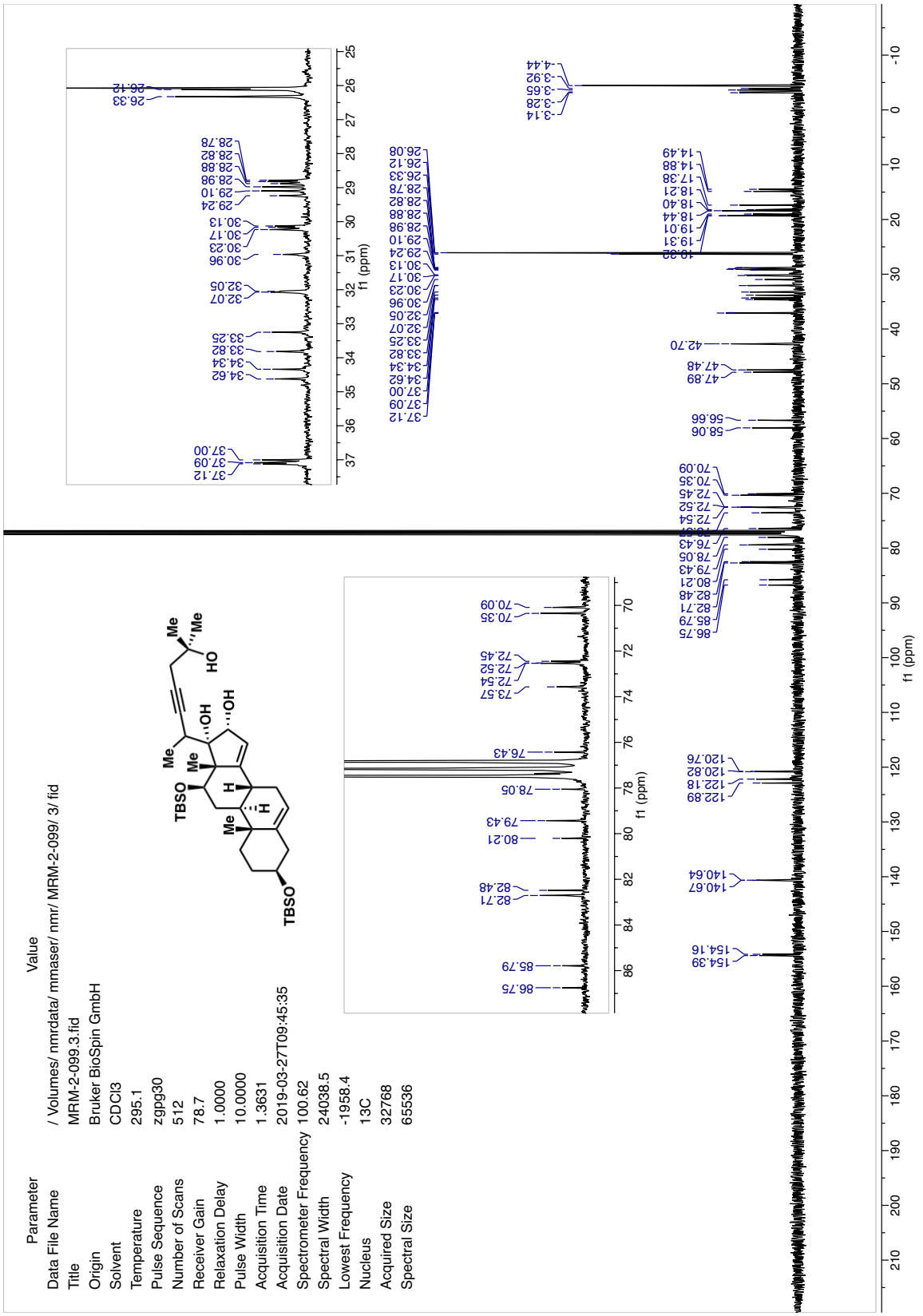

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-089/ 6/ ser
Title MRM-3-089.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence cosygpppqf
Number of Scans 1
Receiver Gain 38.0
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.1970
Acquisition Date 2020-08-27T21:16:16
Spectrometer Frequency (400.13, 400.13)
Spectral Width (5197.5, 5197.5)
Lowest Frequency (-198.0, -198.0)
Nucleus (1H, 1H)
Acquired Size (1024, 128)
Spectral Size (1024, 1024)
Me
Me
TMSO
OH
OH
Me
Me
TBSO
H
H
Me
TBSO

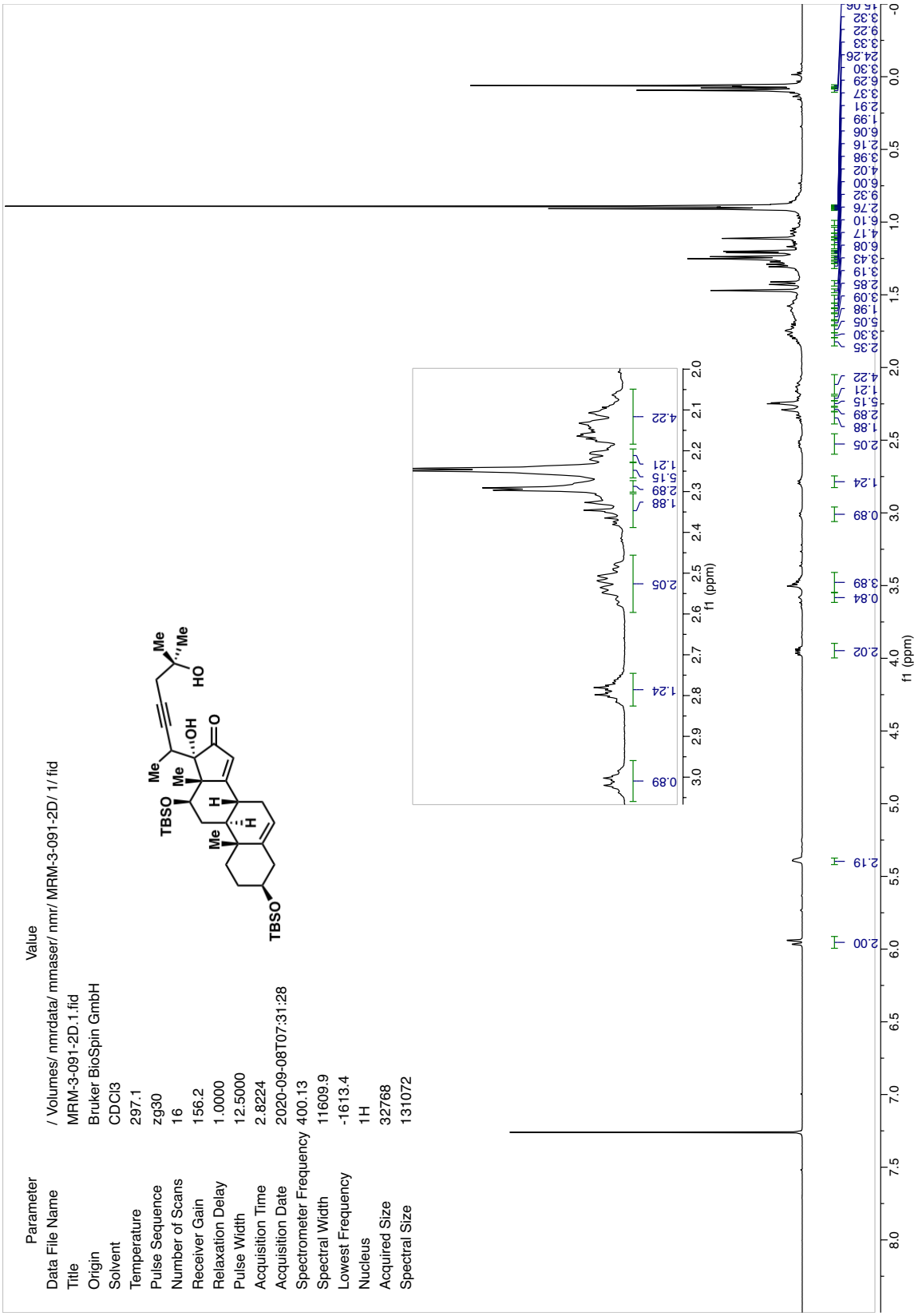

Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-089/ 4/ ser
Title
MRM-3-089.4.ser
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.2
Pulse Sequence
hsqcedetgpsisp2.4
Number of Scans
2
Receiver Gain
197.4
Relaxation Delay
1.5000
Pulse Width
12.5000
Acquisition Time
0.2135
Acquisition Date
2020-08-27T21:04:51
Spectrometer Frequency
(400.13, 100.62)
Spectral Width
(4795.4, 16611.3)
Lowest Frequency
(-517.1, -1262.8)
Nucleus
(1H, 13C)
Acquired Size
(1024, 180)
Spectral Size
(1024, 1024)
f1 (ppm)
f2 (ppm)
TBSO
Me
Me
H
H
Me
TBSO
Me
Me
OH
TMSO
OH

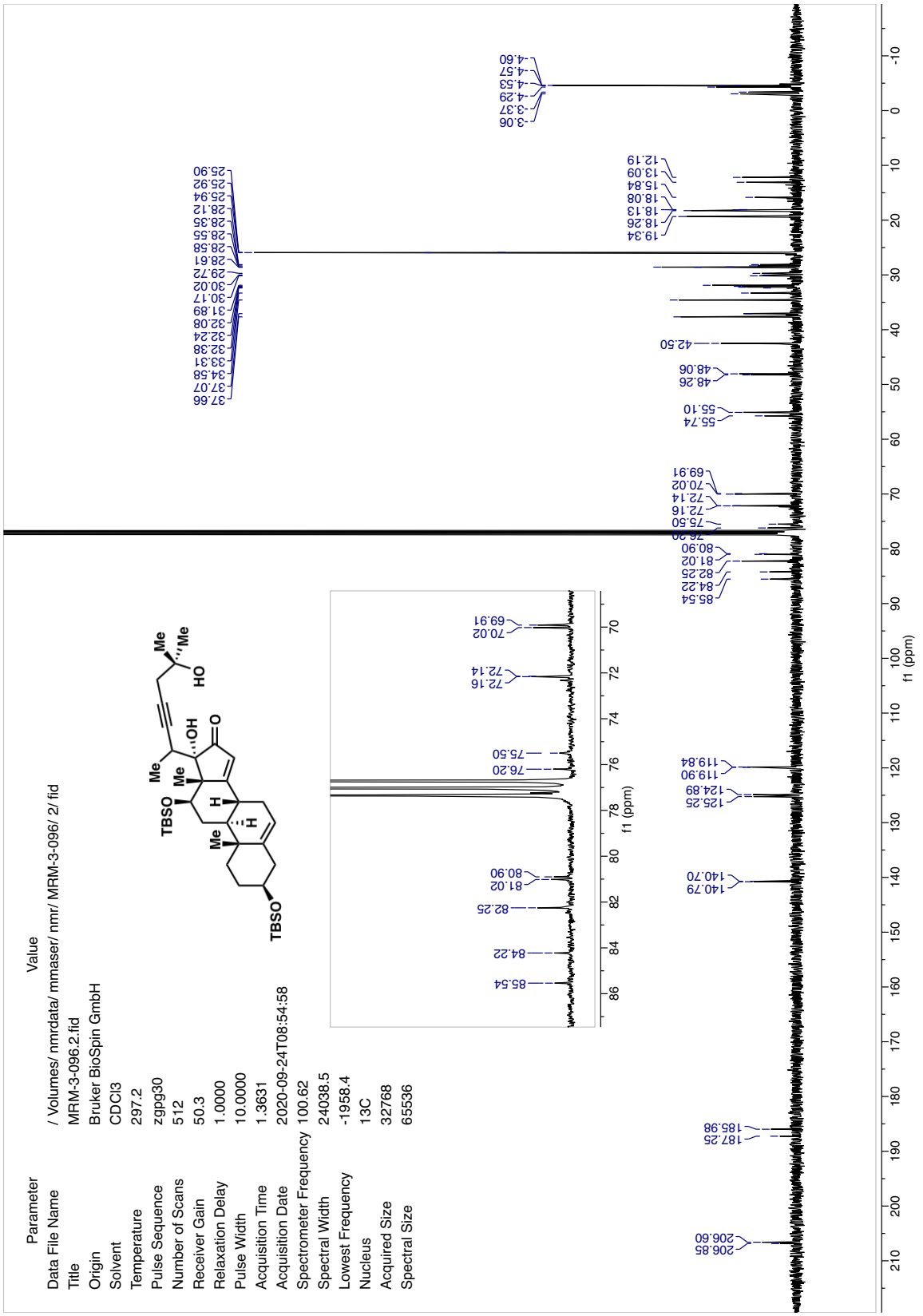

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-089/ 8/ ser
Title MRM-3-089.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-08-27T21:22:07
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-163.1, -1116.4)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
Me
Me
H
H
Me
TBSO
Me
Me
OH
TMSO
OH

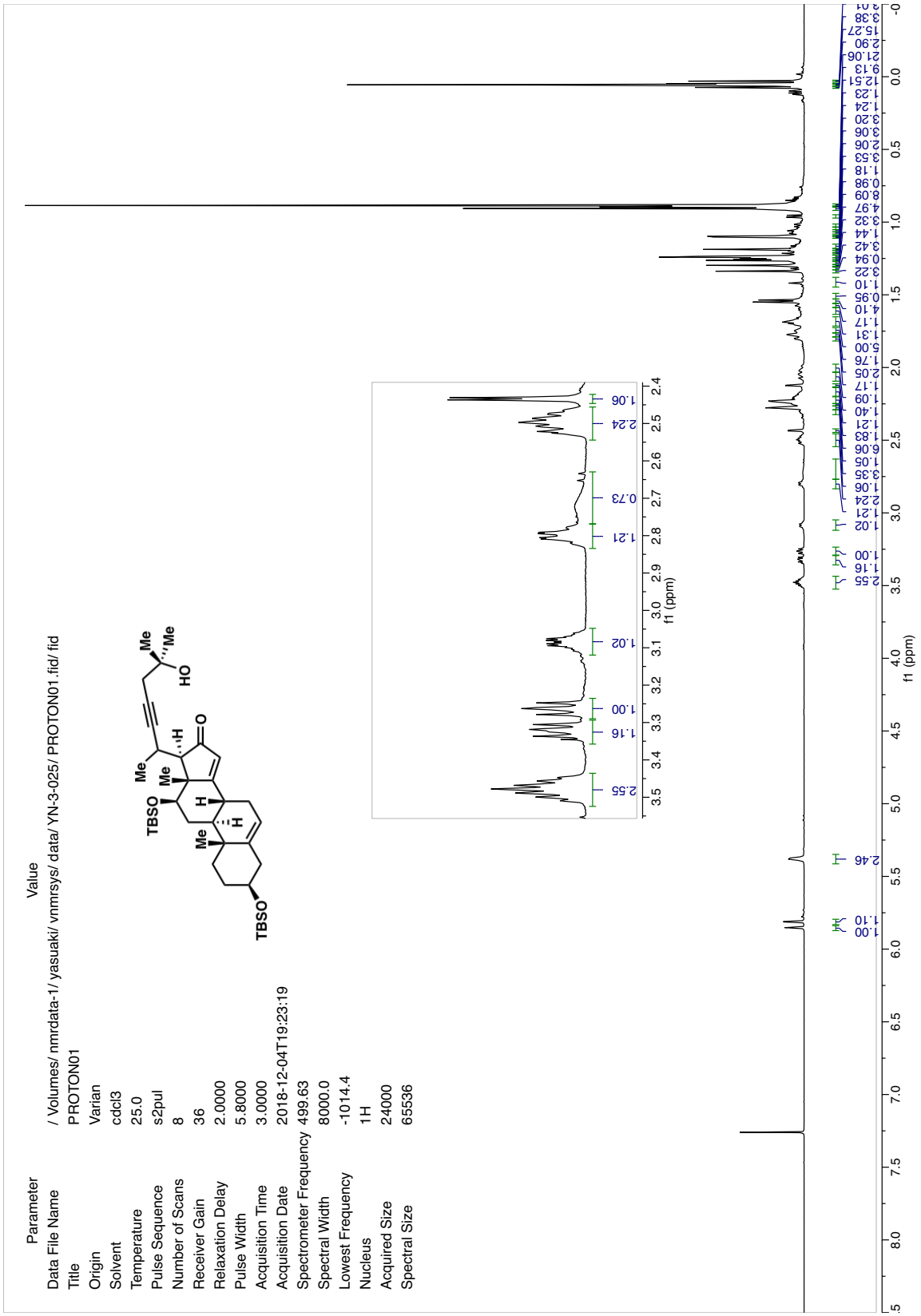

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-089/ 10/.ser
Title MRM-3-089.10.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence noesygpphzs
Number of Scans 4
Receiver Gain 64.2
Relaxation Delay 1.8000
Pulse Width 12.5000
Acquisition Time 0.2560
Acquisition Date 2020-08-27T22:06:35
Spectrometer Frequency (400.13, 400.13)
Spectral Width (4000.0, 4000.0)
Lowest Frequency (-158.1, -158.1)
Nucleus (1H, 1H)
Acquired Size (1024, 256)
Spectral Size (1024, 1024)
Me
Me
TMSO
OH
OH
Me
Me
TBSO
H
H
Me
TBSO

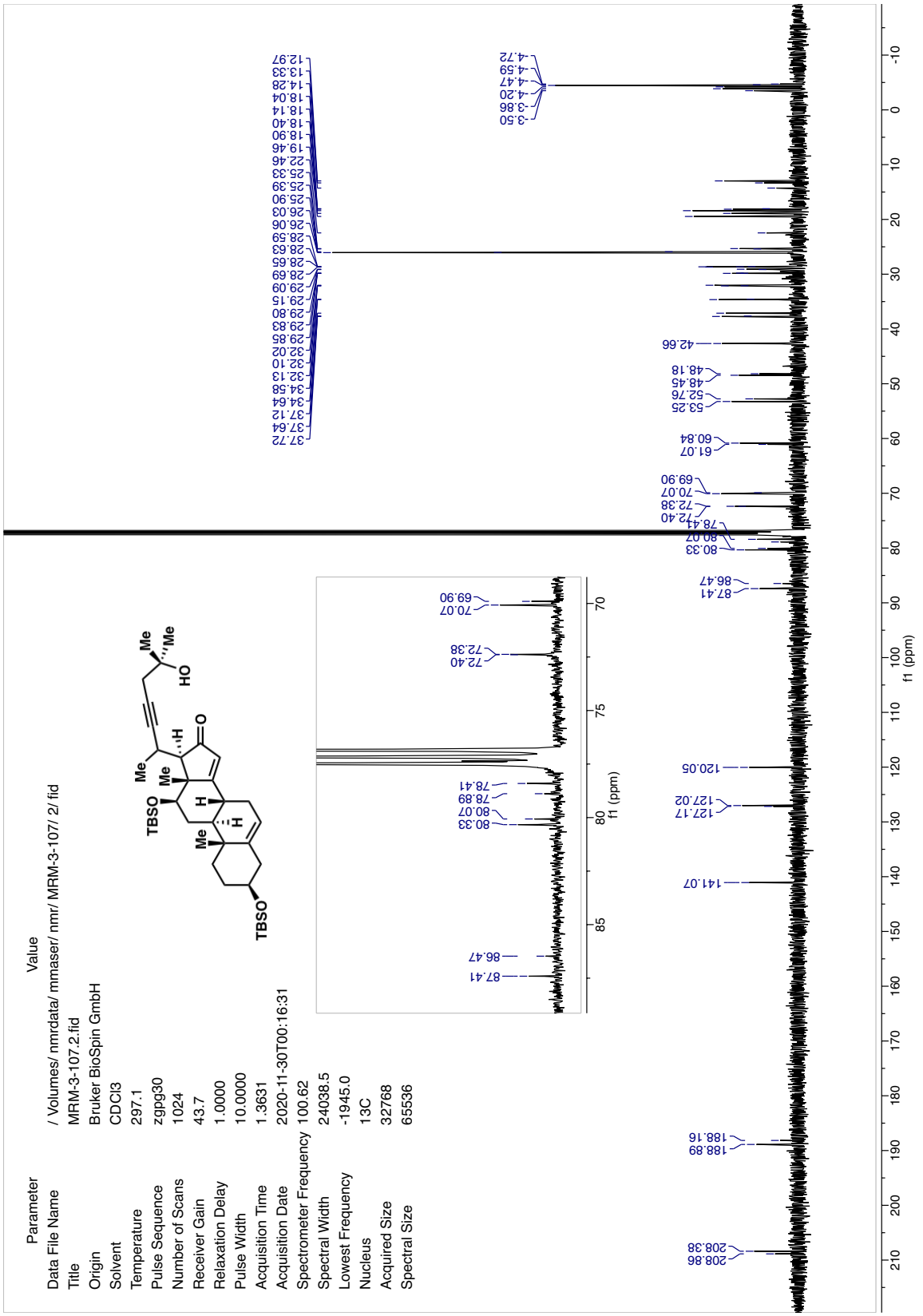

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-099/ 2/ fid
Title MRM-2-099.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 112.8
Relaxation Delay 1.0000
Pulse Width 11.7000
Acquisition Time 4.0894
Acquisition Date 2019-03-27T09:23:47
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1568.7
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
TBSO
Me
Me
H
H
Me
TBSO
OH
OH
Me
Me
HO
1.13
1.00
2.11
1.15
1.03
2.20
f1 (ppm)
3.02
3.04
3.02
3.38
12.00
18.33
9.29
9.02
3.41
3.32
0.96
2.02
0.88
3.26
3.01
1.94
3.63
9.57
3.02
3.03
3.05
5.05
0.90
2.00
2.08
1.22
0.80
1.02
1.22
2.15
1.32
2.80
1.24
0.95
4.12
0.82
1.12
0.97
3.44
0.94
1.13
1.00
2.11
1.15
1.03
2.20
f1 (ppm)

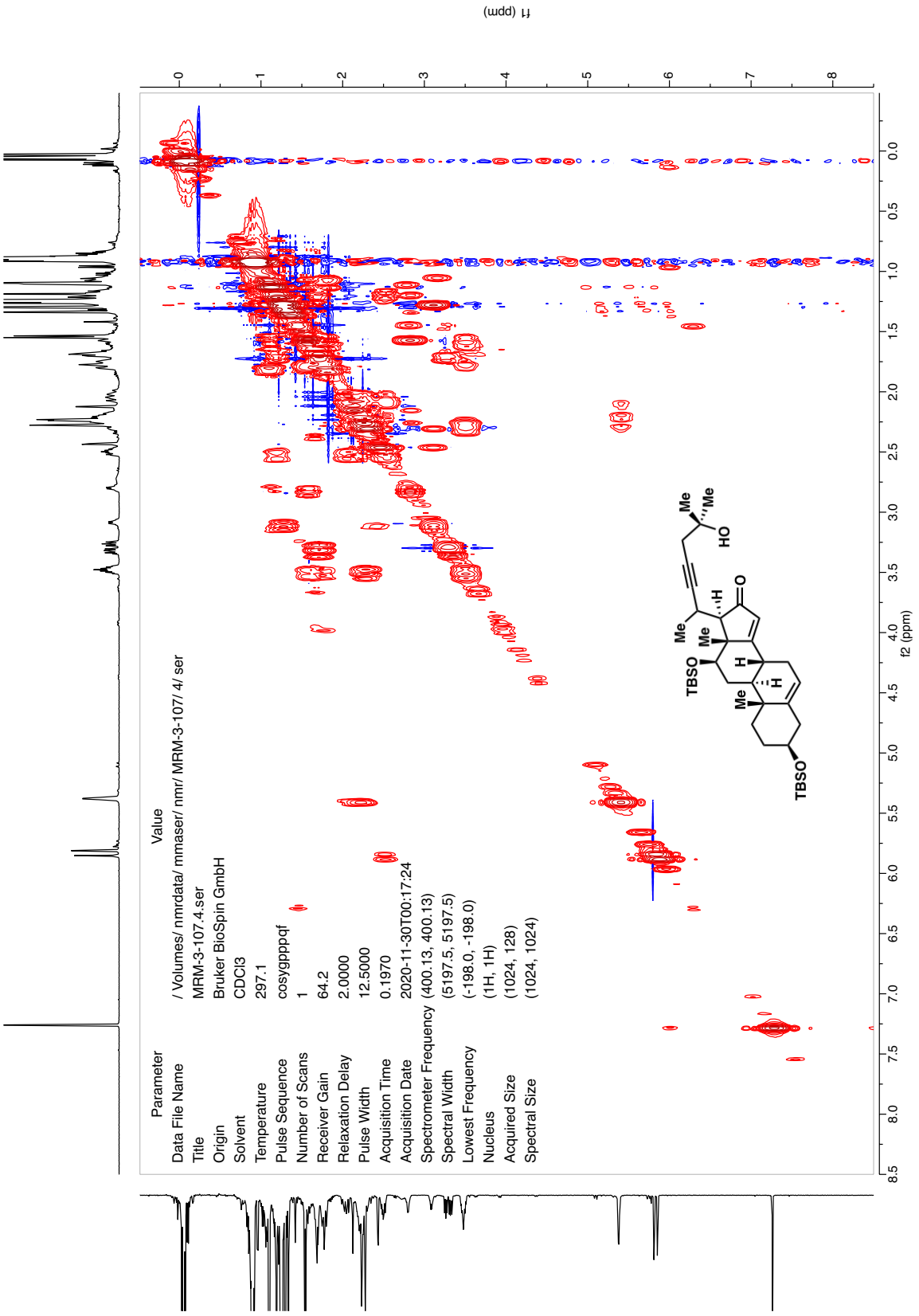

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-099/ 3/ fid
Title MRM-2-099.3.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 295.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-03-27T09:45:35
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1958.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
Me
Me
H
H
Me
HO
HO
Me
Me
TBSO
f1 (ppm)

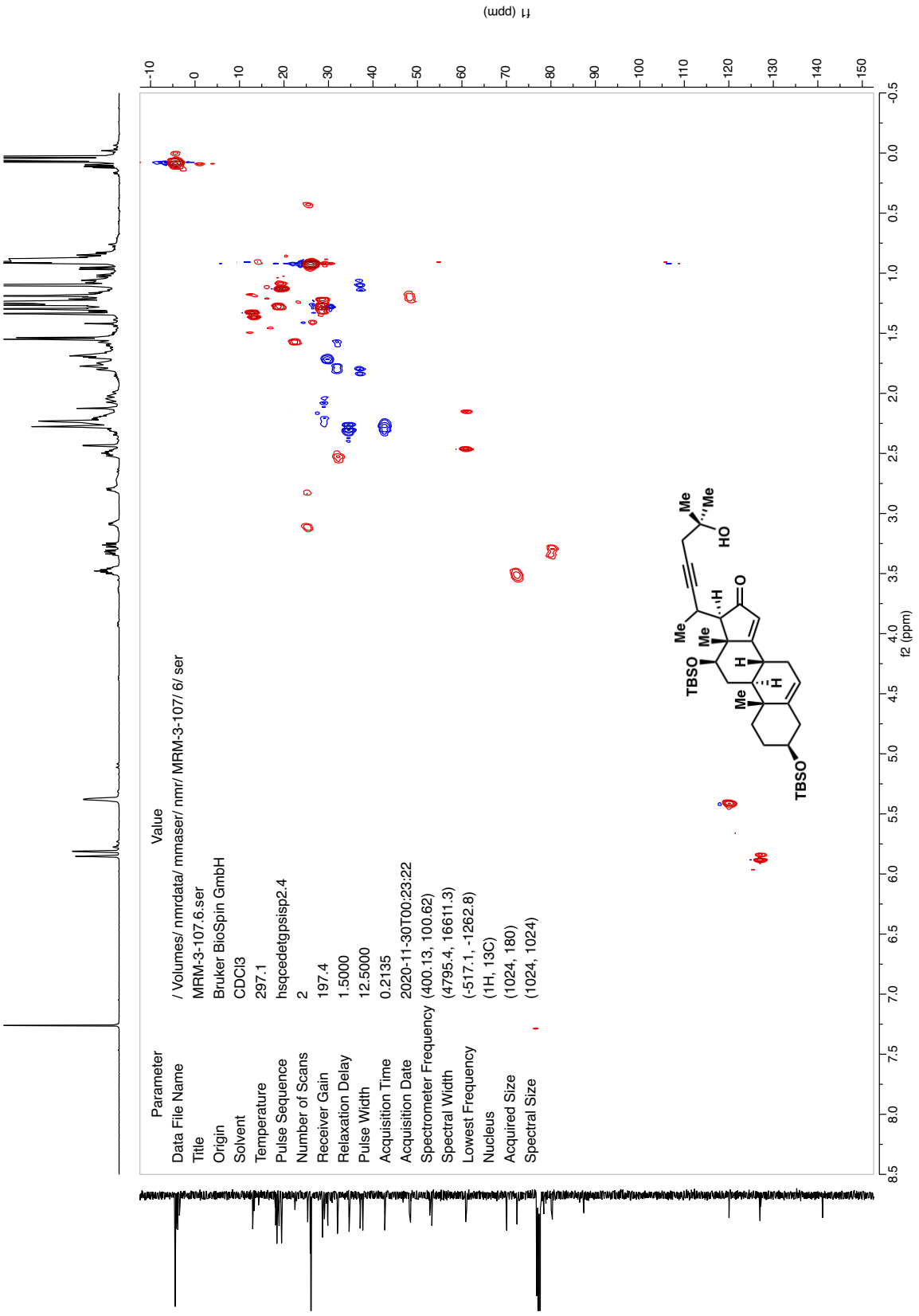

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-091-2D/ 1/ fid
Title MRM-3-091-2D.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-09-08T07:31:28
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.4
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
Me
Me
HO
Me
TBSO
Me
OH
O
H
Me
H
TBSO
f1 (ppm)

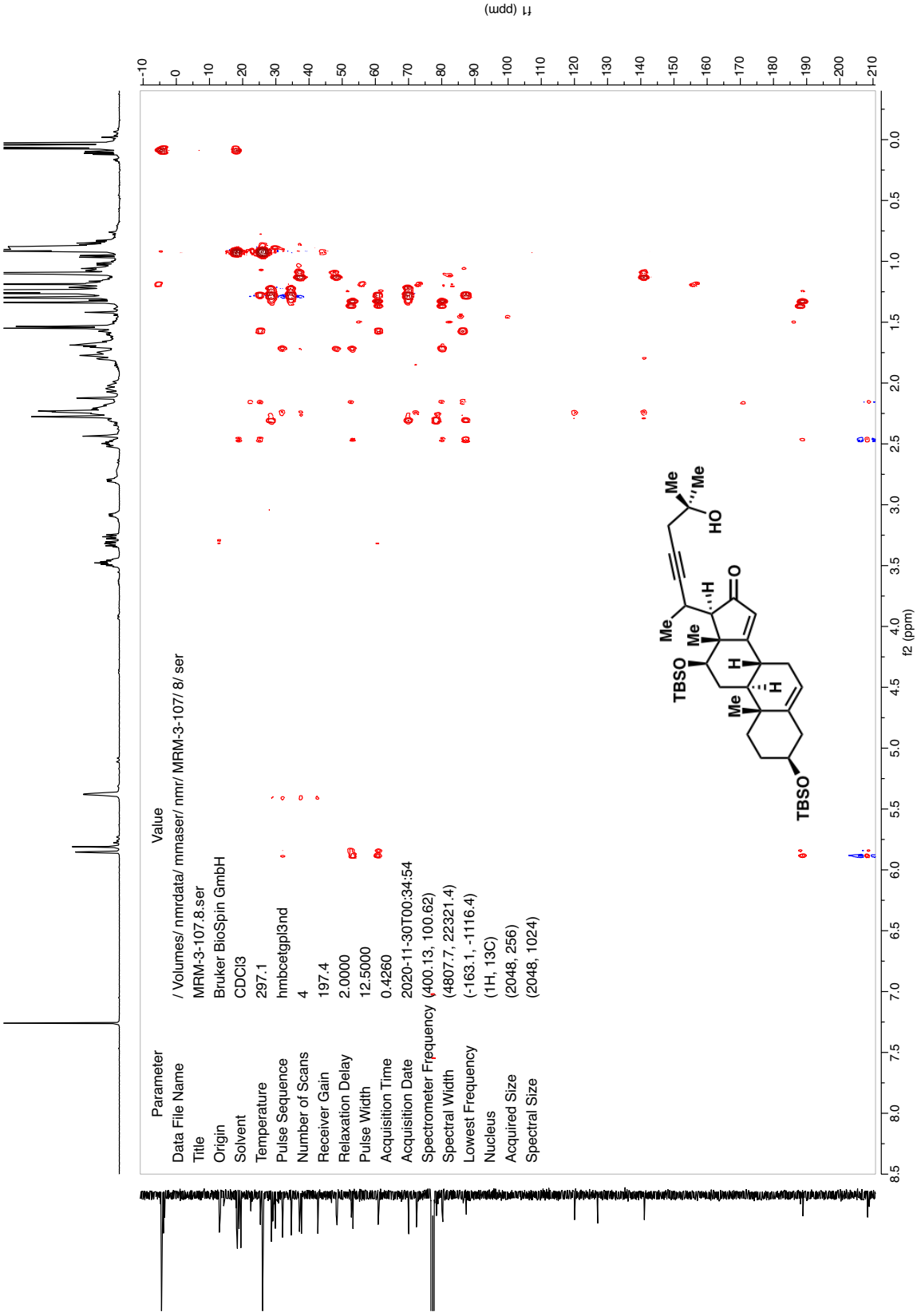

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-096/ 2/ fid
Title MRM-3-096.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-09-24T08:54:58
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1958.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
Me
Me
H
H
Me
OH
O
HO
Me
Me
TBSO
f1 (ppm)

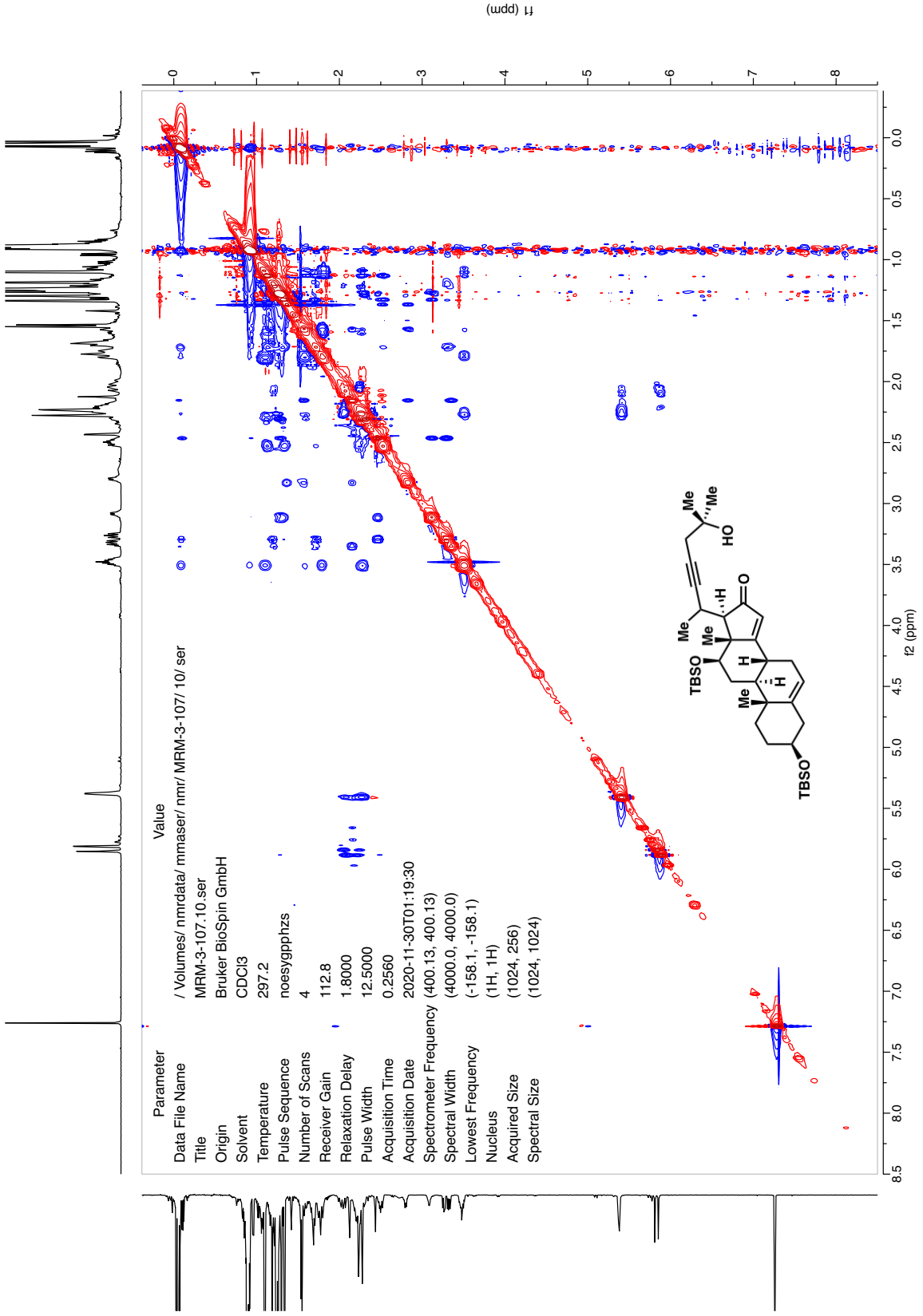

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ yasuaki/ vnmrsys/ data/ YN-3-025/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 8
Receiver Gain 36
Relaxation Delay 2.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2018-12-04T19:23:19
Spectrometer Frequency 499.63
Spectral Width 8000.0
Lowest Frequency -1014.4
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
TBSO
Me
Me
H
H
Me
TBSO
Me
Me
HO
O
H
f1 (ppm)

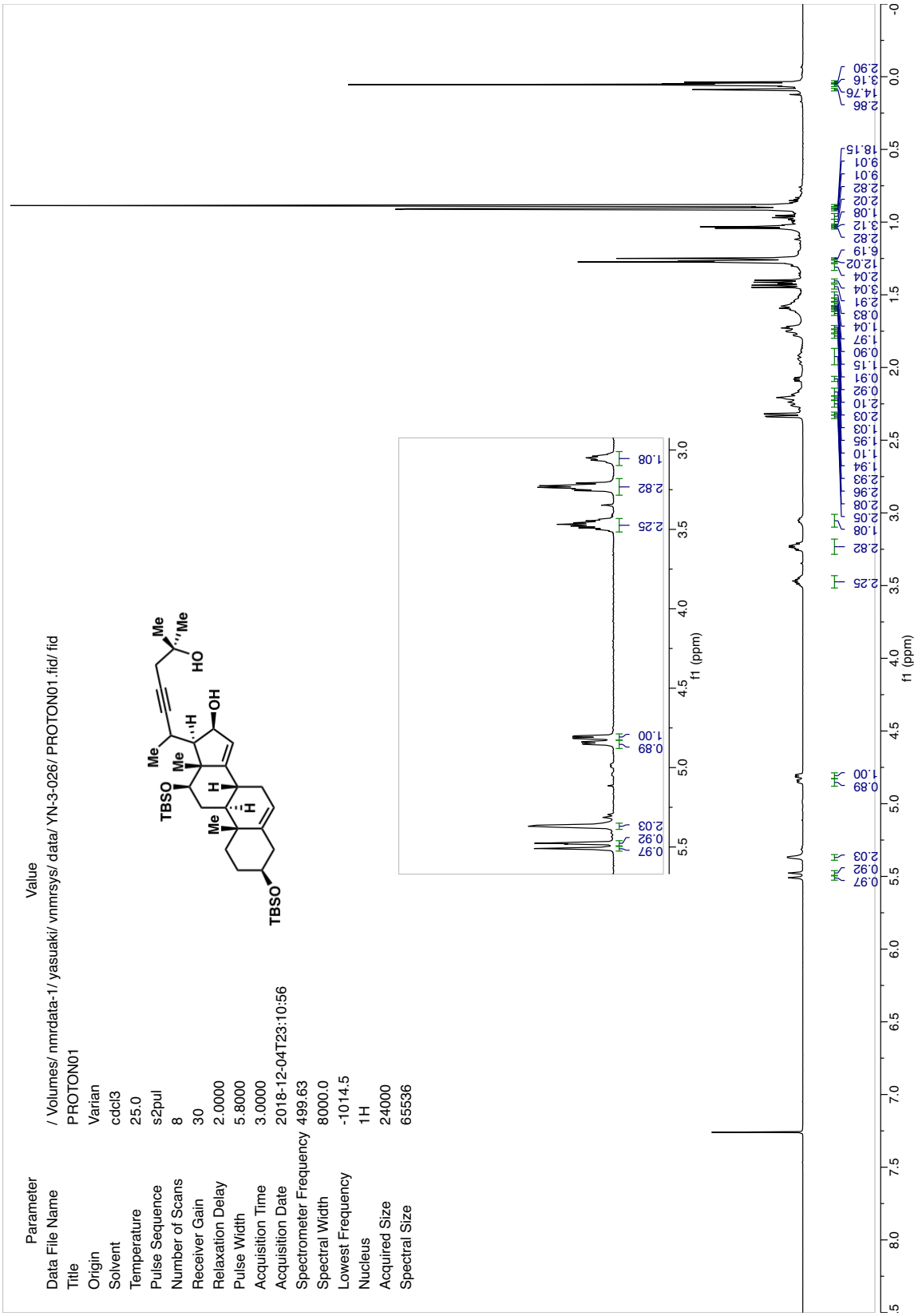

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-107/ 2/ fid
Title MRM-3-107.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 43.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-11-30T00:16:31
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
Me
Me
H
H
Me
H
O
Me
Me
HO
TBSO
f1 (ppm)

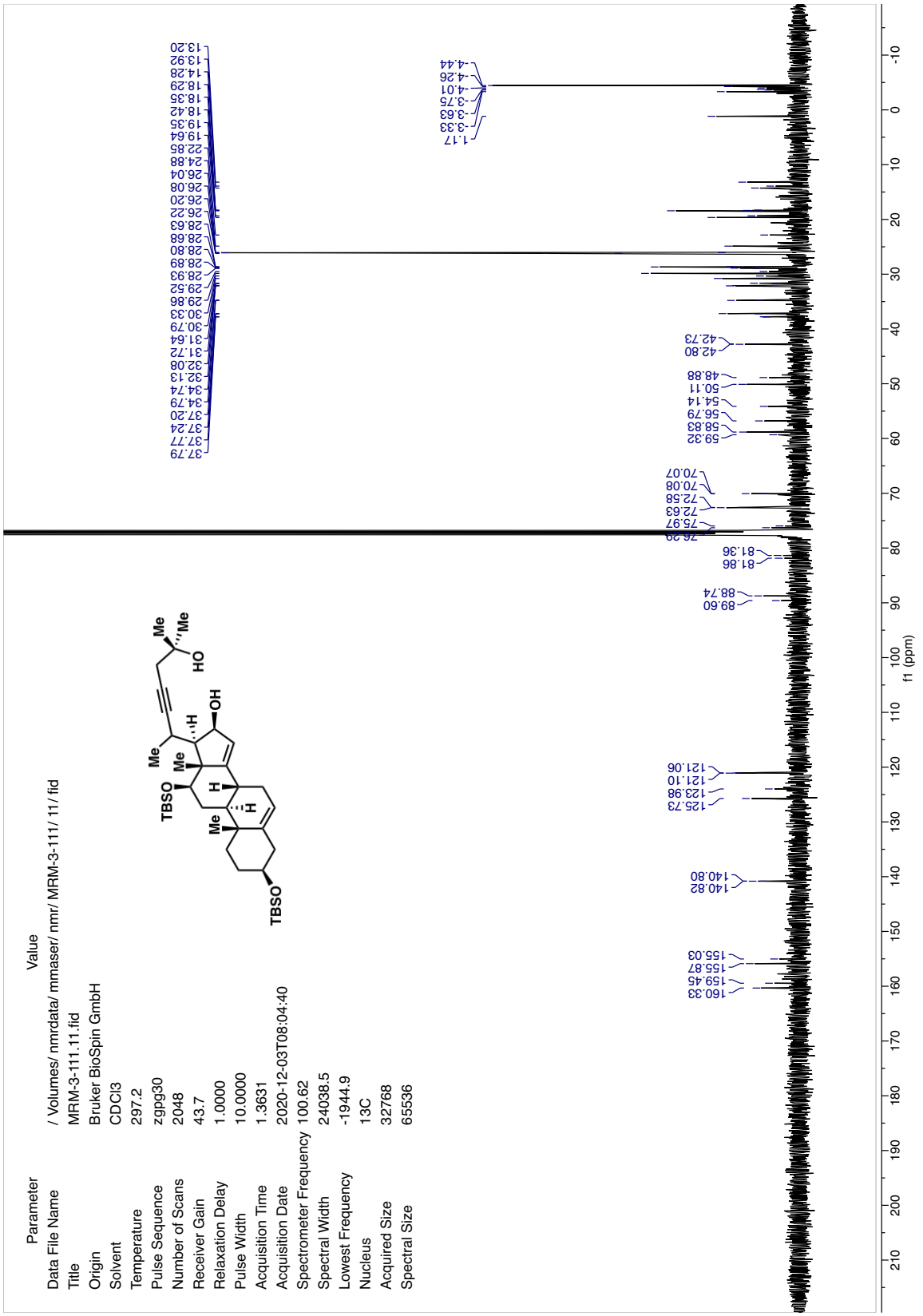

f1 (ppm)
f2 (ppm)
Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-107/ 4/ ser
Title
MRM-3-107.4.ser
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
cosygpppqf
Number of Scans
1
Receiver Gain
64.2
Relaxation Delay
2.0000
Pulse Width
12.5000
Acquisition Time
0.1970
Acquisition Date
2020-11-30T00:17:24
Spectrometer Frequency
(400.13, 400.13)
Spectral Width
(5197.5, 5197.5)
Lowest Frequency
(-198.0, -198.0)
Nucleus
(1H, 1H)
Acquired Size
(1024, 128)
Spectral Size
(1024, 1024)
Me
Me
HO
O
H
Me
Me
TBSO
H
H
Me
TBSO

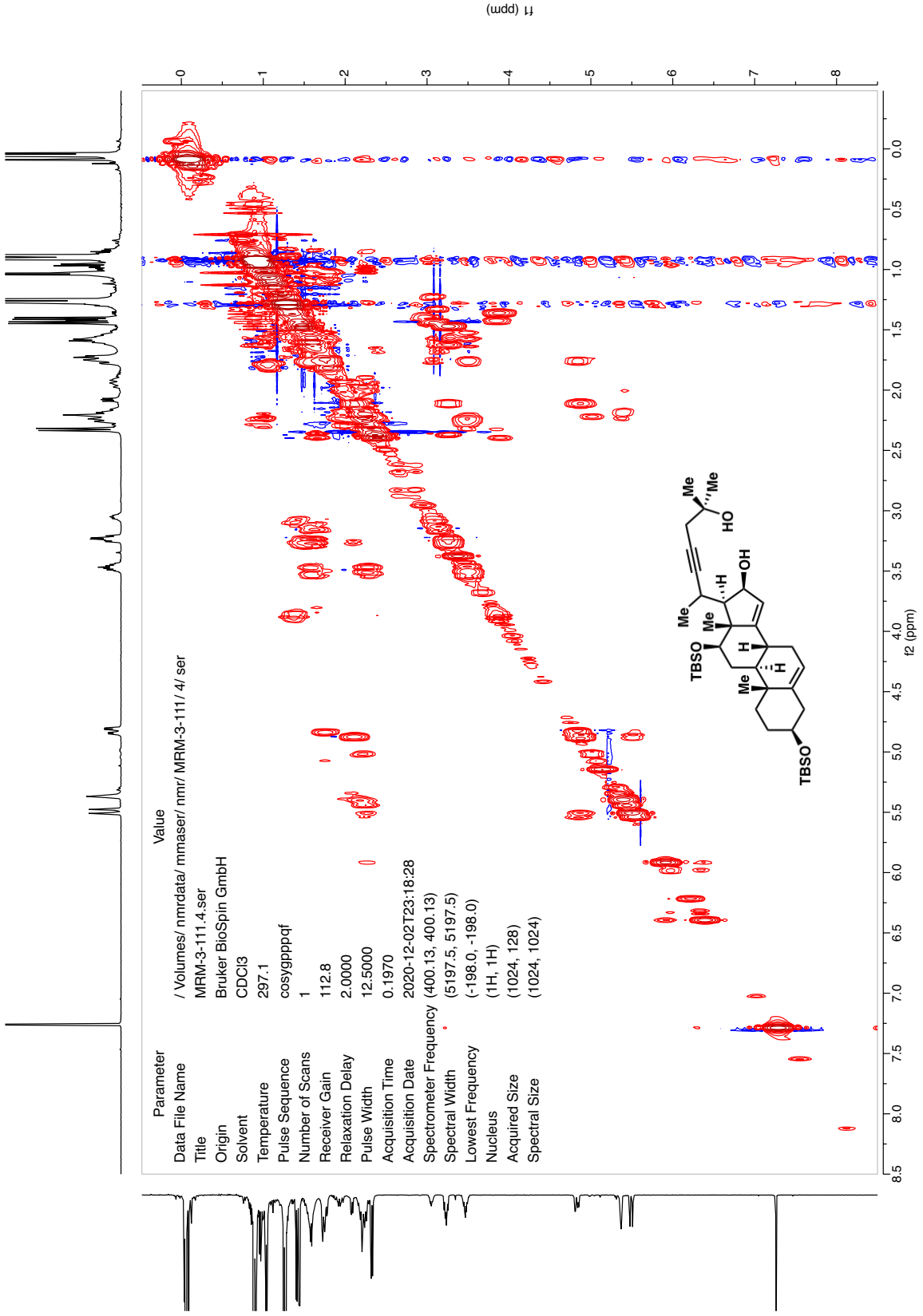

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-107/ 6/ ser
Title MRM-3-107.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-11-30T00:23:22
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-517.1, -1262.8)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
Me
Me
H
H
Me
TBSO
Me
Me
HO
H
O

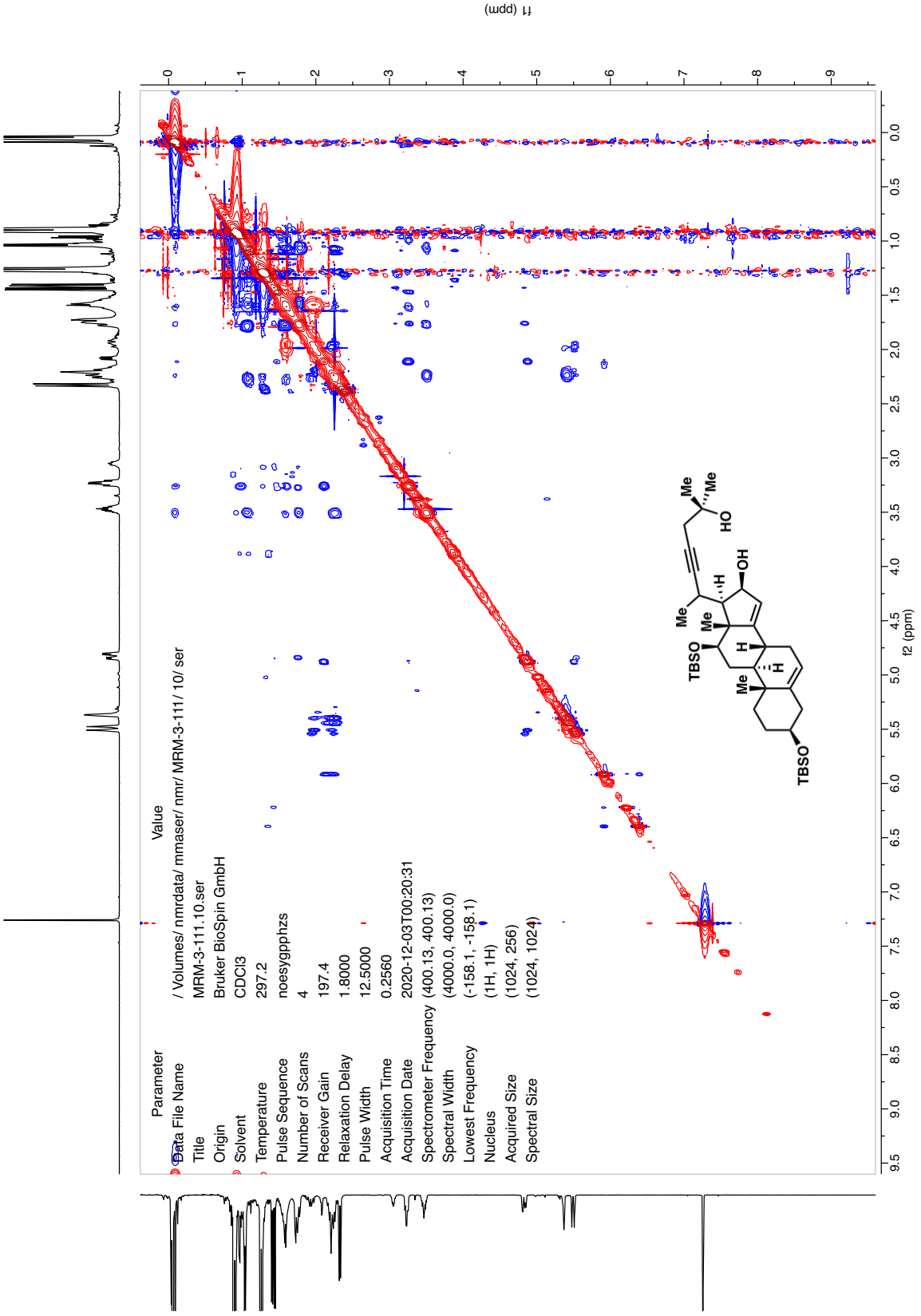

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-107/ 8/ ser
Title MRM-3-107.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-11-30T00:34:54
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-163.1, -1116.4)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
Me
Me
H
O
HO
Me
Me
TBSO

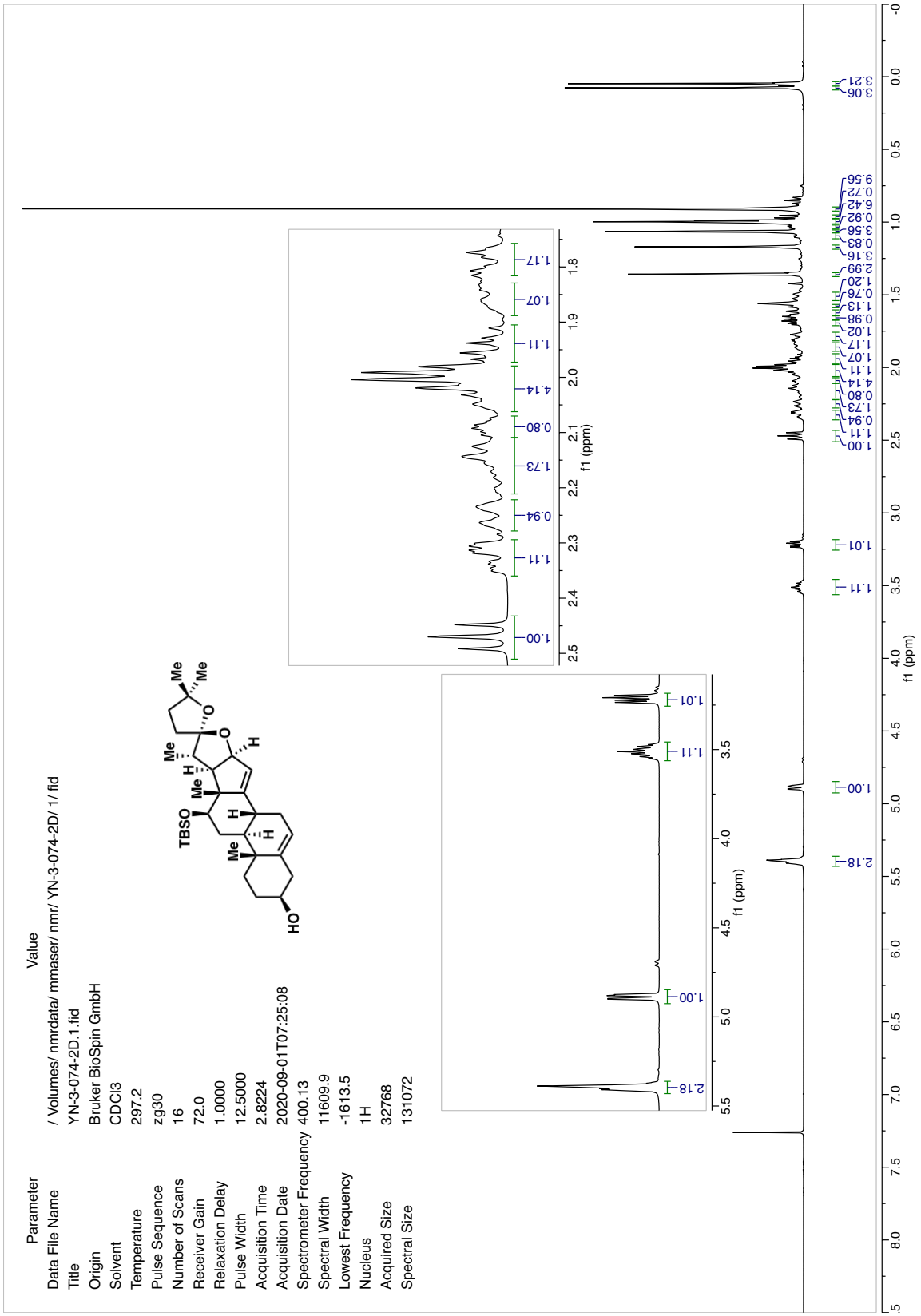

f1 (ppm)
f2 (ppm)
Parameter | Value
Data File Name | / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-107/ 10/ ser
Title | MRM-3-107.10.ser
Origin | Bruker BioSpin GmbH
Solvent | CDCl3
Temperature | 297.2
Pulse Sequence | noesygpphzs
Number of Scans | 4
Receiver Gain | 112.8
Relaxation Delay | 1.8000
Pulse Width | 12.5000
Acquisition Time | 0.2560
Acquisition Date | 2020-11-30T01:19:30
Spectrometer Frequency | (400.13, 400.13)
Spectral Width | (4000.0, 4000.0)
Lowest Frequency | (-158.1, -158.1)
Nucleus | (1H, 1H)
Acquired Size | (1024, 256)
Spectral Size | (1024, 1024)
Me
Me
HO
Me
Me
TBSO
H
H
Me
O
TBSO

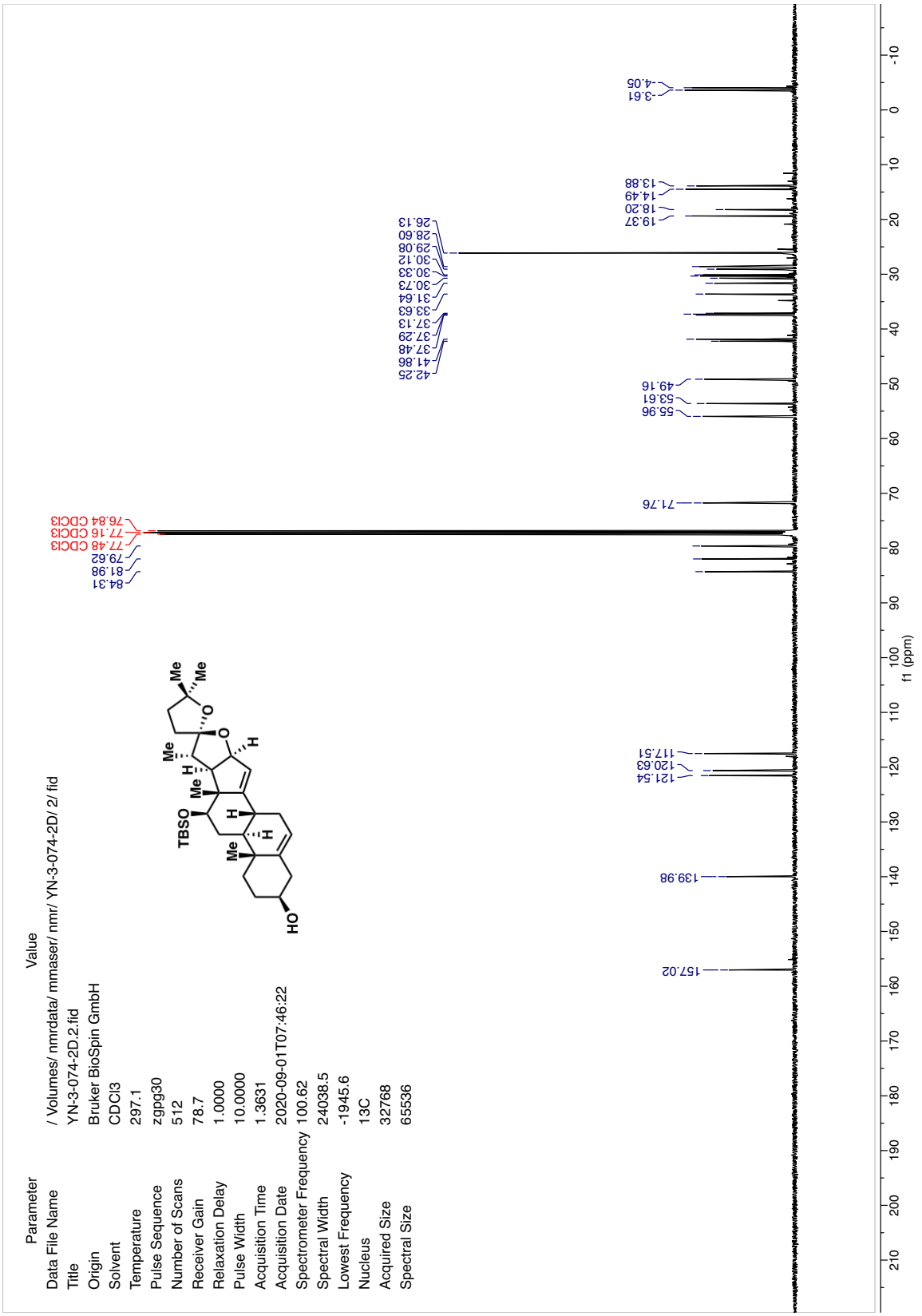

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ yasuaki/ vnmrsys/ data/ YN-3-026/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 8
Receiver Gain 30
Relaxation Delay 2.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2018-12-04T23:10:56
Spectrometer Frequency 499.63
Spectral Width 8000.0
Lowest Frequency -1014.5
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
Me
Me
HO
H
OH
Me
Me
TBSO
H
Me
H
TBSO
f1 (ppm)

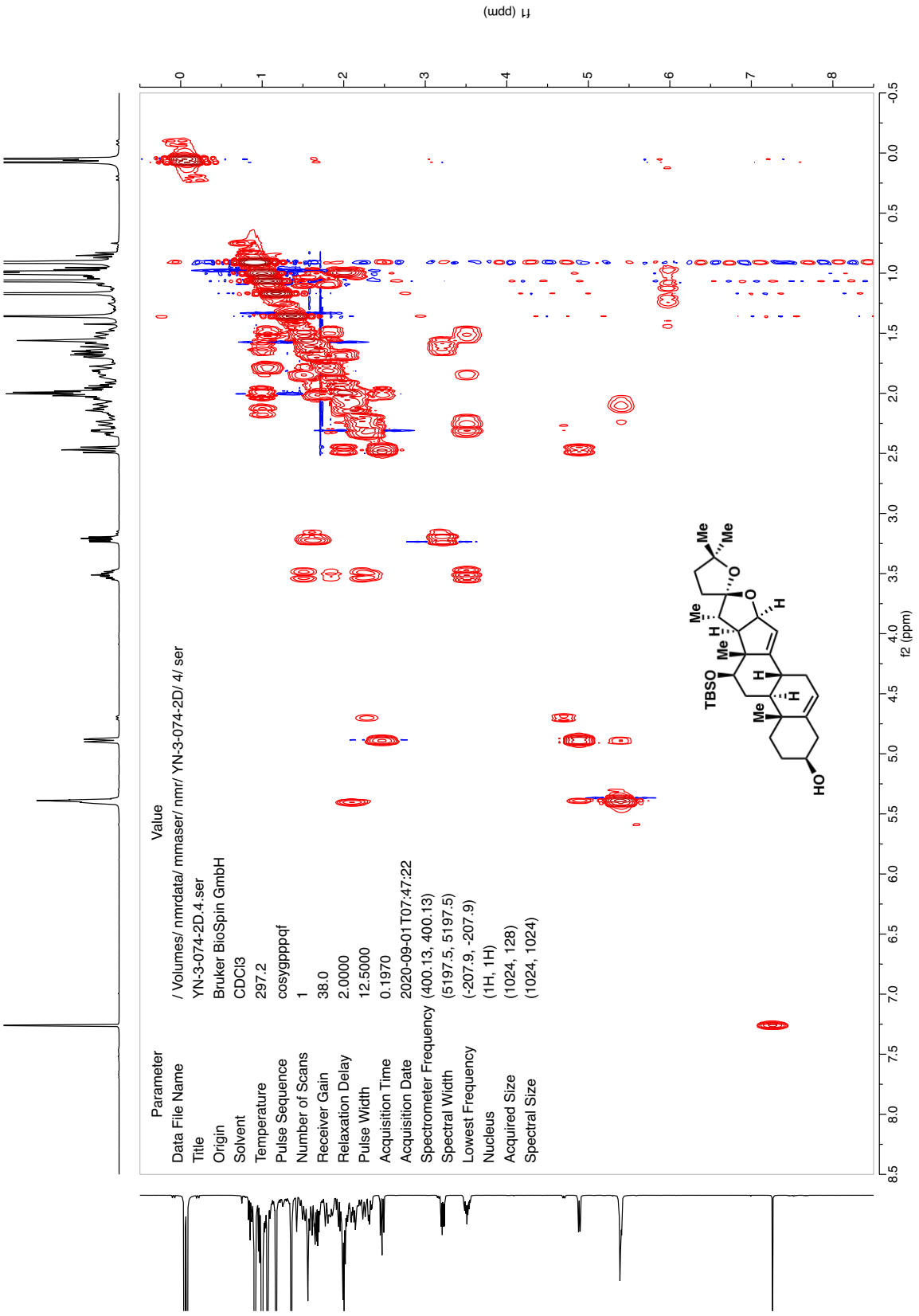

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-111/ 11/ fid
Title MRM-3-111.11.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 2048
Receiver Gain 43.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-12-03T08:04:40
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
Me
Me
HO
H
OH
Me
Me
TBSO
H
H
Me
TBSO
f1 (ppm)

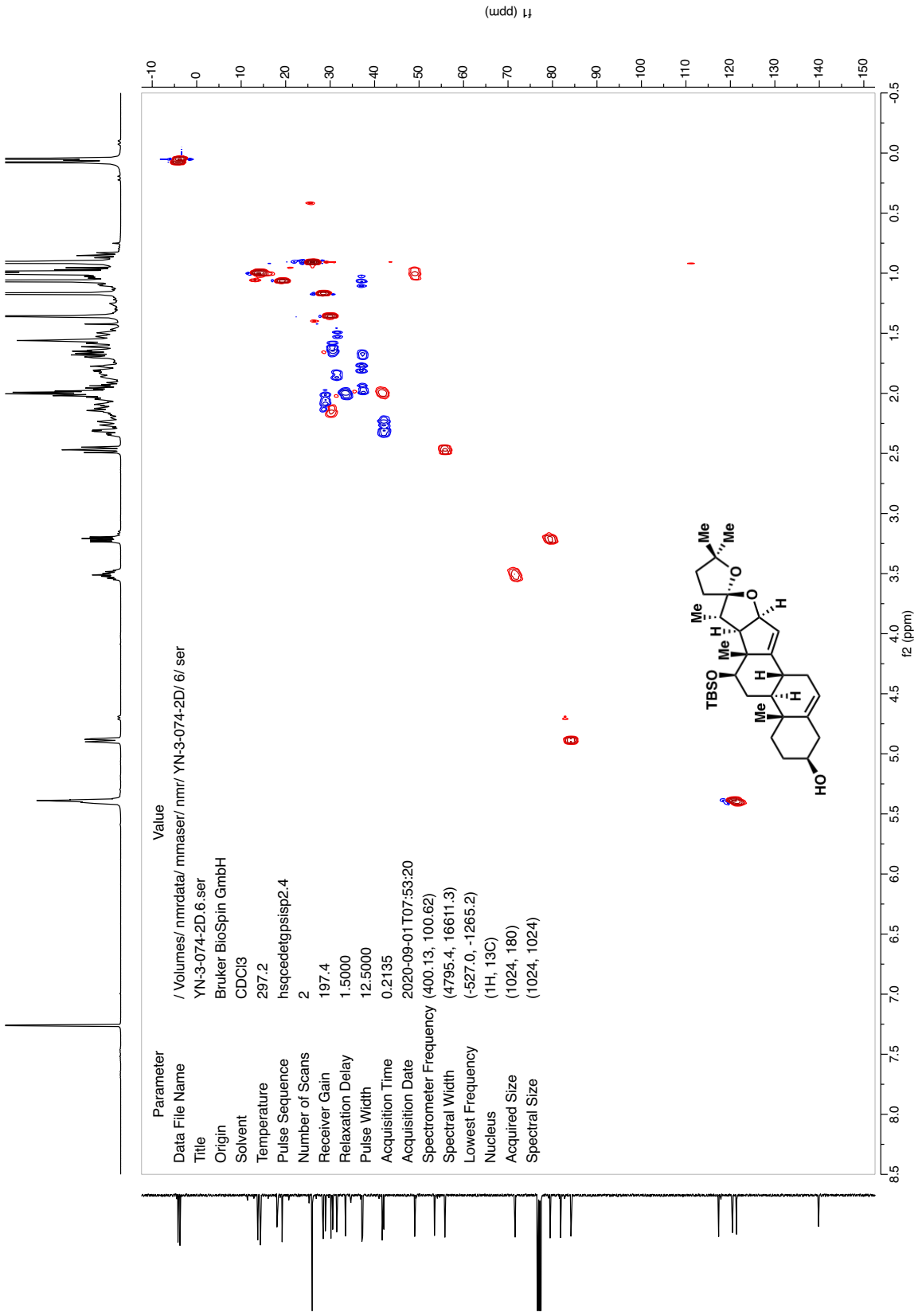

f1 (ppm)
f2 (ppm)
Parameter | Value
Data File Name | / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-111/ 4/ ser
Title | MRM-3-111.4.ser
Origin | Bruker BioSpin GmbH
Solvent | CDCl3
Temperature | 297.1
Pulse Sequence | cosygpppqf
Number of Scans | 1
Receiver Gain | 112.8
Relaxation Delay | 2.0000
Pulse Width | 12.5000
Acquisition Time | 0.1970
Acquisition Date | 2020-12-02T23:18:28
Spectrometer Frequency | (400.13, 400.13)
Spectral Width | (5197.5, 5197.5)
Lowest Frequency | (-198.0, -198.0)
Nucleus | (1H, 1H)
Acquired Size | (1024, 128)
Spectral Size | (1024, 1024)
Me
Me
HO
OH
H
Me
Me
TBSO
H
H
Me
TBSO

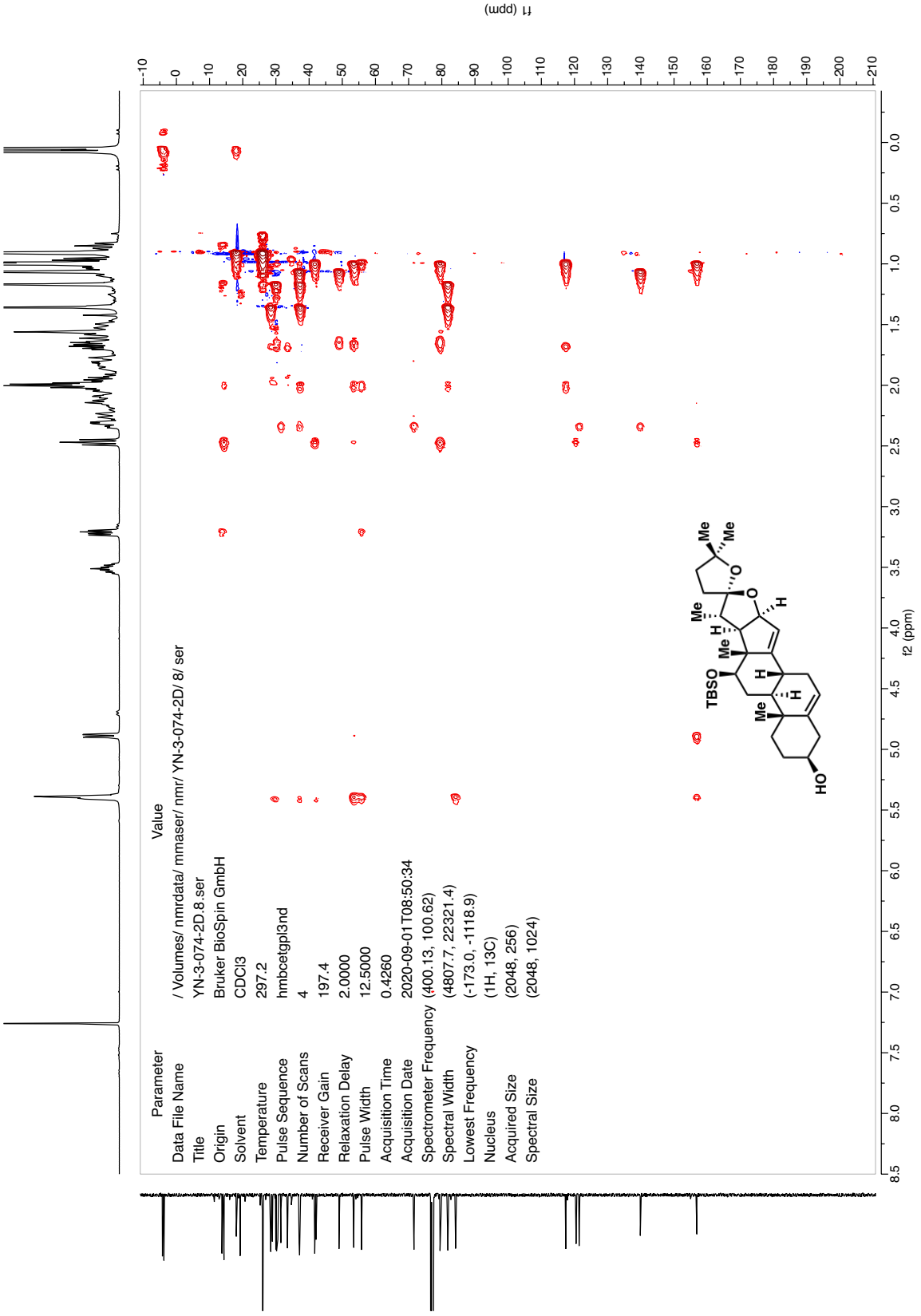

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-111/ 10/ ser
Title MRM-3-111.10.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence noesygpphzs
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 1.8000
Pulse Width 12.5000
Acquisition Time 0.2560
Acquisition Date 2020-12-03T00:20:31
Spectrometer Frequency (400.13, 400.13)
Spectral Width (4000.0, 4000.0)
Lowest Frequency (-158.1, -158.1)
Nucleus (1H, 1H)
Acquired Size (1024, 256)
Spectral Size (1024, 1024)
Me
Me
HO
OH
H
Me
Me
TBSO
H
H
Me
H
TBSO

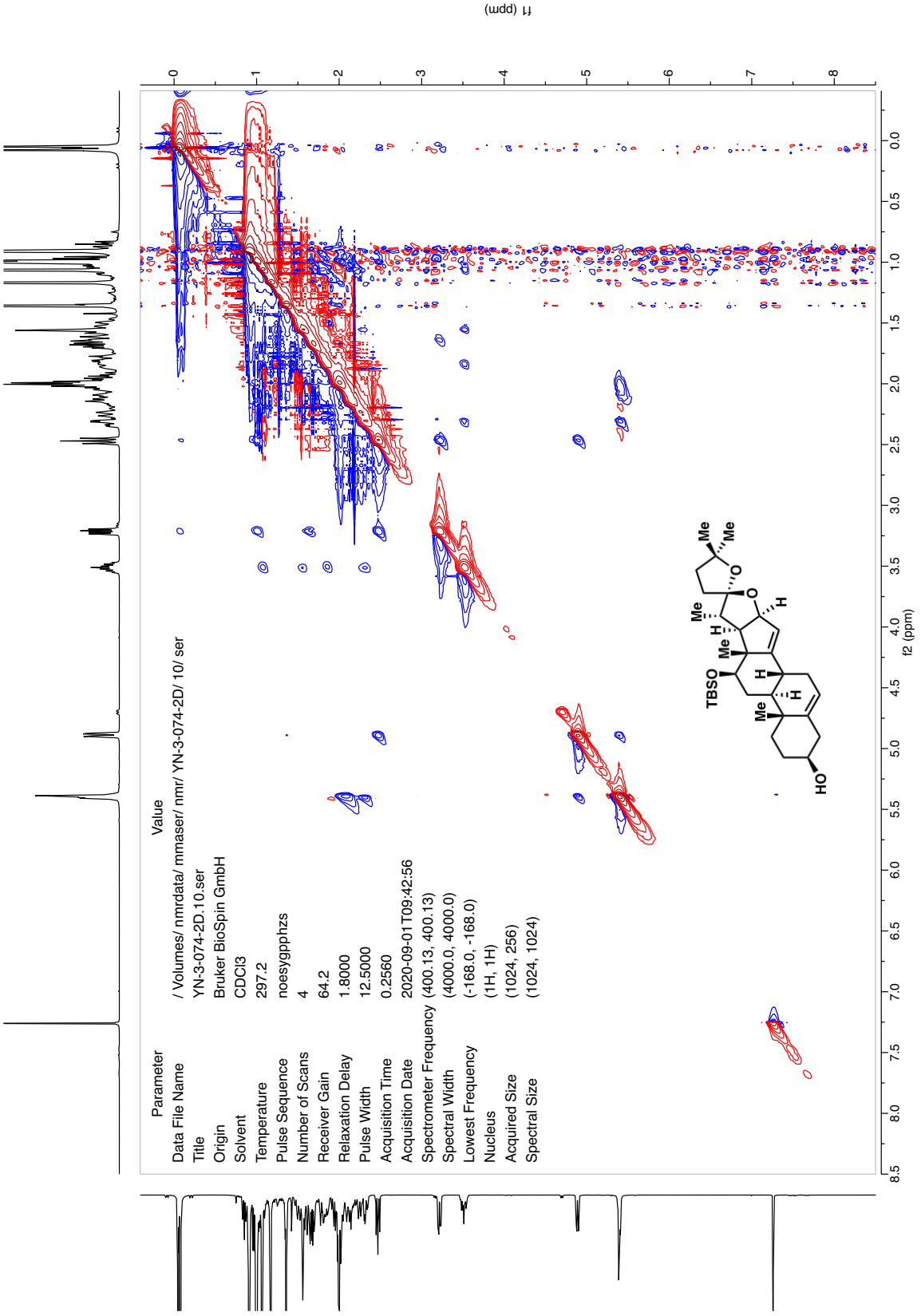

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-074-2D/ 1/ fid
Title YN-3-074-2D.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-09-01T07:25:08
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.5
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
Me
Me
Me
Me
H
H
H
H
O
O
HO
f1 (ppm)

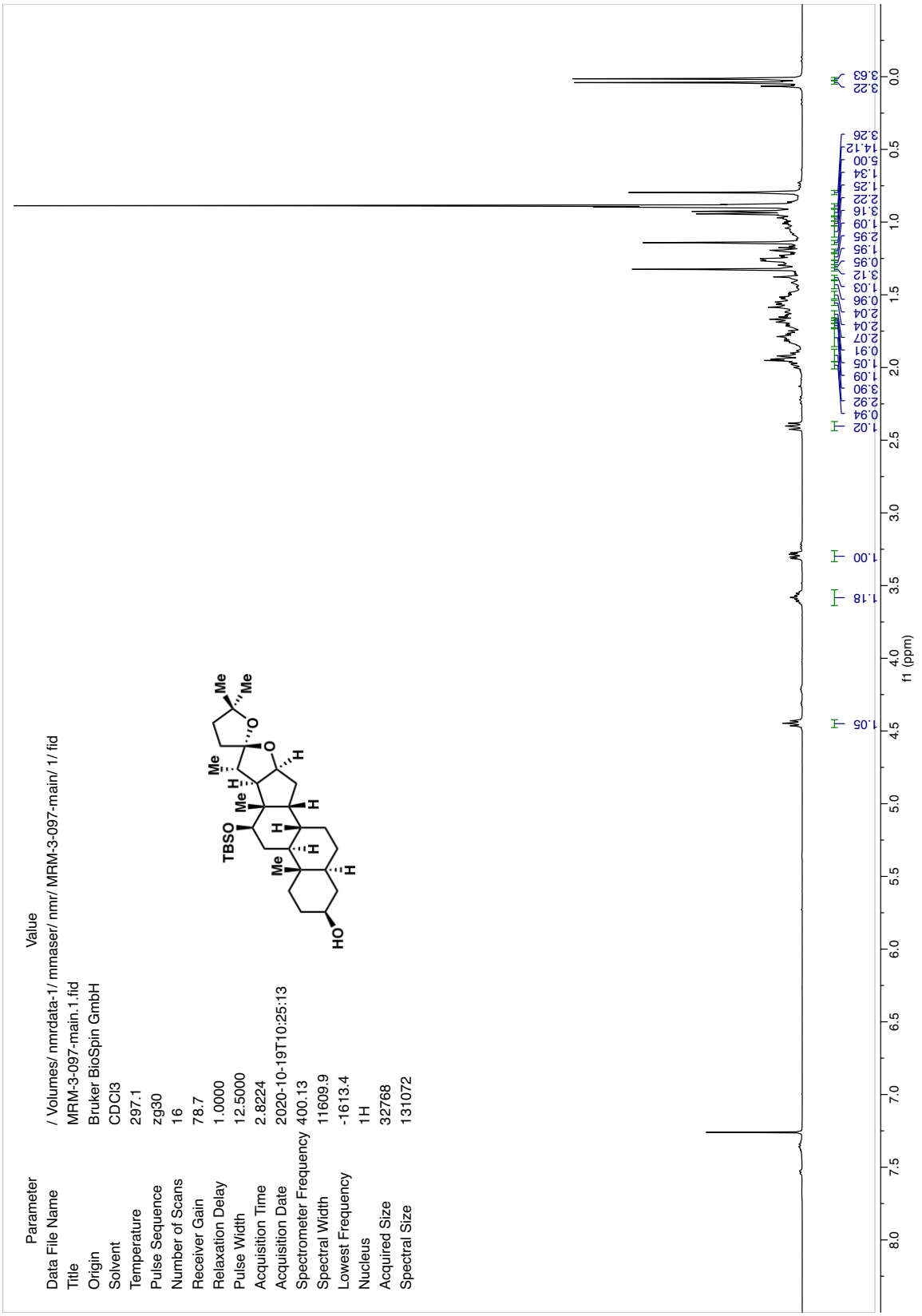

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-074-2D/ 2/ fid
Title YN-3-074-2D.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-09-01T07:46:22
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.6
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
HO
Me
H
O
84.31
81.98
79.62
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
157.02
139.98
121.54
120.63
117.51
71.76
55.96
53.61
49.16
42.25
41.86
37.48
37.29
37.13
33.63
31.64
30.73
30.33
30.12
29.08
28.60
26.13
19.37
18.20
14.49
13.88
-3.61
-4.05
f1 (ppm)

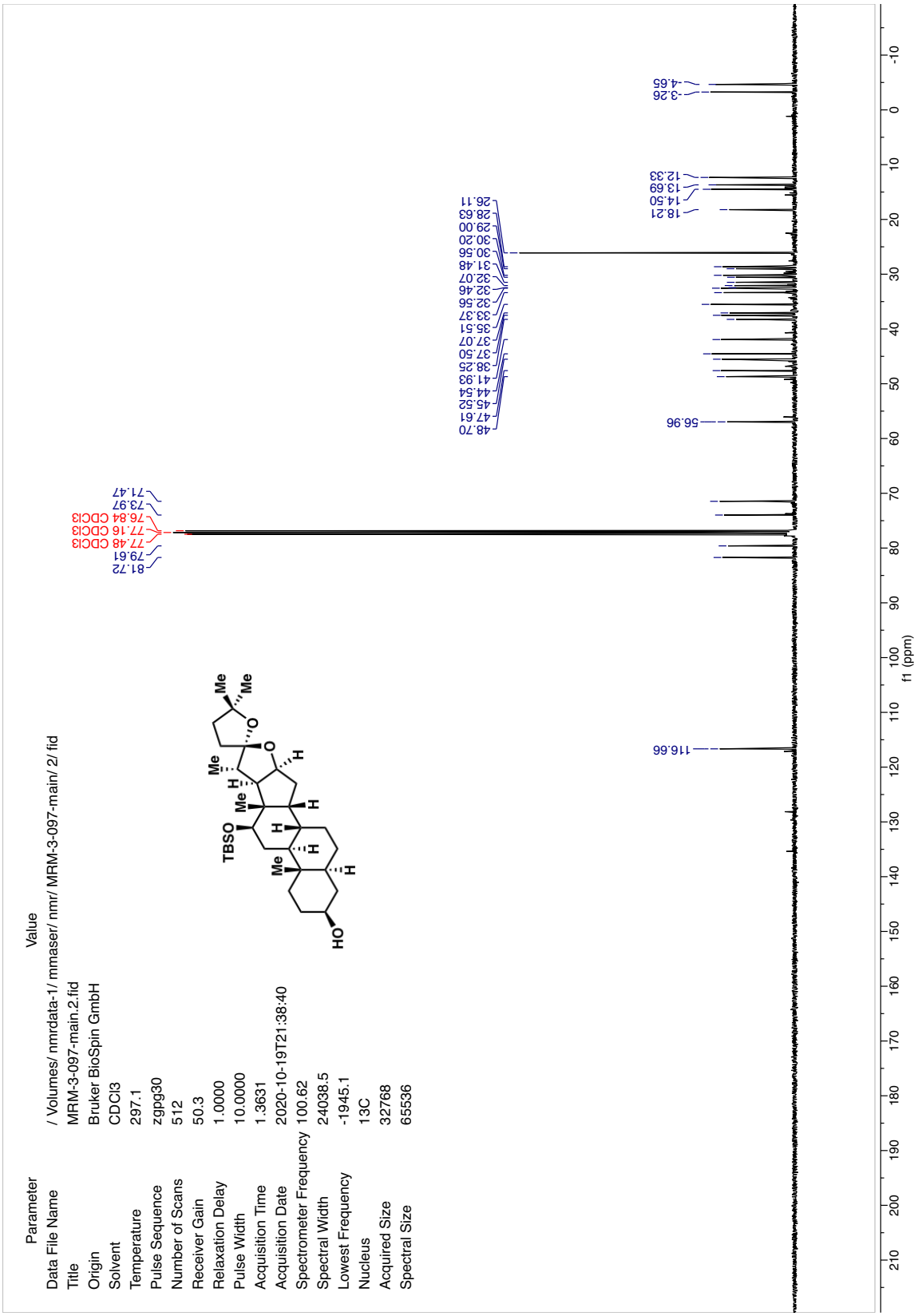

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-074-2D/ 4/ ser
Title YN-3-074-2D.4.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence cosygpppqf
Number of Scans 1
Receiver Gain 38.0
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.1970
Acquisition Date 2020-09-01T07:47:22
Spectrometer Frequency (400.13, 400.13)
Spectral Width (5197.5, 5197.5)
Lowest Frequency (-207.9, -207.9)
Nucleus (1H, 1H)
Acquired Size (1024, 128)
Spectral Size (1024, 1024)
Me
Me
O
O
H
Me
H
Me
TBSO
H
H
Me
HO

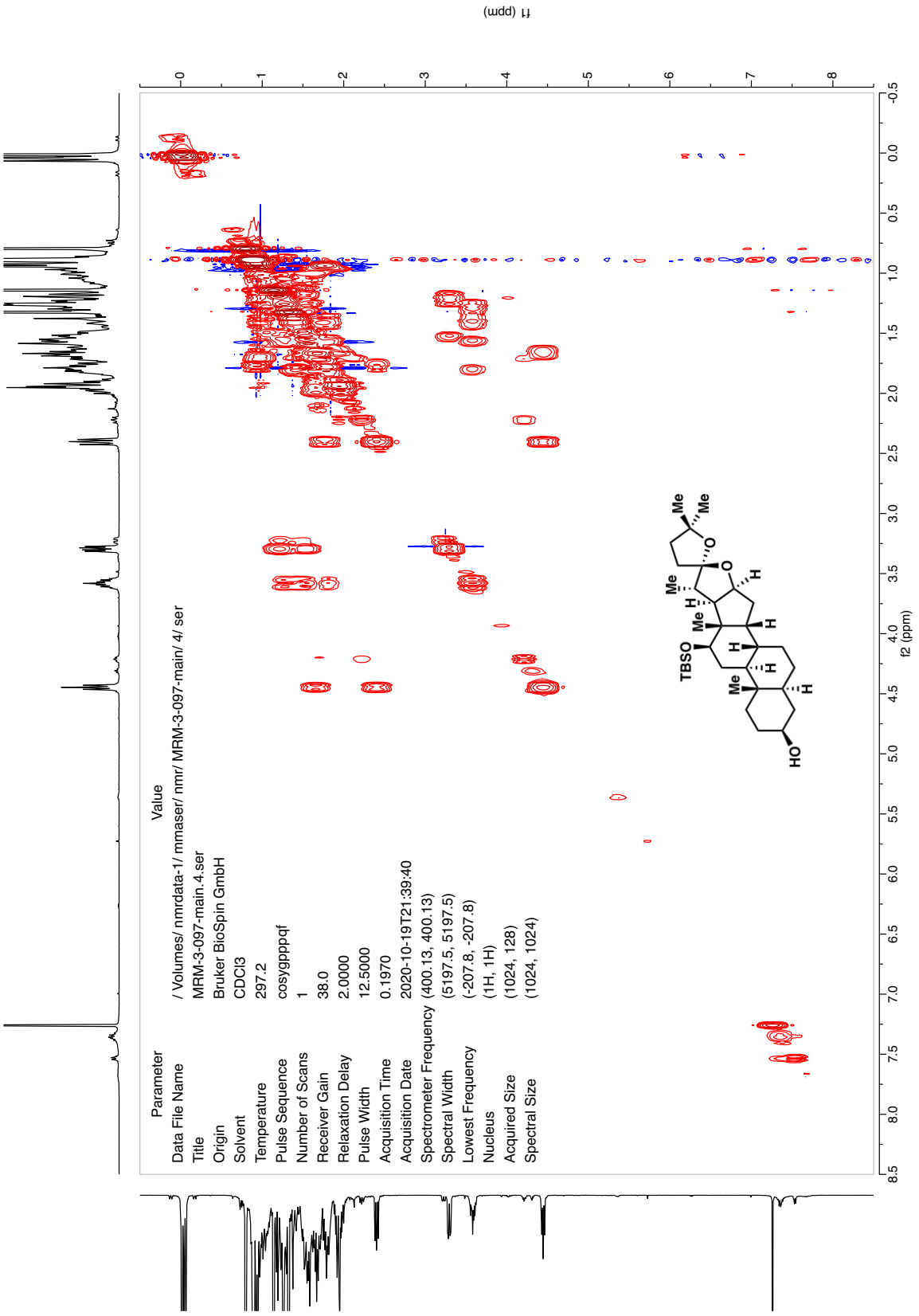

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-074-2D/ 6/ ser
Title YN-3-074-2D.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-09-01T07:53:20
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-527.0, -1265.2)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
HO
Me
H

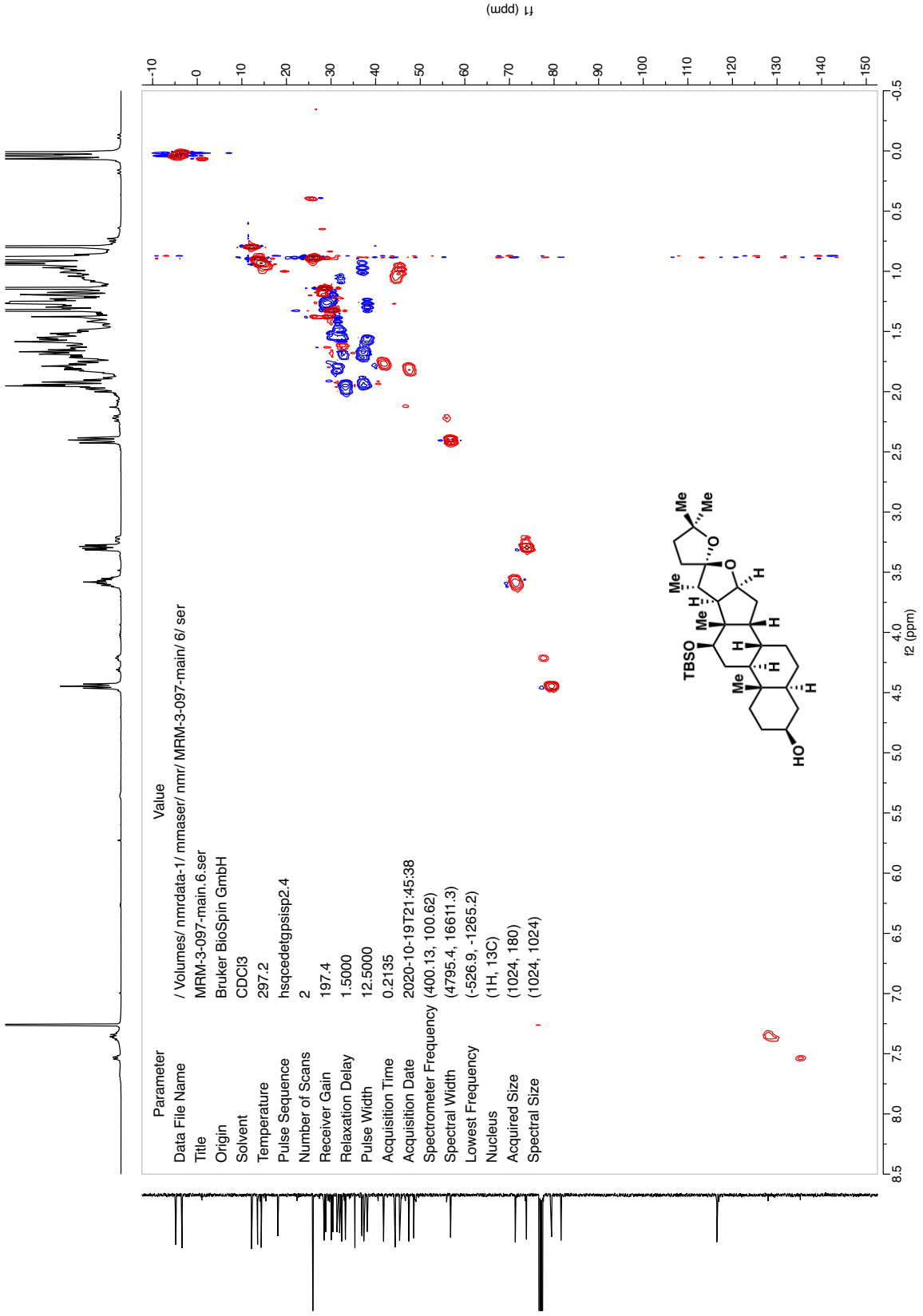

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-074-2D/ 8/ ser
Title YN-3-074-2D.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-09-01T08:50:34
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-173.0, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
Me
H
HO

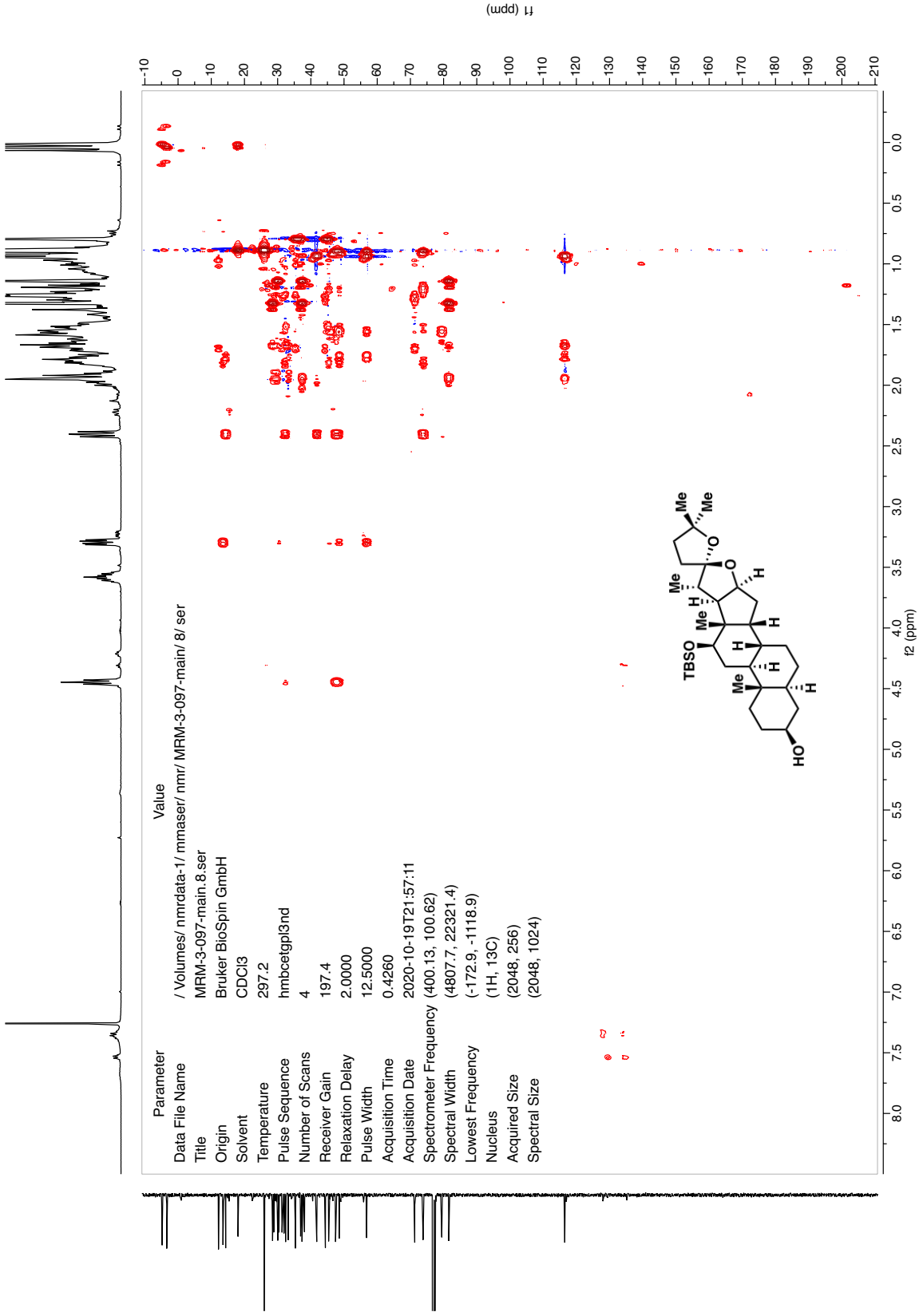

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ YN-3-074-2D/ 10/ ser
Title YN-3-074-2D.10.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence noesygpphzs
Number of Scans 4
Receiver Gain 64.2
Relaxation Delay 1.8000
Pulse Width 12.5000
Acquisition Time 0.2560
Acquisition Date 2020-09-01T09:42:56
Spectrometer Frequency (400.13, 400.13)
Spectral Width (4000.0, 4000.0)
Lowest Frequency (-168.0, -168.0)
Nucleus (1H, 1H)
Acquired Size (1024, 256)
Spectral Size (1024, 1024)
TBSO
HO
Me
H

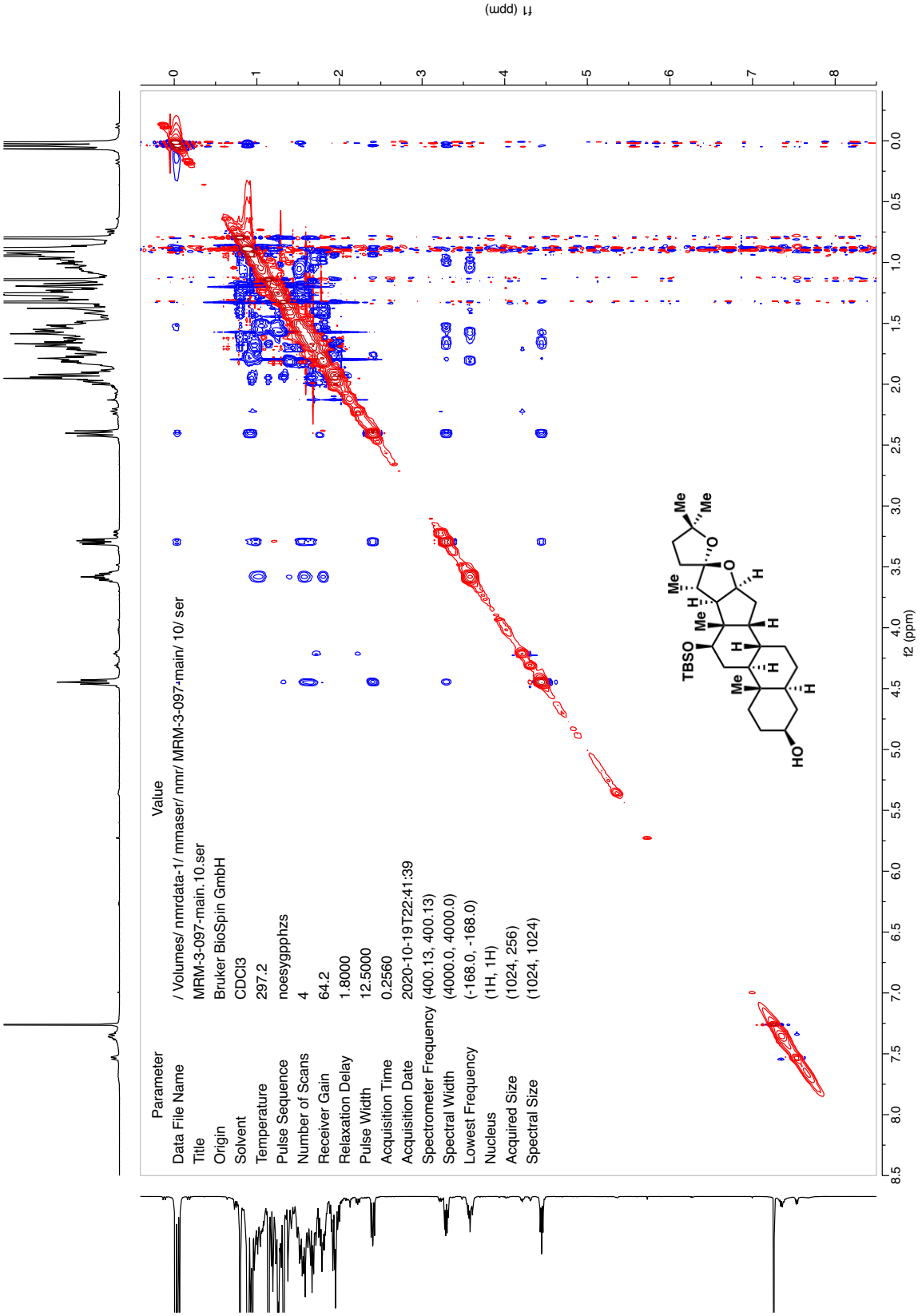

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-097-main/ 1/ fid
Title MRM-3-097-main.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-10-19T10:25:13
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.4
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
Me
HO
f1 (ppm)

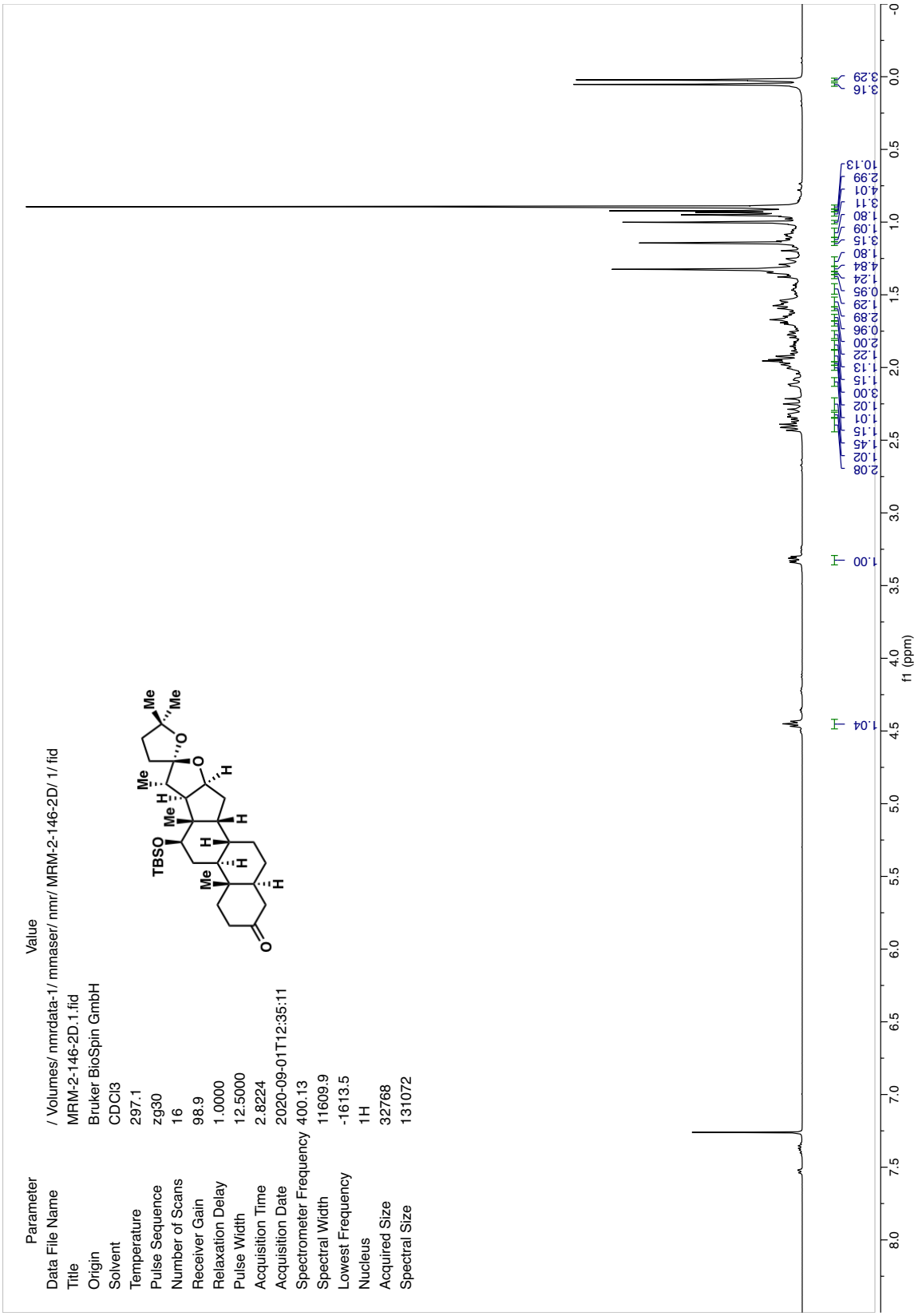

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-097-main/ 2/ fid
Title MRM-3-097-main.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-10-19T21:38:40
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.1
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
HO
Me
H
O
81.72
79.61
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
73.97
71.47
56.96
48.70
47.61
45.52
44.54
41.93
38.25
37.50
37.07
35.51
33.37
32.56
32.46
32.07
31.48
30.56
30.20
29.00
28.63
26.11
18.21
14.50
13.69
12.33
-3.26
-4.65
116.66
f1 (ppm)

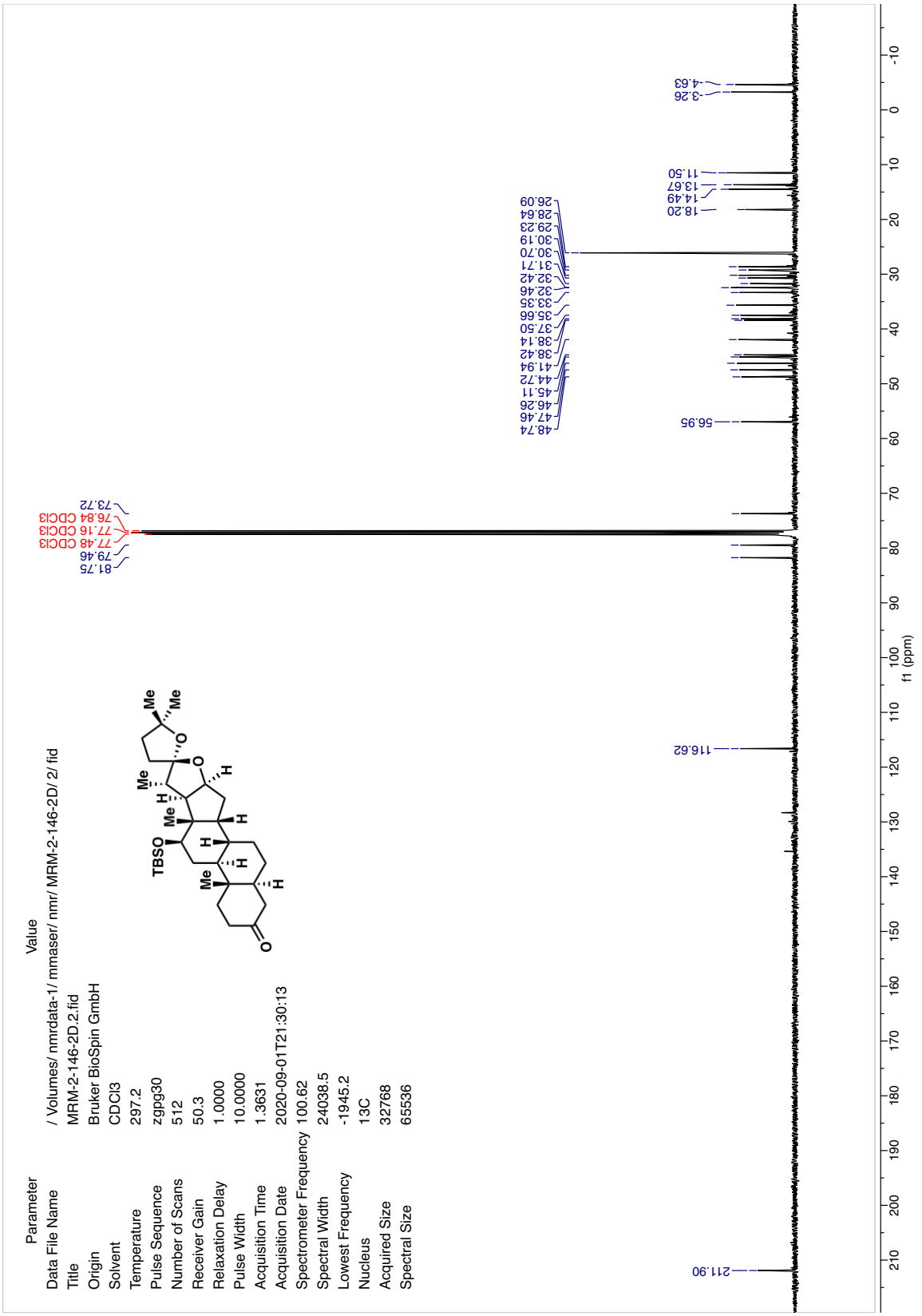

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-097-main/ 4/ ser
Title MRM-3-097-main.4.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence cosygpppqf
Number of Scans 1
Receiver Gain 38.0
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.1970
Acquisition Date 2020-10-19T21:39:40
Spectrometer Frequency (400.13, 400.13)
Spectral Width (5197.5, 5197.5)
Lowest Frequency (-207.8, -207.8)
Nucleus (1H, 1H)
Acquired Size (1024, 128)
Spectral Size (1024, 1024)
TBSO
HO
Me
H

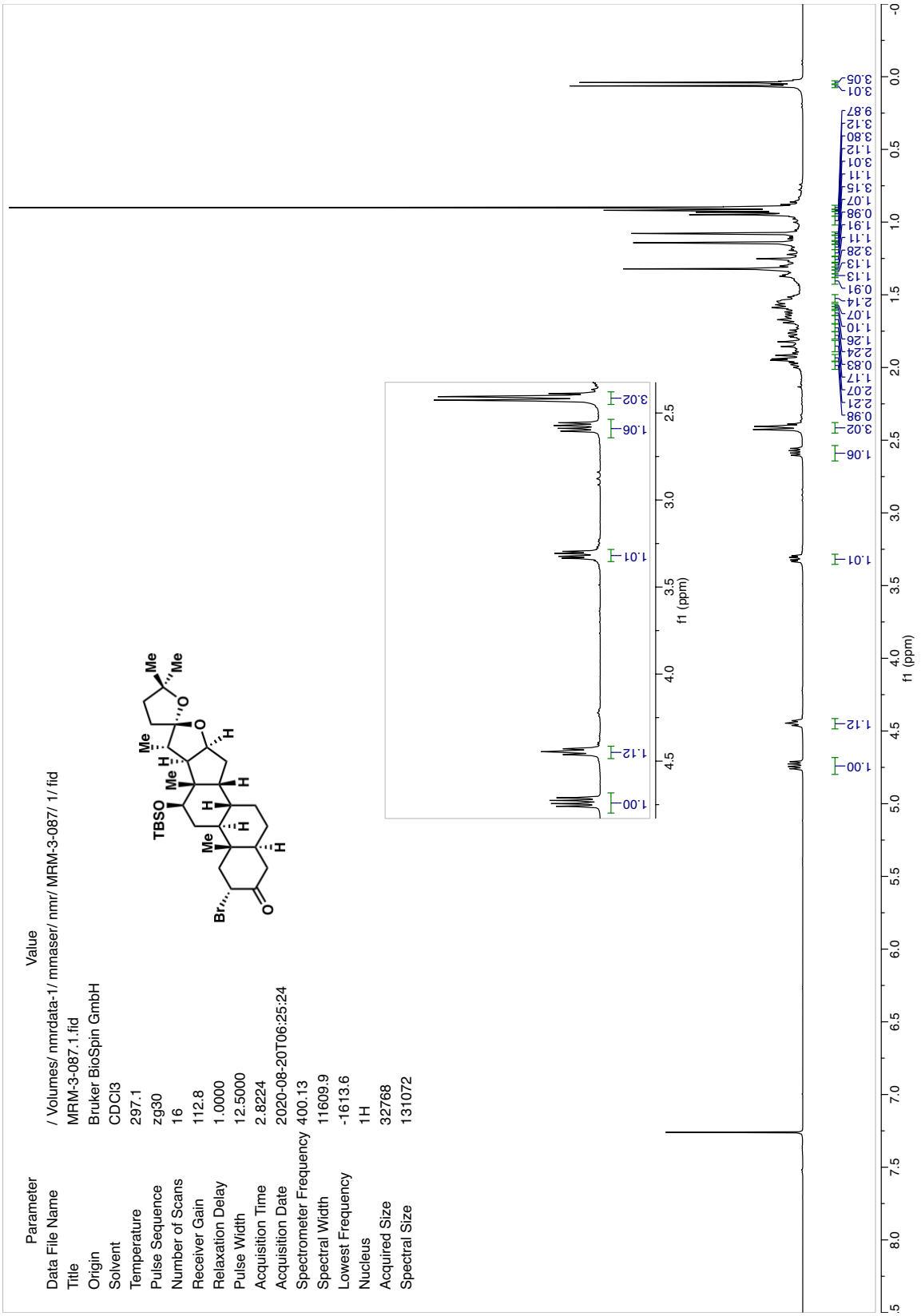

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-097-main/ 6/ ser
Title MRM-3-097-main.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-10-19T21:45:38
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-526.9, -1265.2)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
HO
Me
H

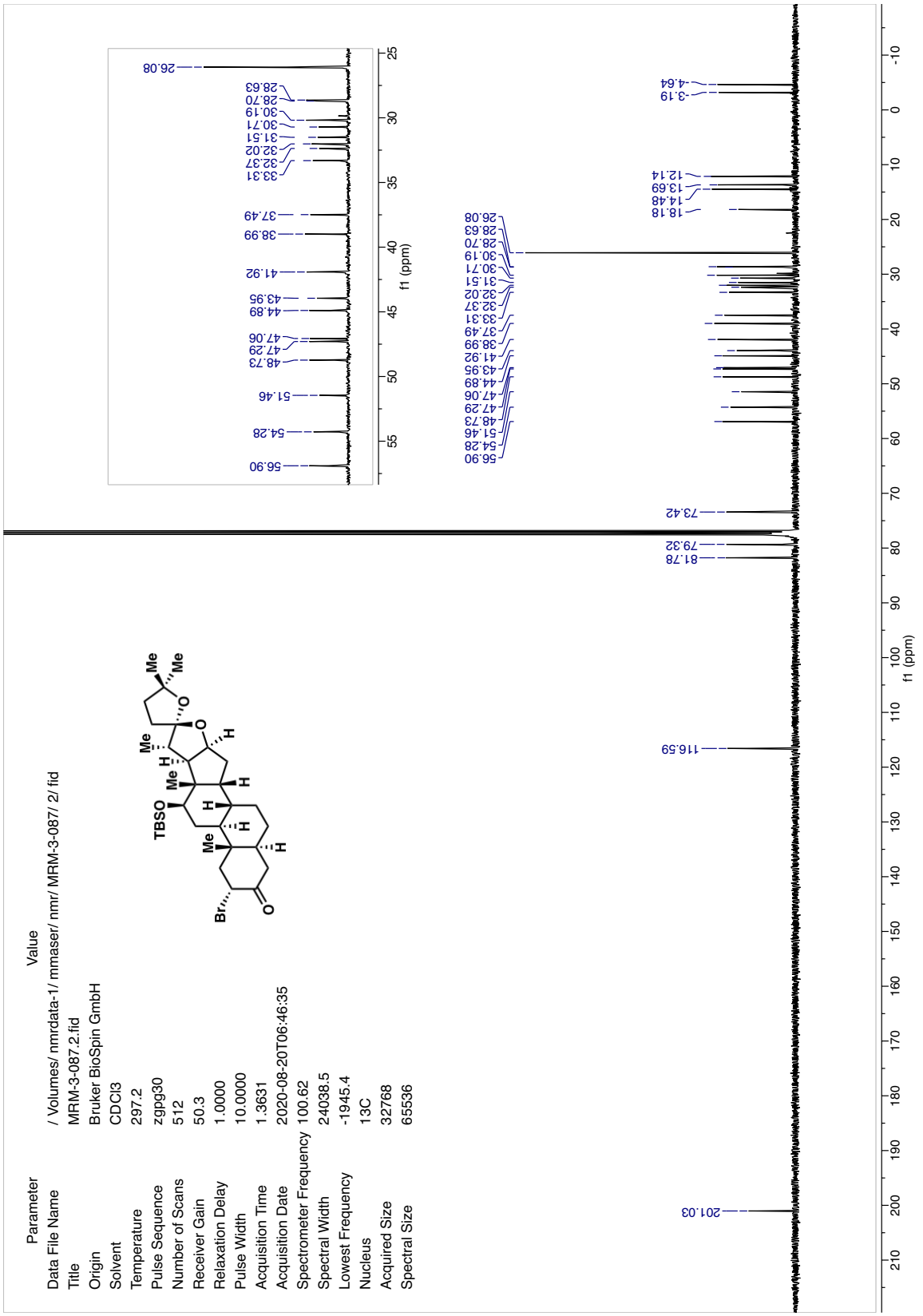

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-097-main/ 8/ ser
Title MRM-3-097-main.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-10-19T21:57:11
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-172.9, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
HO
Me
H

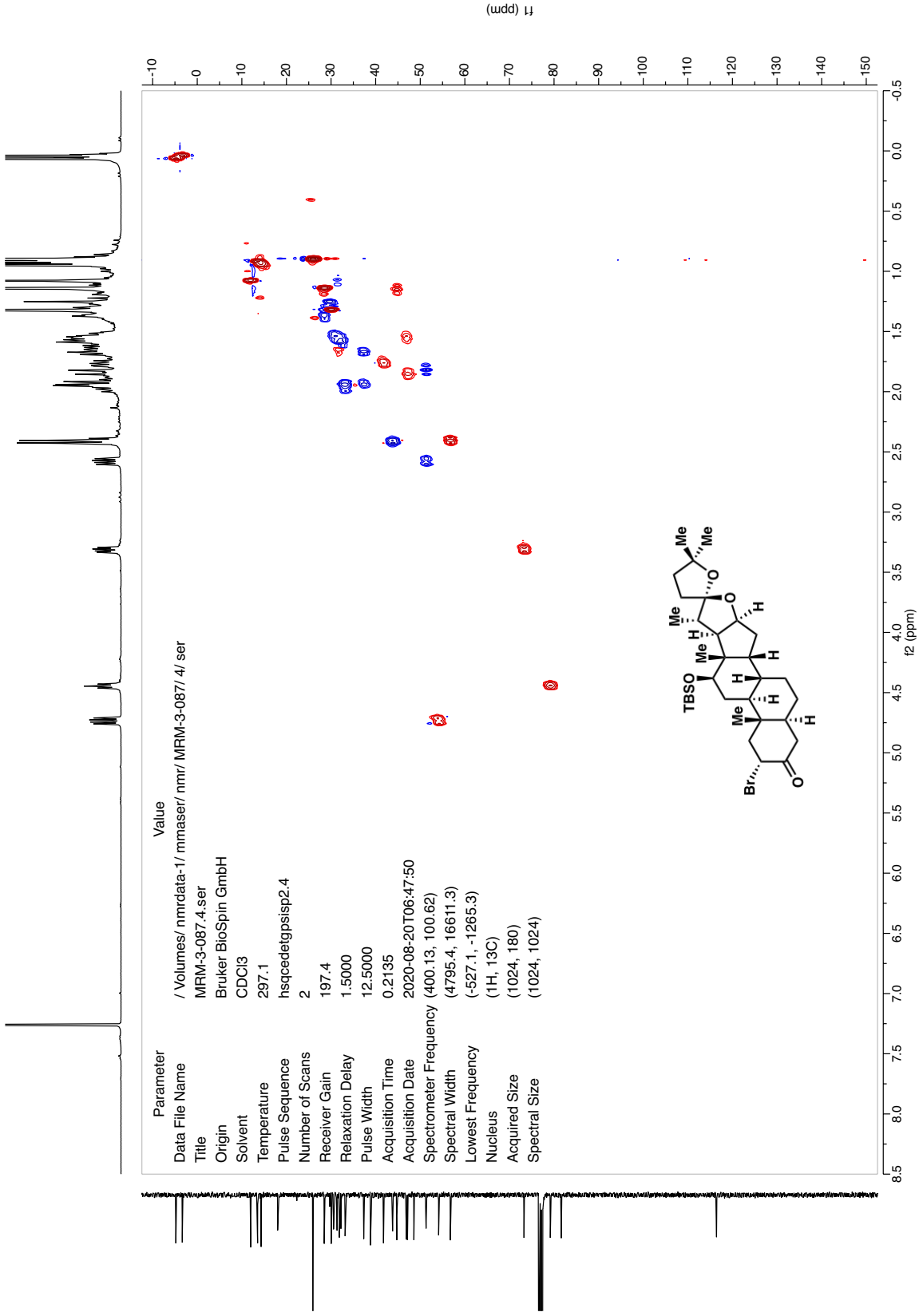

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-097-main/ 10/ ser
Title MRM-3-097-main.10.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence noesygpphzs
Number of Scans 4
Receiver Gain 64.2
Relaxation Delay 1.8000
Pulse Width 12.5000
Acquisition Time 0.2560
Acquisition Date 2020-10-19T22:41:39
Spectrometer Frequency (400.13, 400.13)
Spectral Width (4000.0, 4000.0)
Lowest Frequency (-168.0, -168.0)
Nucleus (1H, 1H)
Acquired Size (1024, 256)
Spectral Size (1024, 1024)
Me
Me
O
O
H
Me
H
Me
H
TBSO
H
H
Me
H
HO

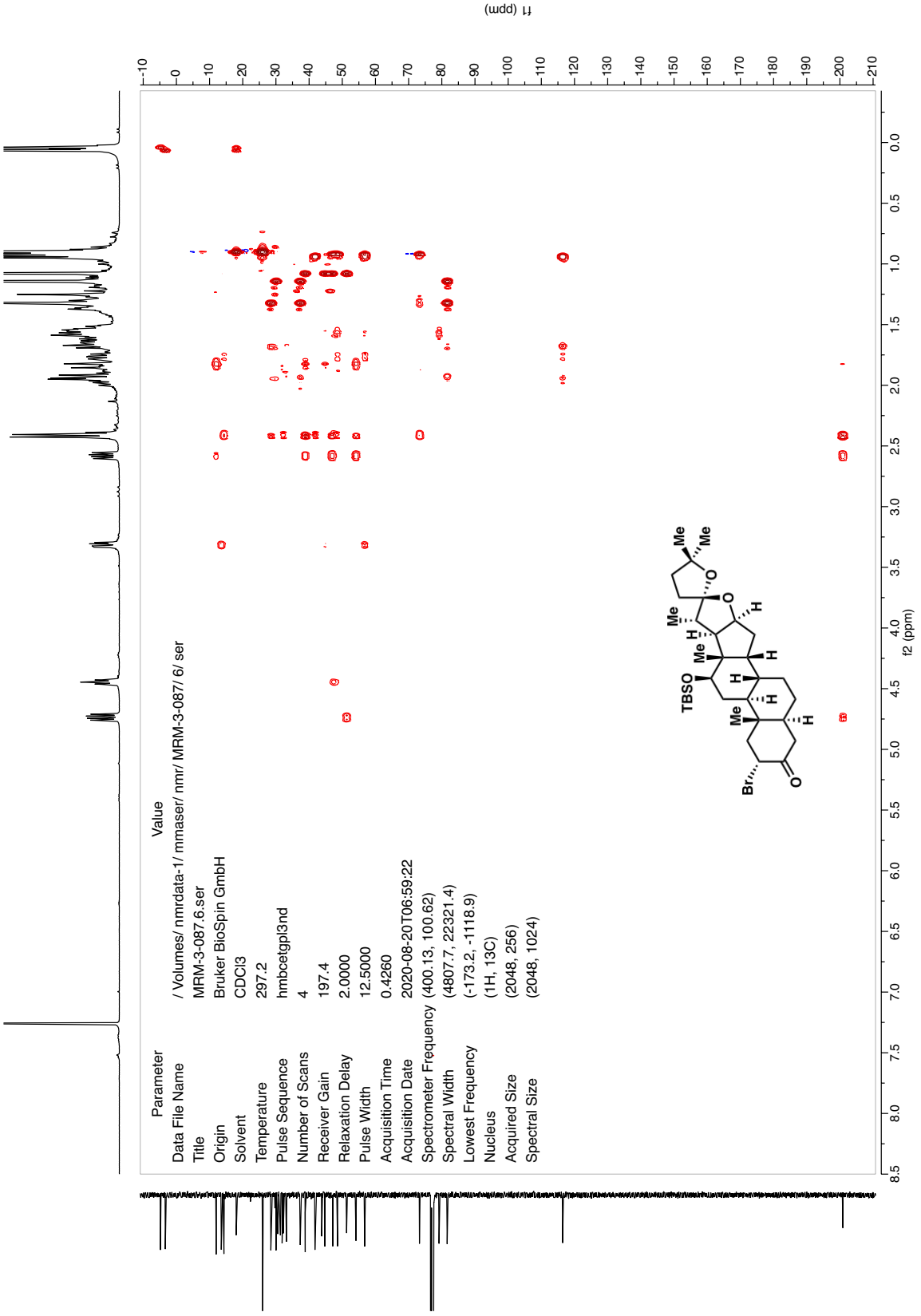

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-2-146-2D/ 1/ fid
Title MRM-2-146-2D.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 98.9
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-09-01T12:35:11
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.5
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
Me
H
O
1.04
1.00
3.16
3.29
f1 (ppm)

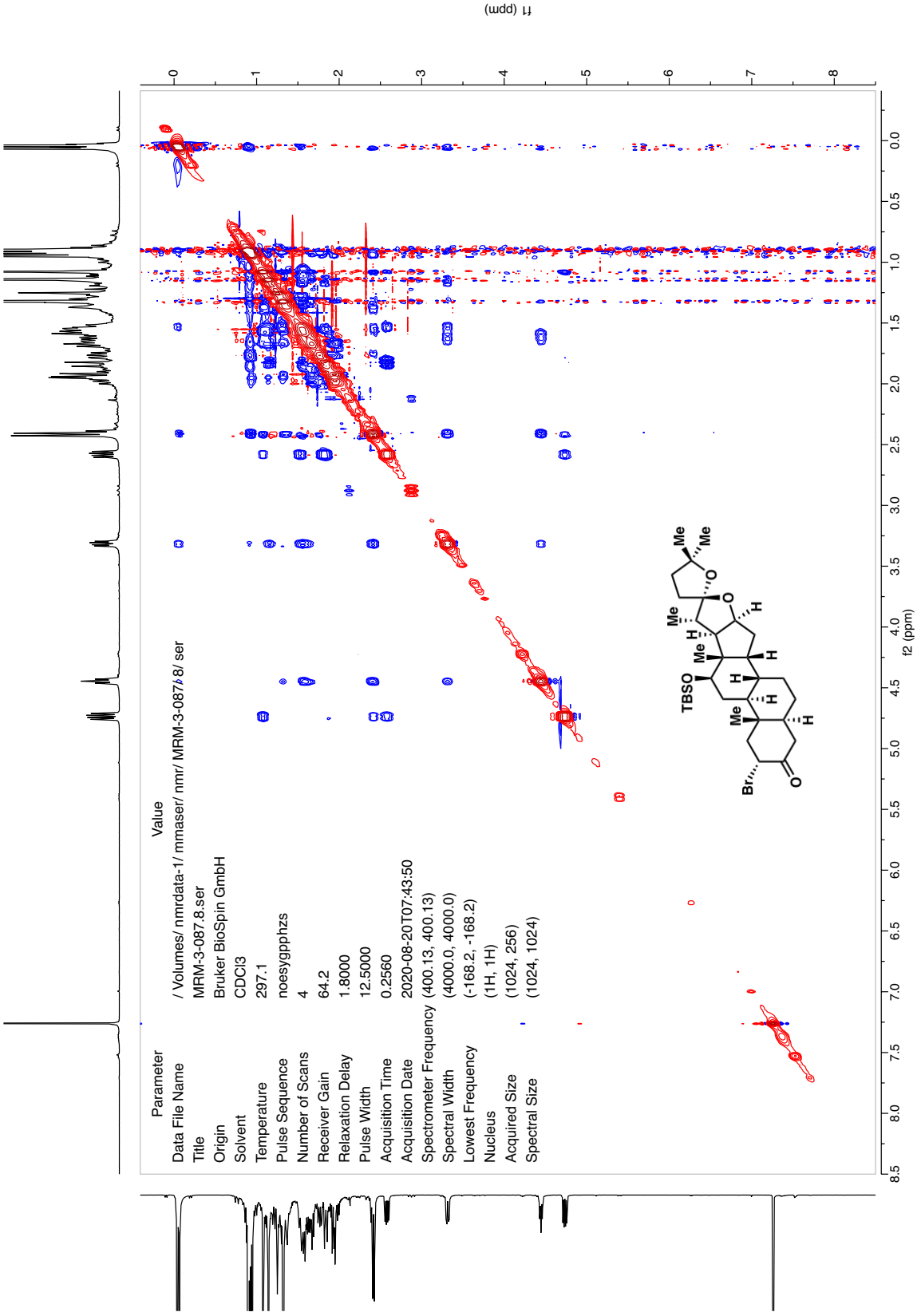

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-2-146-2D/ 2/ fid
Title MRM-2-146-2D.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-09-01T21:30:13
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.2
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
H
O
81.75
79.46
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
73.72
56.95
48.74
47.46
46.26
45.11
44.72
41.94
38.42
38.14
37.50
35.66
33.35
32.46
32.42
31.71
30.70
30.19
29.23
28.64
26.09
18.20
14.49
13.67
11.50
-3.26
-4.63
116.62
211.90
f1 (ppm)

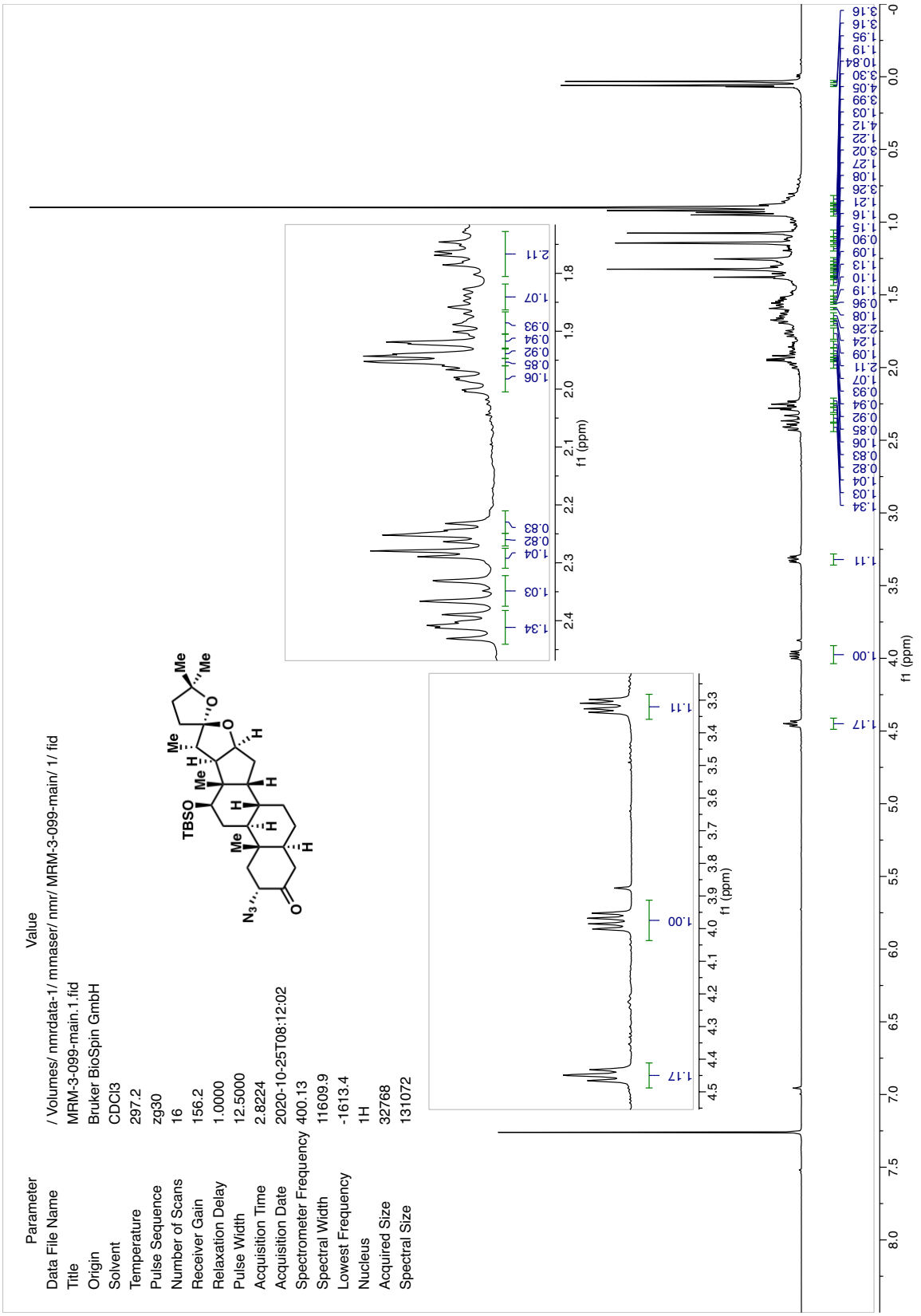

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-087/ 1/ fid
Title MRM-3-087.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 112.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-08-20T06:25:24
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.6
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
Me
Br
O
H
f1 (ppm)
3.02
1.06
1.01
1.12
1.00

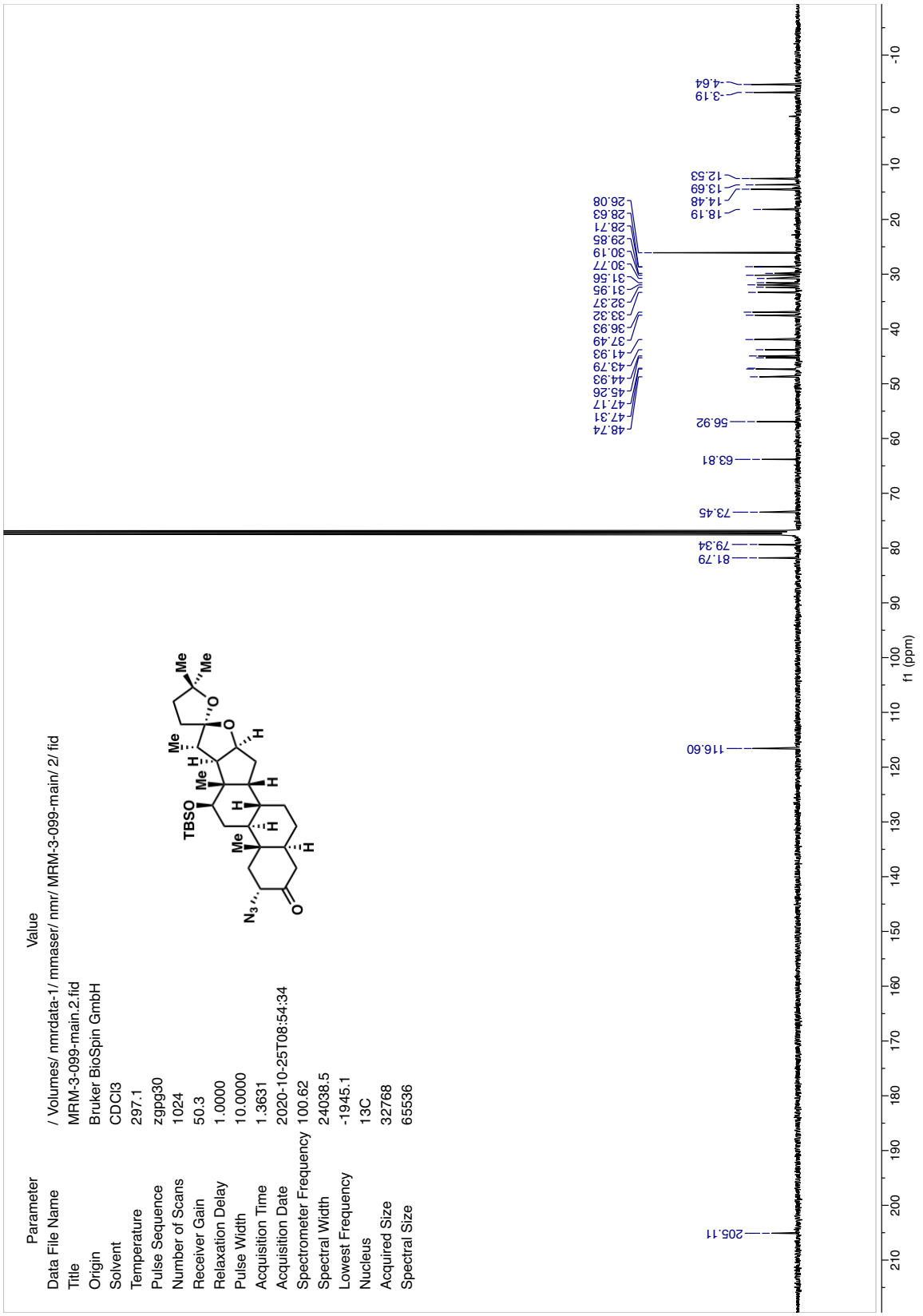

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-087/ 2/ fid
Title MRM-3-087.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-08-20T06:46:35
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
Br
O
H
f1 (ppm)

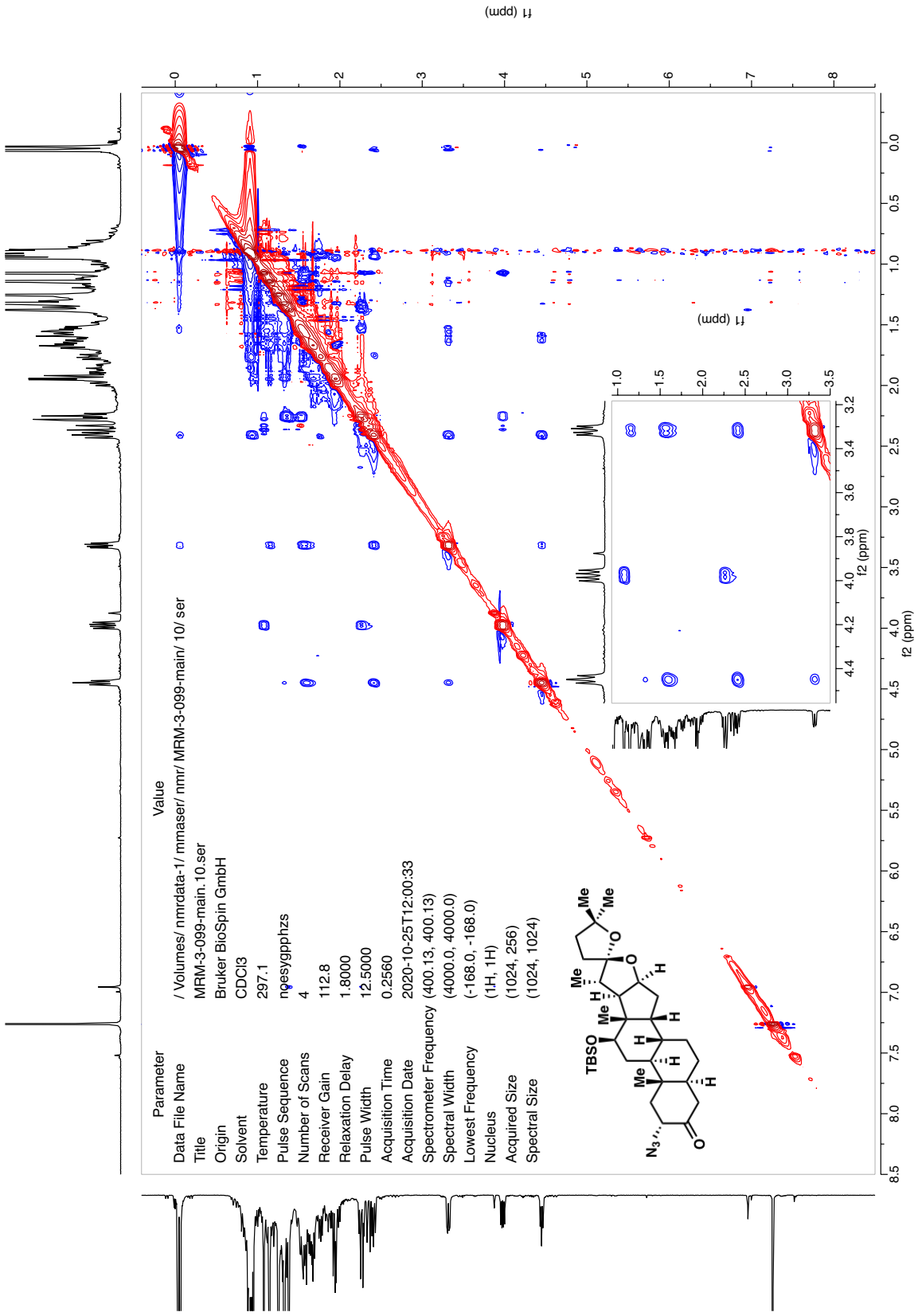

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-087/ 4/ ser
Title MRM-3-087.4.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-08-20T06:47:50
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-527.1, -1265.3)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
Me
Br
O
H

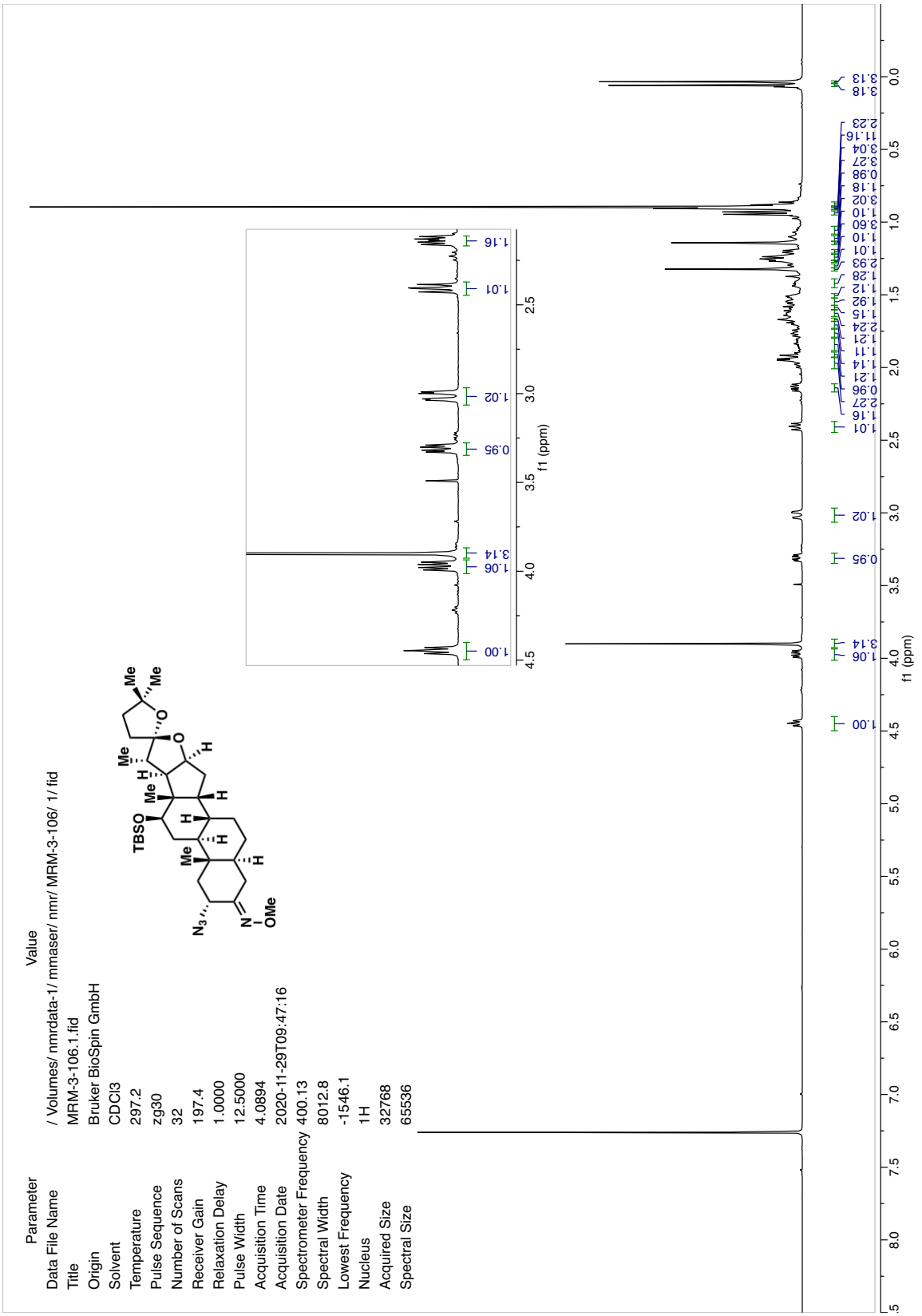

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-087/ 6/ ser
Title MRM-3-087.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-08-20T06:59:22
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-173.2, -1118.9)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
Me
Me
Me
Me
H
H
H
H
H
H
Br
O
O
O

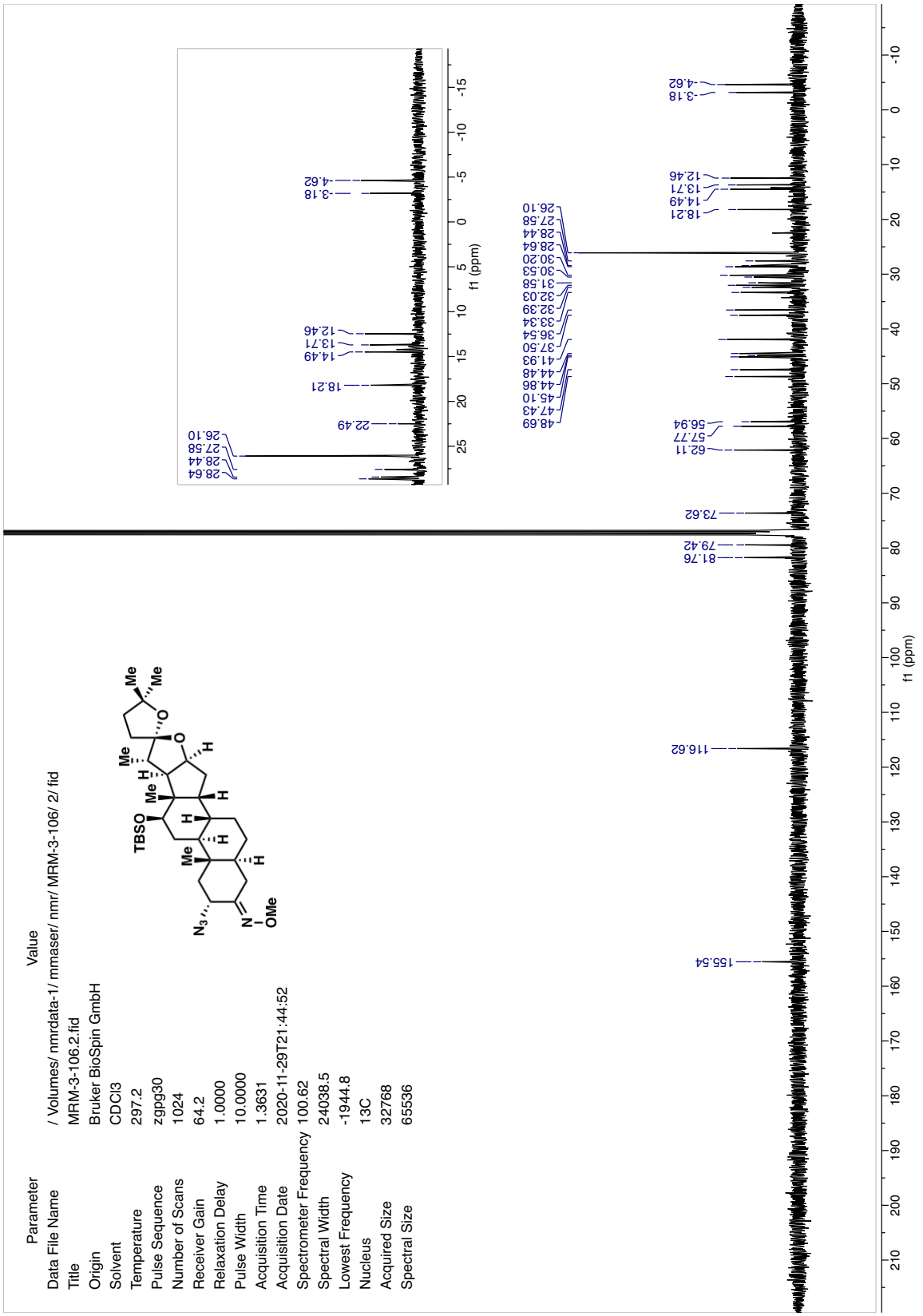

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-087/8/ ser
Title MRM-3-087.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence noesygpphzs
Number of Scans 4
Receiver Gain 64.2
Relaxation Delay 1.8000
Pulse Width 12.5000
Acquisition Time 0.2560
Acquisition Date 2020-08-20T07:43:50
Spectrometer Frequency (400.13, 400.13)
Spectral Width (4000.0, 4000.0)
Lowest Frequency (-168.2, -168.2)
Nucleus (1H, 1H)
Acquired Size (1024, 256)
Spectral Size (1024, 1024)
TBSO
Me
H
Br
O

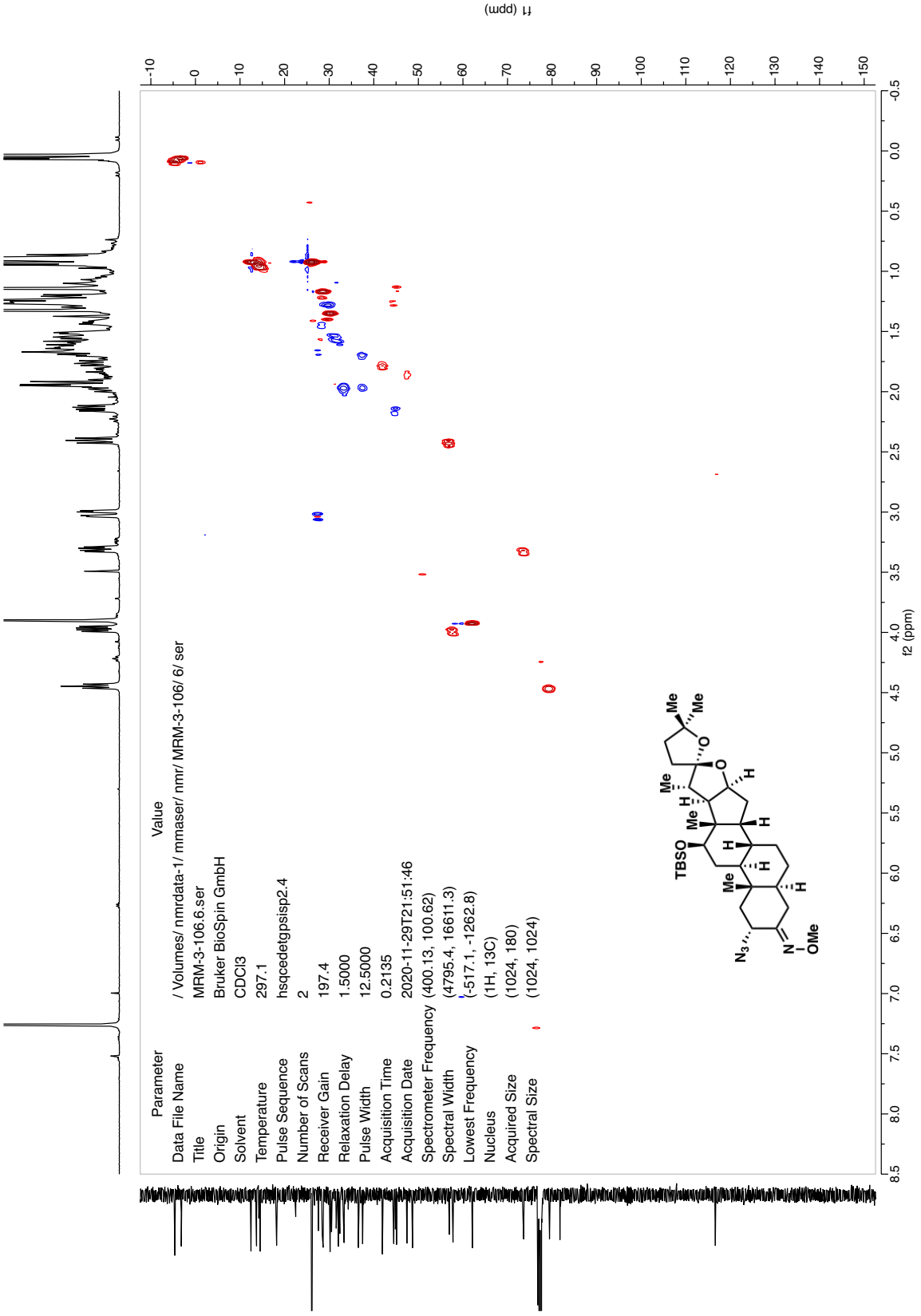

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-099-main/ 1/ fid
Title MRM-3-099-main.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 2.8224
Acquisition Date 2020-10-25T08:12:02
Spectrometer Frequency 400.13
Spectral Width 11609.9
Lowest Frequency -1613.4
Nucleus 1H
Acquired Size 32768
Spectral Size 131072
TBSO
Me
H
O
N3
f1 (ppm)

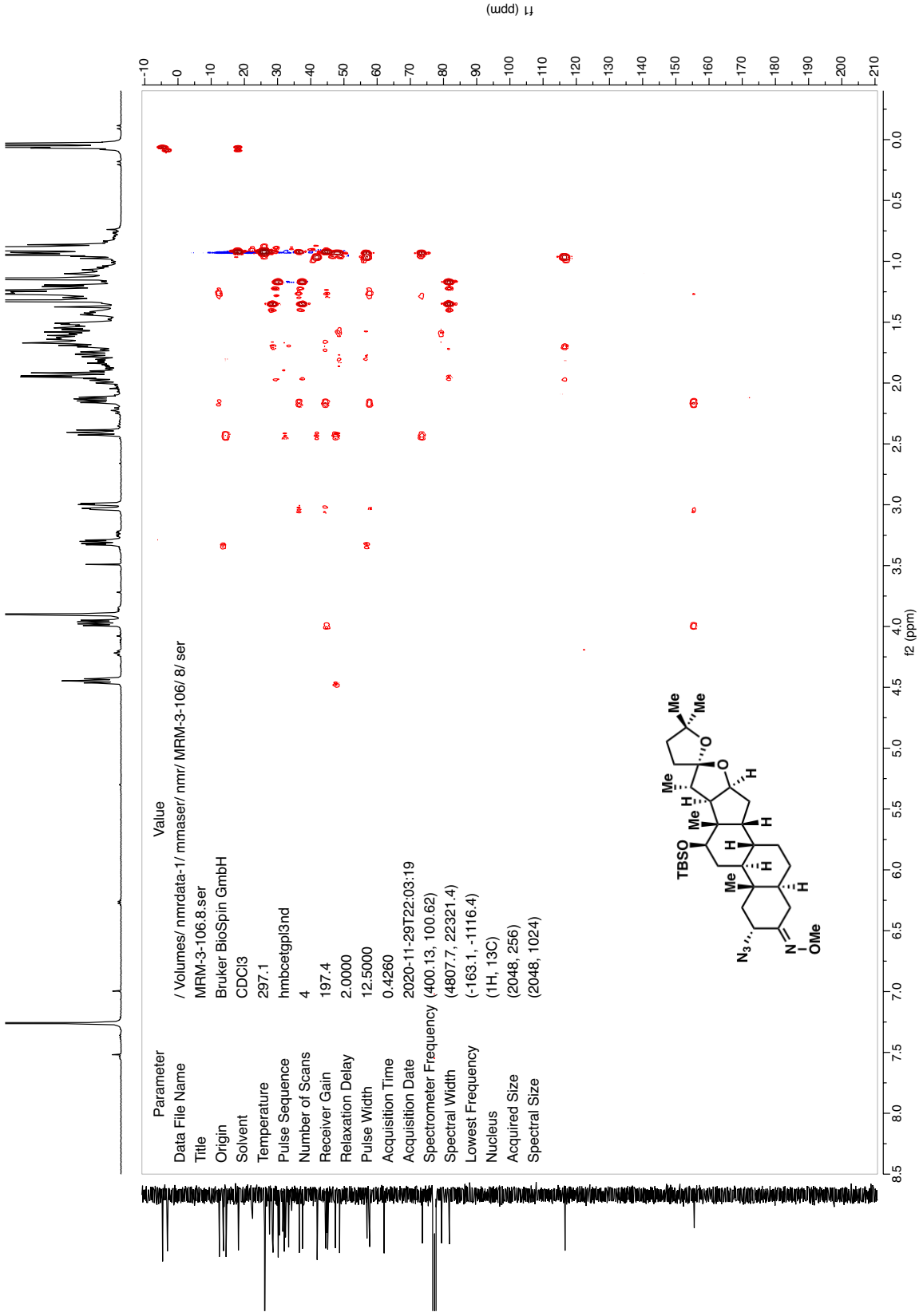

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-099-main/ 2/ fid
Title MRM-3-099-main.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-10-25T08:54:34
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.1
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
H
O
N3
205.11
116.60
81.79
79.34
73.45
63.81
56.92
48.74
47.31
47.17
45.26
44.93
43.79
41.93
37.49
36.93
33.32
32.37
31.95
31.56
30.77
30.19
29.85
28.71
28.63
26.08
18.19
14.48
13.69
12.53
-3.19
-4.64
f1 (ppm)

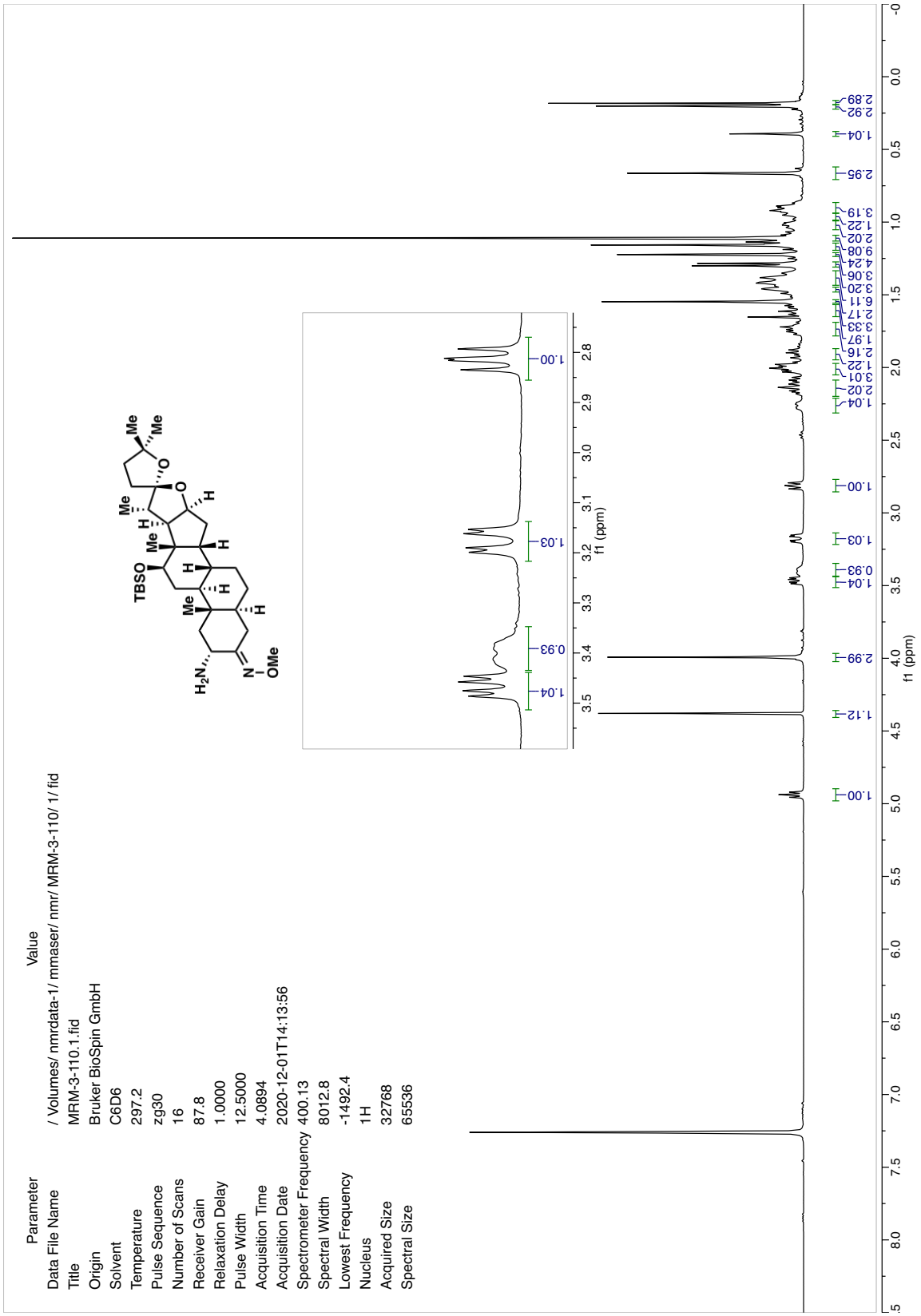

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-099-main/ 10/ ser
Title MRM-3-099-main.10.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence ngesygpphzs
Number of Scans 4
Receiver Gain 112.8
Relaxation Delay 1.8000
Pulse Width 12.5000
Acquisition Time 0.2560
Acquisition Date 2020-10-25T12:00:33
Spectrometer Frequency (400.13, 400.13)
Spectral Width (4000.0, 4000.0)
Lowest Frequency (-168.0, -168.0)
Nucleus (1H, 1H)
Acquired Size (1024, 256)
Spectral Size (1024, 1024)
TBSO
Me
H
N3
O

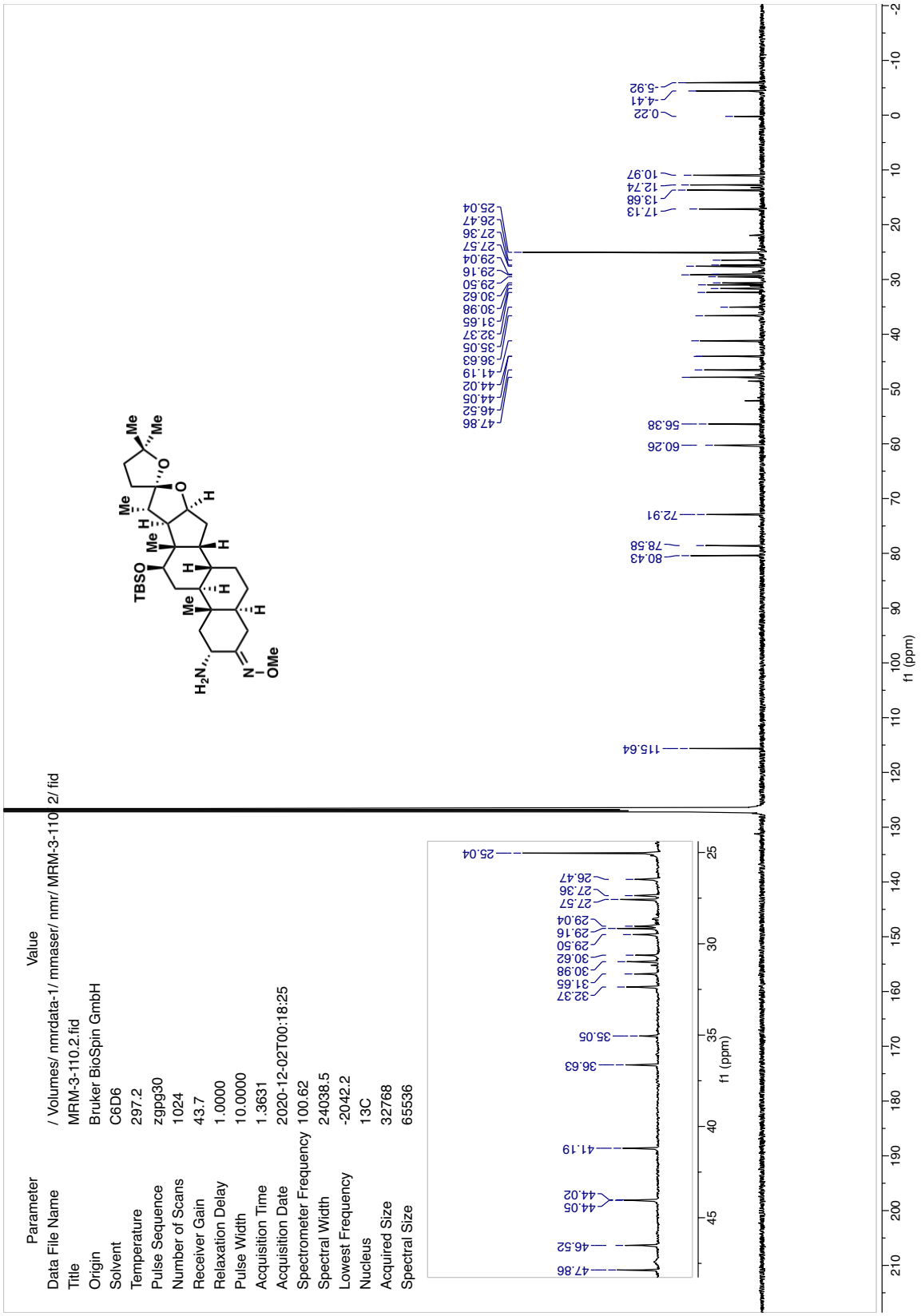

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-106/ 1/ fid
Title MRM-3-106.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-11-29T09:47:16
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1546.1
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
TBSO
OMe
f1 (ppm)

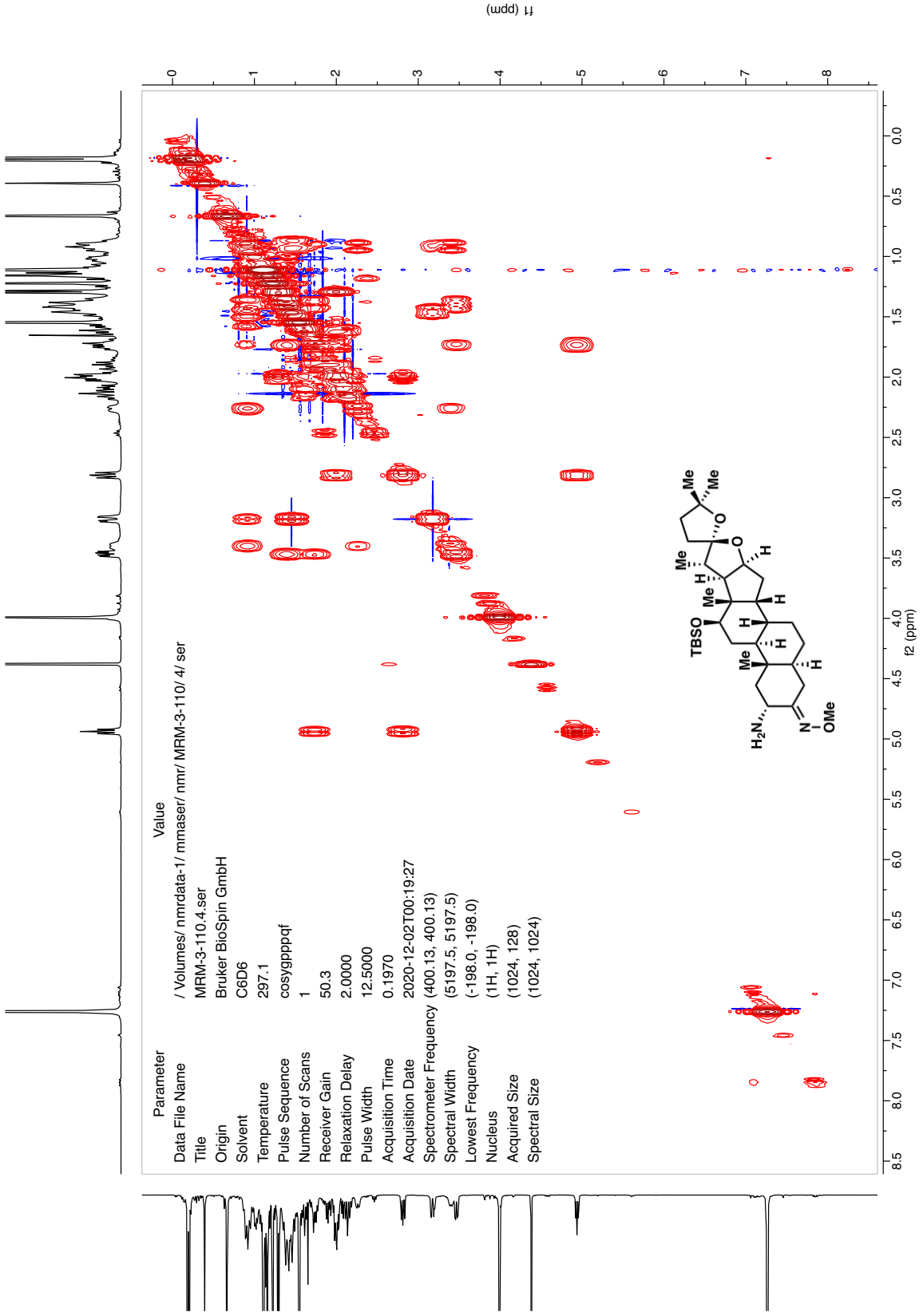

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-106/ 2/ fid
Title MRM-3-106.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-11-29T21:44:52
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1944.8
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
H
N3
N
OMe
f1 (ppm)

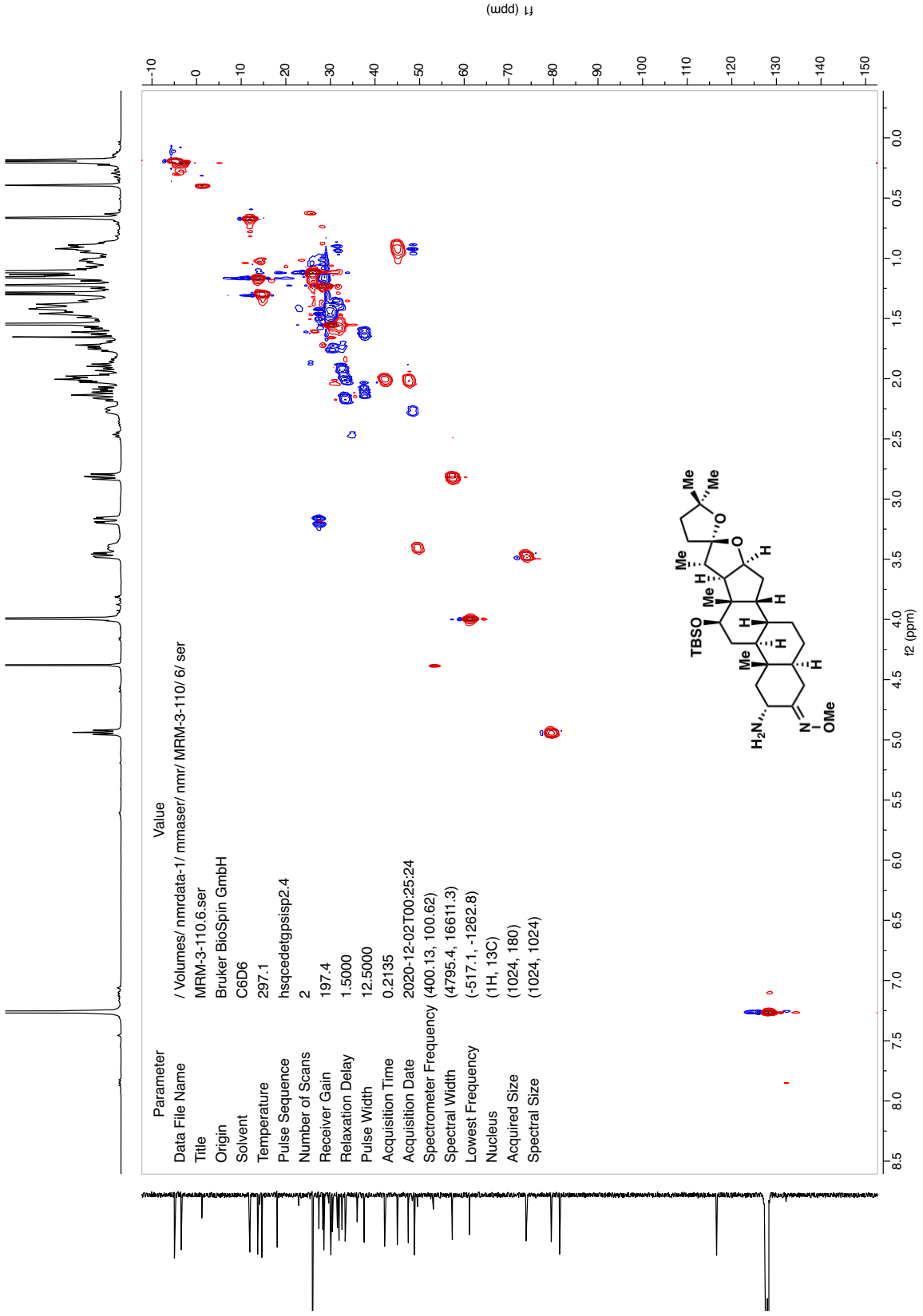

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-106/ 6/ ser
Title MRM-3-106.6.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-11-29T21:51:46
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-517.1, -1262.8)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
Me
H
O
N3
N
OMe

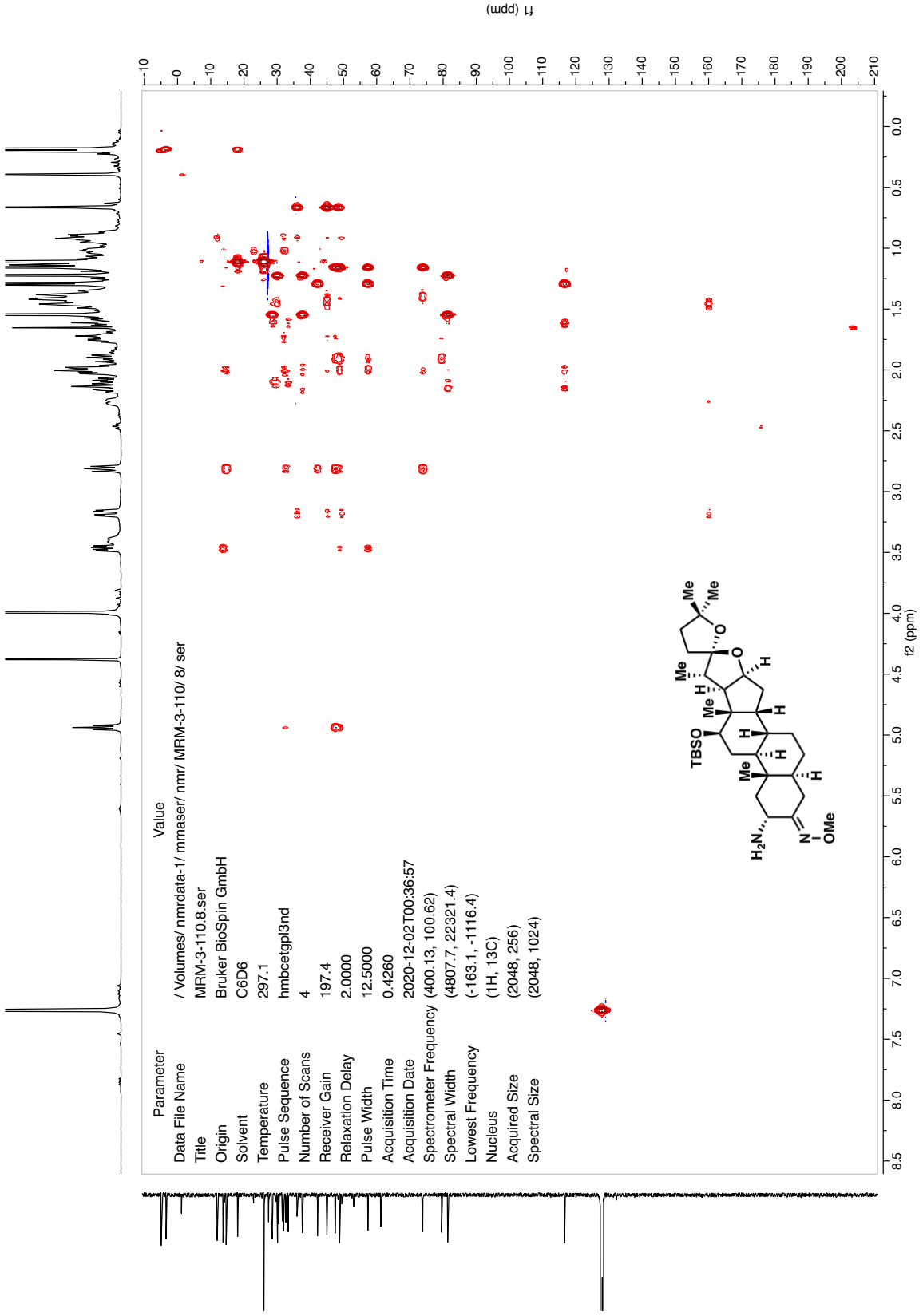

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-106/ 8/ ser
Title MRM-3-106.8.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-11-29T22:03:19
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-163.1, -1116.4)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
Me
H
O
N3
N
OMe

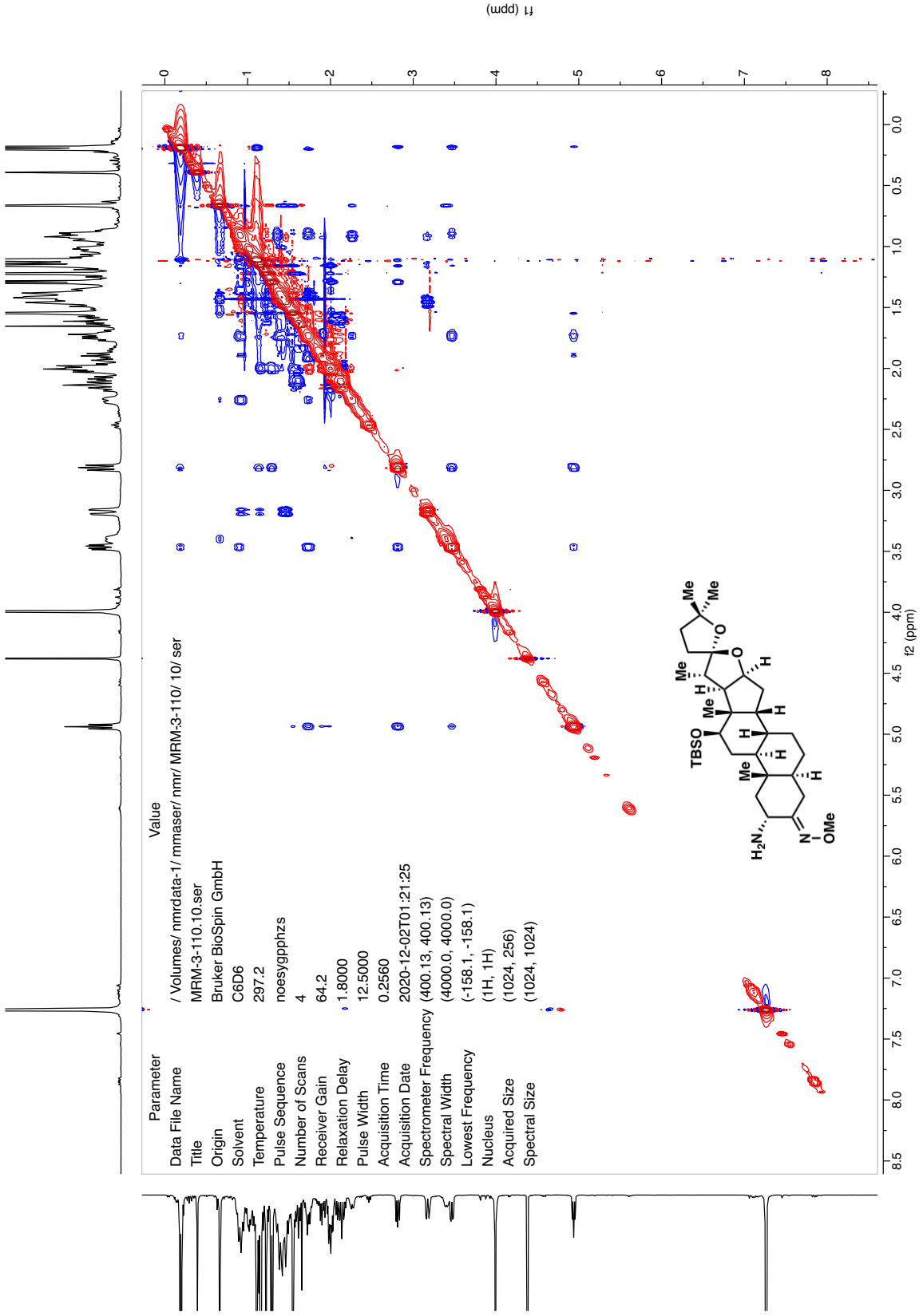

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-110/ 1/ fid
Title MRM-3-110.1.fid
Origin Bruker BioSpin GmbH
Solvent C6D6
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 87.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-12-01T14:13:56
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1492.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
TBSO
OMe
f1 (ppm)

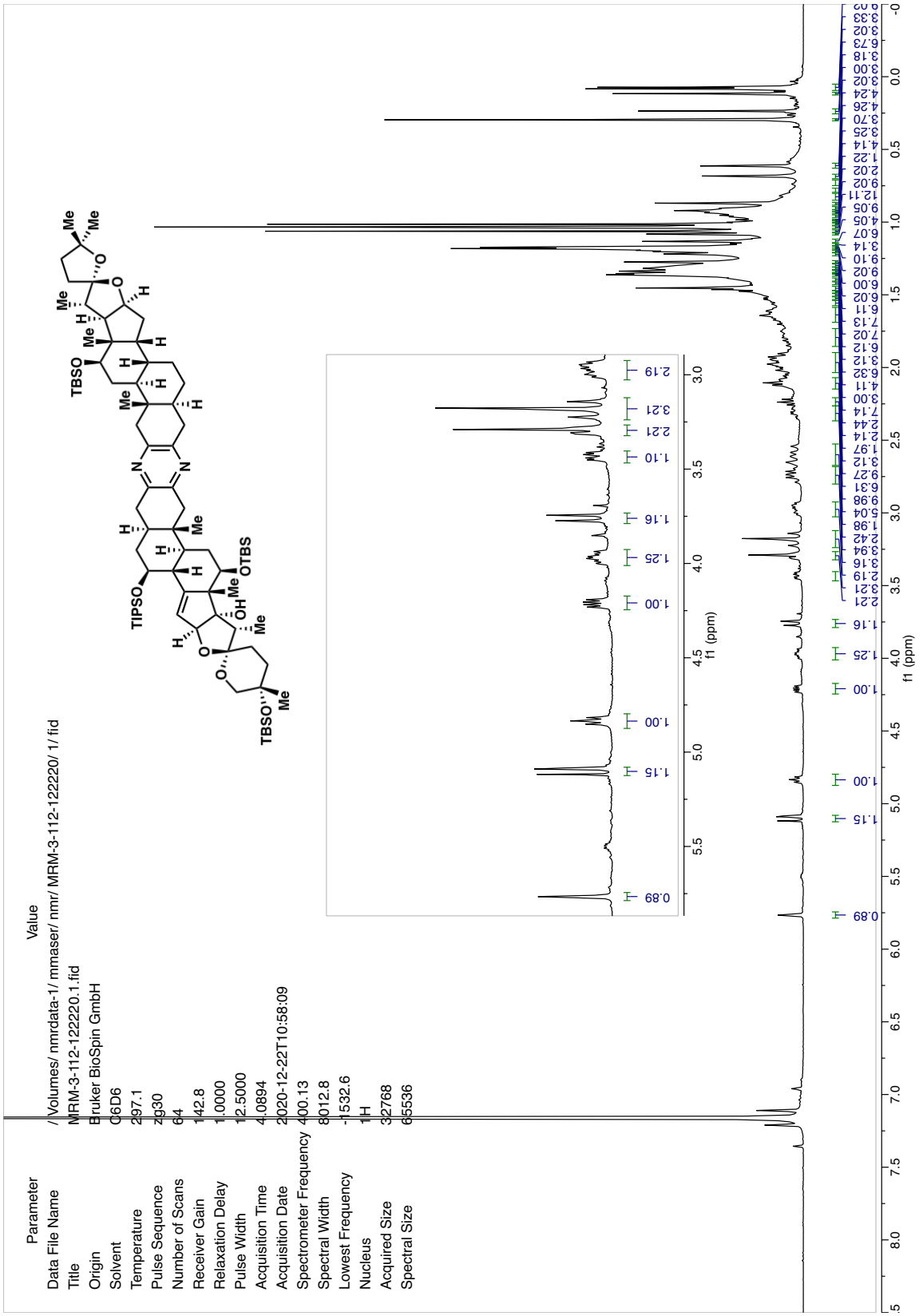

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-110 2/ fid
Title MRM-3-110.2.fid
Origin Bruker BioSpin GmbH
Solvent C6D6
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 43.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-12-02T00:18:25
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -2042.2
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
Me
H
O
N
OMe
H_2N
f1 (ppm)

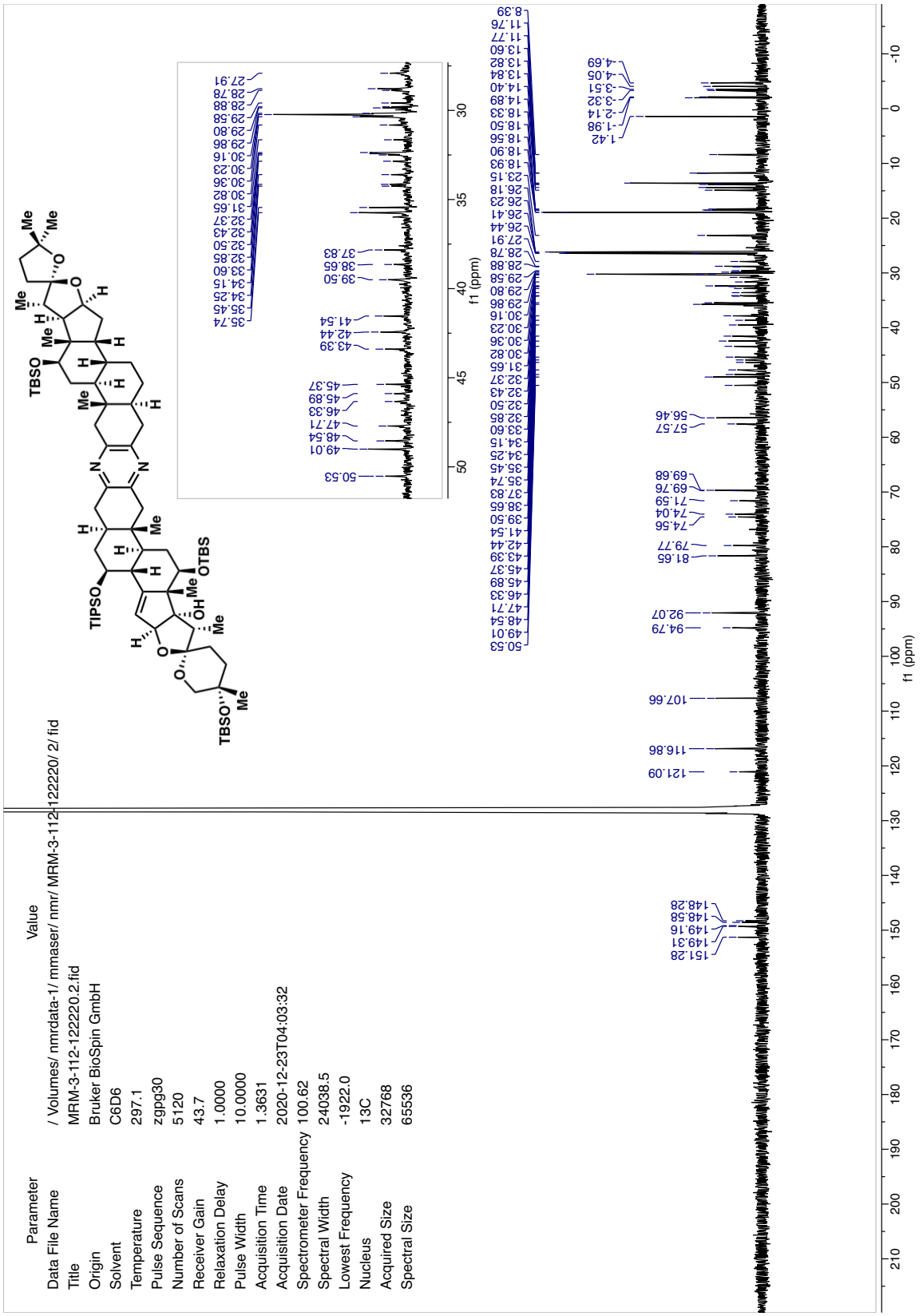

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-110/ 4/ ser
Title MRM-3-110.4.ser
Origin Bruker BioSpin GmbH
Solvent C6D6
Temperature 297.1
Pulse Sequence cosygpppqf
Number of Scans 1
Receiver Gain 50.3
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.1970
Acquisition Date 2020-12-02T00:19:27
Spectrometer Frequency (400.13, 400.13)
Spectral Width (5197.5, 5197.5)
Lowest Frequency (-198.0, -198.0)
Nucleus (1H, 1H)
Acquired Size (1024, 128)
Spectral Size (1024, 1024)
TBSO
Me
H
H_2N
N
OMe

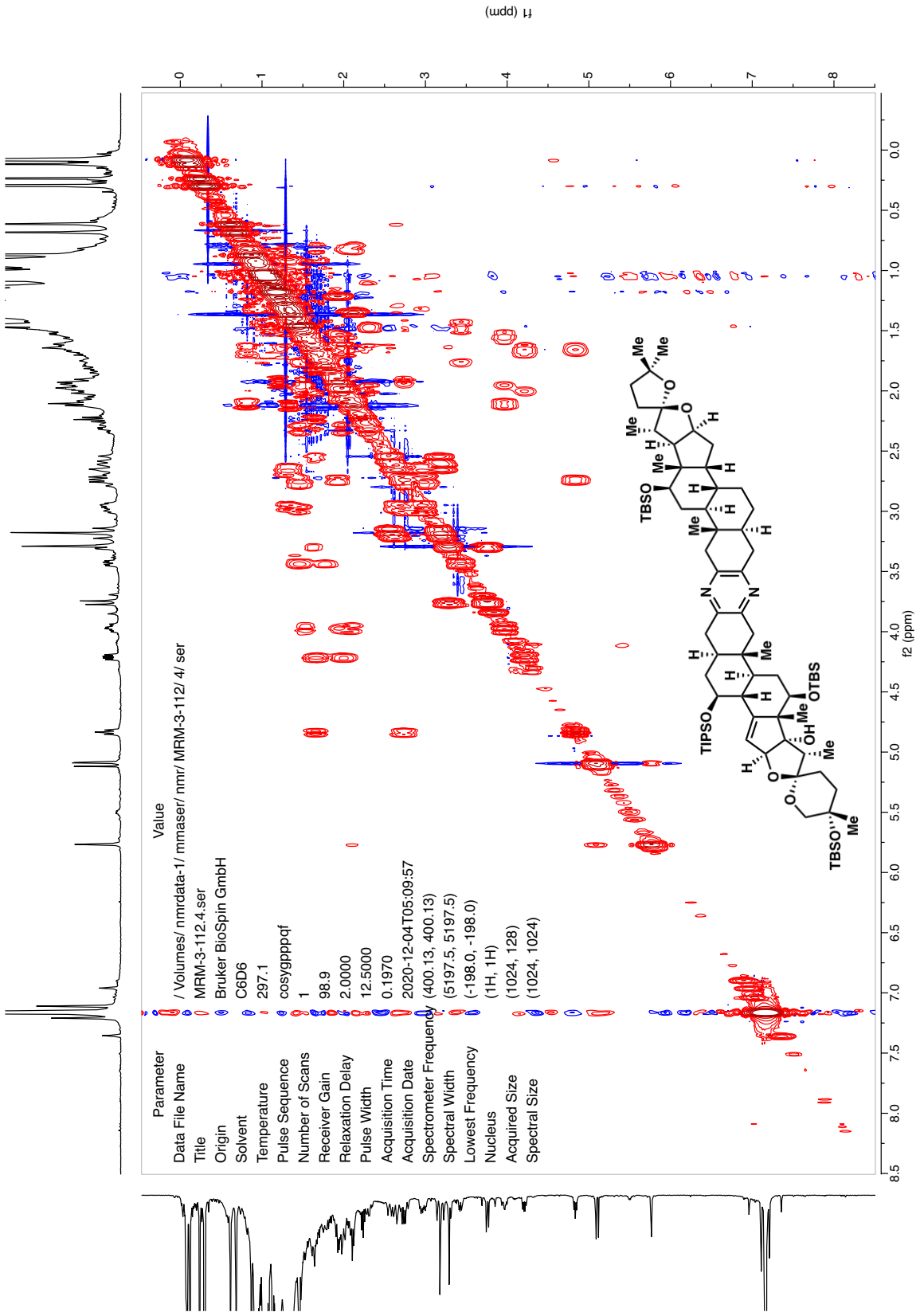

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-110/ 6/ ser
Title MRM-3-110.6.ser
Origin Bruker BioSpin GmbH
Solvent C6D6
Temperature 297.1
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-12-02T00:25:24
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-517.1, -1262.8)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
Me
H
O
OMe
N
H_2N

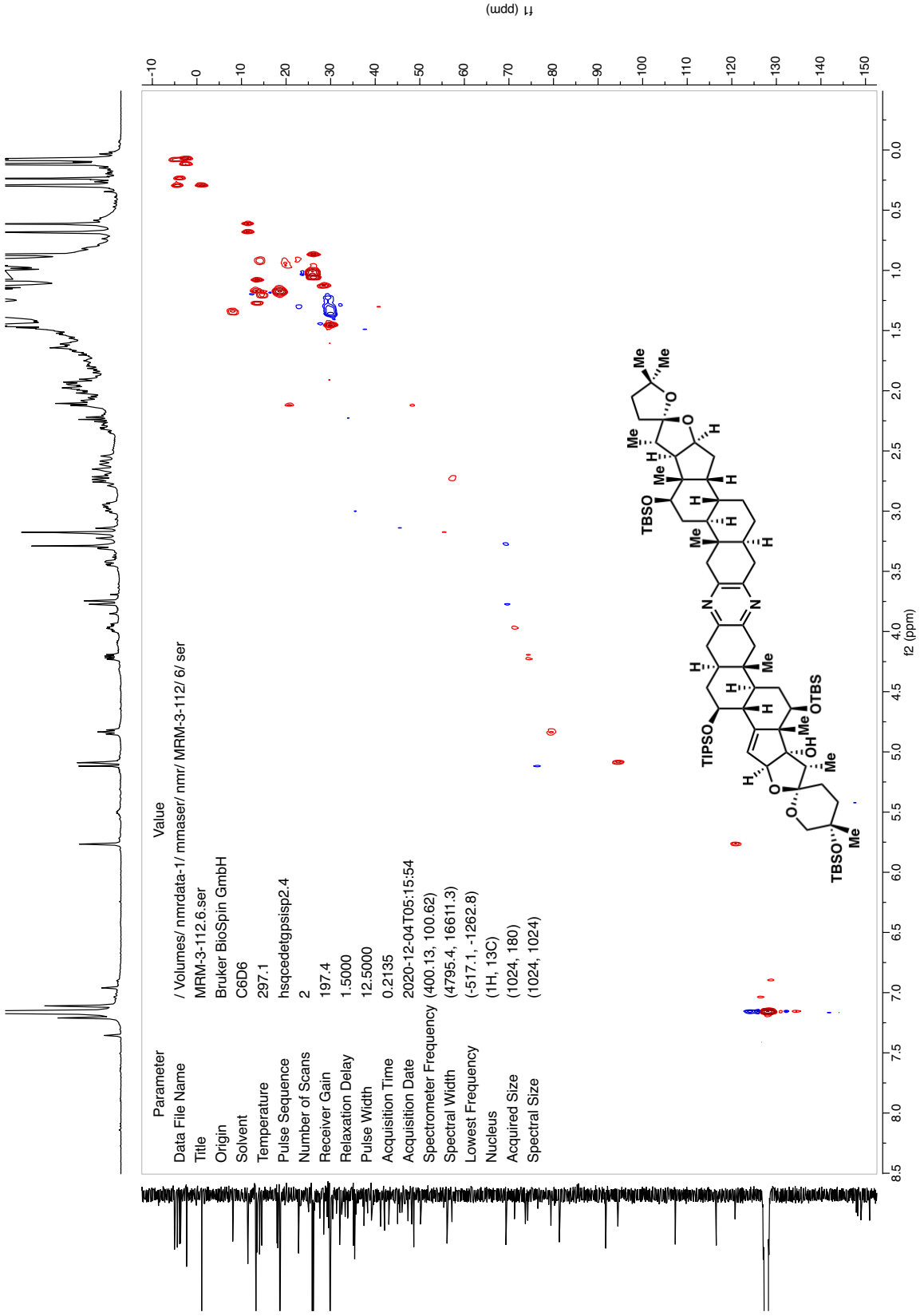

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-110/ 8/ ser
Title MRM-3-110.8.ser
Origin Bruker BioSpin GmbH
Solvent C6D6
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-12-02T00:36:57
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-163.1, -1116.4)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
Me
H
O
N
OMe
H_2N

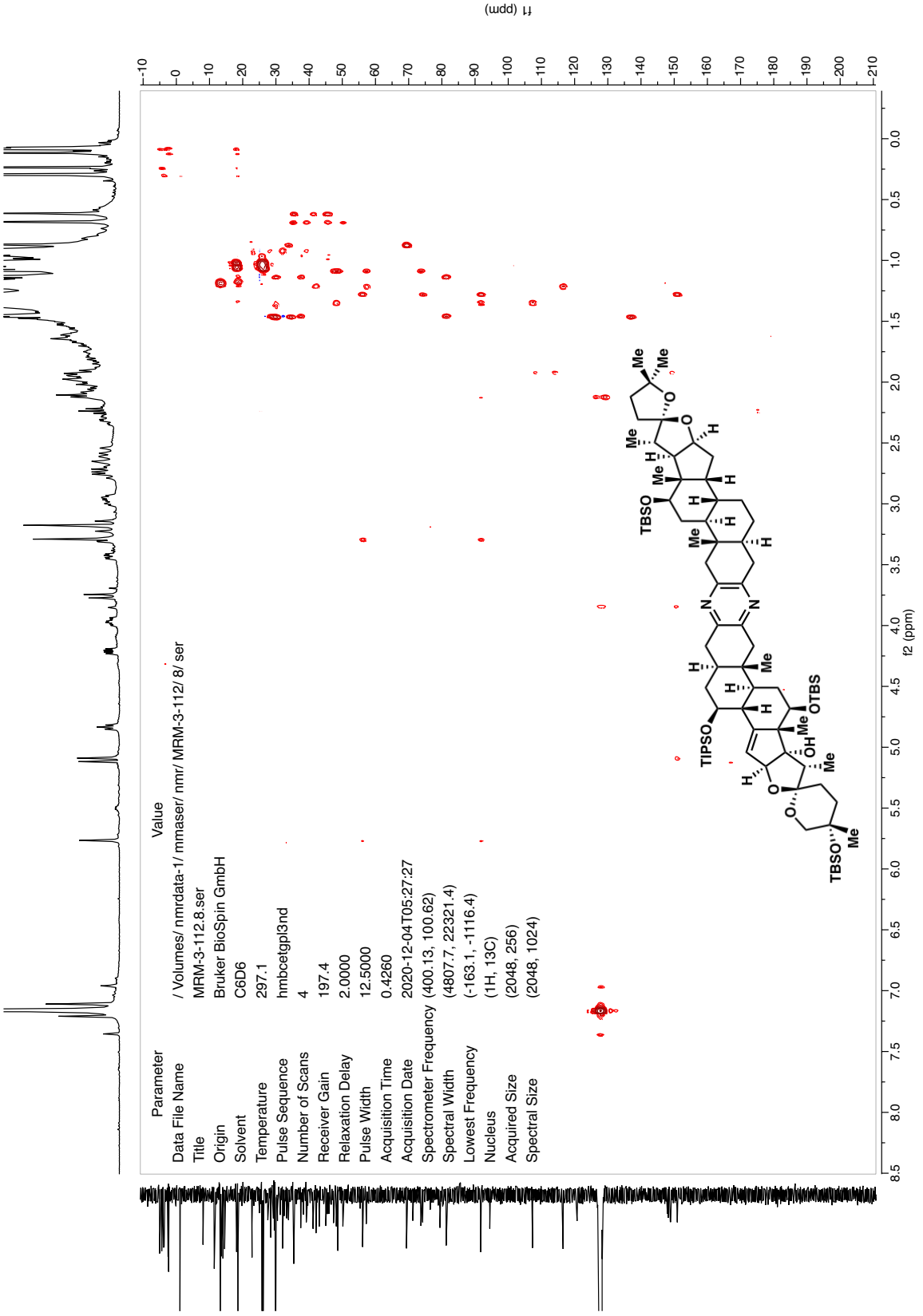

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-110/ 10/ ser
Title MRM-3-110.10.ser
Origin Bruker BioSpin GmbH
Solvent C6D6
Temperature 297.2
Pulse Sequence noesygpphzs
Number of Scans 4
Receiver Gain 64.2
Relaxation Delay 1.8000
Pulse Width 12.5000
Acquisition Time 0.2560
Acquisition Date 2020-12-02T01:21:25
Spectrometer Frequency (400.13, 400.13)
Spectral Width (4000.0, 4000.0)
Lowest Frequency (-158.1, -158.1)
Nucleus (1H, 1H)
Acquired Size (1024, 256)
Spectral Size (1024, 1024)
TBSO
Me
H
O
N
OMe
H_2N

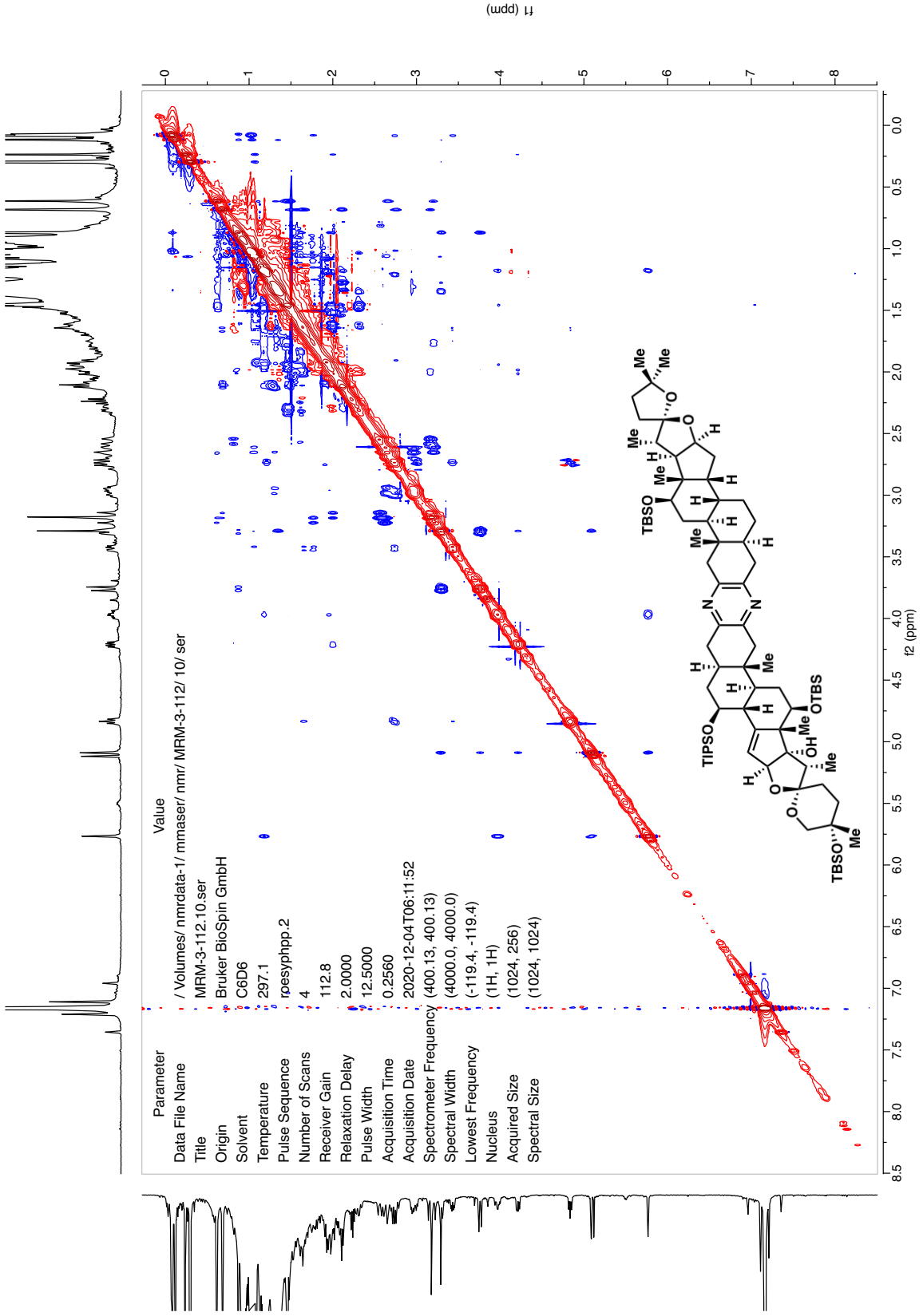

Parameter
Value
Data File Name
/ Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-112-122220/ 1/ fid
Title
MRM-3-112-122220.1.fid
Origin
Bruker BioSpin GmbH
Solvent
C6D6
Temperature
297.1
Pulse Sequence
zg30
Number of Scans
64
Receiver Gain
142.8
Relaxation Delay
1.0000
Pulse Width
12.5000
Acquisition Time
4.0894
Acquisition Date
2020-12-22T10:58:09
Spectrometer Frequency
400.13
Spectral Width
8012.8
Lowest Frequency
-1532.6
Nucleus
1H
Acquired Size
32768
Spectral Size
65536
TBSO
TIPSO
OTBS
OH
f1 (ppm)

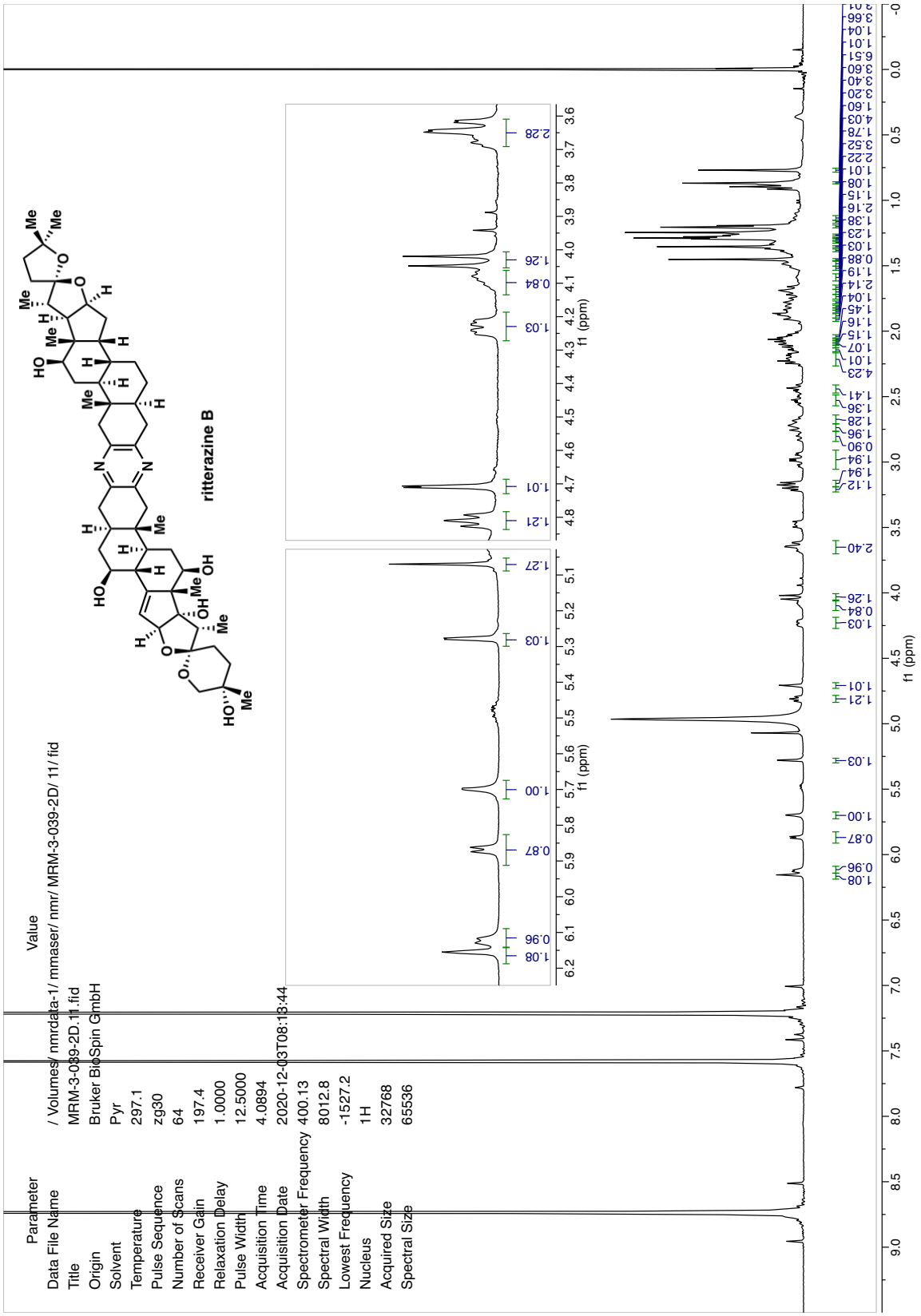

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-112-122220/ 2/ fid
Title MRM-3-112-122220.2.fid
Origin Bruker BioSpin GmbH
Solvent C6D6
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 5120
Receiver Gain 43.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-12-23T04:03:32
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1922.0
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
TIPSO
OTBS
OH
f1 (ppm)

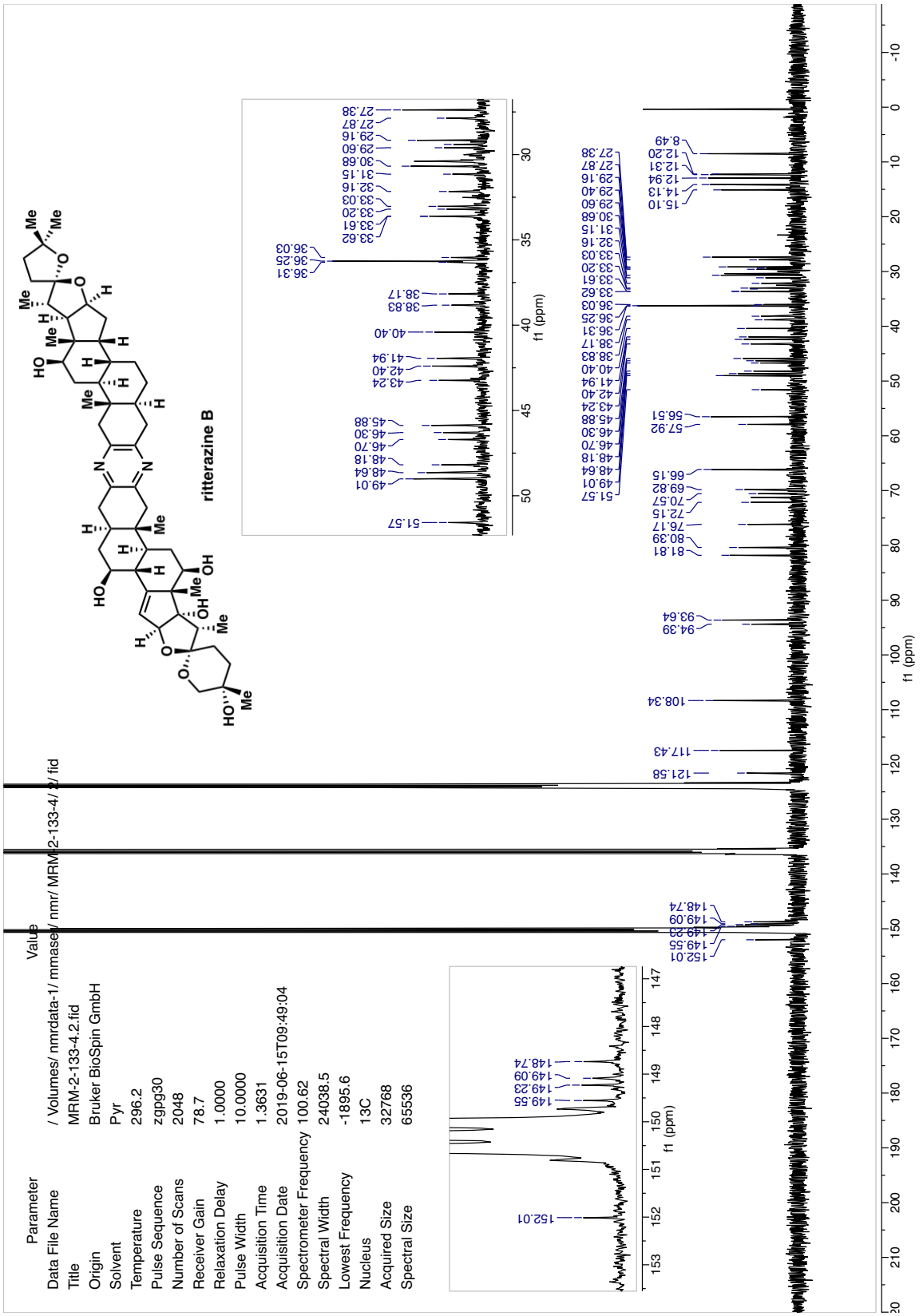

f1 (ppm)
f2 (ppm)
Parameter | Value
Data File Name | / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-112/ 4/ ser
Title | MRM-3-112.4.ser
Origin | Bruker BioSpin GmbH
Solvent | C6D6
Temperature | 297.1
Pulse Sequence | cosygpppqf
Number of Scans | 1
Receiver Gain | 98.9
Relaxation Delay | 2.0000
Pulse Width | 12.5000
Acquisition Time | 0.1970
Acquisition Date | 2020-12-04T05:09:57
Spectrometer Frequency | (400.13, 400.13)
Spectral Width | (5197.5, 5197.5)
Lowest Frequency | (-198.0, -198.0)
Nucleus | (1H, 1H)
Acquired Size | (1024, 128)
Spectral Size | (1024, 1024)
TBSO
TIPSO
OTBS
OH
Me
H
N
O

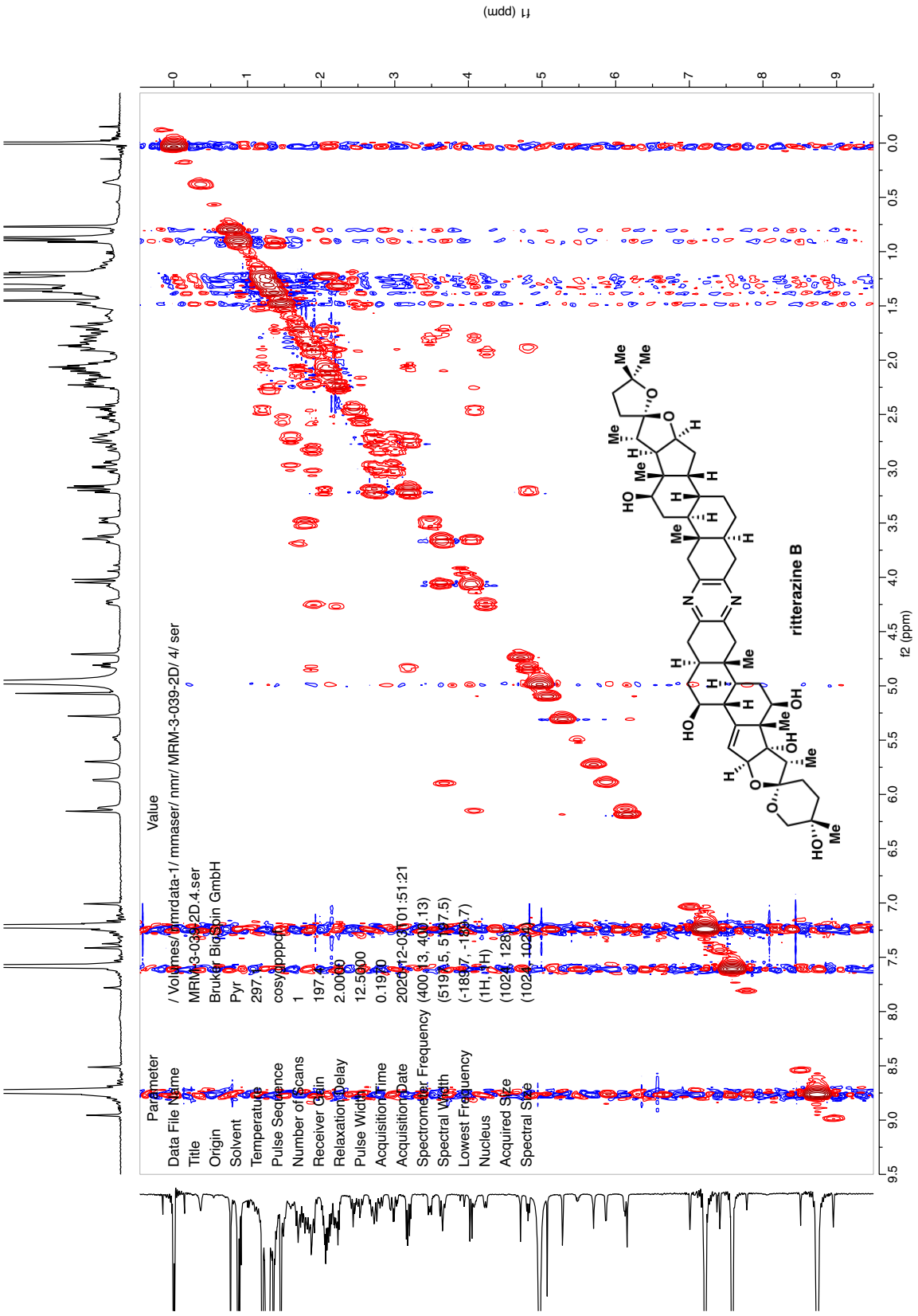

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-112/ 6/ ser
Title MRM-3-112.6.ser
Origin Bruker BioSpin GmbH
Solvent C6D6
Temperature 297.1
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-12-04T05:15:54
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-517.1, -1262.8)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
TBSO
TIPSO
OTBS
TBSO
Me
OH

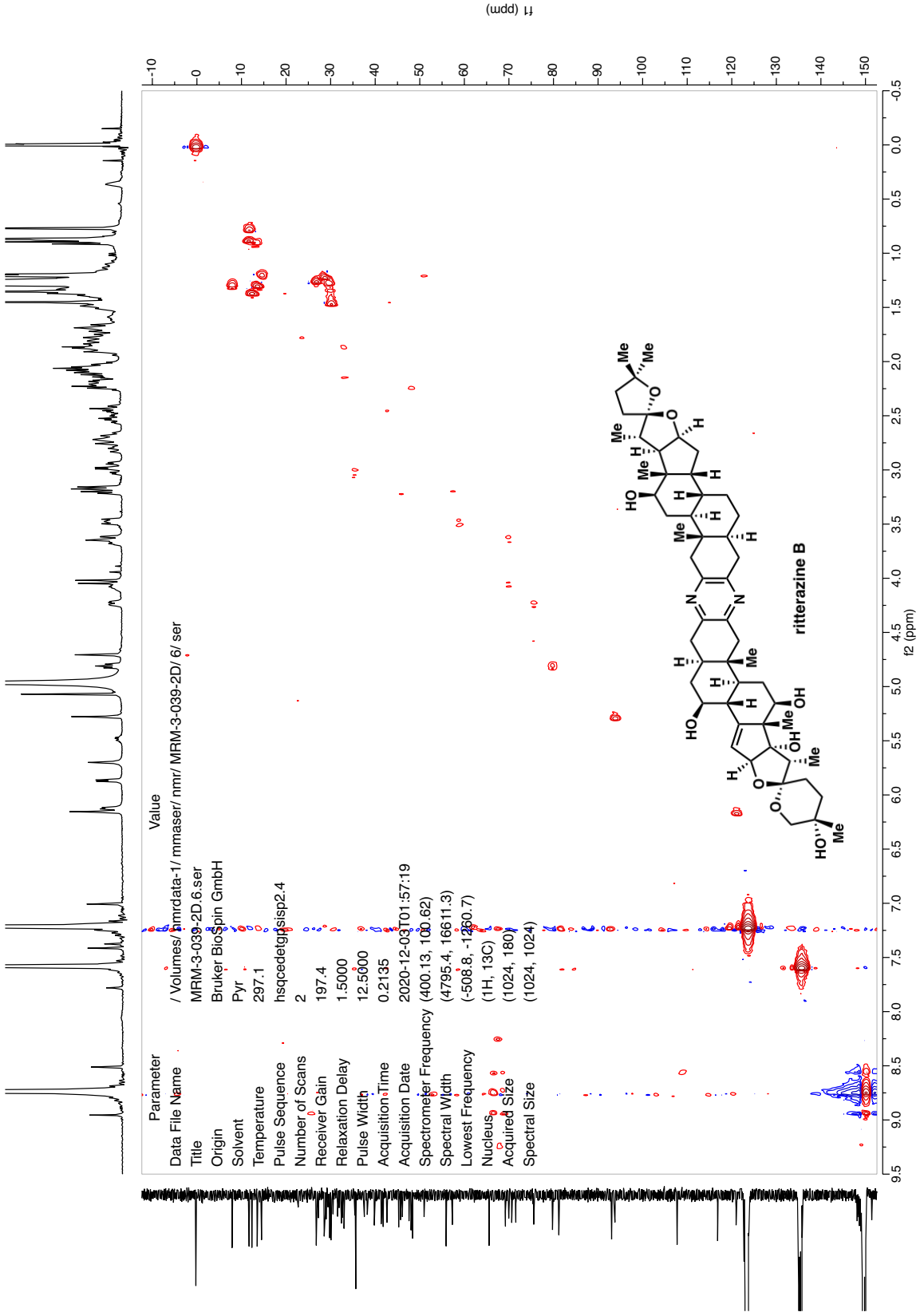

f1 (ppm)
f2 (ppm)
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-112/ 8/ ser
Title MRM-3-112.8.ser
Origin Bruker BioSpin GmbH
Solvent C6D6
Temperature 297.1
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-12-04T05:27:27
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-163.1, -1116.4)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
TBSO
TIPSO
OTBS
TBSO
Me
OH

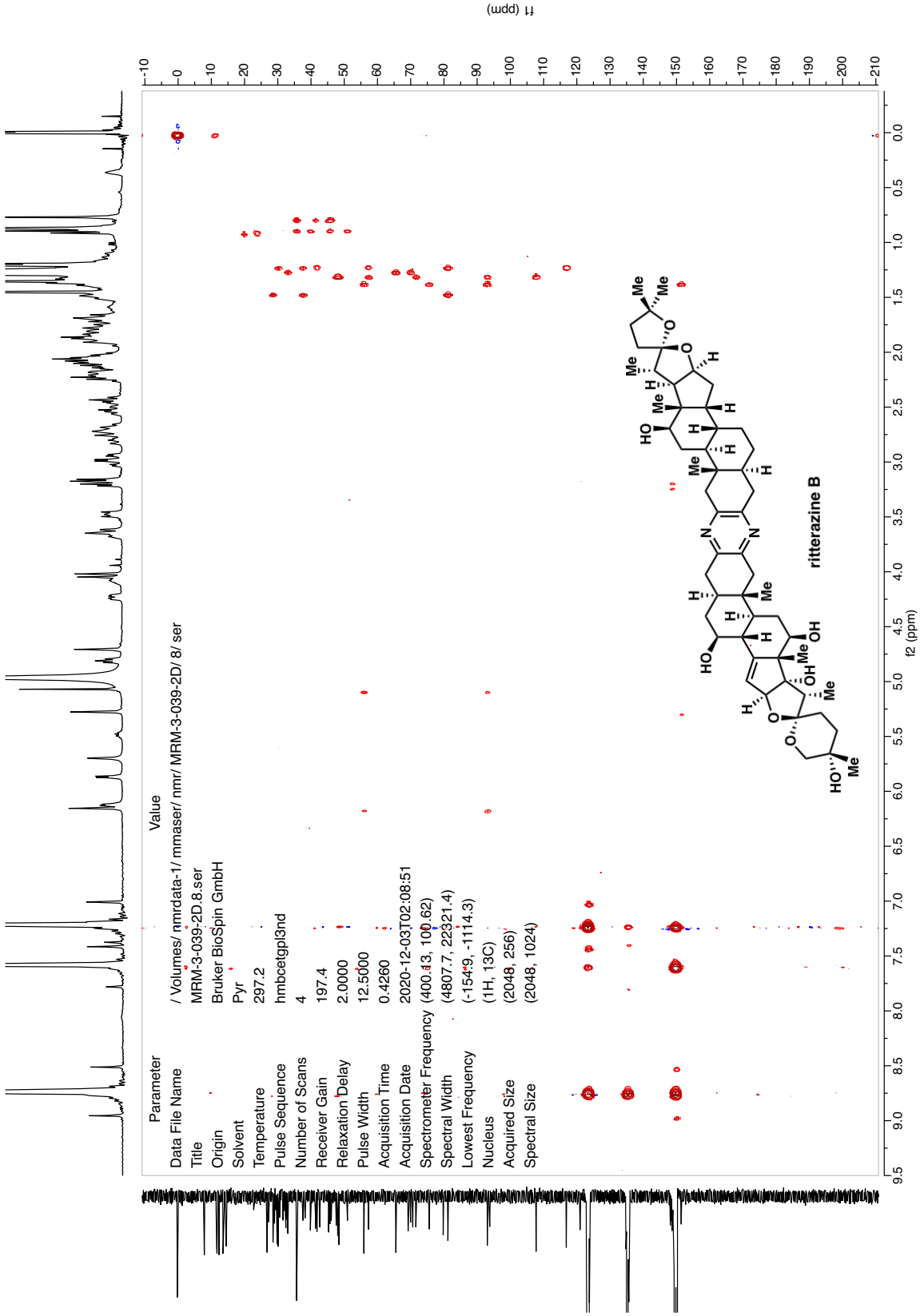

f1 (ppm)
f2 (ppm)
Parameter | Value
Data File Name | / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-112/ 10/ ser
Title | MRM-3-112.10.ser
Origin | Bruker BioSpin GmbH
Solvent | C6D6
Temperature | 297.1
Pulse Sequence | rpesyphpp.2
Number of Scans | 4
Receiver Gain | 112.8
Relaxation Delay | 2.0000
Pulse Width | 12.5000
Acquisition Time | 0.2560
Acquisition Date | 2020-12-04T06:11:52
Spectrometer Frequency | (400.13, 400.13)
Spectral Width | (4000.0, 4000.0)
Lowest Frequency | (-119.4, -119.4)
Nucleus | (1H, 1H)
Acquired Size | (1024, 256)
Spectral Size | (1024, 1024)
TBSO
TIPSO
OTBS
TBSO
Me
OH
H
N
O

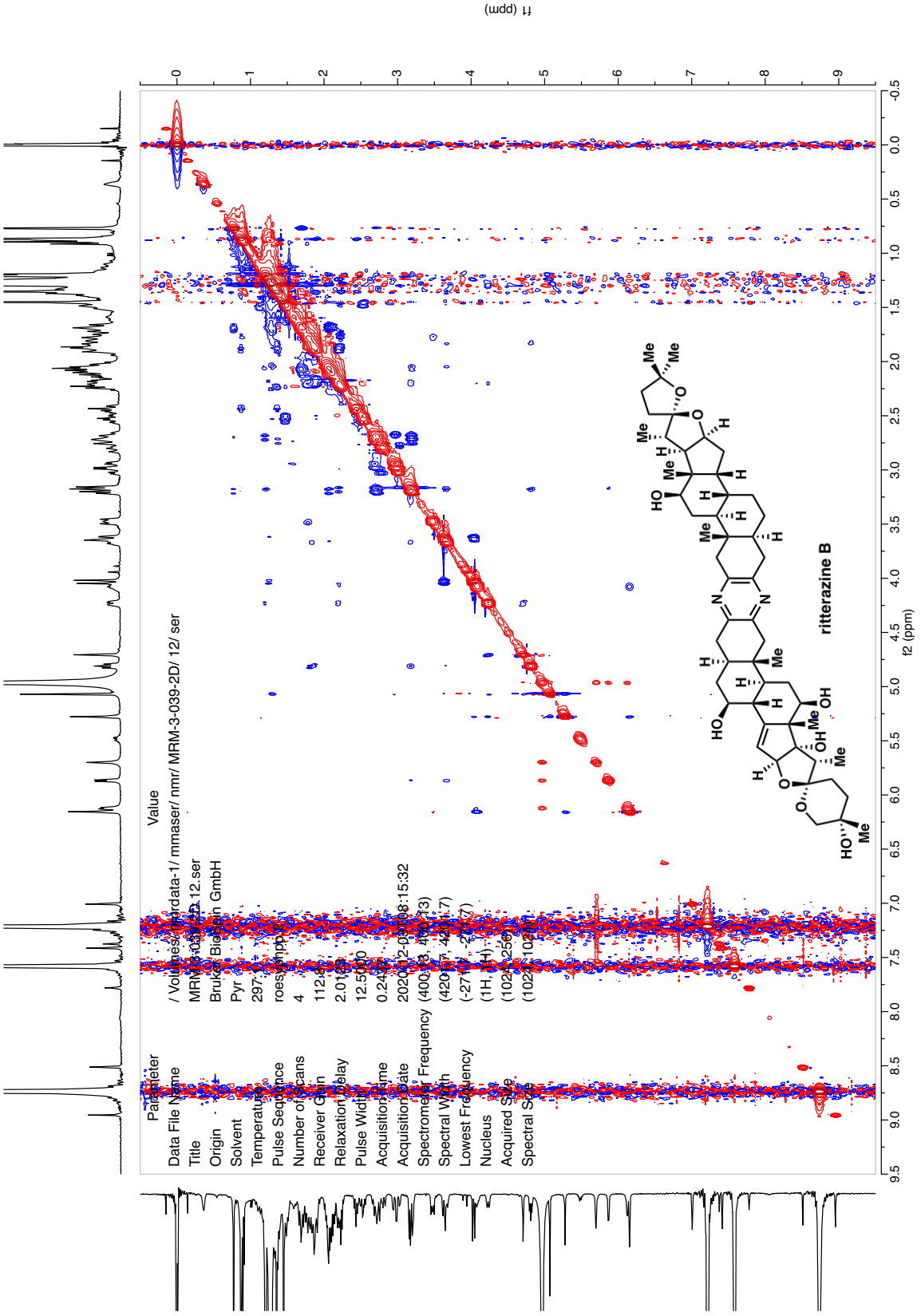

ritterazine B
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-039-2D/ 11/ fid
Title MRM-3-039-2D.11.fid
Origin Bruker BioSpin GmbH
Solvent Pyr
Temperature 297.1
Pulse Sequence zg30
Number of Scans 64
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-12-03T08:13:44
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1527.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
f1 (ppm)

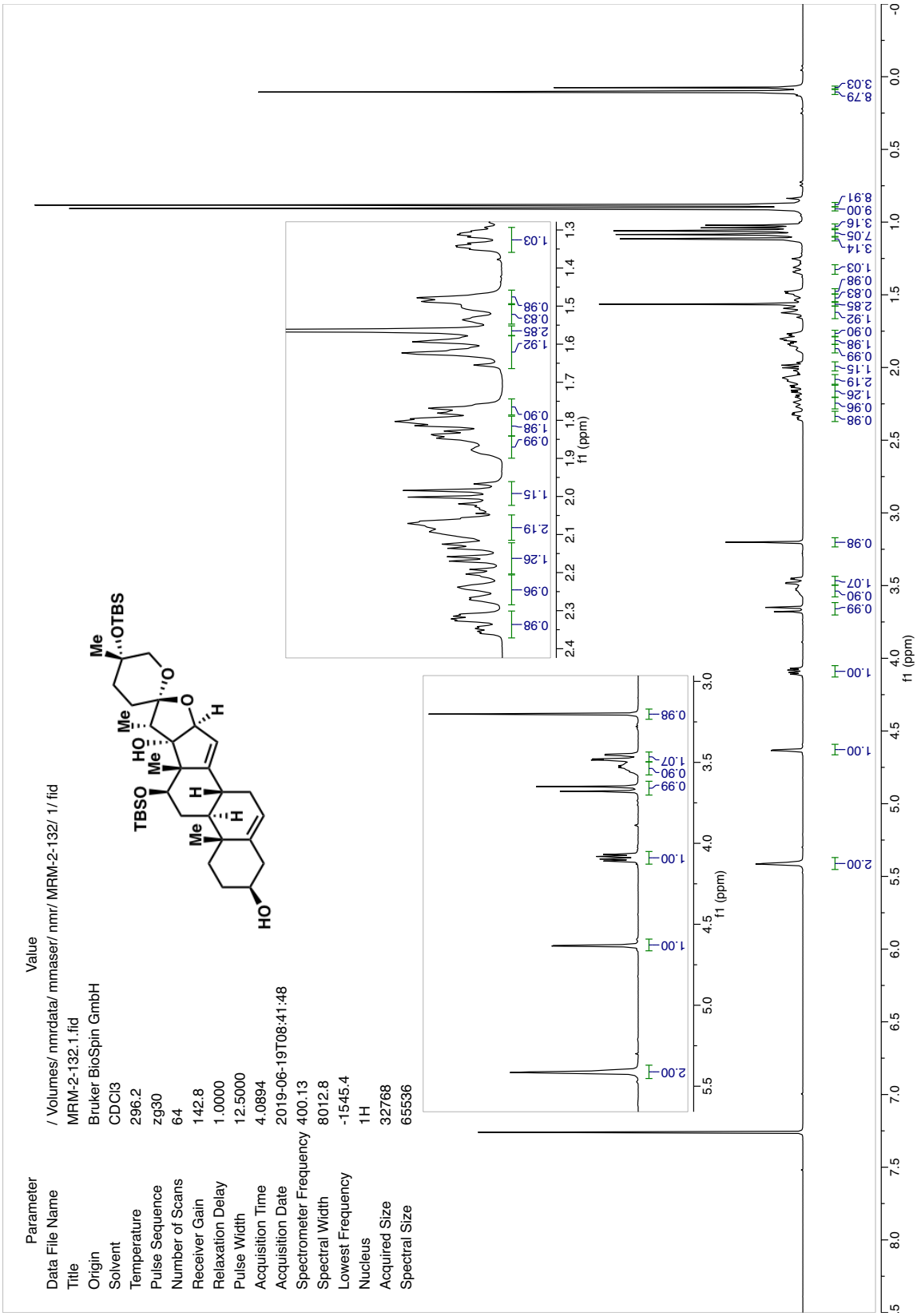

ritterazine B
Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmase/ nmr/ MRM-2-133-4/ 2/ fid
Title MRM-2-133-4.2.fid
Origin Bruker BioSpin GmbH
Solvent Pyr
Temperature 296.2
Pulse Sequence zgpg30
Number of Scans 2048
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-06-15T09:49:04
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1895.6
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
f1 (ppm)

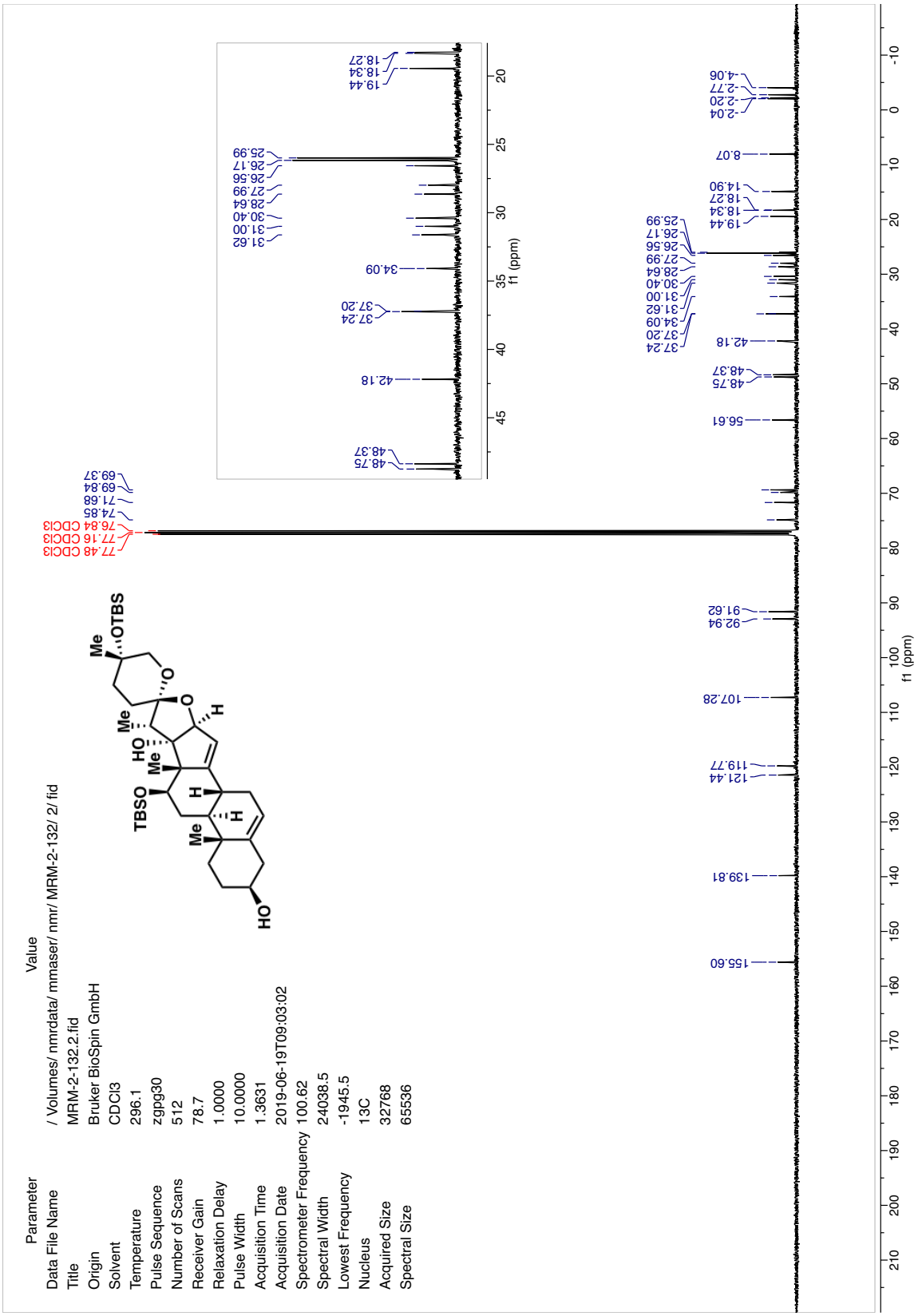

f1 (ppm)
f2 (ppm)
Parameter
Value
Data File Name
/ Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-039-2D/ 4/ ser
Title
MRM-3-039-2D 4.ser
Origin
Bruker BioSpin GmbH
Solvent
Pyr
Temperature
297.
Pulse Sequence
cosygpppqf
Number of Scans
1
Receiver Gain
197.
Relaxation Delay
2.0000
Pulse Width
12.5000
Acquisition Time
0.1970
Acquisition Date
2020
T01:51:21
Spectrometer Frequency
(400.13, 400.13)
Spectral Width
(5194.5, 5194.5)
Lowest Frequency
(-1859.7, -1859.7)
Nucleus
(1H, 1H)
Acquired Size
(1024, 128
Spectral Size
(1024, 1024
ritterazine B

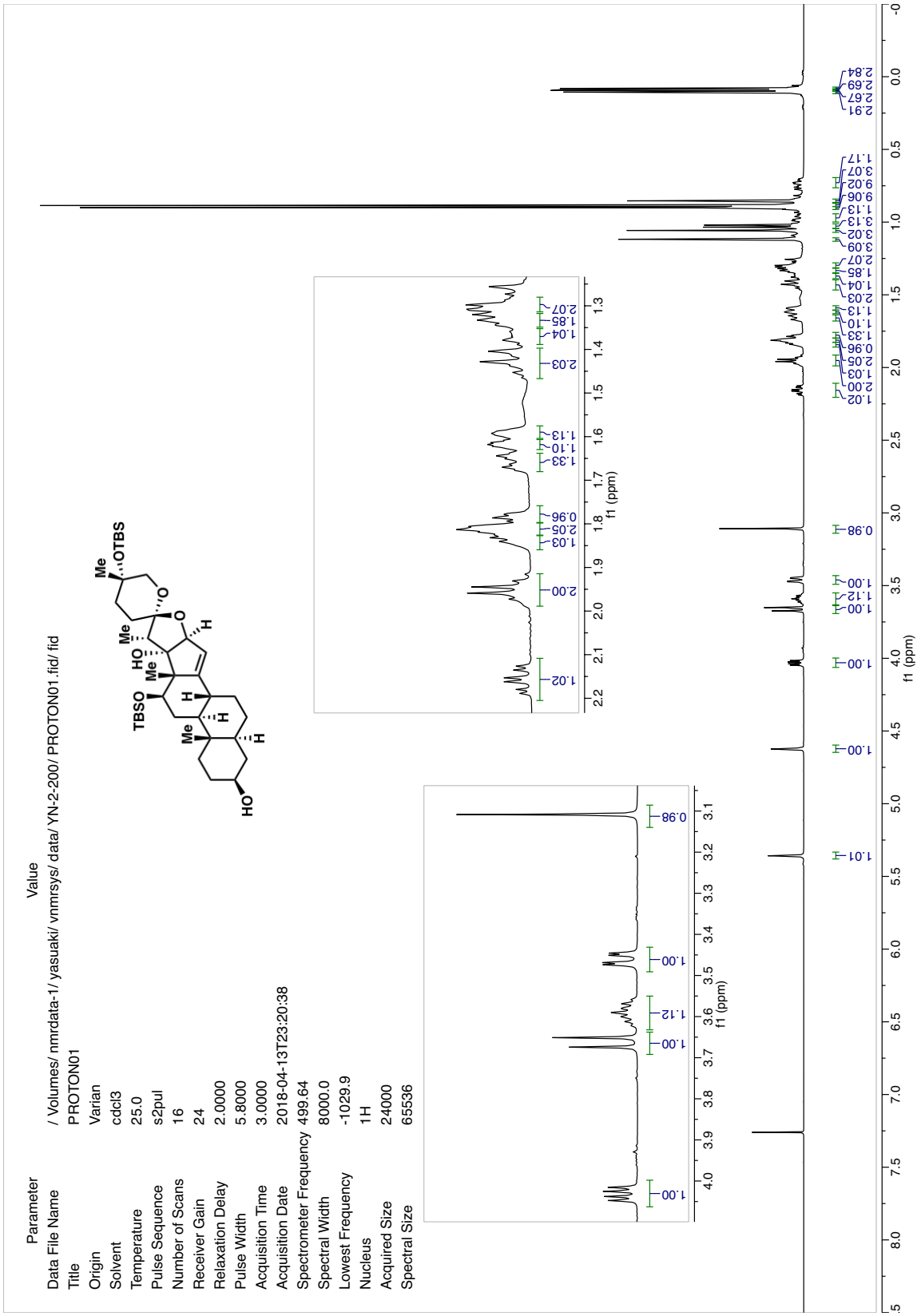

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ nmrdata/ nmr/ MRM-3-039-2D/ 6/ ser
Title MRM-3-039-2D 6.ser
Origin Bruker Biospin GmbH
Solvent Pyr
Temperature 297.1
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 2
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-12-03T01:57:19
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-508.8, -1260.7)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)
ritterazine B
f1 (ppm)
f2 (ppm)

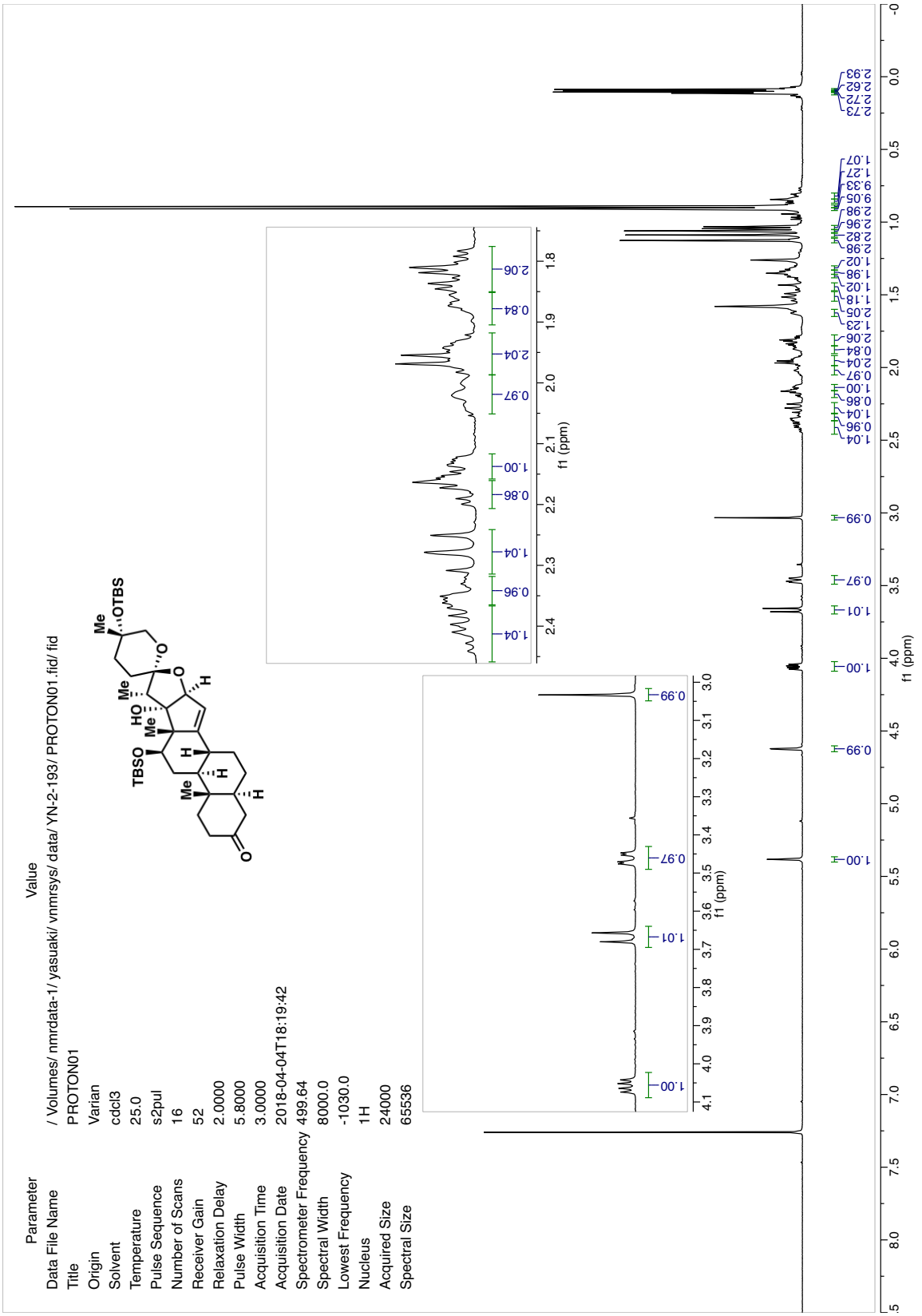

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-039-2D/ 8/ ser
Title MRM-3-039-2D 8.ser
Origin Bruker BioSpin GmbH
Solvent Pyr
Temperature 297.2
Pulse Sequence hmbcetgpl3nd
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 2.0000
Pulse Width 12.5000
Acquisition Time 0.4260
Acquisition Date 2020-12-03T02:08:51
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4807.7, 22321.4)
Lowest Frequency (-154.9, -1114.3)
Nucleus (1H, 13C)
Acquired Size (2048, 256)
Spectral Size (2048, 1024)
ritterazine B
f1 (ppm)
f2 (ppm)

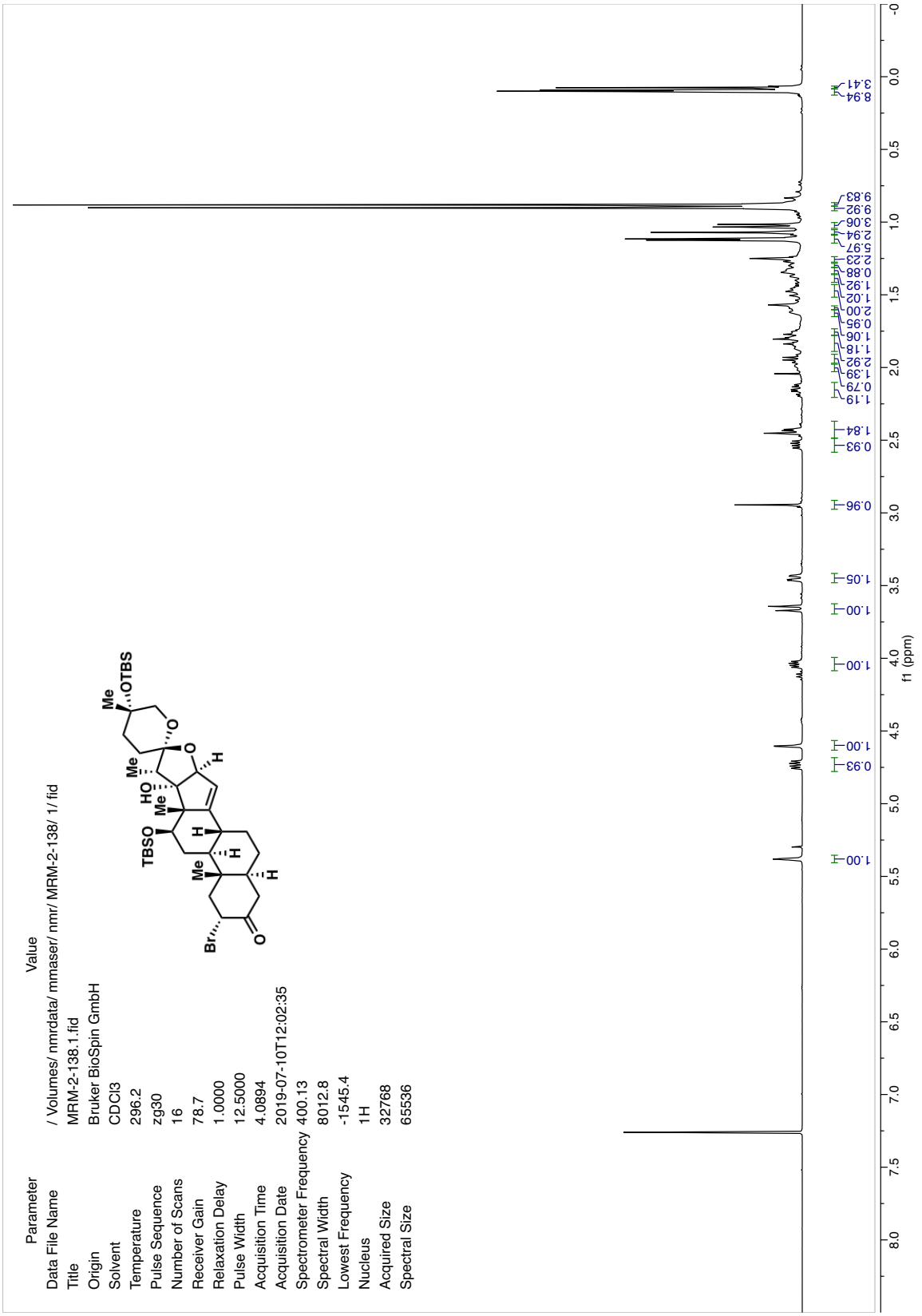

f1 (ppm)
f2 (ppm)
Parameter
Value
Data File Name
/ Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-039-2D/ 12/ ser
Title
MRM-3-039-2D 12.ser
Origin
Bruker BioSpin GmbH
Solvent
Pyr
Temperature
297
Pulse Sequence
roesyph
Number of Scans
4
Receiver Gain
112.6
Relaxation Delay
2.0100
Pulse Width
12.5000
Acquisition Time
0.2457
Acquisition Date
2020-12-03T08:15:32
Spectrometer Frequency
(400.13, 400.13)
Spectral Width
(4206.7, 4206.7)
Lowest Frequency
(-274.7, -274.7)
Nucleus
(1H, 1H)
Acquired Size
(1024, 256)
Spectral Size
(1024, 1024)
ritterazine B

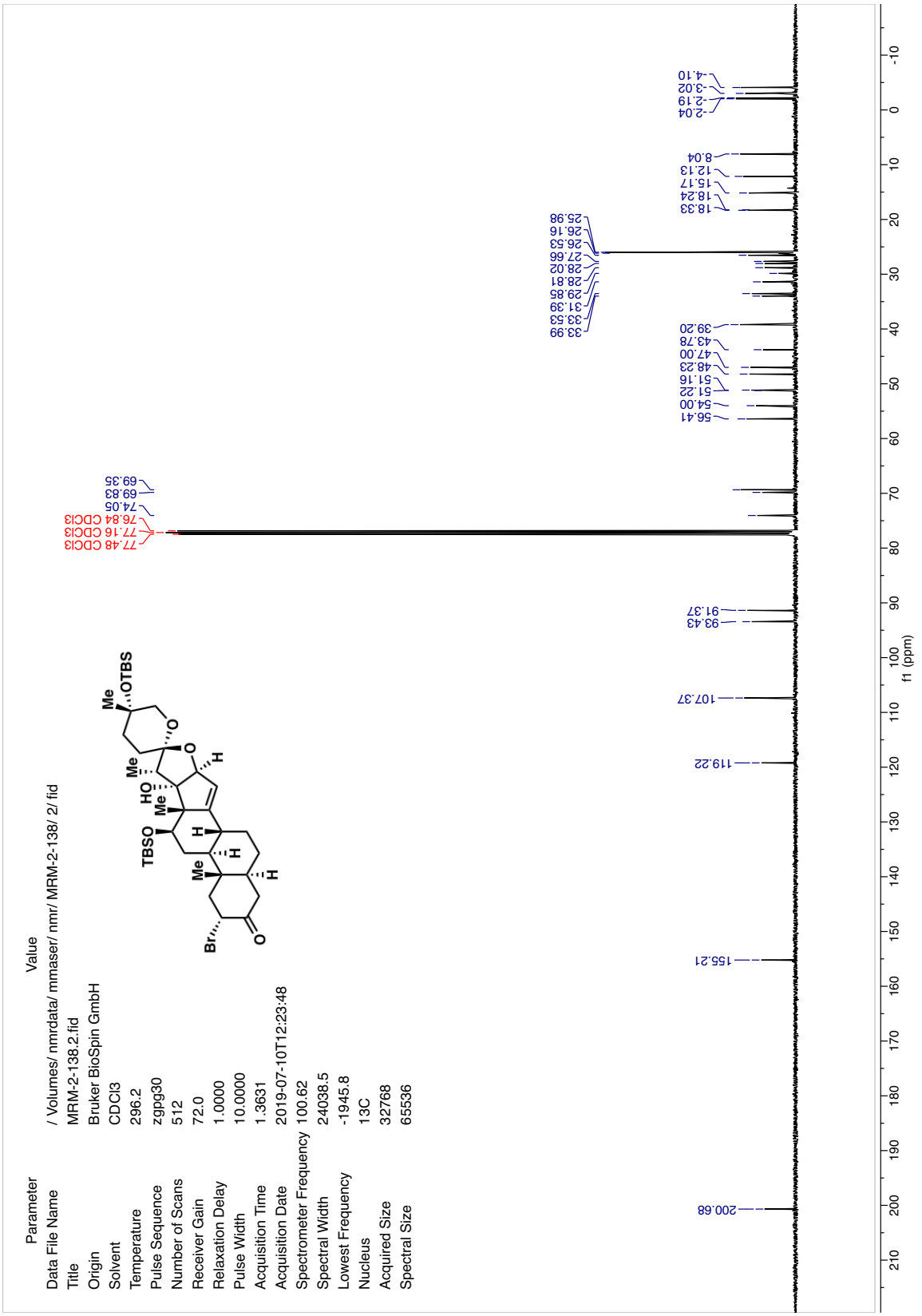

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-132/ 1/ fid
Title MRM-2-132.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.2
Pulse Sequence zg30
Number of Scans 64
Receiver Gain 142.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-06-19T08:41:48
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
f1 (ppm)

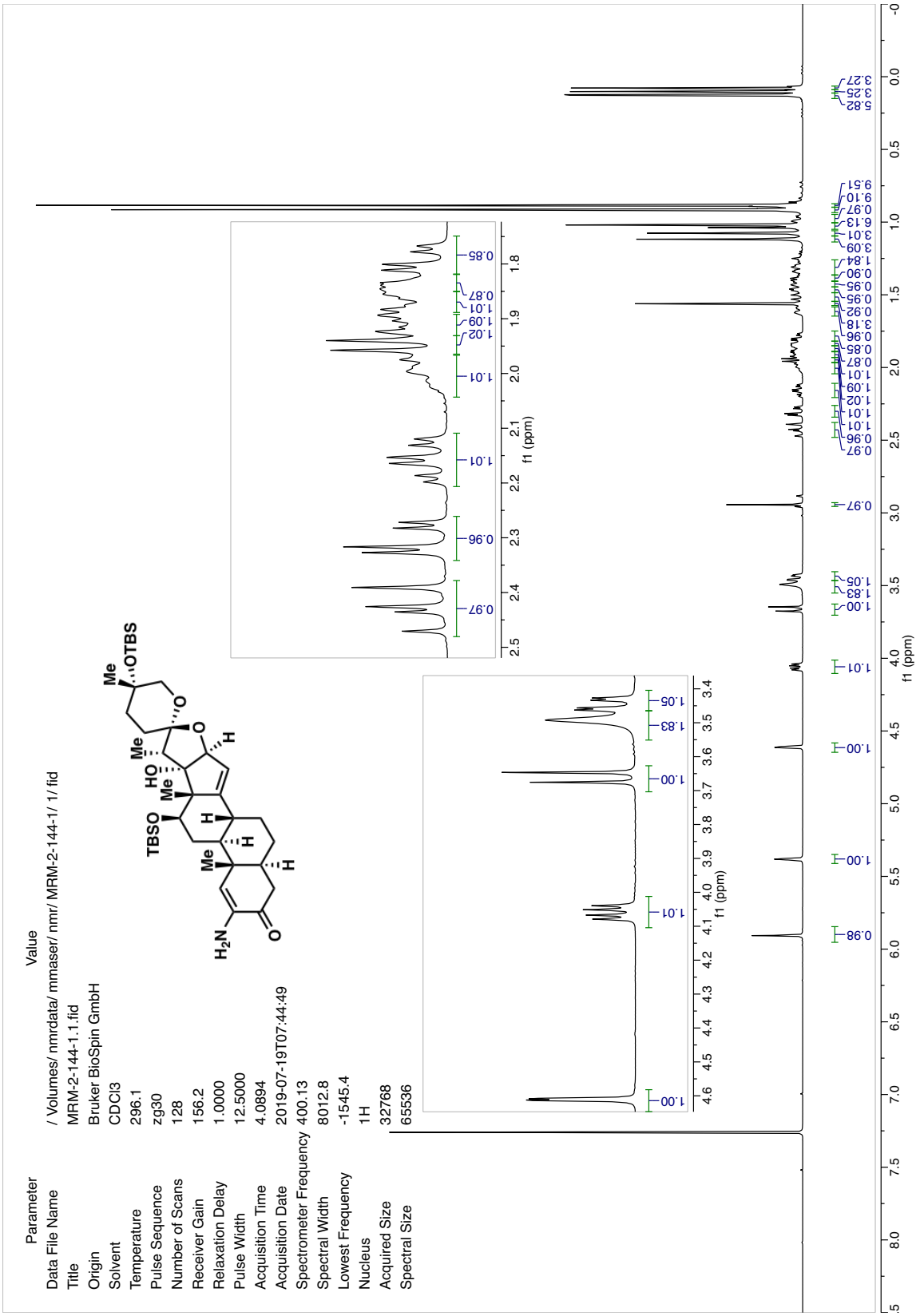

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-132/ 2/ fid
Title MRM-2-132.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-06-19T09:03:02
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.5
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
HO
Me
OTBS
H
O
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
74.85
71.68
69.84
69.37
48.75
48.37
42.18
37.24
37.20
34.09
31.62
31.00
30.40
28.64
27.99
26.56
26.17
25.99
19.44
18.34
18.27
155.60
139.81
121.44
119.77
107.28
92.94
91.62
56.61
48.75
48.37
42.18
37.24
37.20
34.09
31.62
31.00
30.40
28.64
27.99
26.56
26.17
25.99
19.44
18.34
18.27
14.90
8.07
-2.04
-2.20
-2.77
-4.06
f1 (ppm)

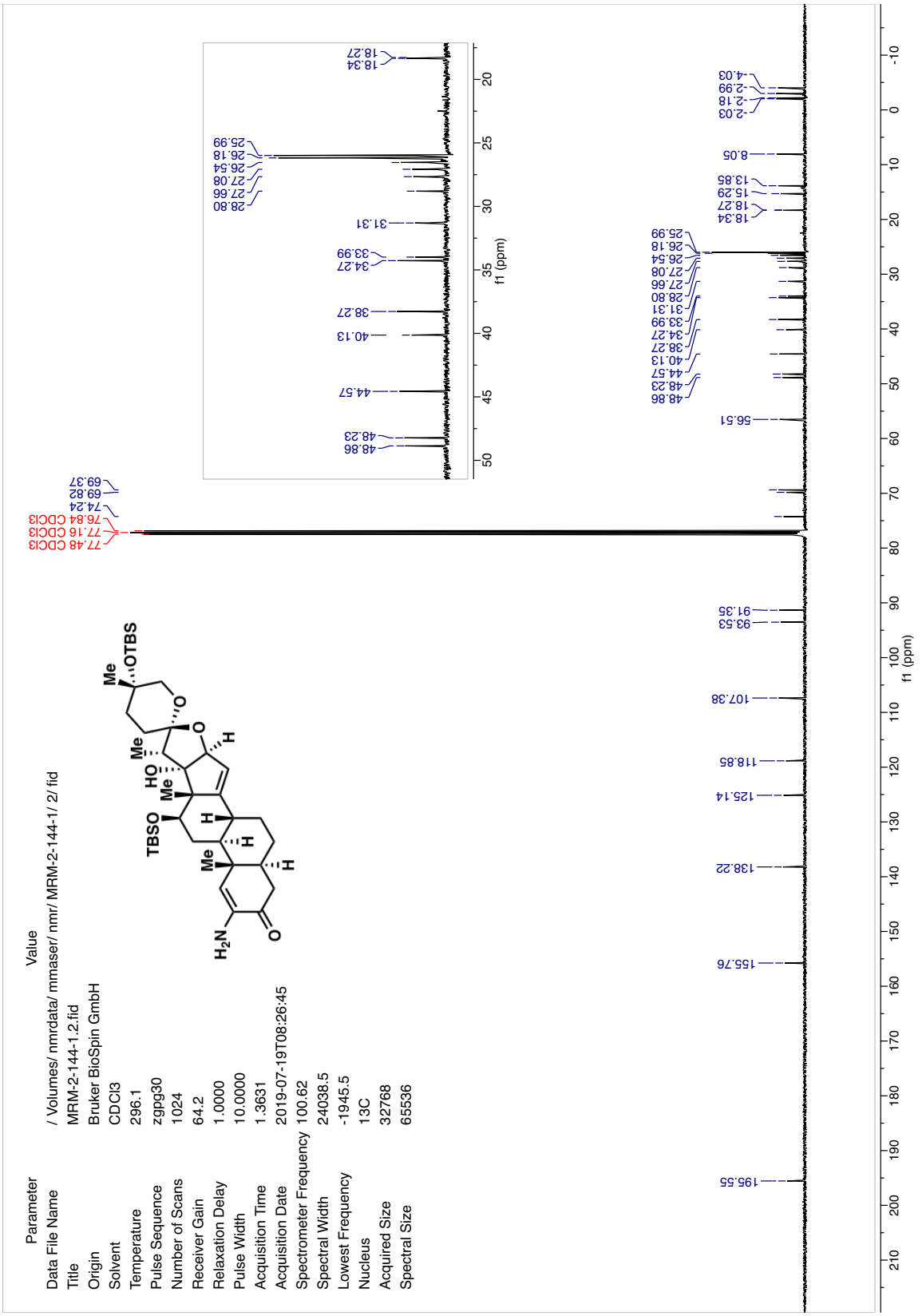

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ yasuaki/ vnmrsys/ data/ YN-2-200/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 16
Receiver Gain 24
Relaxation Delay 2.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2018-04-13T23:20:38
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1029.9
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
TBSO
HO
OTBS
f1 (ppm)

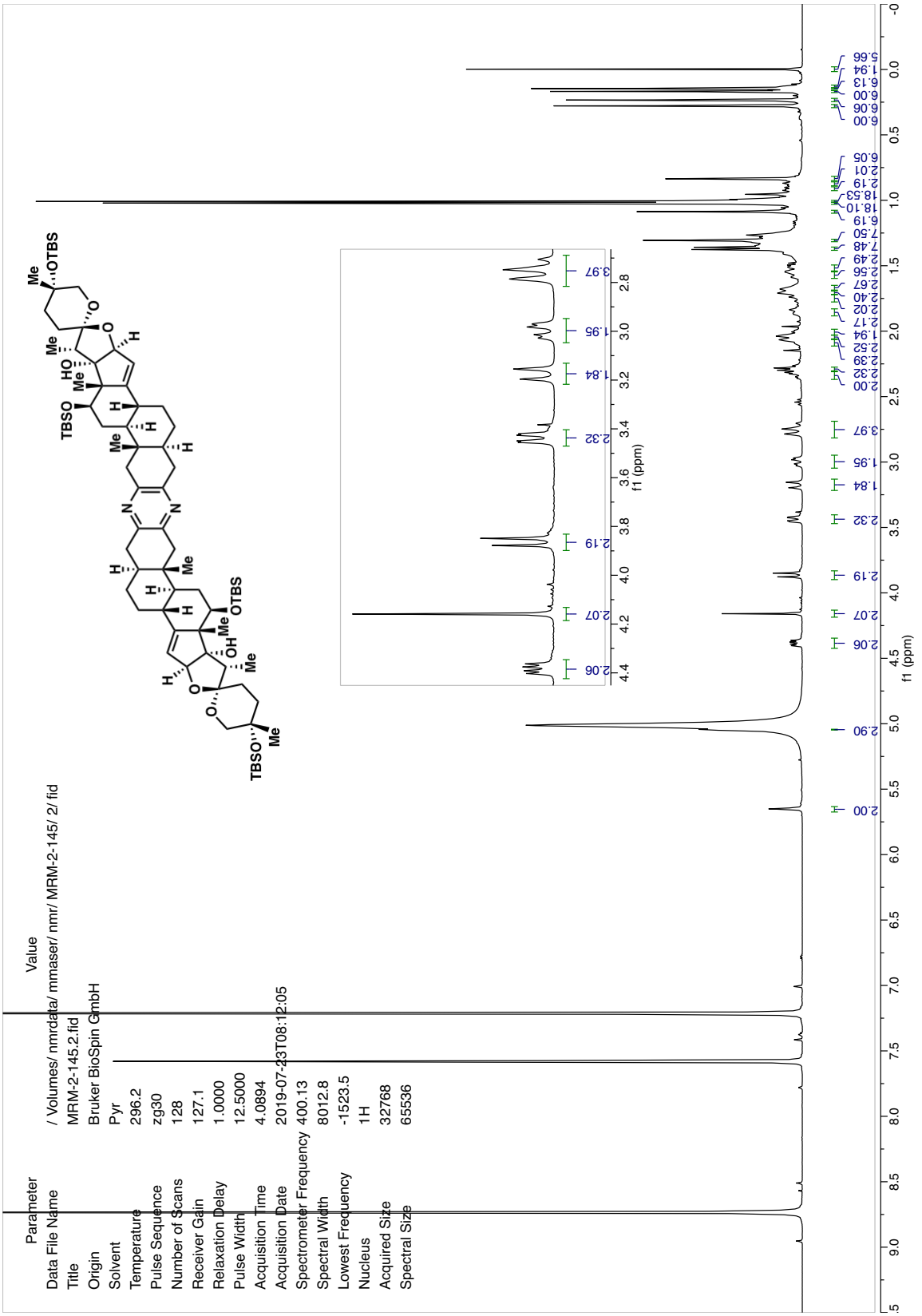

Parameter Value
Data File Name / Volumes/ nmrdata-1/ yasuaki/ vnmrsys/ data/ YN-2-193/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 16
Receiver Gain 52
Relaxation Delay 2.0000
Pulse Width 5.8000
Acquisition Time 3.0000
Acquisition Date 2018-04-04T18:19:42
Spectrometer Frequency 499.64
Spectral Width 8000.0
Lowest Frequency -1030.0
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
Me
OTBS
HO
Me
TBSO
Me
H
H
H
H
O
O
O
f1 (ppm)

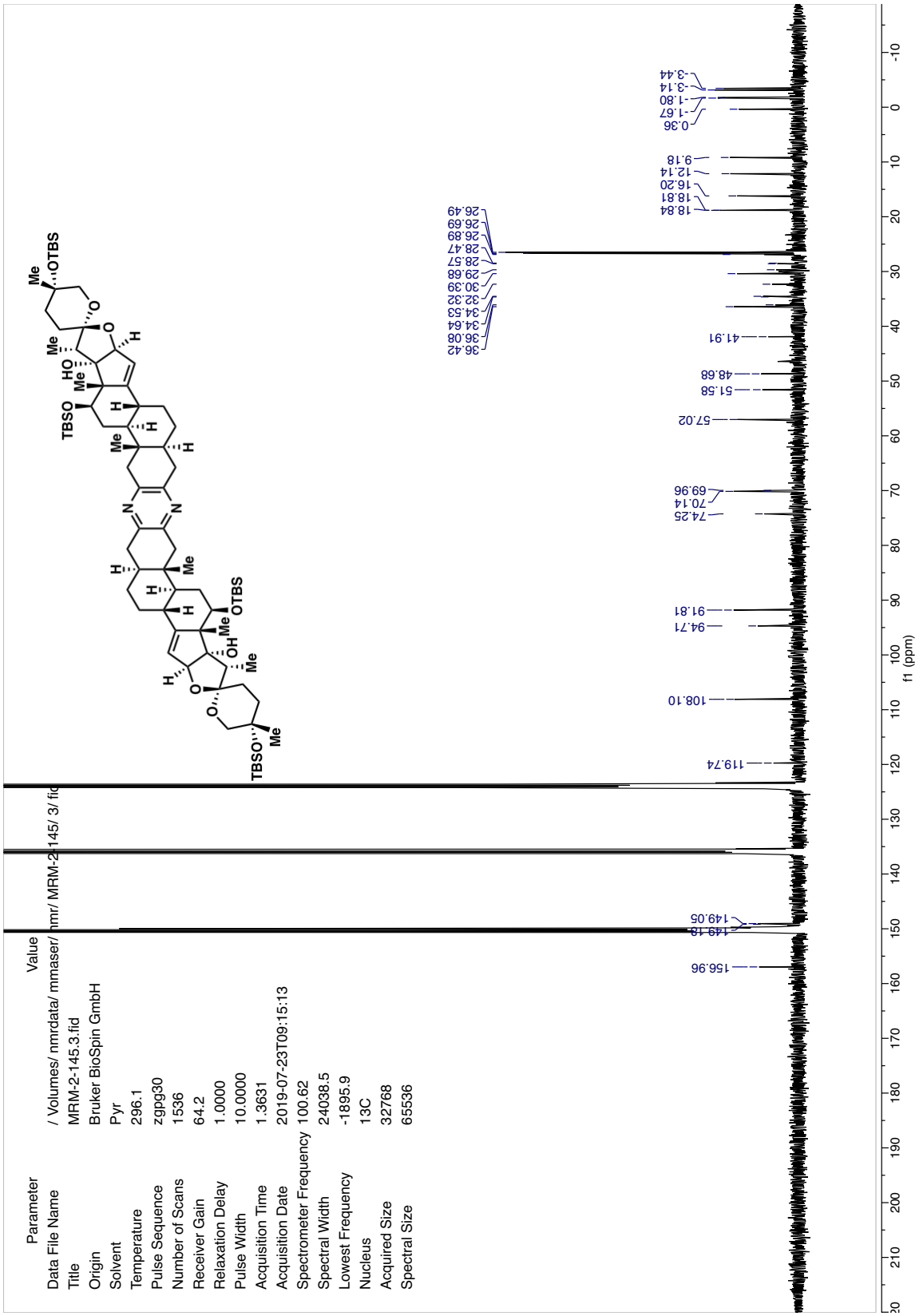

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-138/ 1/ fid
Title MRM-2-138.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-07-10T12:02:35
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
OTBS
HO
TBSO
Br
O
H
f1 (ppm)

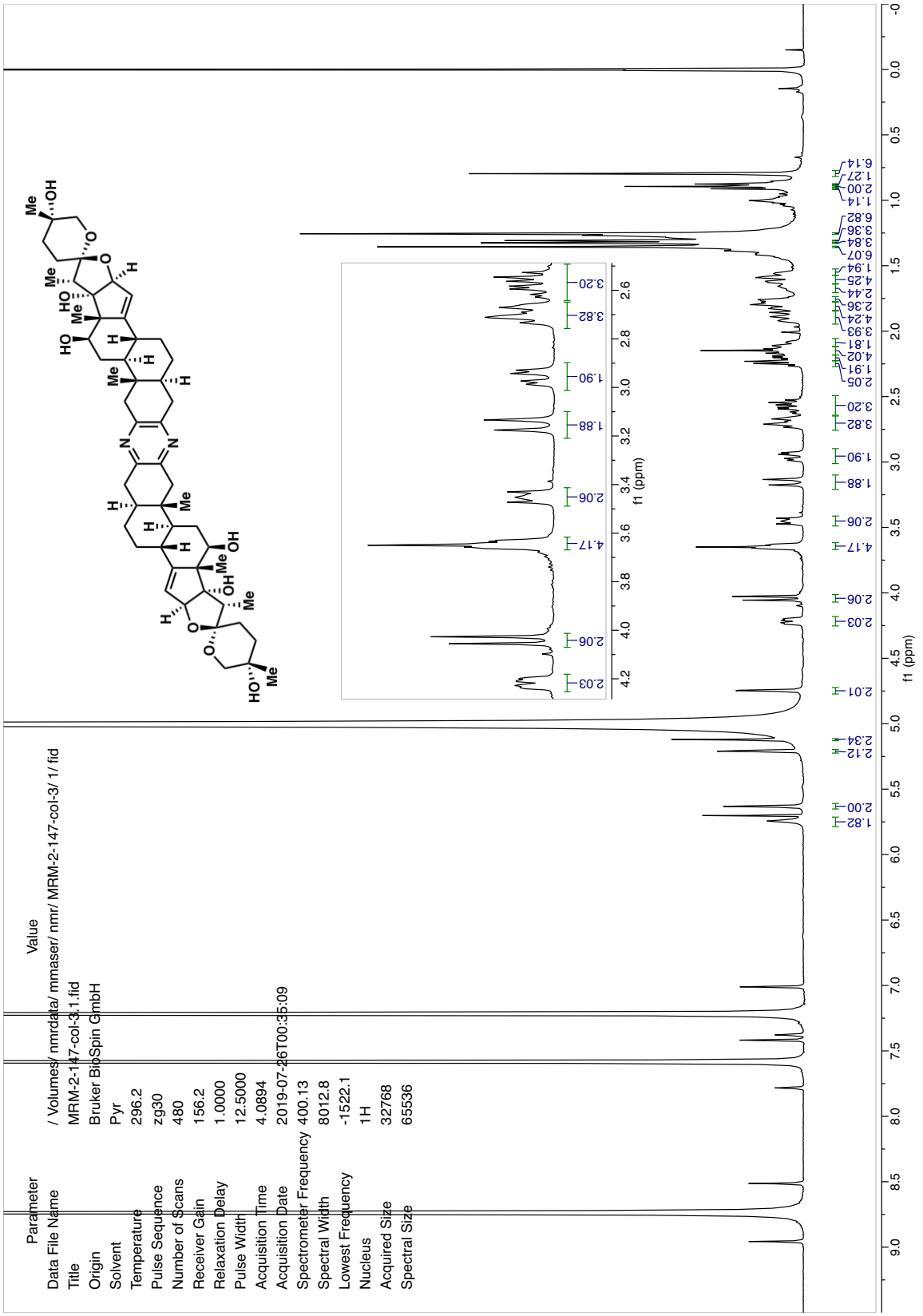

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-138/ 2/ fid
Title MRM-2-138.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.2
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 72.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-07-10T12:23:48
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.8
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
HO
OTBS
Br
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
74.05
69.83
69.35
200.68
155.21
119.22
107.37
93.43
91.37
56.41
54.00
51.22
51.16
48.23
47.00
43.78
39.20
33.99
33.53
31.39
29.85
28.81
28.02
27.66
26.53
26.16
25.98
18.33
18.24
15.17
12.13
8.04
-2.04
-2.19
-3.02
-4.10
f1 (ppm)

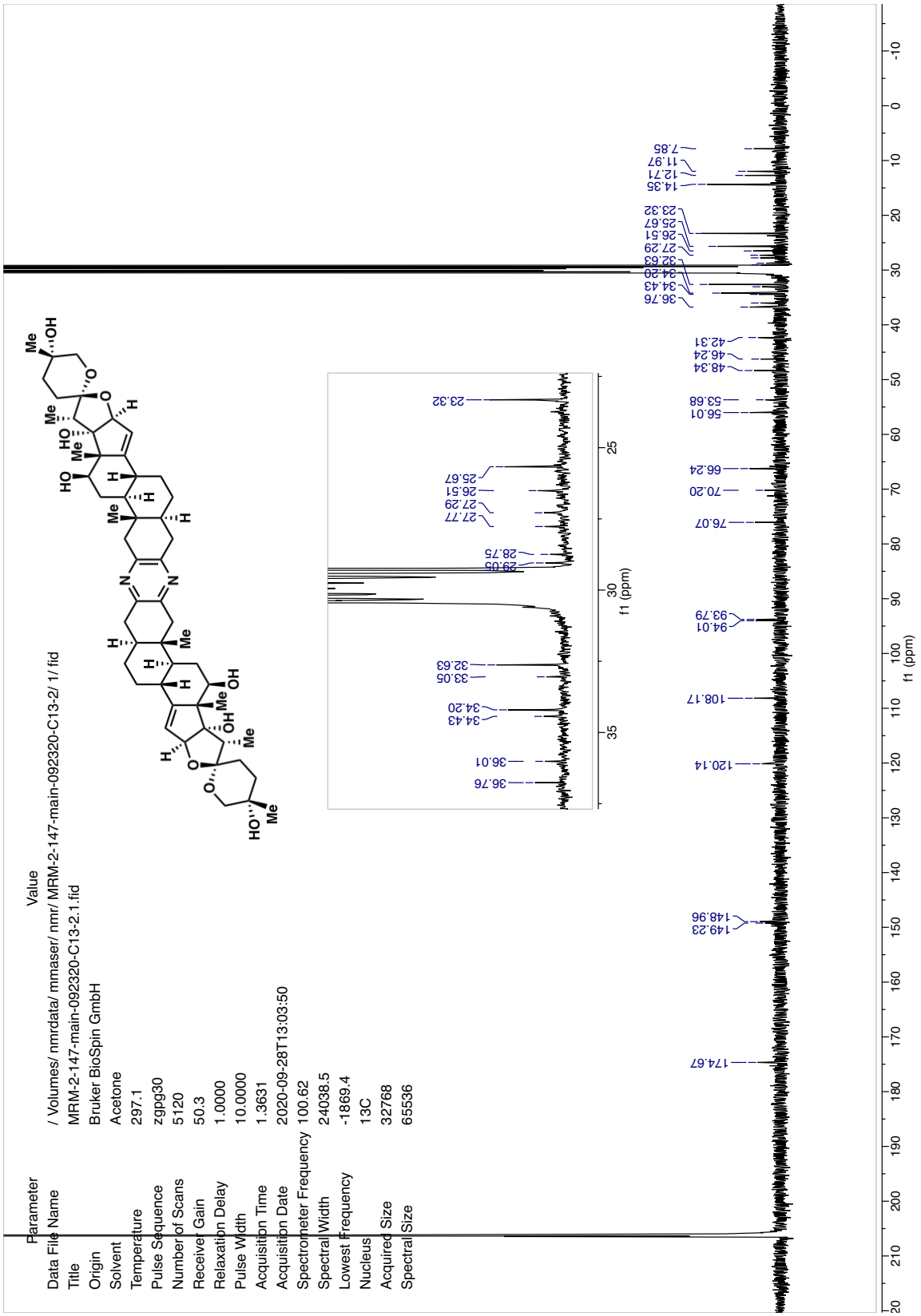

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-144-1/ 1/ fid
Title MRM-2-144-1.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.1
Pulse Sequence zg30
Number of Scans 128
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-07-19T07:44:49
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.4
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Me
OTBS
HO
Me
TBSO
Me
H
H
H
H
Me
H2N
O
O
O
f1 (ppm)

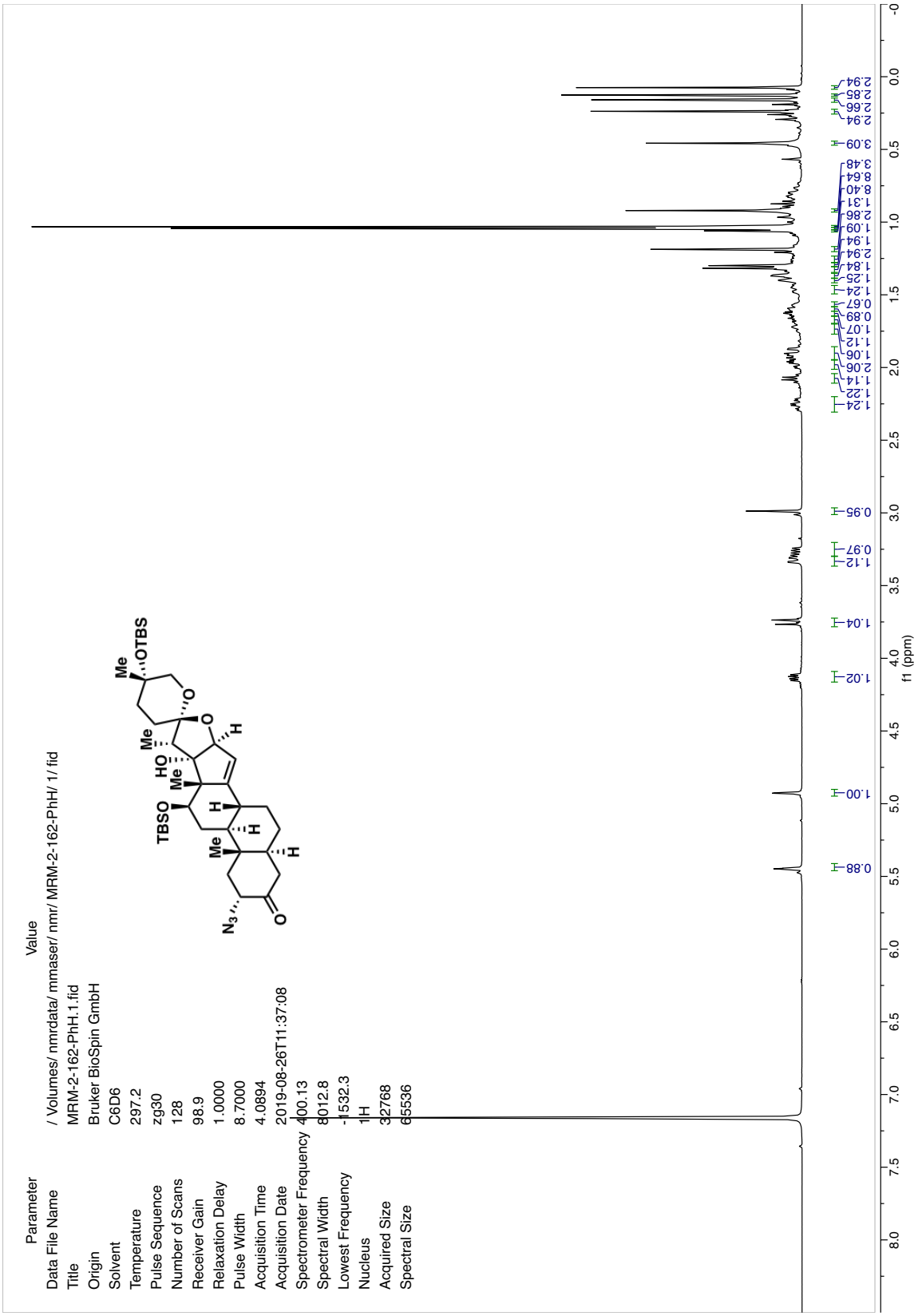

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-144-1/ 2/ fid
Title MRM-2-144-1.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 296.1
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-07-19T08:26:45
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.5
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OTBS
TBSO
HO
Me
H
H2N
O
77.48 CDCl3
77.16 CDCl3
76.84 CDCl3
74.24
69.82
69.37
195.55
155.76
138.22
125.14
118.85
107.38
93.53
91.35
56.51
48.86
48.23
44.57
40.13
38.27
34.27
33.99
31.31
28.80
27.66
27.08
26.54
26.18
25.99
18.34
18.27
15.29
13.85
8.05
-2.03
-2.18
-2.99
-4.03
f1 (ppm)

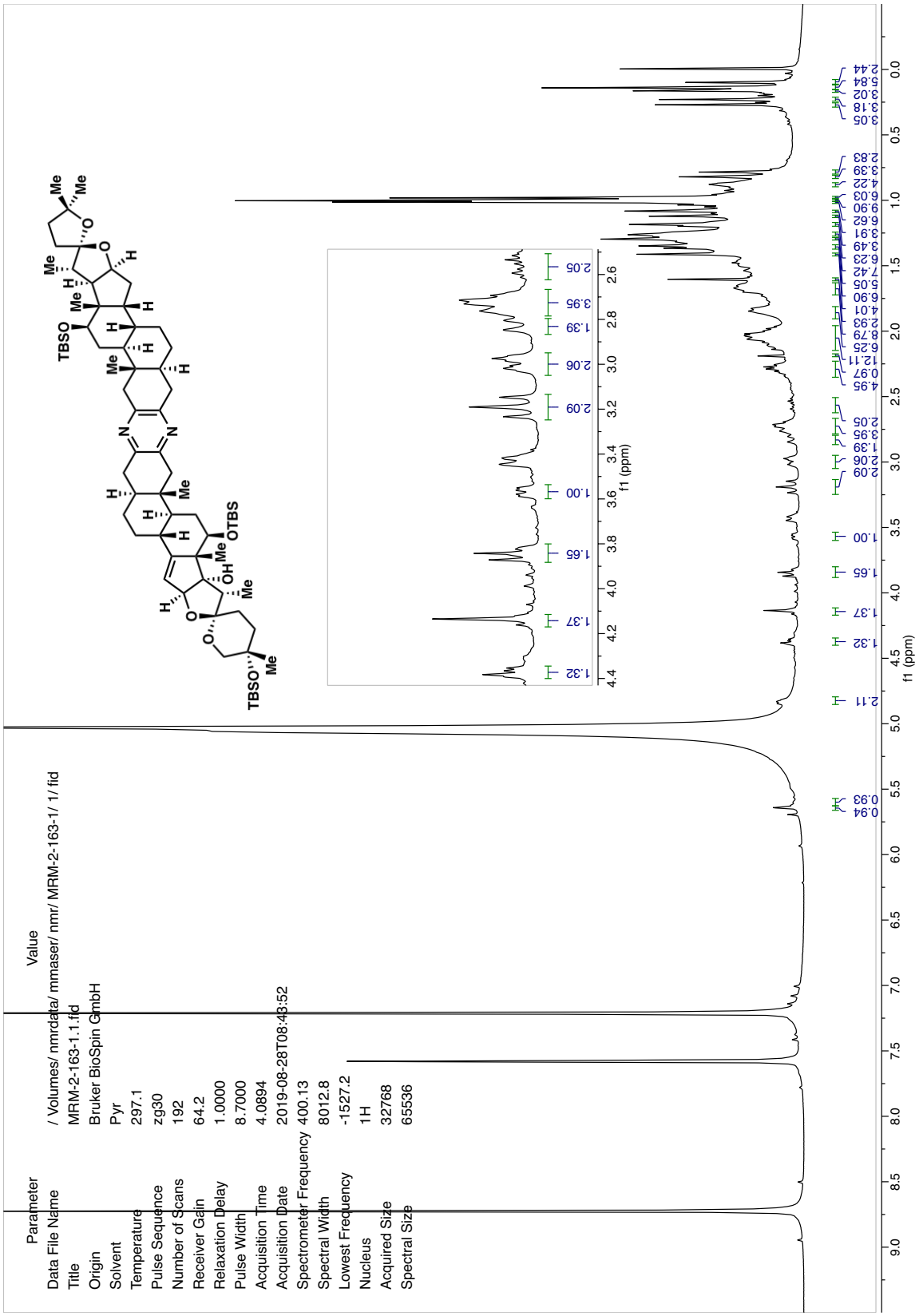

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-145/ 2/ fid
Title MRM-2-145.2.fid
Origin Bruker BioSpin GmbH
Solvent Pyr
Temperature 296.2
Pulse Sequence zg30
Number of Scans 128
Receiver Gain 127.1
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-07-23T08:12:05
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1523.5
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
f1 (ppm)

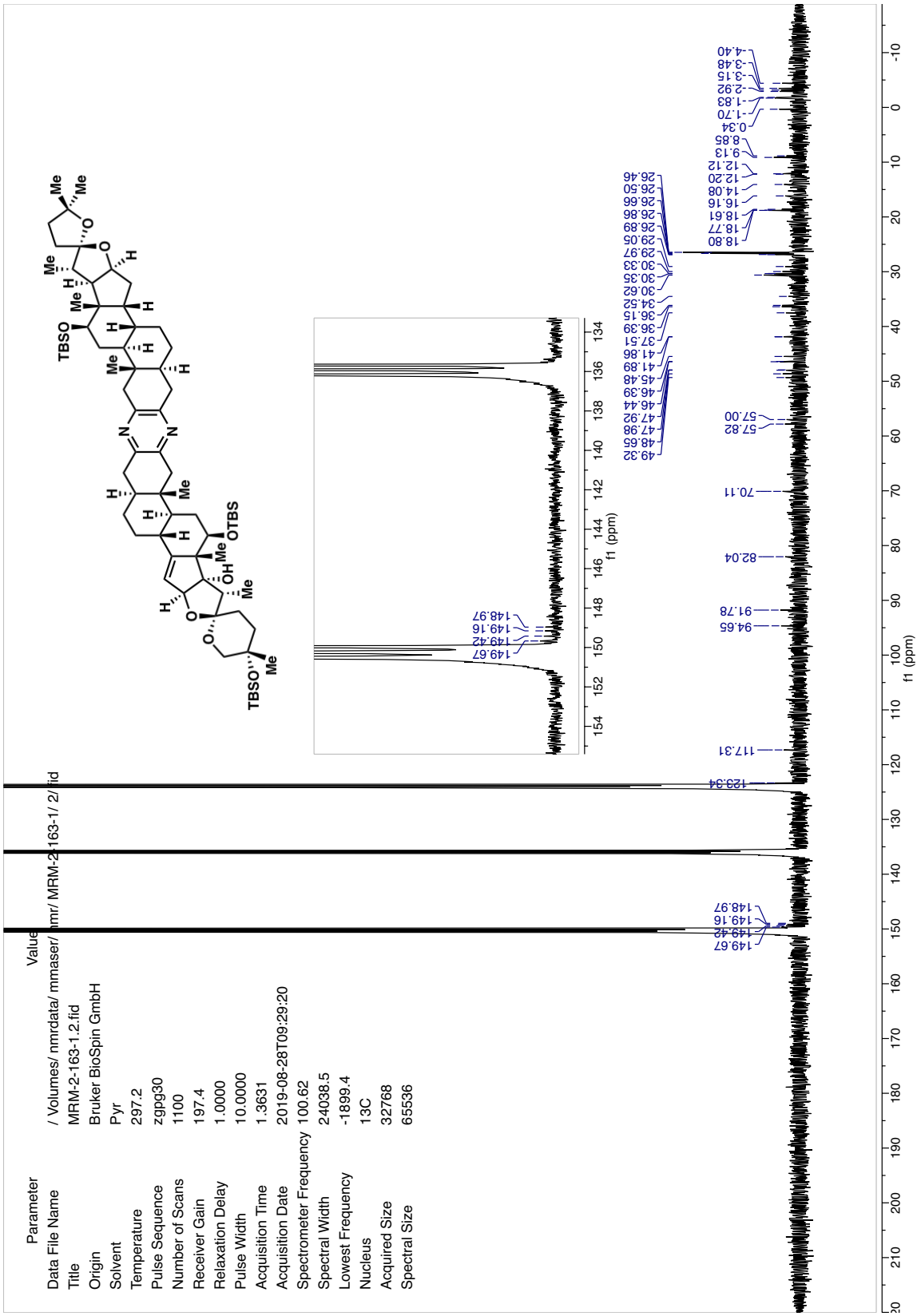

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ hmr/ MRM-2-145/ 3/ fid
Title MRM-2-145.3.fid
Origin Bruker BioSpin GmbH
Solvent Pyr
Temperature 296.1
Pulse Sequence zgpg30
Number of Scans 1536
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-07-23T09:15:13
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1895.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
HO
OTBS
OH
Me
H
N
O
156.96
149.18
149.05
119.74
108.10
94.71
91.81
74.25
70.14
69.96
57.02
51.58
48.68
41.91
36.42
36.08
34.64
34.53
32.32
30.39
29.68
28.57
28.47
26.89
26.69
26.49
18.84
18.81
16.20
12.14
9.18
0.36
-1.67
-1.80
-3.14
-3.44
f1 (ppm)

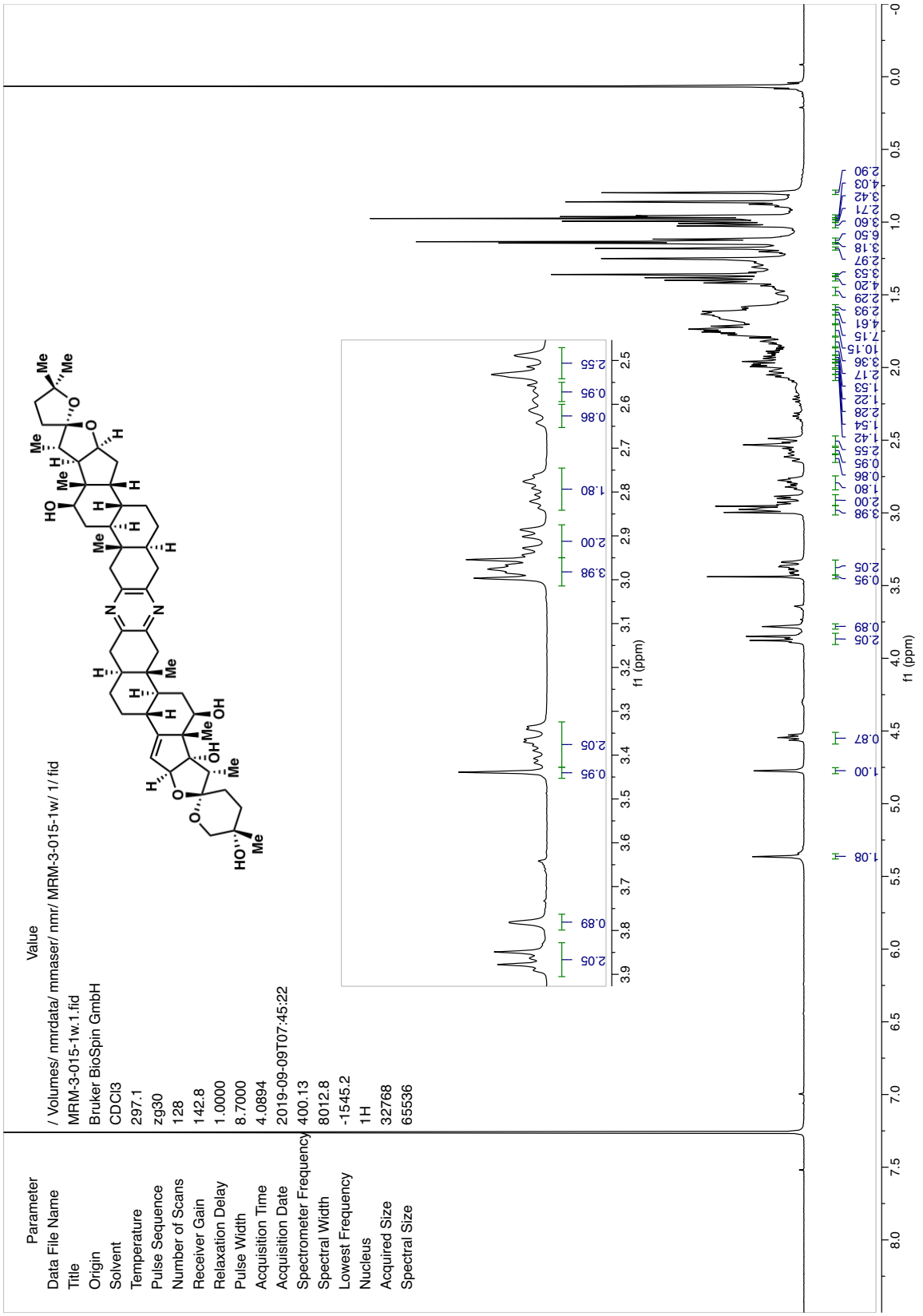

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-147-col-3/ 1/ fid
Title MRM-2-147-col-3.1.fid
Origin Bruker BioSpin GmbH
Solvent Pyr
Temperature 296.2
Pulse Sequence zg30
Number of Scans 480
Receiver Gain 156.2
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-07-26T00:35:09
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1522.1
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
f1 (ppm)

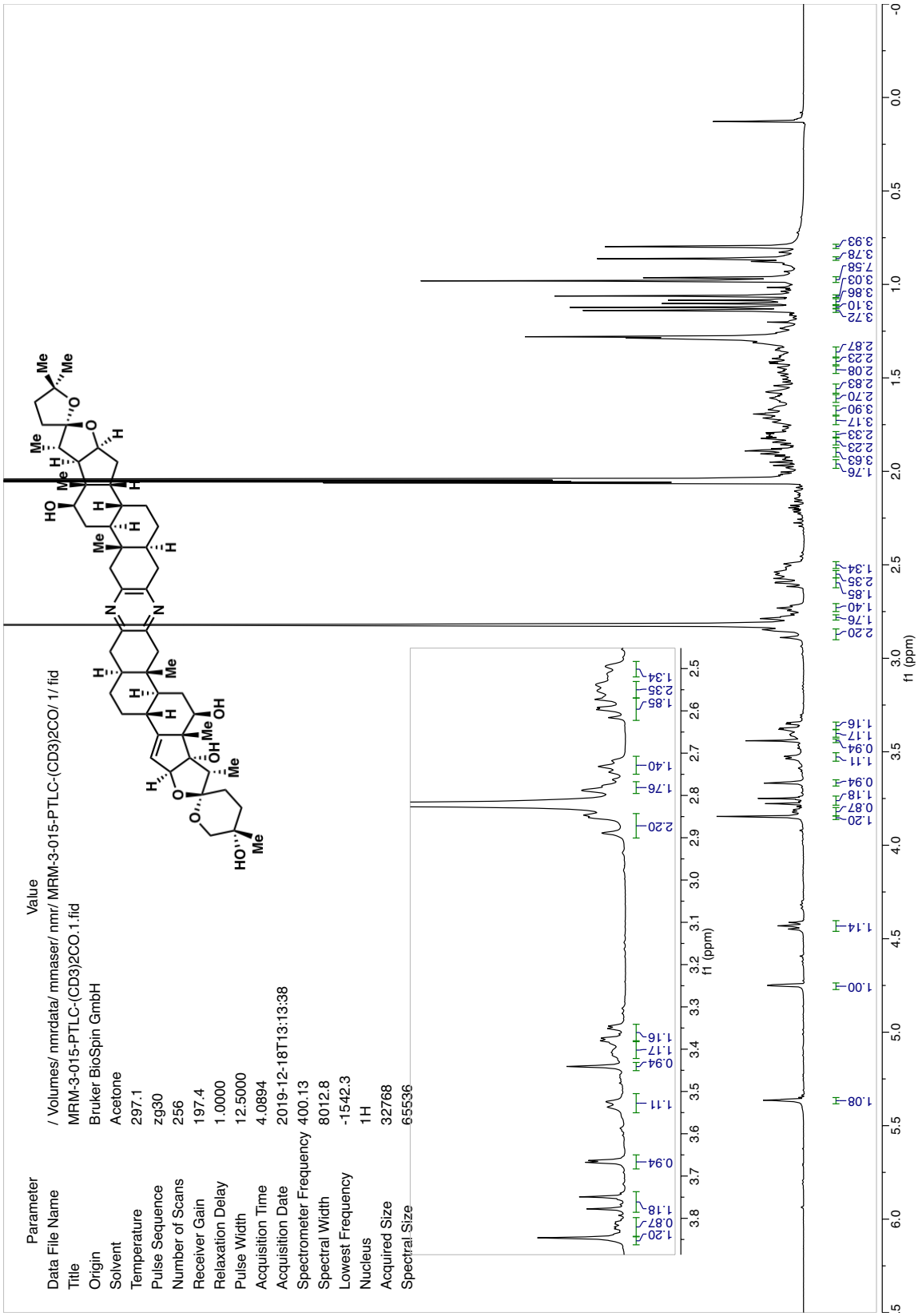

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-147-main-092320-C13-2/ 1/ fid
Title MRM-2-147-main-092320-C13-2.1.fid
Origin Bruker BioSpin GmbH
Solvent Acetone
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 5120
Receiver Gain 50.3
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-09-28T13:03:50
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1869.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
174.67
149.23
148.96
120.14
108.17
94.01
93.79
76.07
70.20
66.24
56.01
53.68
48.34
46.24
42.31
36.76
34.43
34.20
32.63
27.29
26.51
25.67
23.32
14.35
12.71
11.97
7.85
f1 (ppm)

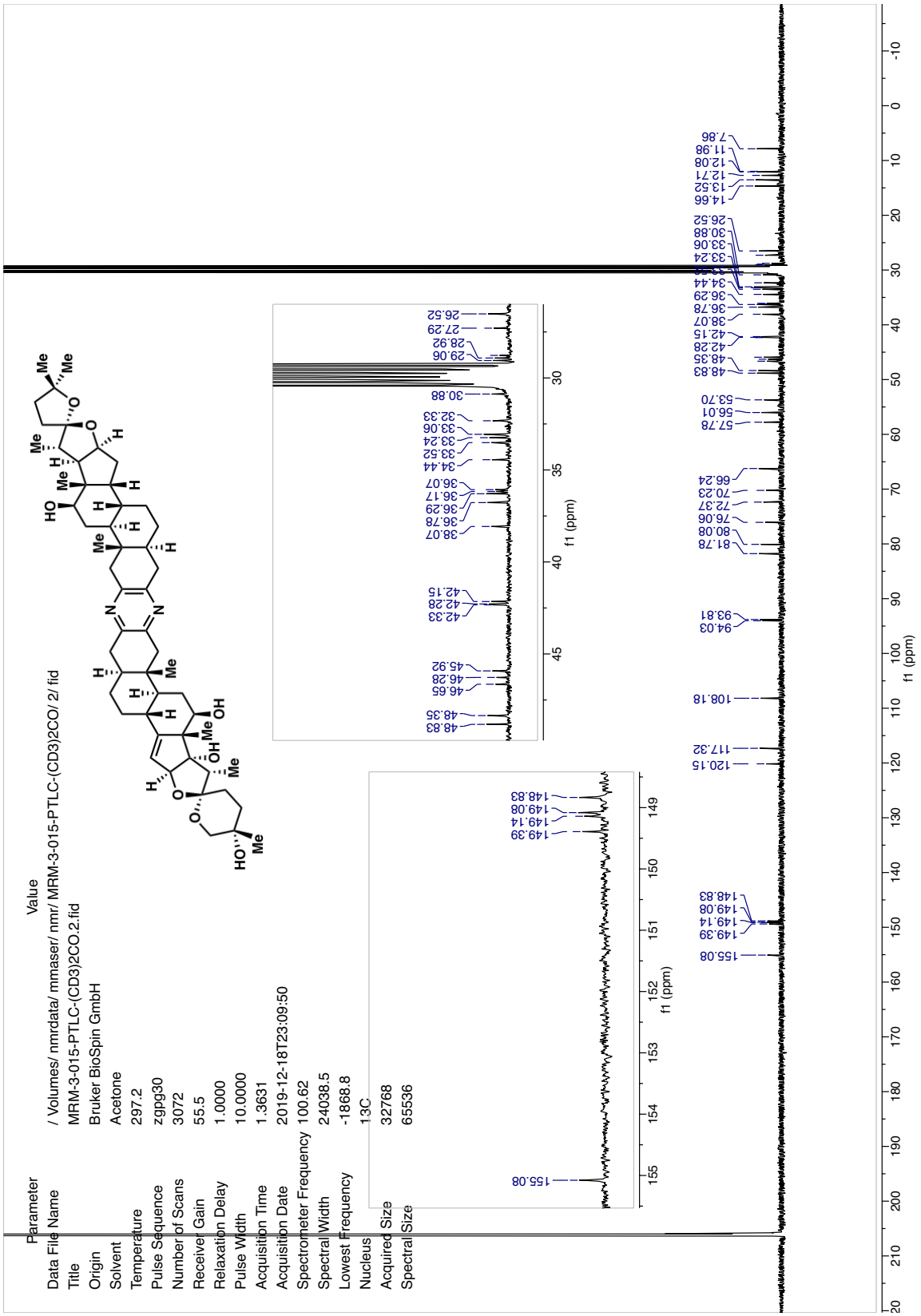

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-162-PhH/ 1/ fid
Title MRM-2-162-PhH.1.fid
Origin Bruker BioSpin GmbH
Solvent C6D6
Temperature 297.2
Pulse Sequence zg30
Number of Scans 128
Receiver Gain 98.9
Relaxation Delay 1.0000
Pulse Width 8.7000
Acquisition Time 4.0894
Acquisition Date 2019-08-26T11:37:08
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1532.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
f1 (ppm)


Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-163-1/ 1/ fid
Title MRM-2-163-1.1.fid
Origin Bruker BioSpin GmbH
Solvent Pyr
Temperature 297.1
Pulse Sequence zg30
Number of Scans 192
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 8.7000
Acquisition Time 4.0894
Acquisition Date 2019-08-28T08:43:52
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1527.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
TBSO
OTBS
OH
f1 (ppm)


Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-2-163-1/ 2/ fid
Title MRM-2-163-1.2.fid
Origin Bruker BioSpin GmbH
Solvent Pyr
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1100
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-08-28T09:29:20
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1899.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
TBSO
OTBS
TBSO
f1 (ppm)


Parameter
Value
Data File Name
/ Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-015-1w/ 1/ fid
Title
MRM-3-015-1w.1.fid
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.1
Pulse Sequence
zg30
Number of Scans
128
Receiver Gain
142.8
Relaxation Delay
1.0000
Pulse Width
8.7000
Acquisition Time
4.0894
Acquisition Date
2019-09-09T07:45:22
Spectrometer Frequency
400.13
Spectral Width
8012.8
Lowest Frequency
-1545.2
Nucleus
1H
Acquired Size
32768
Spectral Size
65536
f1 (ppm)


Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-015-PTLC-(CD3)2CO/ 1/ fid
Title MRM-3-015-PTLC-(CD3)2CO.1.fid
Origin Bruker BioSpin GmbH
Solvent Acetone
Temperature 297.1
Pulse Sequence zg30
Number of Scans 256
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-12-18T13:13:38
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1542.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
f1 (ppm)


Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ nmr/ MRM-3-015-PTLC-(CD3)2CO/ 2/ fid
Title MRM-3-015-PTLC-(CD3)2CO.2.fid
Origin Bruker BioSpin GmbH
Solvent Acetone
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 3072
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-12-18T23:09:50
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1868.8
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
f1 (ppm)

Chapter 3

Multilabel Classification Models for the Prediction of Substrate-Specific Organic Reaction Conditions[1]

3.1 INTRODUCTION

A common roadblock encountered in organic synthesis occurs when canonical conditions for a given reaction type fail in complex molecule settings.[1] Optimizing these reactions frequently requires iterative experimentation that can slow progress, waste material, and add significant costs to research.[2] This is especially prevalent in catalysis, where the substrate-specific nature of reported conditions is often deemed a major drawback, leading to the slow adoption of new methods.[1–3] If, however, a transformation's structure-reactivity relationships (SRRs) were well-known or predictable, this roadblock could be avoided and new reactions could see much broader use in the field.[4]

Machine learning (ML) algorithms have demonstrated great promise as predictive tools for chemistry domain tasks.[5] Strong approaches to molecular property prediction[6–9] and generative design[10–13] have been developed, particularly in the field of medicinal chemistry.[14] Some applications have emerged in organic synthesis, geared mainly towards

[1] Portions of this chapter were adapted from the following communications: Ryou*, S.; Maser*, M. R.; Cui*, A. Y.; DeLano, T. J.; Yue, Y.; Reisman, S. E. Graph Neural Networks for the Prediction of Substrate-Specific Organic Reaction Conditions. *arXiv:2007.04275*, **2020**, *Preprint*. Copyright © 2020 The Authors; and Maser*, M. R.; Cui*, A. Y.; Ryou*, S.; DeLano, T. J.; Yue, Y.; Reisman, S. E. Multilabel Classification Models for the Prediction of Cross-Coupling Reaction Conditions. *J. Chem. Inf. Model.* **2021**, *61*, 156–166. DOI: 10.1021/acs.jcim.0c01234. Copyright © 2021 American Chemical Society. The research discussed in this chapter was completed in collaboration with former graduate student Dr. Serim Ryou of the Perona Group, former undergraduate Alexander Y. Cui of the Yue Group, and current graduate student Travis J. DeLano of the Reisman Lab.

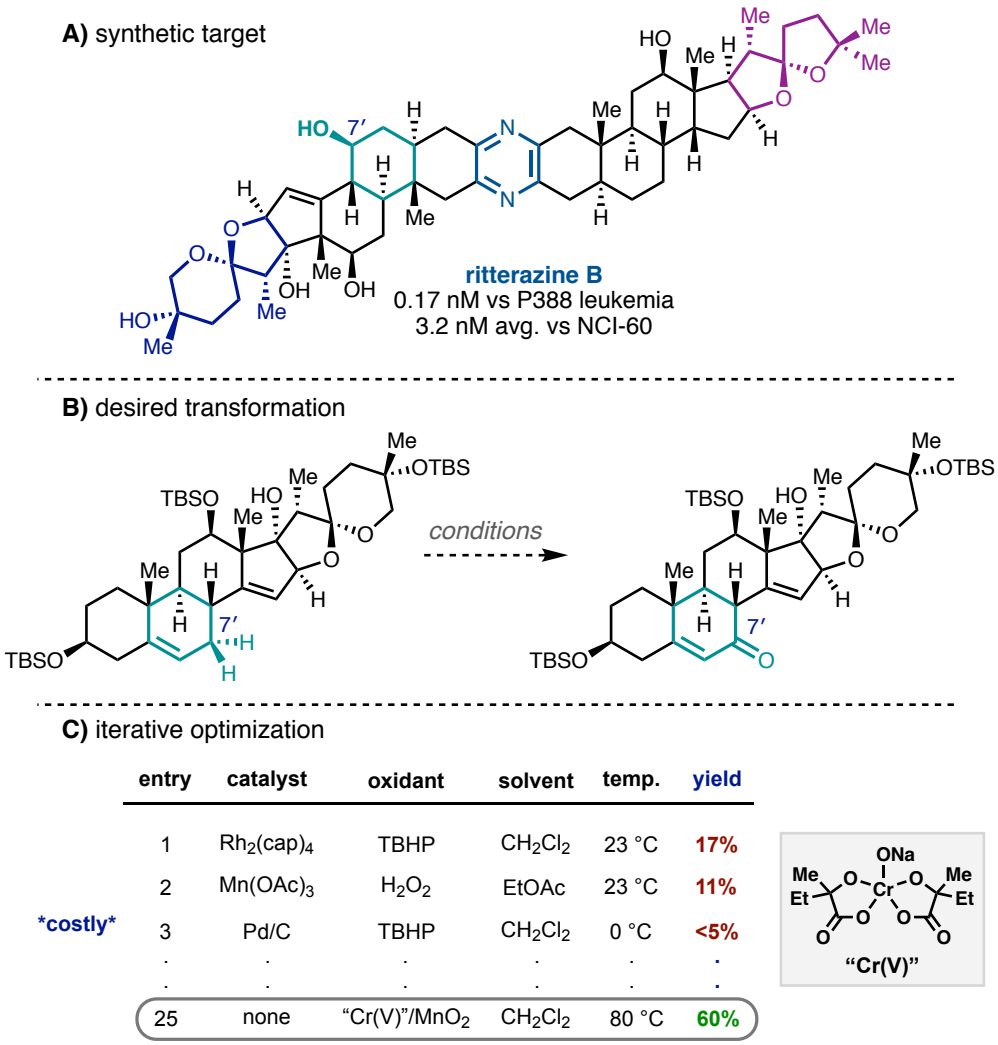
predicting reaction products,[15,16] yield,[17–20] and selectivity.[21–25] Significant effort has also been invested in computer-aided synthesis planning (CASP)[26] and the development of retrosynthetic design algorithms.[27–30]

To supplement these tools, initial attempts have been made to predict reaction conditions in the forward direction based on the substrates and products involved.[31] Thus far, studies have focused on global datasets with millions of data points of mixed reaction types. Advantages of this approach include ample training data and the ability to query any transformation with a single model. However, the sparse representation of individual reactions is a major drawback, in that reliable predictions can likely only be expected for the most common reactions and conditions within. This precludes the ability to distinguish subtle variations in substrate structures that lead to different condition requirements, which is critical for SRR modeling.

In recent years, it has become a goal of ours to develop predictive tools to overcome challenges in selecting substrate-specific reaction conditions. Towards this end, we initiated a collaboration with researchers in Computing and Mathematical Sciences at Caltech. We identified a challenging reaction from our group's recent synthesis of the anticancer marine natural product ritterazine B[32] (see Chapter 2) as a starting point for ML investigations (**Figure 3.1A**). The route relies on a critical allylic oxidation to install oxygenation at C7′ of the natural product, which had previously not been achieved (**Figure 3.1B**). Though a seemingly straightforward reaction, a slew of canonical allylic oxidation conditions failed to provide the desired product in appreciable yields (**Figure 3.1C**). After months of iterative experimentation, a rather contrived oxidation reagent ("**Cr(V)**") was

eventually identified for successful conversion (~60% isolated yield). This process is highly representative of optimization challenges frequented in complex synthesis.

Figure 3.1 *Project motivation. A) Synthetic target ritterazine B. B) Desired allylic oxidation at C7′ en route to target. C) Costly, iterative experimental optimization.*



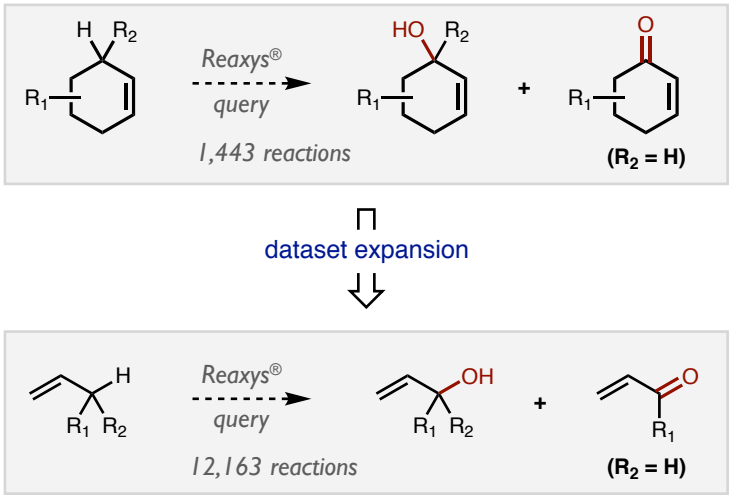
entry	catalyst	oxidant	solvent	temp.	yield
1	$Rh_2(cap)_4$	TBHP	CH_2Cl_2	23 °C	17%
2	$Mn(OAc)_3$	H_2O_2	EtOAc	23 °C	11%
3	Pd/C	TBHP	CH_2Cl_2	0 °C	<5%
.	.	.	.	.	.
.	.	.	.	.	.
25	none	"Cr(V)"/MnO_2	CH_2Cl_2	80 °C	60%

3.2 INITIAL STUDIES

We hypothesized that there would be sufficient precedent in the synthetic literature to develop data-driven SRR models of allylic oxidation reactions that would allow us to streamline the optimization phase. Indeed, a query of the Reaxys® database returned nearly

1,500 data points for allylic oxidations of cyclohexenes alone (**Figure 3.2**, top). In preliminary experiments, we found this dataset rather challenging to model (data not shown) and elected to expand the chemical space to include all allylic and benzylic oxidations (**Figure 3.2**, bottom). This dataset contained over 16,000 reactions, and was filtered to remove reactions of salts, with multiple products, and without yields or reagents. The resulting dataset contained 12,163 reactions.

Figure 3.2 *Dataset acquisition for cyclohexene (top) and general (bottom) allylic oxidations.*



We desired an ML model that would predict reaction yield based on the oxidation substrate and/or product structure and reaction conditions. To construct reaction encodings, we tested a hybrid 1D molecular representation concatenating tokenized SMILES strings and descriptor vectors (*vide infra*). These were further concatenated with binary vectors of the reagents and conditions utilized. The raw data from Reaxys® featured over 1200 unique reagents; we manually assigned bin labels to each reagent in order to reduce the dimensionality and noise of this encoding region. The full set of reagent class bins is shown in Table 3.1 below.

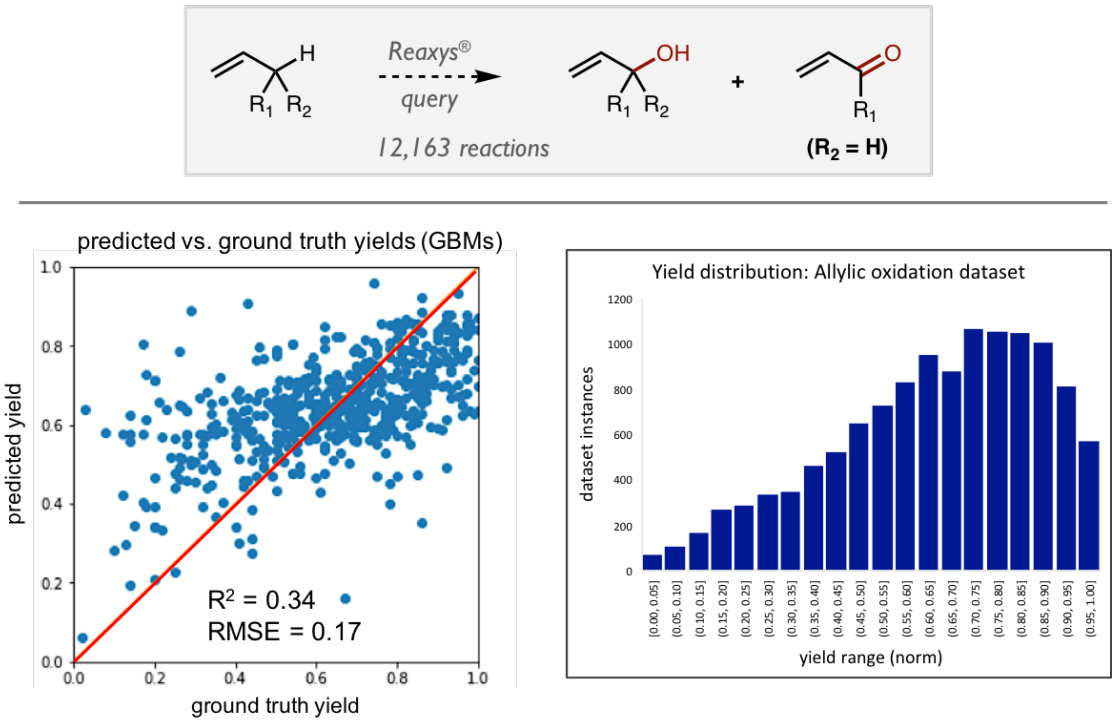
Table 3.1 *Reagent class bins for allylic oxidation dataset*

bin #	reagent	bin #	reagent	bin #	reagent	bin #	reagent
1	peroxides	18	(hydr/alk)oxide base	35	lead	52	enzyme/bio
2	chromium	19	manganese	36	rhodium	53	other
3	oxygen	20	strong acid	37	ozone	54	quinone
4	weak acid	21	dehydrating agent	38	light sensitizer	55	P ligand
5	selenium	22	copper	39	ionic liquid	56	reducing agent
6	amine base	23	silica	40	phosphite	57	nitrite
7	pyrazole	24	hypervalent Cl/Br	41	MX salt	58	silver
8	acid anhydride	25	DMSO	42	iron	59	bipyridine
9	permanganate	26	ruthenium	43	NHP	60	crown ether
10	DDQ	27	peracid	44	molybdenum	61	iridium
11	water	28	radical initiator	45	phosphate base	62	amide
12	carbonate base	29	CAN	46	(thio)sulfate/ites	63	surfactant
13	celite	30	TBAX salts	47	metal oxide	64	alcohol
14	halogenating	31	palladium	48	zinc	65	N ligand
15	acetate base	32	peroxysulfate/ites	49	N-oxide	66	Lewis acid
16	TEMPO	33	strong base	50	O-electrophile	67	tungsten
17	hypervalent iodine	34	cobalt	51	cerium	68	nickel

We tested a variety of model types for the task of predicting numerical yields, which were normalized between 0–1 (data not shown). We found gradient-boosting machines (GBMs, *vide infra*)[33] to be most effective (**Figure 3.3**, bottom-left), giving a root mean squared error (RMSE) of 0.17 (i.e., 17% yield) and a correlation coefficient (R^2) of 0.34. These metrics approach those of state-of-the-art yield predictors trained on uniform and rigorously obtained high-throughput datasets.[17-20] This perhaps speaks to the challenge of modeling literature reaction data, about which others have provided perspectives in the years surrounding our reports.[34] However, we were unsatisfied with the results obtained and sought a more accurate system. When analyzing the target distribution in the dataset, we noticed an unsurprising bias towards high-yielding reactions (**Figure 3.3**, bottom-right, *vide infra*). In line with this skewness, we also recognized a tendency in our models to

predict higher yields than the ground truth (**Figure 3.3**, bottom-left). We thus hypothesized that increasing the number of data points in the low yield regime would benefit modeling.

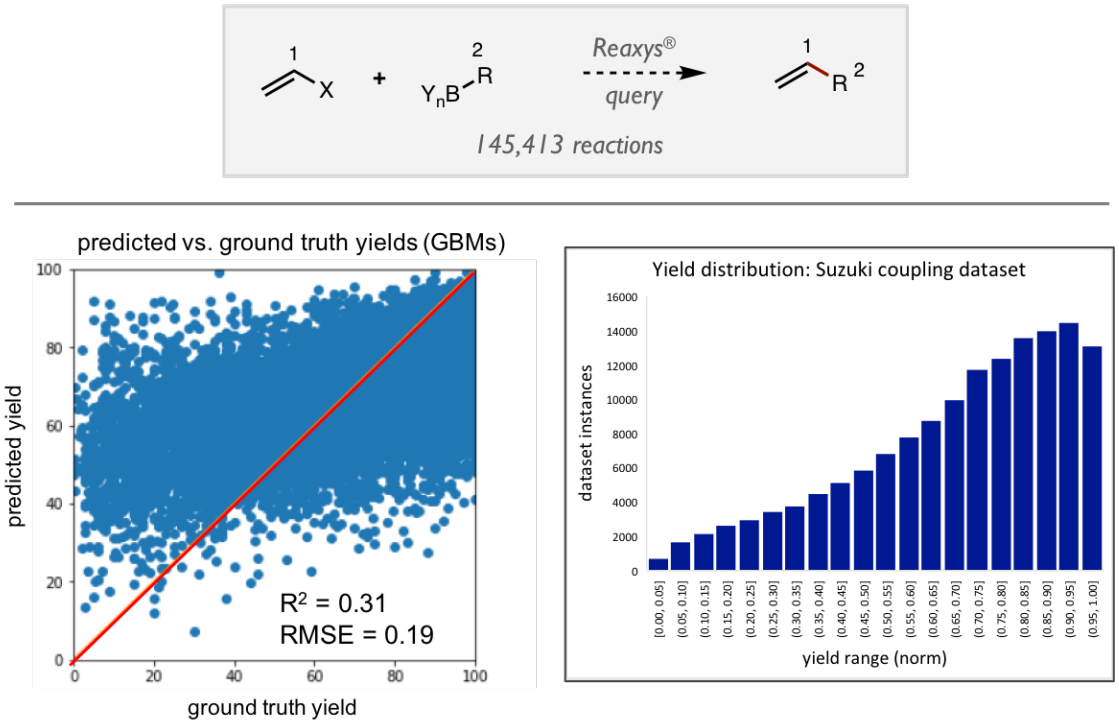
Figure 3.3 *Modeling results (left) and reaction yield distribution (right) for the allylic oxidation dataset (top).*



Having already expanded the chemical space of the allylic oxidation dataset, we turned our attention to another reaction type we expected would have much more data, namely, Suzuki cross-couplings.[35a] Indeed, a Reaxys® query for the reaction of organohalides and organoboron species returned a dataset of 145,413 reactions after the same pruning procedure described above (**Figure 3.4**, top). This dataset contained nearly 200 reagent classes after binning (*vide infra*). Unfortunately, the yields in this dataset were perhaps even more skewed than those of the allylic oxidation, centered around 85% (**Figure 3.4**, bottom-right). As expected, we observed a similar bias in GBM models for

predicting higher yields than the ground truth and obtained similar metrics as before, with a RMSE of 0.19 and R^2 of 0.31 (**Figure 3.4**, bottom-left).

Figure 3.4 *Modeling results (bottom-left) and reaction yield distribution (bottom-right) for the Suzuki coupling dataset (top).*



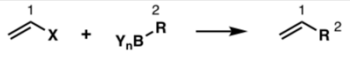
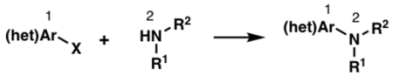
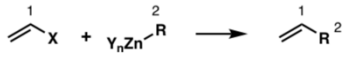
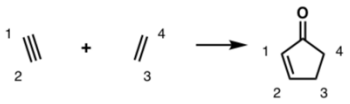
We do not rule out the possibility that other model types and reaction encodings could be more suitable for this modeling task. For example, (possibly recurrent) deep neural networks (DNNs) or graph convolutional networks (GCNs)[36] could provide enough capacity to overcome the apparent underfitting observed in the GBM experiments (*vide infra*). We further expect that careful data-splitting techniques such as under- or over-sampling in the different yield ranges could lead to more robust models. We leave these topics for future works.

3.3 REVISED APPROACH AND METHODS

Recognizing the challenge of predicting reaction yields from literature data, we elected to pursue a revised modeling approach. Instead of manually (or algorithmically) evaluating combinatorial condition vectors to obtain predicted yields for a given reaction, we anticipated that the direct prediction of suitable condition vectors may actually be more practically useful and could avoid the bias in regression targets. We thus devised a multi-label classification (MLC) approach for this task. We selected four high-value reaction types from the cross-coupling literature as testing grounds: Suzuki,[35a] C–N,[35b] and Negishi couplings,[35c] as well as Pauson-Khand reactions (PKRs).[35d]

3.3.1 DATA ACQUISITION AND PRE-PROCESSING

Table 3.2 *Statistical summary of reaction datasets with Reaxys® queries.*

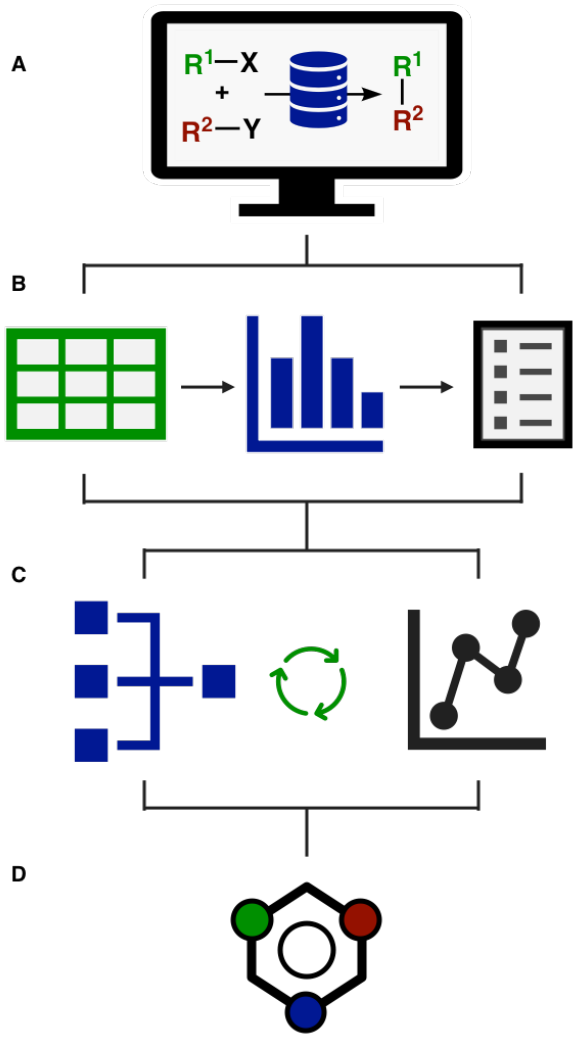
name	depiction	reactions	raw labels	label bins	categories
Suzuki		145,413	3,315	118	5
C–N		36,519	1,528	205	5
Negishi		6,391	492	105	5
PKR		2,749	335	83	8

A summary of the datasets studied here is shown in Table 3.2. Each dataset was manually pre-processed using the following procedure:[37]

1. Reaction data was exported from Reaxys® query results (**Figure 3.5A**);[34b,38]
2. SMILES strings[39] of coupling partners and major products were identified for each reaction entry (i.e., data point);

3. Condition labels including reagents, catalysts, solvents, temperatures, etc. were extracted for each data point (**Figure 3.5B**);
4. All unique labels were enumerated into a dataset dictionary, which was sorted by reaction role and trimmed at a threshold frequency to avoid sparsity;[40] and
5. Labels were re-indexed within categories and applied to the raw data to construct binary condition vectors for each reaction. We refer to this process as binning.

Figure 3.5 *Schematic modeling pipeline. A) Reaction querying and data gathering. B) Tabulation, label thresholding, and dictionary construction. C) Iterative model optimization. D) Inference and Interpretation.*



The reactions studied here were chosen for their ubiquity and value in synthesis, breadth of known conditions, and range of dataset size and chemical space. It should be noted that certain parameters (e.g. temperature, pressure, etc.) were more fully recorded in some datasets than others. In cases where this data was well-represented, reactions with

missing values were simply removed, or in the case of temperature and pressure were assumed to occur ambiently. However, when appropriate, these parameters were dropped from the prediction space to avoid discarding large portions of data.

The Suzuki dataset (**Table 3.2**, line 1) was obtained from a search of C–C-bond-forming reactions between C(sp^2) halides or pseudohalides and organoboron species. Data processing returned 145k reactions with 118 label bins in 5 categories. Similarly, the C–N coupling dataset (line 2) details reactions between aryl (pseudo)halides and amines, with 37k reactions and 205 bins in 5 categories. The Negishi dataset (line 3) contains C–C-bond-forming reactions between organozinc compounds and C(sp^2) (pseudo)halides. After processing, this dataset gave 6.4k reactions with 105 bins in 5 categories. The PKR dataset (line 4) describes couplings of C–C double bonds with C–C triple bonds to form the corresponding cyclopentenones, containing 2.7k reactions with 83 bins in 8 categories. For all datasets, atom mapping was used as depicted in Table 3.2 to ensure only the desired transformation type was obtained.[41] Samples of the C–N and Negishi label dictionaries are included in Table 3.3, and full dictionaries for all reactions are provided in the Experimental Section 3.6.

Table 3.3 *Samples of categorized reaction dictionaries for C–N and Negishi datasets.*

C-N coupling dictionary sample		
agent	**label**	**category**
CuI	M1	metal
$Pd_2(dba)_3$	M2	metal
$Pd(OAc)_2$	M3	metal
—	—	—
BINAP	L1	ligand
P(*t*-Bu)$_3$	L2	ligand
Xantphos	L3	ligand
—	—	—
NaO*t*-Bu	B1	base
K_2CO_3	B2	base
Cs_2CO_3	B3	base
—	—	—
toluene	S1	solvent
1,4-dioxane	S2	solvent
DMF	S3	solvent
—	—	—
18-crown-6	A1	additive
Bu_4NBr	A2	additive
8-quinolinol	A3	additive

Negishi coupling dictionary sample		
agent	**label**	**category**
$Pd(PPh_3)_4$	M1	metal
$Pd_2(dba)_3$	M2	metal
$Pd(PPh_3)_2Cl_2$	M3	metal
—	—	—
dppf	L1	ligand
Sphos	L2	ligand
Xphos	L3	ligand
—	—	—
LiCl	A1	additive
Zn(0)	A2	additive
CuI	A3	additive
—	—	—
THF	S1	solvent
DMF	S2	solvent
NMP	S3	solvent
—	—	—
T<18	T1	temp
18≤T<23	T2	temp
23≤T<50	T3	temp

3.3.2 PROBLEM SETUP

For each dataset, an 80/10/10 train/validation/test split was used in modeling. Training and test sets were kept consistent between model types for sake of comparability. Model inputs were prepared as reactant/product structure tuples, with encodings tailored to each learning method. Models were trained using binary cross-entropy loss to output probability scores for all reagent/condition labels in the reaction dictionary (**Figure 3.5C**).[42] The top-k ranked labels in each dictionary category were selected as the final prediction, where k is user-determined.

We define an accurate prediction as one where the ground truth label appears in the top-k predicted labels. Given the variable class imbalance in each dictionary category,[43] accuracy is evaluated at the categorical level as follows:

$$A_c = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left[\hat{Y}_i \cap Y_i\right], \qquad (3.1)$$

where $\hat{Y}_i$ and Y_i are the sets of top-k predicted and ground truth labels for the i-th sample in category c, respectively. The correct instances are summed and divided by the number of samples in the test set (N) to give the overall test accuracy in the category, or A_c.[44]

As a general measure of a model's performance, we calculate its average error reduction (AER) from a baseline predictor (dummy) that always predicts the top-k most frequently occurring labels in each category:

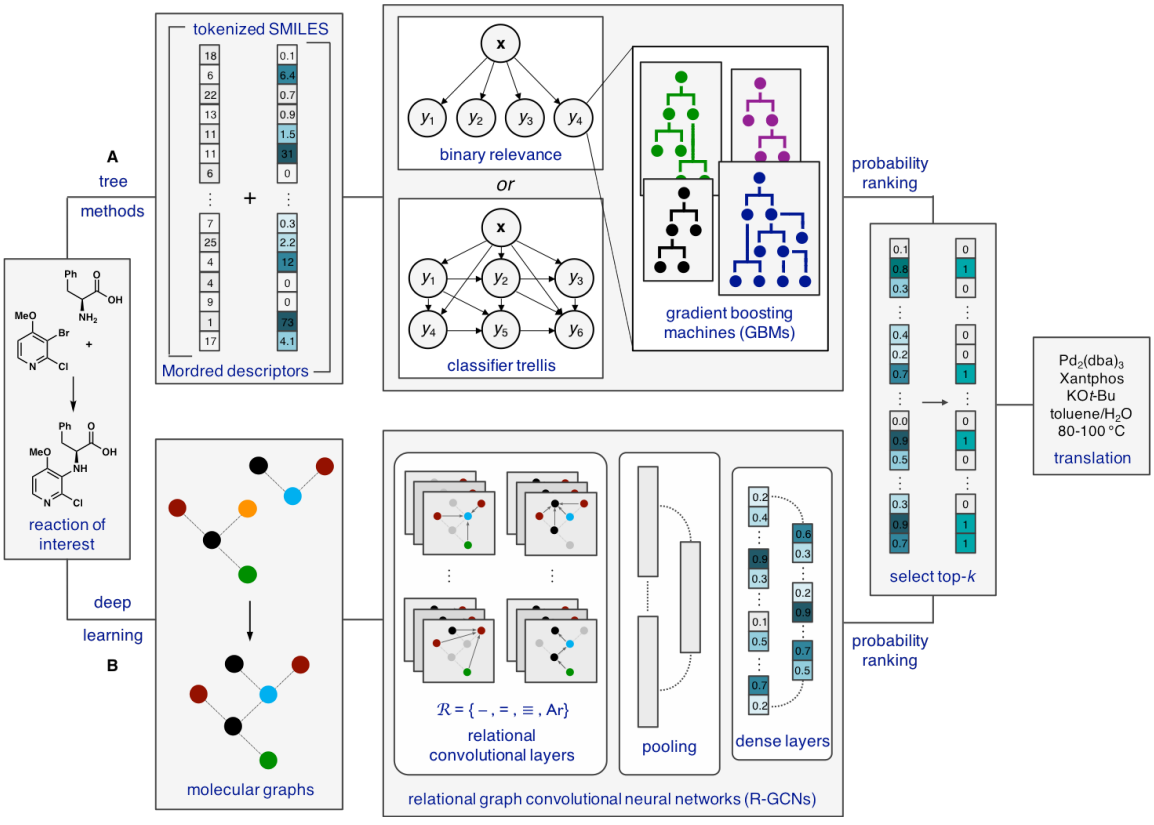
$$\text{AER} = \frac{1}{C}\sum_{c=1}^{C} \frac{A_c^g - A_c^d}{1 - A_c^d}, \qquad (3.2)$$

where A_c^g and A_c^d are the accuracies of the GNN and dummy model in the c-th category, respectively, and C is the number of categories in the dataset dictionary. AER represents a model's average improvement over the naïve approach that one might use as a starting point for experimental optimization. In other words, AER is the percent of the gap closed between the naïve model and a perfect predictor of accuracy 1.0.

3.3.3 MODEL CONSTRUCTION

Both tree- and deep learning methods were explored for this MLC task (**Figure 3.6**), and their individual development is discussed below.

Figure 3.6 *Schematic modeling workflow. Reactions are modeled independently by: A) Tree-based methods, where string and descriptor vectors for each molecule in a reaction are concatenated and used as inputs to gradient-boosting machines (GBMs); and B) Deep learning methods, where molecular graphs are constructed for each molecule in a reaction and are passed as inputs to a graph convolutional neural network (GCNN). Both model types predict probability rankings for the full reaction dictionary, which are sorted by reaction role and translated to their final outputs.*



3.3.3.1 Gradient-Boosting Machines (GBMs)

GBMs are decision-tree-based learning algorithms that are popular in the ML literature for their performance in modeling numerical data.[33,45] We explored several string and

descriptor-based encodings as numerical inputs (see Experimental Section) and found that a hybrid encoding scheme provided the greatest learnability (**Figure 3.6A**).[42] The hybrid inputs are a concatenation of tokenized SMILES strings for each molecule in a reaction (coupling partners and products), further concatenated with molecular property vectors obtained from the Mordred descriptor calculator.[46] GBMs consistently outperformed other tree-based learners such as random forests (RFs, data not shown),[47] perhaps owing to their use of sequential ensembling to improve in poor-performance regions.[45,48]

In our GBM experiments,[48,49] a separate classifier was trained for all bins in a dataset dictionary, predicting whether or not they should be present in each reaction.[31,50] Two general strategies have been developed for related MLC tasks, known as the binary relevance method (BM) and classifier chaining (CC).[51] The BM approach considers each classifier as an independent model, predicting the label of its bin irrespective of the others. Conversely, CCs make predictions sequentially, taking the output of each label as an additional input for the next one, where the optimal order of chaining is a learned parameter.[52] While the BM approach is significantly simpler from a computational perspective, CCs offer the potential for higher accuracy by modeling interdependencies between labels.[51]

We saw modeling reagent correlations as prudent in our studies since they are frequently observed in synthesis. Some examples relevant to this work include using a polar protic solvent with an inorganic base, excluding exogenous ligand when using a pre-ligated metal source, setting the temperature below the boiling point of the solvent, etc. We decided to explore both methods, testing BM against a modern update to CCs introduced by Read and coworkers known as classifier trellises (CTs).[53] In the CT method, instead of

fully sequential propagation, models are fit in a pre-defined grid structure (the "trellis"), where the output of each prediction is passed to multiple downstream classifiers at once (**Figure 3.6A**, center). This eliminates the cost of chain structure discovery, while still benefiting from nesting predictions.[51]

The ordering of a CT is enforced algorithmically starting from a seed label, chosen randomly or by expert intervention. From Read et al.,[53] the trellis is populated by maximizing the mutual information (MI) between source and target labels (s_ℓ) at each step (ℓ) as follows:

$$s_\ell = \text{argmax}_{k \in S} \sum_{j \in \text{pa}(\ell)} I(y_j; y_k), \qquad (3.3)$$

where S and $\text{pa}(\ell)$ are the set of remaining labels and the available trellis structure at the current step, respectively, and y_j and y_k are the j-th and k-th target labels, respectively. Here, $I(y_j; y_k)$ represents the MI between labels j and k based on their co-occurrences in the dataset. The matrix of *all* pairwise label dependencies $I(Y_j; Y_k)$ is constructed as below:

$$I(y_j; y_k) = \sum_{y_j \in \mathcal{Y}_j} \sum_{y_k \in \mathcal{Y}_k} p(y_j, y_k) \log\left(\frac{p(y_j, y_k)}{p(y_j)p(y_k)}\right), \qquad (3.4)$$

where $p(y_j, y_k)$, and $p(y_j)$ and $p(y_k)$ are the joint and marginal probability mass functions of y_j and y_k, respectively. $\mathcal{Y}_j$ and $\mathcal{Y}_k$ represent the possible values y_j and y_k can each assume, which for our task of binary classification are both $\{0,1\}$. Full MI matrices and optimized trellises for each dataset are included in the Experimental Section, and an example is discussed with the results.

3.3.3.2 Graph Convolutional Networks (GCNs)

To determine the most suitable GCN model for the MLC task, we tested implementations of a variety of GCN architectures from the literature using the ChainerChem library.[54] Each model contains two sub-networks that are jointly trained for the overall task. The first subnet is a graph processing network (GPN) that differs between architectures and forms the basis of their relative performances. The GPNs convert input graphs to learned molecular embeddings, which are concatenated to form the overall reaction vectors. These are passed as input to the second subnet, a multilayer perceptron (MLP), for the ultimate predictions.

GPNs explored in this work include neural fingerprinting networks (**NFP**) by Duvenaud et al.,[55a] gated graph sequence NNs (**GGNN**) by Li et al.,[55b] message passing NNs (**MPNN**) by Gilmer et al.,[55c] Weave module NNs (**Weave**) by Kearnes et al.,[36] relational graph attention networks (**R-GAT**) by Veličković et al.,[55d] relational graph convolutional networks (**R-GCN**) by Schlichtkrull et al.,[55e] and renormalized spectral graph convolutional networks (**RS-GCN**) by Kipf & Welling.[55f] Models are trained for 100 epochs using the Adam optimizer,[56] sigmoid cross entropy loss, and an 80/10/10 train/validation/test split in all experiments. Further general modeling parameters and detailed hyperparameter settings for each model are included in the Experimental Section.

After screening the above architectures (*vide infra*), we focused further on R-GCNs. Originally reported by Schlichtkrull et al.,[55e] R-GCNs are a subclass of message passing neural networks (MPNNs)[55c] that explicitly model relational data such as molecular graphs (**Figure 3.6B**, middle). This is achieved by constructing sets of *relation* operations, where each relation $r \in \mathcal{R}$ is specific to a type and direction of edge between connected nodes. In

our setting, the relations operate on atom-bond-atom triples using a learned, sparse weight matrix $\mathbf{W}_r^{(l)}$ in each layer l.[55e] In a propagation step, each current node representation $h_i^{(l)}$ is transformed with all relation-specific neighboring nodes $h_j^{(l)}$ and summed over all relations such that:

$$h_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_i^r}\frac{1}{c_{i,r}}\mathbf{W}_r^{(l)}h_j^{(l)} + \mathbf{W}_0^{(l)}h_i^{(l)}\right), \qquad (3.5)$$

where $\mathcal{N}_i^r$ is the set of applicable neighbors and σ is an element-wise non-linearity, for us the tanh. The self-relation term $\mathbf{W}_0^{(l)}h_i^{(l)}$ is added to preserve local node information, and $c_{i,r}$ is a normalization constant. Unlike traditional GCNs, R-GCNs intuitively model edge-based messages in local sub-graph transformations.[55e] This is potentially very powerful for reaction learning in that information on edge types (i.e., single, double, triple, aromatic, and cyclic bonds) is crucial for modeling reactivity.

Here, we extend the R-GCN architecture with an additional graph attention layer (GAL) at the final readout step inspired by graph attention networks (GATs) from Veličković[55d] and Busbridge.[57] As described by Veličković et al., GALs compute pairwise node attention coefficients α_{ij} for each node h_i in a graph and its neighbors h_j. Two nodes' features are first transformed via a shared weight matrix $\mathbf{W}$, the results of which are concatenated before applying a learned weight vector and softmax normalization. The final update rule is simply a linear combination of α_{ij} with the newly transformed node vectors ($\mathbf{W}h_j$), summed over all neighboring nodes and averaged over a set of parallel attention mechanisms.[55d]

We observed that existing relational GATs (R-GATs)[55d] using atom-level attention layers were less effective for our task than simple R-GCNs (*vide infra*).[55e,58] Inspired nonetheless by the chemical intuition of graph attention, we adapted existing GALs to construct a *reaction-level* attention mechanism. Instead of pairwise α_{ij}, we construct self-attention coefficients α_i^m for all nodes h_i^m in a molecular graph $\boldsymbol{h}^m = \{h_0^m, h_1^m, \dots, h_L^m\}$. As in GATs, we take a linear combination of α_i^m for all L nodes in $\boldsymbol{h}^m$ after further transformation by matrix $\mathbf{W}^g$:

$$\alpha_i^m = \sigma(\mathbf{W}^s h_i^m)\,, \forall\, i \in \{1, 2, \dots, L\}\,, \tag{3.6}$$

$$h_i^a = \alpha_i^m \mathbf{W}^g h_i^m\,, \tag{3.7}$$

where $\mathbf{W}^s$ is the learned attention weight matrix, σ is the sigmoid activation function, and h_i^a is the updated node representation. The convolved graphs $\boldsymbol{h}^a = \{h_0^a, h_1^a, \dots, h_L^a\}$ for each molecule m are then concatenated on the node feature axis to give an overall reaction representation $\boldsymbol{h}^r$ that we term the attended reaction graph (ARG):

$$\text{ARG} = \boldsymbol{h}^r = \left[\Big\Vert_{m=1}^{M} \boldsymbol{h}_m^a \right], \tag{3.8}$$

where M is the number of molecules in the reaction (reactants and products) and ‖ denotes concatenation. Similar to the attention mechanism above, reaction-level attention coefficients α_i^r are then constructed and linearly combined with the ARG nodes h_i^r after transformation with $\mathbf{W}^v$. The final readout vector $\boldsymbol{v}_r$ is obtained from the attention layer by summative pooling over the nodes:

$$\alpha_i^r = \sigma(\mathbf{W}^r h_i^r)\,, \forall\, i \in \{1, 2, \ldots, H\}\,, \tag{3.9}$$

$$\boldsymbol{v}_r = \sum_{i=1}^{H} \alpha_i^r \mathbf{W}^v h_i^r\,, \tag{3.10}$$

where H is the total number of nodes and $\mathbf{W}^r$ is the reaction attention weight matrix. This construction differs from standard R-GCNs, which output readout vectors for individual molecules and concatenate them to form the ultimate reaction representation. Altogether, we term our hybrid architecture as an *attended relational graph convolutional network*, or AR-GCN.

In all deep learning experiments, with or without attention, the reaction vector readouts were passed to a multi-layer perceptron (MLP) of $\mathrm{depth} = 2$. The final prediction is made as a single output vector with one entry for each label in the reaction dictionary, and the result is translated as described in Section 3.3.2.

3.4 RESULTS AND DISCUSSION

3.4.1 MODEL PERFORMANCE

Our modeling pipeline was first tested on the Suzuki coupling dataset, the largest of the four. Figure 3.7 and Table 3.4 summarize top-1 and top-3 categorical AERs (Equation 3.2) and accuracies (Equation 3.1), respectively, for each GCN architecture. Figure 3.8 and Table 3.5 summarize the same metrics for the following "optimized" models: GBMs with no trellising (**BM-GBM**), GBMs with trellising (**CT-GBM**), standard R-GCNs as reported by Schlichtkrull et al. (**R-GCN**),[55e] our AR-GCNs developed here (**AR-GCN**), and the dummy predictor as a baseline (**dummy**).

Figure 3.7 *GCN modeling results for Suzuki coupling dataset.*

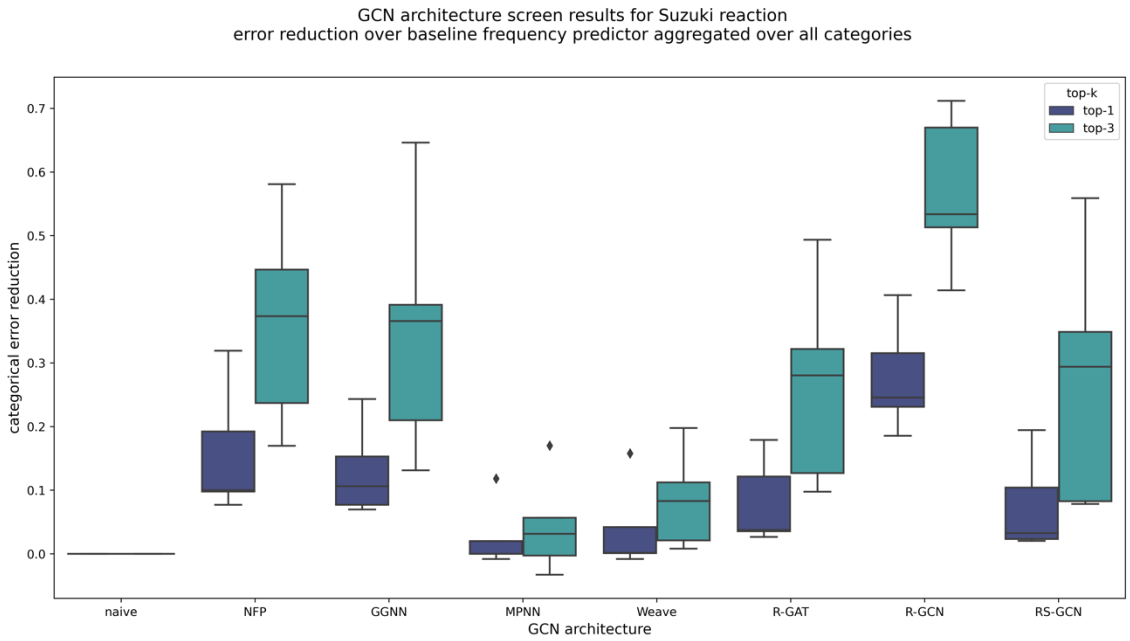


Table 3.4 *GCN modeling results for Suzuki coupling dataset.*

Suzuki coupling condition prediction by GCN models (categorical accuracies)									
category (num)	**top-k**	**naïve (freq)**	**NFP**	**GGNN**	**MPNN**	**Weave**	**R-GAT**	**R-GCN**	**RS-GCN**
AER	**1**	–	0.1572	0.1297	0.0259	0.0388	0.0801	**0.2767**	0.0750
	3	–	0.3615	0.3491	0.0451	0.0847	0.2641	**0.4936**	0.2732
metal (28)	1	0.3777	0.5763	0.5291	0.4513	0.4759	0.4891	**0.6306**	0.4987
	3	0.6744	0.8198	0.7935	0.7298	0.7388	0.7792	**0.8482**	0.7701
ligand (23)	1	0.8722	0.8847	0.8811	0.8722	0.8724	0.8770	**0.9036**	0.8752
	3	0.9269	0.9542	0.9555	0.9292	0.9351	0.9474	**0.9644**	0.9524
base (35)	1	0.3361	0.4637	0.4377	0.3494	0.3640	0.4167	**0.5455**	0.4052
	3	0.7344	0.7795	0.7693	0.7337	0.7366	0.7603	**0.8123**	0.7564
solvent (10)	1	0.6377	0.6656	0.6656	0.6377	0.6381	0.6506	**0.7049**	0.6495
	3	0.8013	0.8484	0.8430	0.7948	0.8055	0.8265	**0.8836**	0.8169
additive (17)	1	0.9511	0.9560	0.9563	0.9507	0.9507	0.9524	**0.9624**	0.9521
	3	0.9771	0.9904	0.9919	0.9784	0.9790	0.9884	**0.9934**	0.9899

Figure 3.8 *Optimized modeling results for Suzuki coupling dataset.*

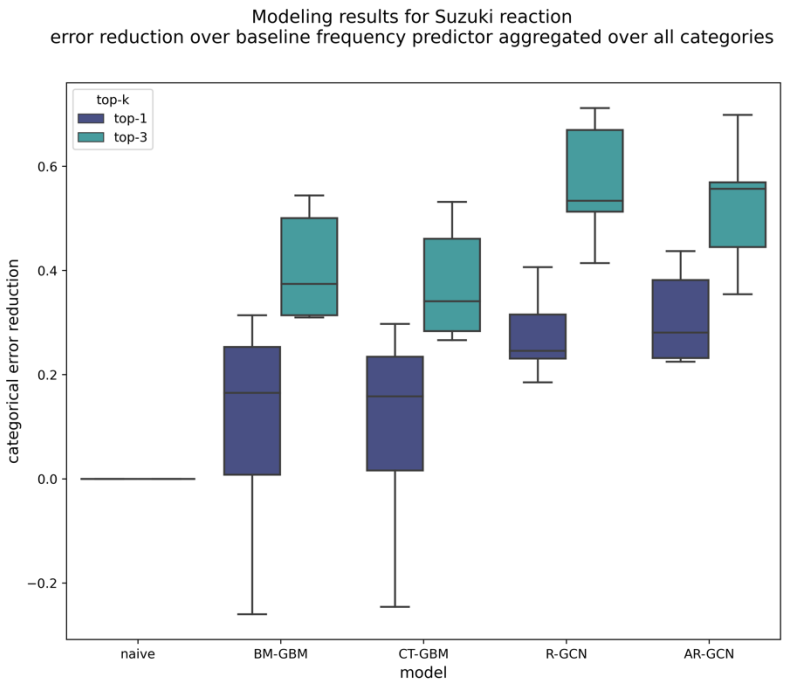


Table 3.5 *Optimized modeling results for Suzuki coupling dataset.*

Suzuki coupling condition prediction by optimized models (categorical accuracies)						
category (num)	**top-k**	**naïve (freq)**	**BM-GBM**	**CT-GBM**	**R-GCN**	**AR-GCN**
AER	**1**	–	-0.0263	-0.0554	0.2767	**0.3115**
	3	–	0.4088	0.3774	0.4936	**0.5246**
metal (28)	1	0.3777	0.5732	0.5629	0.6306	**0.6499**
	3	0.6744	0.8516	0.8475	0.8482	**0.8597**
ligand (23)	1	0.8722	0.8390	0.8408	0.9036	**0.9081**
	3	0.9269	0.9635	0.9606	0.9644	**0.9676**
base (35)	1	0.3361	0.4908	0.4777	0.5455	**0.5896**
	3	0.7344	**0.8338**	0.8250	0.8123	0.8285
solvent (10)	1	0.6377	0.6729	0.6751	0.7049	**0.7217**
	3	0.8013	0.8637	0.8577	0.8836	**0.8897**
additive (17)	1	0.9511	0.9259	0.9196	**0.9624**	0.9621
	3	0.9771	0.9842	0.9832	**0.9934**	0.9931

For this dataset, GCN models significantly outperformed GBMs across categories for both top-1 and top-3 predictions. While GBMs actually gave negative top-1 AERs over baseline, these scores were dominated by the additive contribution; excluding this category,

the BM- and CT-GBMs gave modest 10% and 9% AERs, respectively. Despite struggling with top-1 predictions, GBMs gave significant AERs for top-3, with BM-GBMs at 41% and CT-GBMs at 38%. The AR-GCNs gave the best accuracy of all models, providing 31% and 52% top-1 and top-3 AERs, respectively. AR-GCNs gave roughly 3% AER gain (ΔAER) over the R-GCN in both top-1 and top-3 predictions, demonstrating the value of the added attention layer.

A few interesting categorical trends can be seen across model types. For instance, models provide the best error reduction ($\mathrm{ER} = \frac{A_c^g - A_c^d}{1 - A_c^d}$, see Equation 3.2) in the *metal* category, with the AR-GCN at 44% and 57% for top-1 and top-3, respectively. Similarly, models perform well in the *base* category, where the AR-GCN gave the best top-1 ER and BM-GBMs gave the best top-3 ER. Less consistent ERs between top-1 and top-3 predictions were obtained for the remaining three categories. For example, with *solvents*, the AR-GCN improved baseline by 23% in top-1 predictions, but 44% in top-3. Likewise, for AR-GCN *ligand* predictions, a 28% ER was obtained for top-1 versus a 56% gain in top-3. Finally, although the baseline *additive* accuracy is high as the majority of reactions are `Null` in this category, the AR-GCN still gave a 23% top-1 ER and a 70% top-3 ER.

The trends and differences between top-1 and top-3 performance gains are reflective of the frequency distributions in each label category. These intuitively resemble long-tail or Pareto-type distributions,[59] with the bulk of the cumulative density contained in a small number of bins and the remaining bins supporting smaller frequencies. The distribution shapes are likely to influence the relative top-1 and top-3 AERs, where the highly skewed distributions could be more difficult to improve over baseline.

Having demonstrated the utility of our predictive framework, we turned to the remaining datasets to assess its scope. Modeling results for C–N, Negishi, and PKRs are detailed in Figures 3.9–3.14 and Tables 3.6–3.11. Notable observations for each dataset are discussed below.

Figure 3.9 *GCN modeling results for C–N coupling dataset.*

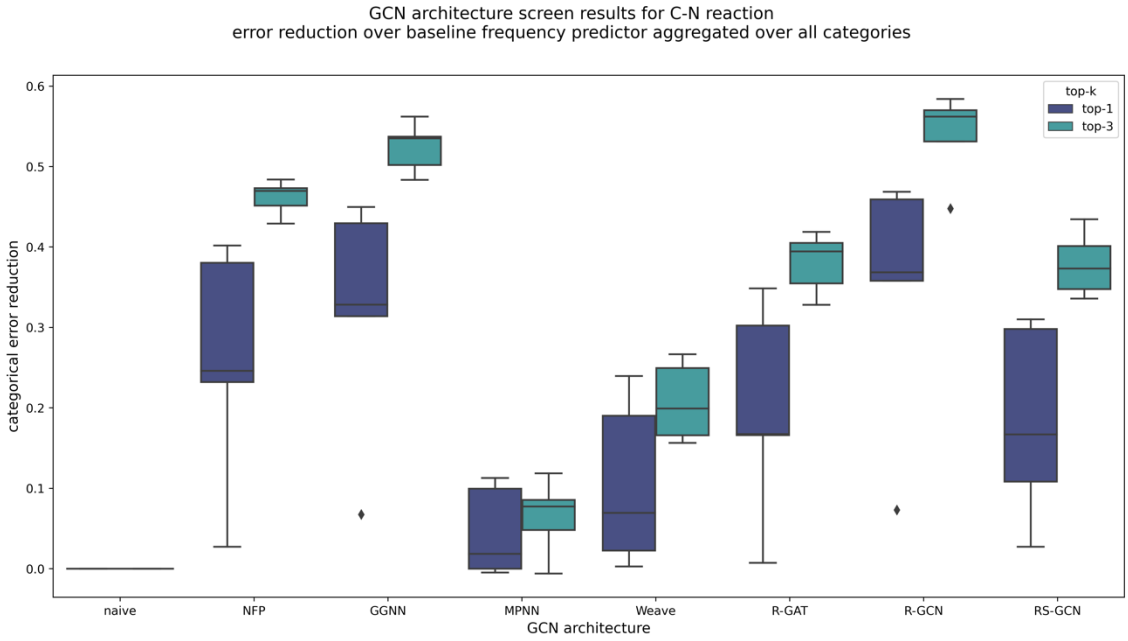


Table 3.6 *GCN modeling results for C–N coupling dataset.*

C–N coupling condition prediction by GCN models (categorical accuracies)									
category (num)	**top-k**	**naïve (freq)**	**NFP**	**GGNN**	**MPNN**	**Weave**	**R-GAT**	**R-GCN**	**RS-GCN**
AER	**1**	–	0.2575	0.3178	0.0453	0.1048	0.1983	**0.3453**	0.1821
	3	–	0.4615	0.5240	0.0648	0.2077	0.3802	**0.5391**	0.3785
metal (44)	1	0.2452	0.5485	0.5847	0.3304	0.4261	0.5082	**0.5989**	0.4792
	3	0.6526	0.8170	0.8392	0.6795	0.7393	0.7981	**0.8479**	0.7734
ligand (47)	1	0.5219	0.6395	0.6789	0.5197	0.5327	0.6019	**0.6981**	0.5737
	3	0.6647	0.8222	0.8532	0.6934	0.7203	0.7970	**0.8605**	0.7992
base (13)	1	0.2479	0.5340	0.5710	0.3227	0.3909	0.4753	**0.5932**	0.4721
	3	0.6400	0.8142	0.8326	0.6827	0.7360	0.7858	**0.8452**	0.7964
solvent (22)	1	0.3219	0.4792	0.5348	0.3345	0.3690	0.4345	**0.5647**	0.4351
	3	0.5677	0.7532	0.7847	0.5885	0.6538	0.7211	**0.7973**	0.7129
additive (74)	1	0.8904	0.8934	0.8978	0.8904	0.8907	0.8912	**0.8984**	0.8934
	3	0.9156	0.9537	**0.9564**	0.9151	0.9288	0.9433	0.9534	0.9471

Figure 3.10 *Optimized modeling results for C–N coupling dataset.*

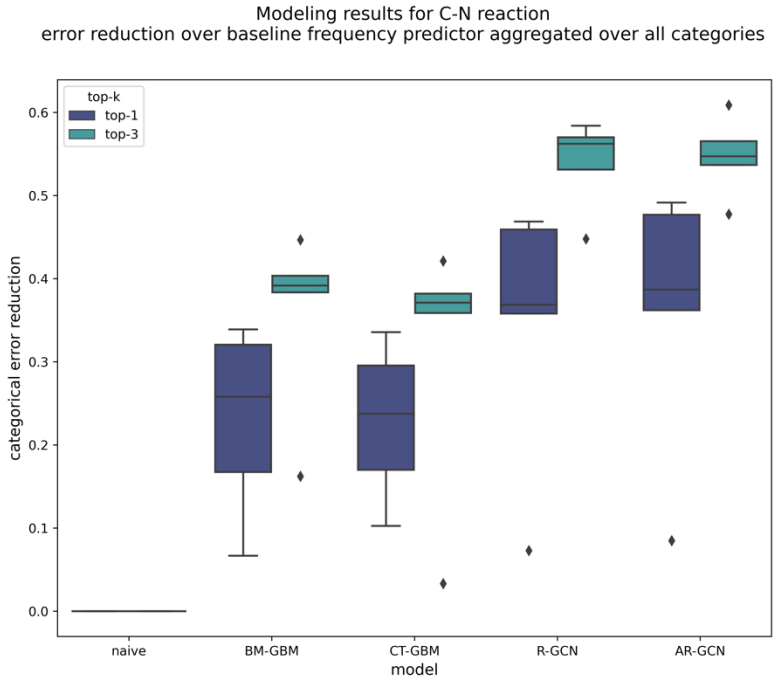


Table 3.7 *Optimized modeling results for C–N coupling dataset.*

C–N coupling condition prediction by optimized models (categorical accuracies)						
category (num)	**top-k**	**naïve (freq)**	**BM-GBM**	**CT-GBM**	**R-GCN**	**AR-GCN**
AER	**1**	–	-0.0413	-0.1593	0.3453	**0.3604**
	3	–	0.3568	0.3131	0.5391	**0.5471**
metal (44)	1	0.2452	0.4825	0.4582	0.5989	**0.6162**
	3	0.6526	0.7928	0.7772	0.8479	**0.8490**
ligand (47)	1	0.5219	0.5538	0.5710	0.6981	**0.7068**
	3	0.6647	0.7933	0.7928	0.8605	**0.8688**
base (13)	1	0.2479	0.5028	0.5003	0.5932	**0.6066**
	3	0.6400	0.8008	0.7916	**0.8452**	0.8370
solvent (22)	1	0.3219	0.4582	0.4524	0.5647	**0.5674**
	3	0.5677	0.7370	0.7281	0.7973	**0.7997**
additive (74)	1	0.8904	0.7669	0.7031	0.8984	**0.8997**
	3	0.9156	0.9290	0.9184	0.9534	**0.9559**

C–N coupling. Similar to the Suzuki results, the AR-GCN was the top performer for C–N couplings in almost all categories, and slightly higher AERs were observed

overall. The AR-GCN afforded 36% and 55% top-1 and top-3 AERs, respectively, again providing slight gains over R-GCNs at 35% and 54%. As above, GBMs struggled with this relatively large dataset (36,519 reactions) due to difficulties with the *additive* category. Models again made strong improvements in the *metal* and *base* categories, but also gave consistently strong gains for *ligands*[50] and *solvents*, especially for top-3 predictions. For example, the AR-GCN returned top-3 ERs of 57% for *metals*, 61% for *ligands*, 55% for *bases*, and 54% for *solvents*. Note that these ERs correspond to very high accuracies (A_c) of 85%, 87%, 84%, and 80%, respectively.

Figure 3.11 *GCN modeling results for Negishi coupling dataset.*

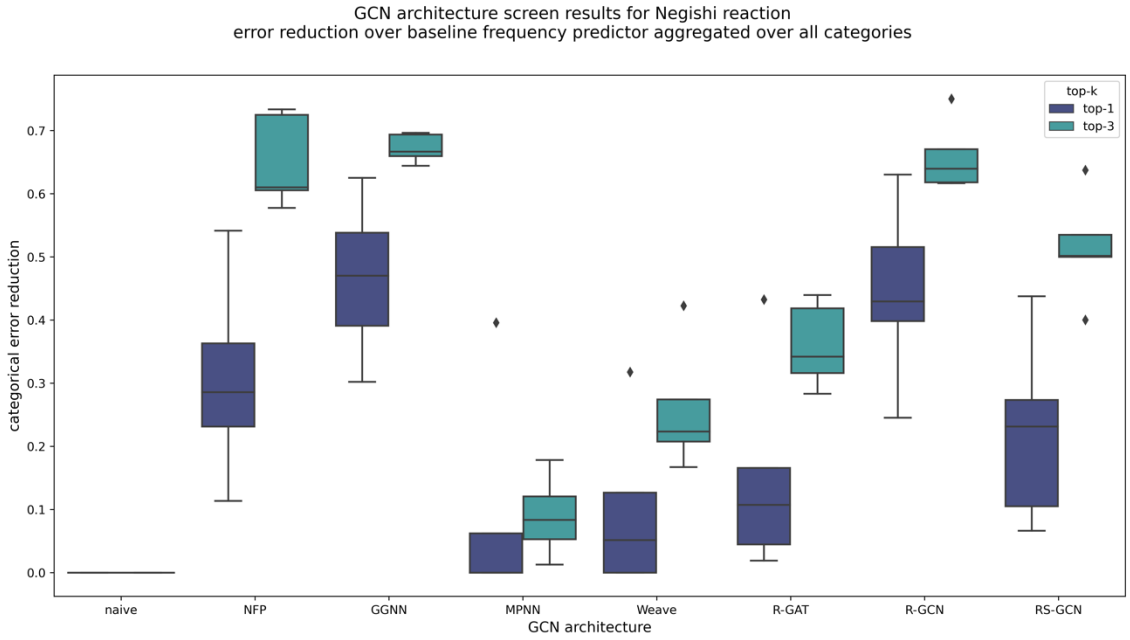


Table 3.8 *GCN modeling results for Negishi coupling dataset.*

Negishi coupling condition prediction by GCN models (categorical accuracies)									
category (num)	**top-k**	**naïve (freq)**	**NFP**	**GGNN**	**MPNN**	**Weave**	**R-GAT**	**R-GCN**	**RS-GCN**
AER	**1**	–	0.3071	**0.4652**	0.0916	0.0992	0.1539	0.4439	0.2228
	3	–	0.6503	**0.6722**	0.0896	0.2590	0.3598	0.6590	0.5148
metal (32)	1	0.2887	0.5470	**0.6715**	0.2887	0.3254	0.4067	0.6555	0.4833
	3	0.5008	0.8054	**0.8485**	0.5072	0.6045	0.6715	0.8086	0.7512
ligand (20)	1	0.7879	0.8485	0.8708	0.7879	0.7879	0.7974	**0.8724**	0.8102
	3	0.8549	**0.9601**	0.9506	0.8724	0.8947	0.9187	0.9522	0.9474
temperature (8)	1	0.3317	0.4864	**0.6459**	0.3732	0.4163	0.4035	0.6188	0.4864
	3	0.5885	0.8262	**0.8740**	0.6619	0.7624	0.7608	0.8517	0.8086
solvent (10)	1	0.6938	0.8596	0.8852	0.8150	0.7911	0.8262	**0.8868**	0.8278
	3	0.8788	0.9522	**0.9569**	0.8852	0.9059	0.9171	0.9537	0.9394
additive (30)	1	0.8309	0.8501	**0.8820**	0.8309	0.8309	0.8341	0.8724	0.8421
	3	0.9043	0.9745	0.9681	0.9123	0.9203	0.9314	**0.9761**	0.9426

Figure 3.12 *Optimized modeling results for Negishi coupling dataset.*

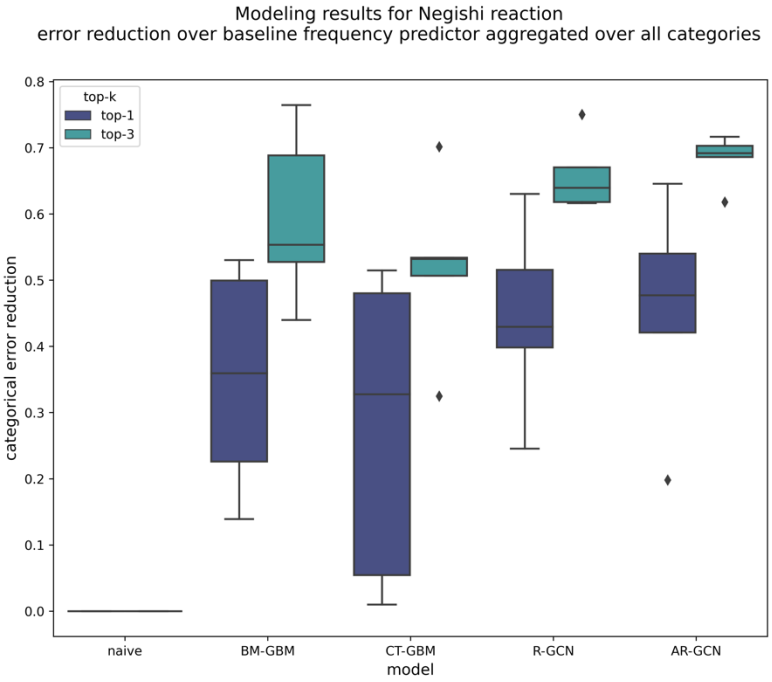


Table 3.9 *Optimized modeling results for Negishi coupling dataset.*

Negishi coupling condition prediction by optimized models (categorical accuracies)						
category (num)	**top-k**	**naïve (freq)**	**BM-GBM**	**CT-GBM**	**R-GCN**	**AR-GCN**
AER	**1**	–	0.3510	0.2773	0.4439	**0.4565**
	3	–	0.5947	0.5199	0.6590	**0.6833**
metal (32)	1	0.2887	0.5444	0.5218	0.6555	**0.6730**
	3	0.5008	0.7771	0.7674	0.8086	**0.8517**
ligand (20)	1	0.7879	0.8174	0.7900	0.8724	**0.8772**
	3	0.8549	0.9548	0.9321	0.9522	**0.9553**
temperature (8)	1	0.3317	**0.6656**	0.6527	0.6188	0.6507
	3	0.5885	**0.9031**	0.8772	0.8517	0.8708
solvent (10)	1	0.6938	0.8562	0.8514	0.8868	**0.8915**
	3	0.8788	0.9321	0.9402	**0.9537**	**0.9537**
additive (30)	1	0.8309	0.8691	0.8401	**0.8724**	0.8644
	3	0.9043	0.9548	0.9354	**0.9761**	0.9729

Negishi coupling. The highest AERs of all modeling experiments came with the Negishi dataset. The AR-GCN again gave the strongest performance, with top-1 and top-3 AERs of 46% and 68%, respectively. In this case, standard GGNNs gave similarly strong performance, with AERs of 47% and 67%. However, the R-GCN and even GBM models gave the highest accuracies in some categories. Interestingly, BM- and CT-GBMs performed significantly better than most GCNs for *temperature* predictions, though the strongest ER for most models came from the *solvent* category.

Figure 3.13 *GCN modeling results for PKR dataset.*

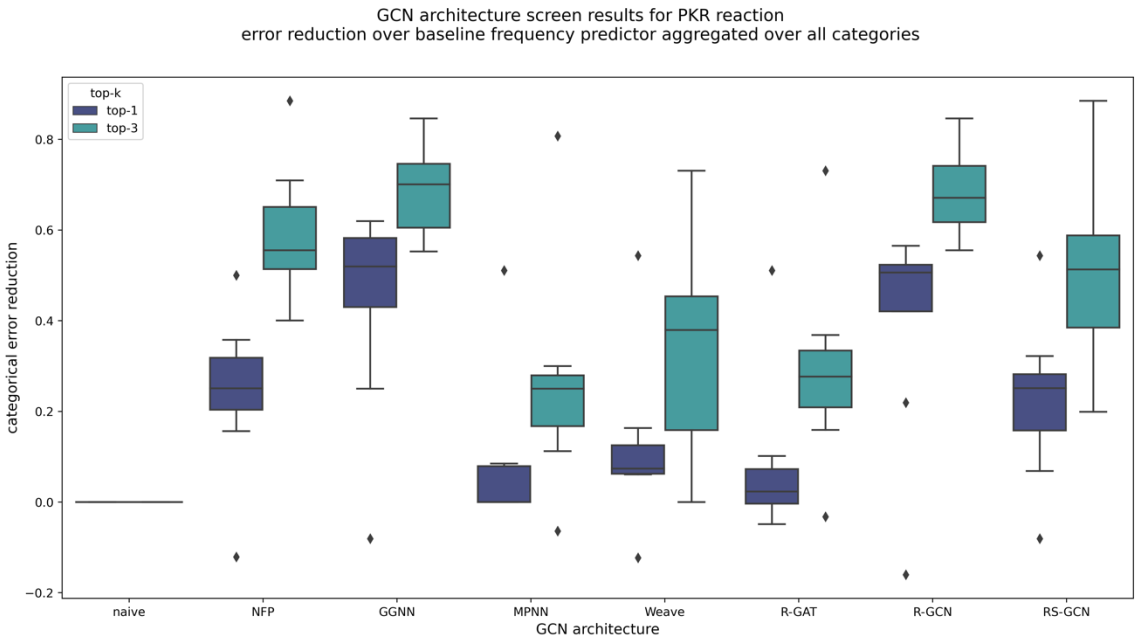


Table 3.10 *GCN modeling results for PKR dataset.*

PKR coupling condition prediction by GCN models (categorical accuracies)									
category (num)	**top-k**	**naïve (freq)**	**NFP**	**GGNN**	**MPNN**	**Weave**	**R-GAT**	**R-GCN**	**RS-GCN**
AER	**1**	–	0.2400	**0.4377**	-0.0294	0.1209	0.0825	0.3973	0.2265
	3	–	0.5957	**0.6861**	0.2695	0.3336	0.2947	0.6844	0.5063
metal (18)	1	0.4302	0.6340	0.7094	0.4302	0.4943	0.4566	**0.7132**	0.5774
	3	0.7132	0.8604	0.8717	0.7849	0.8302	0.8189	**0.9057**	0.8604
ligand (6)	1	0.8792	0.8981	**0.9094**	0.8792	0.8868	0.8792	0.9057	0.9019
	3	0.9019	**0.9887**	0.9849	0.9811	0.9736	0.9736	0.9849	**0.9887**
temperature (7)	1	0.2830	0.4415	**0.6642**	0.3358	0.4000	0.3283	0.6528	0.5755
	3	0.5962	0.8038	**0.8792**	0.6415	0.6981	0.6604	0.8528	0.7509
solvent (15)	1	0.3321	0.5358	**0.7396**	0.3887	0.3774	0.4000	0.6792	0.5472
	3	0.5925	0.8340	**0.8981**	0.6981	0.7472	0.6981	0.8679	0.8226
activator (11)	1	0.6906	0.7774	**0.8679**	0.6906	0.7094	0.6755	0.8415	0.7660
	3	0.8830	0.9660	0.9698	0.8755	0.8906	0.8792	**0.9774**	0.9283
CO (g) (2)	1	0.7245	0.7849	0.8642	0.4755	0.6906	0.7208	**0.8717**	0.7434
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
additive (13)	1	0.9057	0.8943	0.8981	0.9057	**0.9132**	0.9057	0.8906	0.8981
	3	0.9321	0.9698	**0.9736**	0.9472	0.9660	0.9509	0.9698	**0.9736**
pressure (4)	1	0.6528	0.8264	**0.8679**	0.8302	0.8415	0.8302	0.8491	0.8415
	3	0.9623	0.9774	**0.9849**	0.9736	0.9623	0.9736	**0.9849**	0.9698

Figure 3.14 *Optimized modeling results for PKR dataset.*

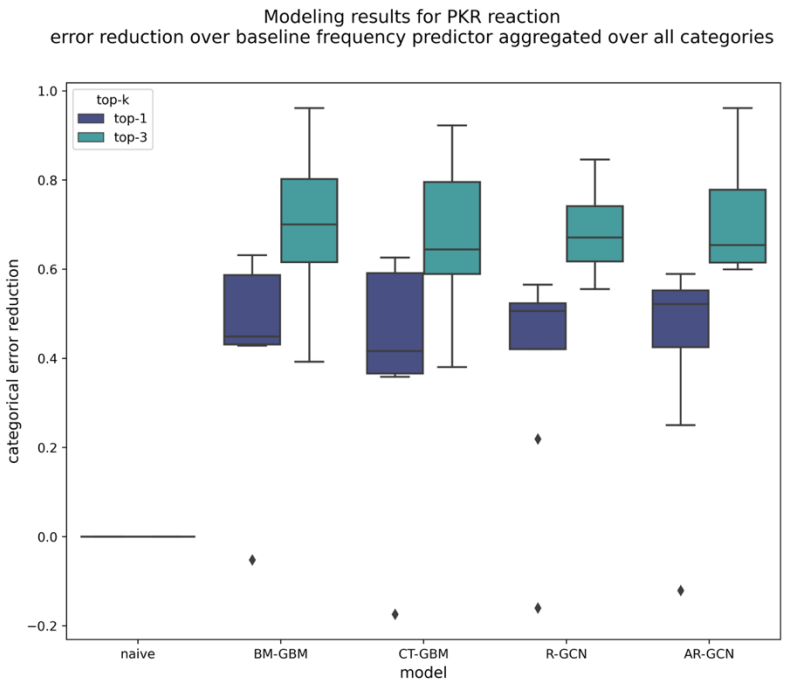


Table 3.11 *Optimized modeling results for PKR dataset.*

PKR coupling condition prediction by optimized models (categorical accuracies)						
category (num)	**top-k**	**naïve (freq)**	**BM-GBM**	**CT-GBM**	**R-GCN**	**AR-GCN**
AER	**1**	–	**0.4396**	0.4010	0.3973	0.4199
	3	–	0.6987	0.6740	0.6844	**0.7145**
metal (18)	1	0.4302	**0.7901**	0.7786	0.7132	0.7057
	3	0.7132	**0.9351**	0.9313	0.9057	0.8906
ligand (6)	1	0.8792	**0.9351**	0.9237	0.9057	0.9094
	3	0.9019	**0.9962**	0.9924	0.9849	**0.9962**
temperature (7)	1	0.2830	0.5954	0.5749	0.6528	**0.6642**
	3	0.5962	**0.8740**	0.8321	0.8528	0.8604
solvent (15)	1	0.3321	0.6183	0.6260	0.6792	**0.6981**
	3	0.5925	0.8779	0.8550	0.8679	**0.8981**
activator (11)	1	0.6906	0.8244	0.8015	0.8415	**0.8491**
	3	0.8830	0.9466	0.9275	**0.9774**	**0.9774**
CO (g) (2)	1	0.7245	0.8855	0.8855	0.8717	**0.8868**
	3	1.0000	1.0000	1.0000	1.0000	1.0000
additive (13)	1	0.9057	**0.9008**	0.8893	0.8906	0.8491
	3	0.9321	**0.9885**	**0.9885**	0.9698	0.9736
pressure (4)	1	0.6528	0.8588	**0.8702**	0.8491	0.8491
	3	0.9623	0.9771	0.9847	**0.9849**	**0.9849**

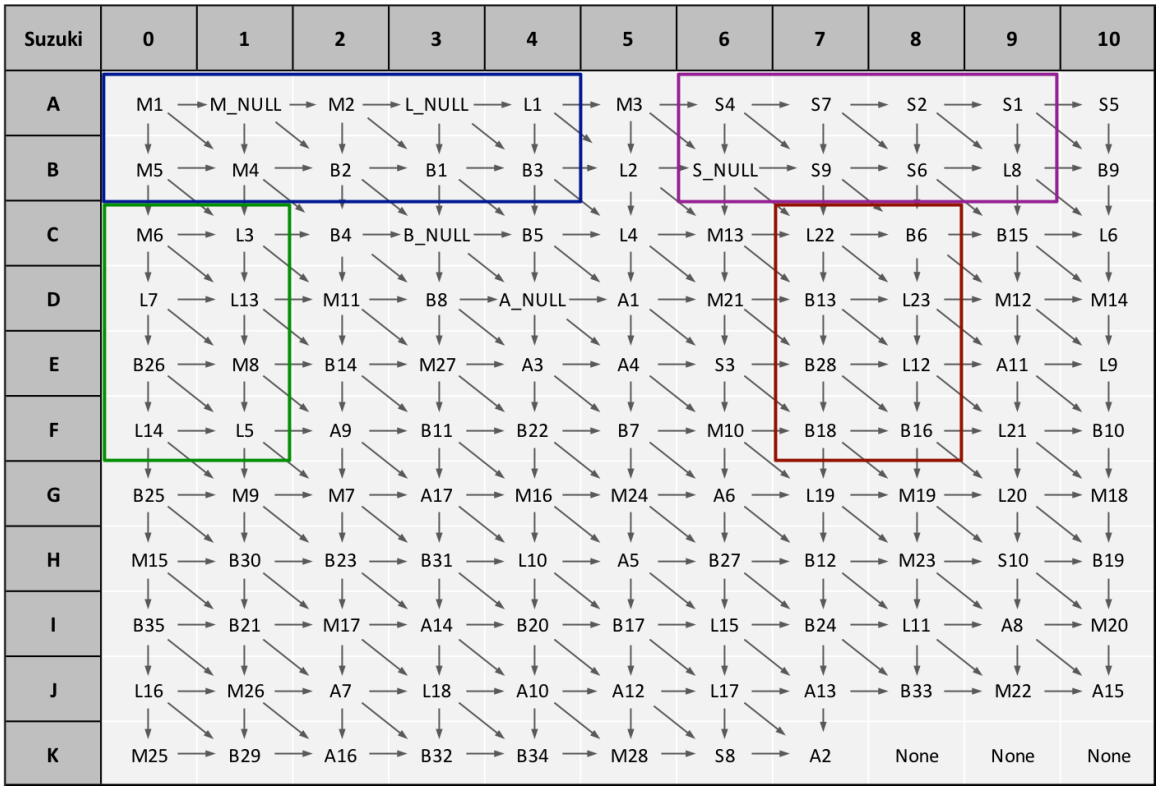
PKR. For the PKR dataset—the smallest of the four—simple BM-GBMs along with GGNNs gave the best top-1 AERs at 44%, followed closely by the AR-GCN at 42%. Similarly, for top-3 predictions, the BM-GBM and AR-GCN models gave AERs of 70% and 71%, respectively. Compared to the other reactions, GCNs are perhaps more prone to overfitting this small of a dataset,[60] making tree-based modeling more suitable. It is interesting to note that in general for PKRs, the GCN models were better at predicting physical parameters like *temperature*, *solvent*, and *CO(g)* atmosphere, whereas GBMs gave better performance for reaction components such as *metal*, *ligand*, and *additive*.

3.4.2 INTERPRETABILITY

3.4.2.1 Tree Methods

Given the results described above, we sought an understanding of the chemical features informing our predictions. Tree-based learning is often favored in this regard in that feature importances (FIs) can be directly extracted from models. As expected, we found that FIs for our BM-GBMs were roughly uniform across the SMILES regions of the encodings. The most informative physical descriptors from the Mordred vectors pertained to two classes: topological charge distributions[61] correlated with local molecular dipoles; and Moreau–Broto autocorrelations[62] weighted by polarizability, ionization potential, and valence electrons (see Experimental Section 3.6 for detailed rankings). The latter class is particularly intriguing as they are calculated from molecular graphs in what have been described as atom-pair convolutions,[63] not unlike the GCN models used here.[36,55]

Figure 3.15 *Optimized predictor trellis for the Suzuki dataset*



An advantage to using CTs is the ability to extract their MI matrices and trellis structures for interpretation.[53] The optimized trellis for the Suzuki CT-GBMs is included in Figure 3.15, where several chemically intuitive features and category blocks can be noted:

1. Block A0–B4 (blue): The result of `M1` ($Pd(PPh_3)_4$) is used to predict three more metals: `M2` ($Pd(OAc)_2$), `M4` ($Pd(dppf)Cl_2 \cdot DCM$), and `M5` ($Pd(PPh_3)_2Cl_2$). Based on these metal complexes, the probability of using exogenous ligand (`L_NULL`) and `L1` (PPh_3) is then predicted.
2. Block C0–F2 (green): The use of unligated `M6` ($Pd_2(dba)_3$) informs the predictions of ligands `L3` (XPhos), `L7` ($[(t\text{-}Bu)_3PH]BF_4$), and `L13` (^{Me}CgPPh). These in turn feed the model of unligated `M8` ($Pd(dba)_2$), which then informs `L5` ($P(o\text{-tolyl})_3$).

3. Block A6–B9 (purple): Several solvents are connected, where the predictions of `S4` (1,4-dioxane) and `S7` (PhMe) propagate through `S9` (H_2O), `S2` (EtOH), and `S6` (MeCN). These additionally feed classifiers of `S1` (THF) and `S_NULL` (neat).
4. Block C7–F8 (red): Four different classes of base are interwoven, including `B6` (CsF) and `B13` (KO*t*-Bu). This informs the prediction of `B28` ($LiOH \cdot H_2O$), which then goes on to feed models of `B18` (DIPEA) and `B16` (NaO*t*-Bu).

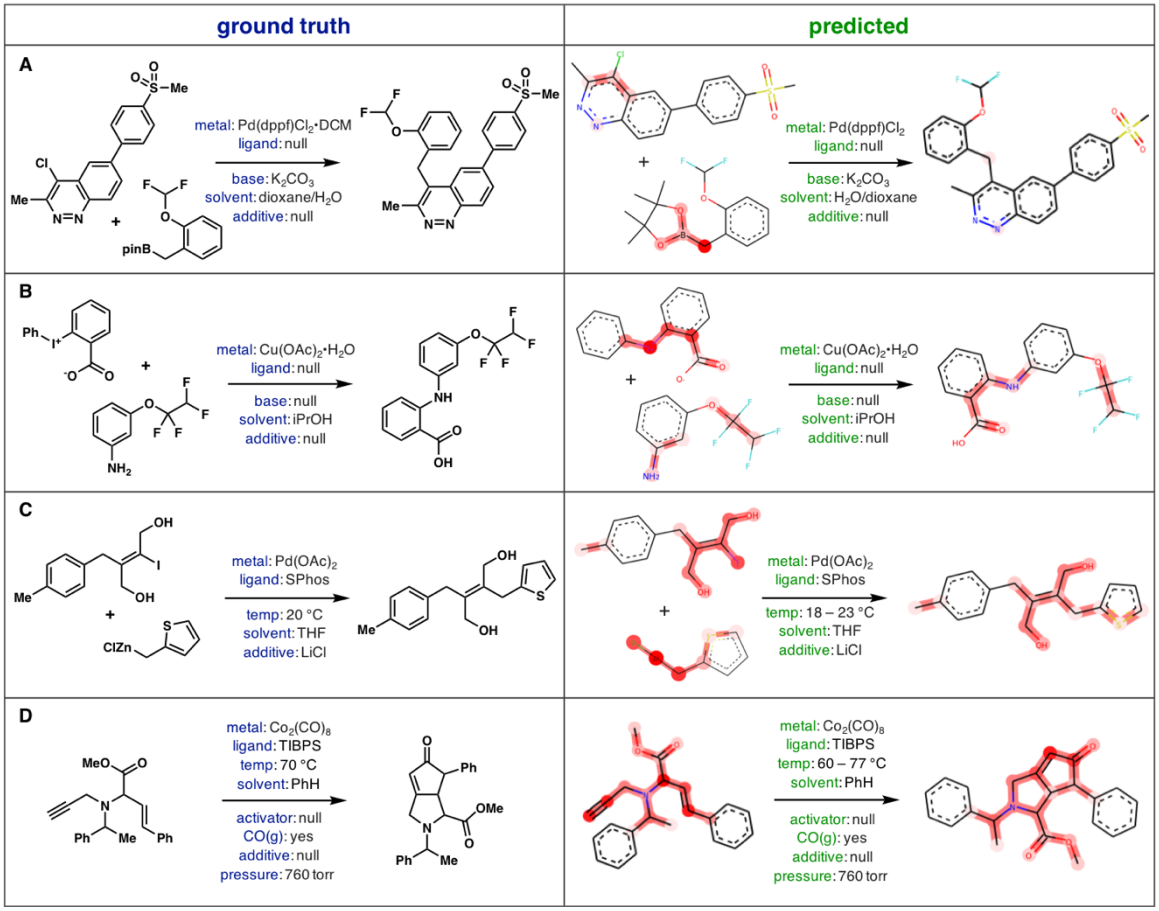
As a control experiment,[64] we withheld the propagated predictions from the CT-GBMs to test whether the MI was actually being used.[65] Indeed, model accuracy dropped off markedly, even below baseline in some categories. While this suggests that CT-GBMs do learn reagent correlations, the sharp performance loss may also indicate overfitting to this information.[53] This is supported by the strong emphasis on the label-encoding region in the CT-GBM FIs. We hypothesize that reducing the dimensionality of the structure-encoding region could help balance model attention and avoid overfitting. Further studies are necessary to uncover the optimal molecule featurization in combination with CTs, though the results here suggest their promise in modeling structured reaction data.

3.4.2.2 Deep Learning Methods

For AR-GCNs, a valuable interpretability feature lies in the learned feature weights α_i^r (Equation 3.9). Intuitively, the weights represent the model's assignment of importance on an atom, as they re-scale node features in the final graph layer before inference. When extracted, the weights can be mapped back onto a molecule's atoms and displayed by color scale using RDKit (**Figure 3.5D**).[66,67] This gives a visual interpretation of the functional groups most heavily informing the predictions. Example visualizations from a random reaction in each dataset and their AR-GCN predictions are included in Figure 3.16, and

several additional random examples for each reaction type can be found in the Experimental Section 3.6.

Figure 3.16 *AR-GCN attention weight visualization and prediction examples from randomly chosen reactions in each dataset. Darker highlighting indicates higher attention.*



In the Suzuki example (**Figure 3.16A**), the attention is dominated by the sp^3 carbon bearing the Bpin group, with additional contributions from the bis-*o*-substituted heteroaryl-chloride and its cinnoline nitrogen, all of which could be reasonably expected to influence reactivity. It is interesting that weights on the *o*-difluoromethoxy group, the sulfone, and the majority of the product are suppressed, perhaps indicating that an alkyl nucleophile is

sufficient to predict the reported conditions. The AR-GCN predictions are correct in each category besides the *metal*, where the model erroneously identifies the metal source Pd(dppf)Cl_2 instead of its ground truth DCM adduct Pd(dppf)Cl_2·DCM. It is worth noting that this prediction is extremely close in context to the ground truth and could very well display productive reactivity. Previous approaches to this task consider this context correct,[31] which improves model metrics and could be reasonable in practice. We assign an incorrect prediction with the knowledge that such subtle changes in reagent/catalyst nature can indeed lead to reactivity cliffs, and we thus ask our models to predict the exact correct context in all cases.

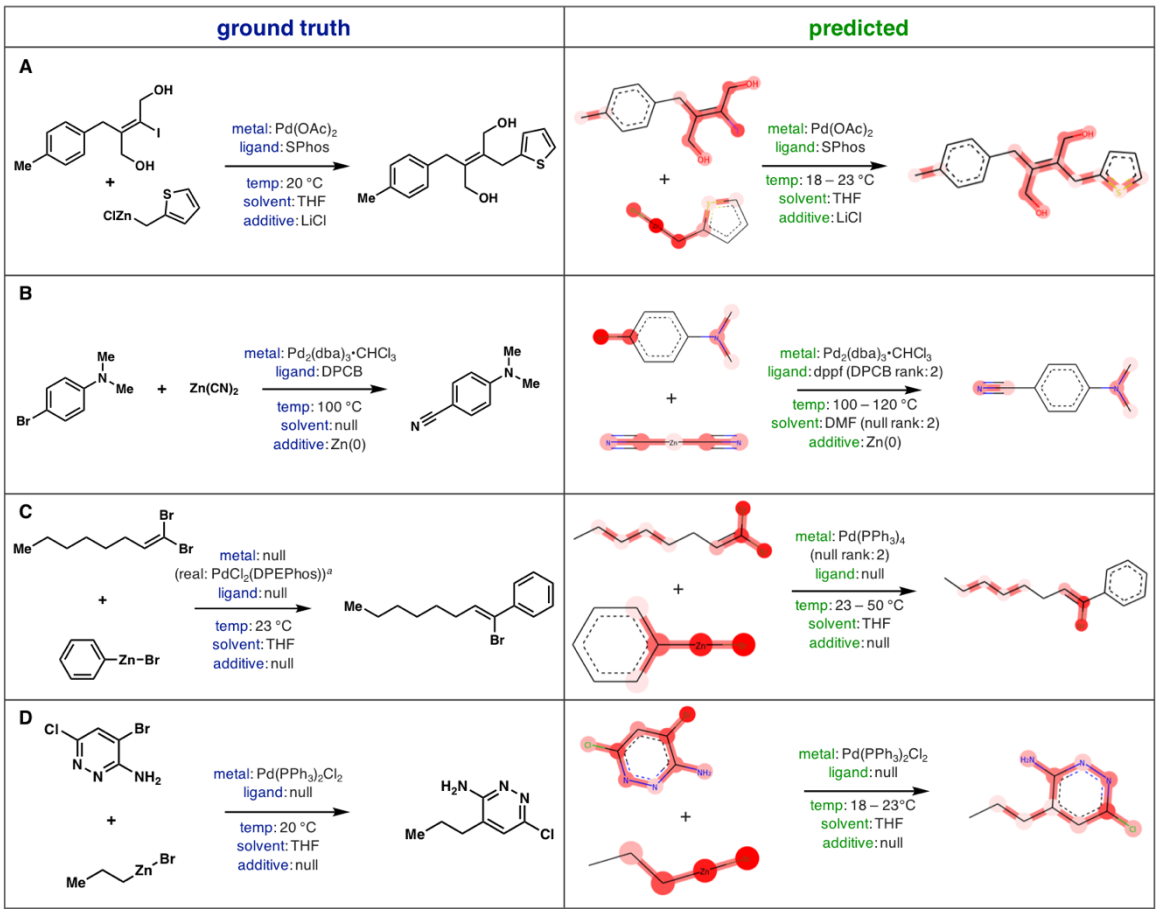
The weights in the C–N coupling example are more evenly distributed than in the Suzuki coupling (**Figure 3.16B**). Intuitively, the chemically active iodonium benzoate is given strong attention in the electrophile, as is the nucleophilic aniline nitrogen. Here, the *m*-tetrafluoroethoxy group is also weighted significantly and these groups are given similar attention in the product. All categories are predicted correctly in this example, though three of them are `null`.

The Negishi example (**Figure 3.16C**) is an interesting C(sp^3)–C(sp^2) coupling of a fully substituted alkenyl-iodide and thiophenyl-methylzinc chloride. Like with A, the strongest weights correspond to the sp^3 nucleophilic carbon, though similarly strong attention is distributed over the electrophilic alkene including the pendant alcohols. These weights are again reflected in the product and all five condition categories are predicted correctly, including *temperature* and use of a LiCl *additive*.

Lastly, an intramolecular PKR (**Figure 3.16D**) showed the most uniformly distributed attention of the four examples. Still, the strongest weights are given to the

participating alkyne and alkene, with additional emphasis on the amino ester bridging group. Weights are similarly distributed in the product, though strongest attention is intuitively assigned to the newly formed enone. Here, all 8 categories are predicted correctly including the use of an ambient carbon monoxide atmosphere (*CO(g)* and *pressure*).

Figure 3.17 *Substrate-specific predictions and attention visualization for four randomly chosen Negishi reactions.*



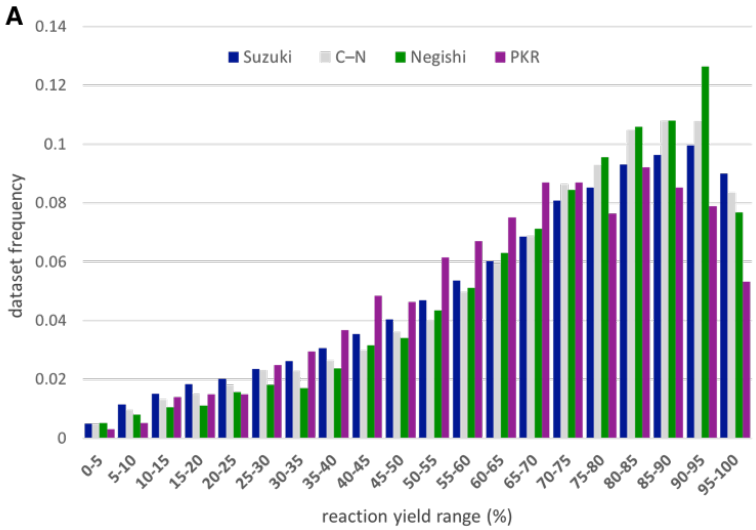
Within a given reaction-type, it is interesting to note our model's ability to distinguish individual substrates for making tailored condition predictions. Figure 3.17 shows four randomly chosen reactions from the Negishi dataset along with their attention

maps and AR-GCN predictions. There is considerable breadth in chemical space between the four examples: reaction A is between a tetrasubstitued alkenyl-iodide and a pseudo-benzyl zincate; reaction B features an electron-rich aryl bromide (*p*-dimethylamino-bromo-benzene) and the sp-nucleophilic salt $Zn(CN)_2$; reaction C couples a di-bromo alkene and the sp^2 phenylzinc bromide; and reaction D features a densely functionalized dihalo-amino pyridazine and a simple alkylzinc species. In all cases, condition predictions are either completely correct or within the top-3 rankings and are almost fully unique from one another. We believe this demonstrates the sought-after substrate specificity.

3.4.3 YIELD ANALYSIS

Having explored our models' chemical feature learning, we lastly investigated the effect of reaction yield, as it is a critical feature of synthesis data. Unsurprisingly, plotting the distribution of reaction yields in each dataset showed a uniformly strong bias towards high-yielding reactions (**Figure 3.18A**). Given the skewness of the data in this regard, we hypothesized that models would perform best at predicting conditions for high-yielding reactions.

Figure 3.18 *Performance dependence on reaction yield. A) Distribution of reaction yields for the four datasets. B) AR-GCN average top-1 A_c values for Suzuki predictions when trained and tested in different yield ranges (top) and dataset quartiles arranged by yield (bottom).*



B

	yield range				
Train\Test	0-25	25-50	50-75	75-100	0-100
0-25	0.7275	0.6701	0.615	0.6072	0.6274
25-50	0.7217	0.7507	0.6781	0.6515	0.6795
50-75	0.6858	0.7309	0.7683	0.7094	0.7285
75-100	0.6552	0.6851	0.723	0.7649	0.7319
0-100	0.7481	0.773	0.7737	0.7586	0.7641

	data quartile				
Train\Test	0-25	25-50	50-75	75-100	0-100
0-25	0.7645	0.7059	0.6685	0.6713	0.7022
25-50	0.7057	0.7589	0.7084	0.6882	0.7153
50-75	0.6684	0.7138	0.7546	0.7224	0.7157
75-100	0.6417	0.6756	0.698	0.7559	0.693
0-100	0.7669	0.7729	0.7643	0.752	0.7641

We divided the dataset into quartiles by reaction yield and re-trained the AR-GCN with each subset, subsequently testing in each region and on the full test set (**Figure 3.18B**). Intuitively, models trained in any yield range tended to give highest accuracy when tested in the same range, occupying the confusion matrix diagonal in Figure 3.18B (top). To our

surprise, however, the standard model trained on the full dataset gave consistently high accuracies, regardless of the test set (bottom row).

Since the yield bins contain varying amounts of data, we re-split the dataset, again ordered by yield but with equal subset sizes (**Figure 3.18B**, bottom). A similar trend was observed, where the highest accuracies were found on the diagonal and bottom row of the confusion matrix. Interestingly, the worst performing model was that trained in the highest yield range and tested in the lowest. We recognize that making "inaccurate" predictions on low-yielding reactions offers an avenue for predictive reaction optimization and future studies will explore this objective.

3.5 CONCLUSION AND OUTLOOK

In summary, we present a multilabel classification approach to predicting experimental reaction conditions for organic synthesis. We successfully model four high-value reaction types using expert-crafted label dictionaries: Suzuki, C–N, and Negishi couplings, and Pauson–Khand reactions. We explore and optimize two model classes: gradient-boosting machines and graph convolutional networks. We find that GCN models perform very well in larger datasets, while GBMs show success for smaller datasets.

We report the first use of classifier trellises in molecular machine learning and find that they are able to incorporate label correlations in modeling. We introduce a novel reaction-level graph-attention mechanism that provides significant accuracy gains when coupled with relational GCNs and construct a hybrid GCN architecture called *attended relational GCNs*, or AR-GCNs. We further provide an analytical framework for the chemical interpretation of our models, extracting the trellis structures and mutual

information matrices of the CT-GBMs and visualizing the attention weights assigned in AR-GCN predictions.

Experimental studies are currently underway assessing the feasibility of model predictions on novel reactions. Additionally, efforts to apply our modeling framework to less-structured reaction types such as oxidations and reductions are ongoing. Future studies will address the interplay between structure representation and classifier chaining, as well as the extension of our reaction attention mechanism to other tasks. We expect the work herein to be very informative for future condition prediction studies, a highly valuable but underexplored learning task.

3.6 EXPERIMENTAL SECTION

3.6.1 DATA PREPARATION AND REACTION DICTIONARIES

All datasets used herein were obtained from queries to the Reaxys® database.[38] Results from both journals and patents are included for all reaction types. For each reaction, the atoms involved in cross-coupling bond formation are enumerated in reactants and products, specified in the Reaxys® queries *via* atom mapping. The number of reactions in Table 3.2 refers to counts after pre-processing. General pre-processing details used for all reactions are included below, followed by reaction-specific procedures. It should be noted that the four reactions modeled here share a similar format, typically containing one or two "coupling partners" (i.e. reactants) that react to form a single product upon treatment with a set of categorized conditions. In principle, the modeling framework described herein can be applied to any reaction format, provided a condition dictionary is written. Studies applying the current framework to reactions of other formats are currently ongoing.

General pre-processing.[37] In all datasets, reactions without reactant or product structures, condition information, or reaction yields are removed unless otherwise specified. Duplicate condition information is removed such that each reaction contained only unique entries for each dictionary category (*vide infra*). It is possible that this duplicate information is meant to signify multiple equivalents of reaction components (i.e. stoichiometry), but given the unstructured nature of this feature in Reaxys®, we do not consider it here. After trimming, condition entries are pooled, and all unique values extracted with their frequency of appearance in the trimmed dataset to construct the full-length reaction dictionary. The dictionary is truncated at a cumulative 95% coverage of instance frequencies. This serves to avoid both sparsity in label bins and unreasonably lengthy curation. All resulting entries

are assigned an identifier for their reaction role, which we term categories. The dictionaries are then sorted into categories for further processing. Reagents that could serve multiple reaction roles are assigned each plausible identifier and copied into each respective category.

Once categorized, the frequencies of each bin are recalculated within the category, and each category is again truncated at 95% total coverage. At the data scale studied here, reagent bins outside of this threshold typically appear in the dataset under 10 times total. We therefore exclude them to avoid unnecessary dimensionality in the label space and reduce overfitting. Even still, class-imbalance within the categories is pronounced,[43] as evidenced by the "long-tail" distributions provided in Section 3.6.5.[59] Categories are combed to identify any bins with unique names that refer to the same reagent, often by misspelling, abbreviation, etc. At this stage, all unique species are assigned a bin label within their category, and the categories are recombined. This constitutes the final reaction dictionary and defines the label space for our prediction task. Appropriate categories were also assigned "null" bins to indicate reactions not specifying labels in that category.

Additional Suzuki pre-processing. The initial export contained 154,634 data points at time of download. Reactions without yields (~1,500) or solvent (~2,700) were removed. Reactions with more or less than 2 reactants, and more than one product (~6,800) were removed. The final dataset contained 145,413 reactions with a dictionary of 118 bins in the categories metal (M), ligand (L), base (B), solvent (S), and additive (A). Roughly 25% of reactions did not have temperatures specified, so this feature was dropped.

Additional C–N pre-processing. The initial export contained 39,902 data points at time of download. Reactions without solvent were retained, as this dataset included non-

catalytic C–N couplings such as S_NAr reactions, often run in the neat amine reactant. Reactions with more or less than 2 reactants and 1 product (~3,000) were removed. Reactions with more than 4 reagents (~250) were removed. The final dataset contained 36,519 reactions with a dictionary of 205 bins in the categories metal (M), ligand (L), base (B), solvent (S), and additive (A). Roughly 30% of reactions did not have temperatures specified, so this feature was dropped.

Additional Negishi pre-processing. The initial export contained 11,388 data points at time of download. Reactions without yields (~3,500) or solvent (~350) were removed. Reactions with more than 2 reactants, 3 solvents, and/or 4 reagents (~1,200) were removed. The final dataset contained 6,391 reactions with a dictionary of 105 bins in the categories metal (M), ligand (L), temperature (T), solvent (S), and additive (A). Almost 90% of reactions had temperatures specified, so this feature was retained. Those with unspecified temperatures were assumed to occur ambiently and assigned as 20 °C.

Additional PKR pre-processing. The initial export contained 4,275 data points at time of download. Both inter- and intramolecular reactions were retained, so reactions contained either 1 or 2 reactants. Reactions without yields (~1,000) were removed. Reactions with more than 2 reactants, 3 solvents, and/or 4 reagents (~500) were removed. The final dataset contained 2,749 reactions with a dictionary of 83 bins in the categories metal (M), ligand (L), temperature (T), solvent (S), additive (O), activator (A), gas (G), and pressure (P). The gas category is a binary identifier for the use of a carbon monoxide (CO) atmosphere.

An example protocol with full code is included in the associated github repository: https://github.com/slryou41/reaction-gcnn.git in the path: `data/data_processing_example.ipynb`. The worked example includes

procedures for sorting reagents into categories by reaction role and aggregating into a full reaction dictionary. Final dictionaries for all four datasets as .csv files can be found in the repository path: `data/all_dictionaries/`, and are tabulated below.

Table 3.12 *Suzuki dataset dictionary.*

category	bin label	condition name	dataset instances
metal	M1	tetrakis(triphenylphosphine) palladium(0)	55829
	M2	palladium diacetate	16927
	M3	(1,1'-bis(diphenylphosphino)ferrocene)palladium(II) dichloride	13723
	M4	dichloro(1,1'-bis(diphenylphosphanyl)ferrocene)palladium(II)*CH2 Cl2	8918
	M5	bis-triphenylphosphine-palladium(II) chloride	8761
	M6	tris-(dibenzylideneacetone)dipalladium(0)	5241
	M7	palladium dichloride	1512
	M8	bis(dibenzylideneacetone)-palladium(0)	1013
	M9	dichloro[1,1'-bis(di-t-butylphosphino)ferrocene]palladium(II)	1074
	M10	bis(tri-t-butylphosphine)palladium(0)	736
	M11	chloro(2-dicyclohexylphosphino-2?,4?,6?-triisopropyl-1,1?-biphenyl)[2-(2?-amino-1,1?-biphenyl?)]palladium(II)	729
	M12	bis(di-tert-?butyl(4-?dimethylaminophenyl)?phosphine)?dichloropalladium(II)	711
	M13	bis(eta3-allyl-mu-chloropalladium(II))	559
	M14	tris(dibenzylideneacetone)dipalladium(0) chloroform complex	509
	M15	palladium 10% on activated carbon	861
	M16	sodium tetrachloropalladate(II)	283
	M17	palladium	280

	M18	(2-dicyclohexylphosphino-2?,4?,6?-triisopropyl-1,1?-biphenyl)[2-(2?-amino-1,1?-biphenyl)]palladium(II) methanesulfonate	191
	M19	bis(benzonitrile)palladium(II) dichloride	179
	M20	(1,2-dimethoxyethane)dichloronickel(II)	158
	M21	bis(1,5-cyclooctadiene)nickel (0)	155
	M22	[1,3-bis(2,6-diisopropylphenyl)imidazol-2-ylidene](3-chloropyridyl)palladium(ll) dichloride	151
	M23	(bis(tricyclohexyl)phosphine)palladium(II) dichloride	148
	M24	dichloro bis(acetonitrile) palladium(II)	143
	M25	Pd EnCat-30TM	137
	M26	nickel(II) nitrate hexahydrate	106
	M27	palladium(II) trifluoroacetate	106
	M28	dichloro{bis[1-(dicyclohexylphosphanyl)piperidine]}palladium(II)	102
ligand	L1	triphenylphosphine	4489
	L2	dicyclohexyl-(2',6'-dimethoxybiphenyl-2-yl)-phosphane	3163
	L3	XPhos	2100
	L4	tricyclohexylphosphine	1808
	L5	tris-(o-tolyl)phosphine	902
	L6	tri-tert-butyl phosphine	694
	L7	tri tert-butylphosphoniumtetrafluoroborate	616
	L8	trisodium tris(3-sulfophenyl)phosphine	556
	L9	1,1'-bis-(diphenylphosphino)ferrocene	486
	L10	4,5-bis(diphenylphos4,5-bis(diphenylphosphino)-9,9-dimethylxanthenephino)-9,9-dimethylxanthene	424
	L11	CyJohnPhos	370
	L12	ruphos	293
	L13	1,3,5,7-tetramethyl-8-phenyl-2,4,6-trioxa-8-phosphatricyclo[3.3.1.13,7]decane	279
	L14	tricyclohexylphosphine tetrafluoroborate	240
	L15	johnphos	223
	L16	4,4'-di-tert-butyl-2,2'-bipyridine	216

	L17	catacxium A	192
	L18	trifuran-2-yl-phosphane	183
	L19	triphenyl-arsane	182
	L20	1,1'-bis(di-tertbutylphosphino)ferrocene	142
	L21	2,2'-bis-(diphenylphosphino)-1,1'-binaphthyl	129
	L22	Tedicyp	218
	L23	bis[2-(diphenylphosphino)phenyl] ether	108
base	B1	potassium carbonate	48981
	B2	sodium carbonate	39769
	B3	potassium phosphate	17799
	B4	caesium carbonate	13345
	B5	sodium hydrogencarbonate	3722
	B6	cesium fluoride	2810
	B7	sodium hydroxide	2156
	B8	potassium hydroxide	2155
	B9	potassium fluoride	2097
	B10	triethylamine	1370
	B11	potassium phosphate tribasic trihydrate	1016
	B12	potassium acetate	931
	B13	potassium tert-butylate	912
	B14	potassium phosphate monohydrate	826
	B15	sodium acetate	418
	B16	sodium t-butanolate	392
	B17	barium dihydroxide	374
	B18	N-ethyl-N,N-diisopropylamine	336
	B19	lithium hydroxide	321
	B20	potassium phosphate tribasic heptahydrate	317
	B21	diisopropylamine	209
	B22	sodium methylate	175
	B23	tetrabutyl ammonium fluoride	173
	B24	barium hydroxide octahydrate	171
	B25	potassium dihydrogenphosphate	166
	B26	potassium fluoride dihydrate	156
	B27	1,4-diaza-bicyclo[2.2.2]octane	154
	B28	lithium hydroxide monohydrate	143
	B29	tetra-butylammonium acetate	137

	B30	sodium phosphate	133
	B31	potassium hydrogencarbonate	131
	B32	dipotassium hydrogenphosphate	127
	B33	tripotassium phosphate "n" hydrate	123
	B34	cesiumhydroxide monohydrate	112
	B35	sodium phosphate dodecahydrate	103
solvent	S1	tetrahydrofuran	18113
	S2	ethanol	24836
	S3	methanol	4374
	S4	1,4-dioxane	39107
	S5	1,2-dimethoxyethane	19131
	S6	acetonitrile	4366
	S7	toluene	28304
	S8	N,N-dimethyl formamide	15110
	S9	water	92175
	S10	1-methyl-pyrrolidin-2-one	472
additive	A1	tetrabutylammomium bromide	3003
	A2	water	1606
	A3	lithium chloride	819
	A4	hydrogenchloride	780
	A5	copper(I) iodide	546
	A6	silver(I) oxide	405
	A7	copper diacetate	183
	A8	dmap	181
	A9	Aliquat 336	169
	A10	cetyltrimethylammonim bromide	167
	A11	copper(I) chloride	164
	A12	potassium bromide	157
	A13	trifluoroacetic acid	151
	A14	oxygen	148
	A15	air	112
	A16	18-crown-6 ether	127
	A17	sodium dodecyl-sulfate	113

Table 3.13 *C–N dataset dictionary.*

category	bin label	condition name	dataset instances
metal	M1	copper(I) iodide	8180
	M2	tris-(dibenzylideneacetone)dipalladium(0)	6995
	M3	palladium diacetate	4668
	M4	copper	1875
	M5	bis(dibenzylideneacetone)-palladium(0)	1292
	M6	copper(I) oxide	932
	M7	copper(II) oxide	402
	M8	copper(I) chloride	386
	M9	copper(I) bromide	348
	M10	bis(eta3-allyl-mu-chloropalladium(II))	433
	M11	copper(II) acetate monohydrate	352
	M12	(1,1'-bis(diphenylphosphino)ferrocene)palladium(II) dichloride	159
	M13	bis(tri-t-butylphosphine)palladium(0)	181
	M14	iron(III) chloride	116
	M15	copper(II) bis(trifluoromethanesulfonate)	91
	M16	copper(II) bromide	88
	M17	bis-triphenylphosphine-palladium(II) chloride	82
	M18	copper(II) sulfate	154
	M19	bis(acetylacetonate)nickel(II)	78
	M20	palladium 10% on activated carbon	71
	M21	tetrakis(triphenylphosphine) palladium(0)	68
	M22	dichlorobis(tri-O-tolylphosphine)palladium	67
	M23	(1,2-dimethoxyethane)dichloronickel(II)	66
	M24	palladium dichloride	63
	M25	copper(I) thiophene-2-carboxylate	58
	M26	cobalt(II) oxalate dihydrate	56
	M27	copper dichloride	52
	M28	dichloro(1,3-bis(2,6-bis(3-pentyl)phenyl)imidazolin-2-ylidene)(3-chloropyridyl)palladium(II)	49

	M29	chloro[2-(dicyclohexylphosphino)-3 ,6-dimethoxy-2?,4?, 6?-triisopropyl- 1,1?-biphenyl] [2-(2-aminoethyl)phenyl]palladium(II)	97
	M30	[2-(di-tert-butylphosphino)-2?,4?,6?-triisopropyl-1,1?-biphenyl][2-((2-aminoethyl)phenyl)]palladium(II) chloride	49
	M31	iron(III) oxide	48
	M32	C36H45Cl2N3OPd	46
	M33	nickel(II) bromide trihydrate	45
	M34	copper acetylacetonate	45
	M35	C36H43Cl2N3Pd	45
	M36	C30H43O2P*C13H12N(1-)*CH3O3S(1-)*Pd(2+)	45
	M37	bis(1,5-cyclooctadiene)nickel (0)	45
	M38	CuPy2Cl2	42
	M39	dichloro(3-chloropyridinyl)(1,3-(diisopropylphenyl)-4,5-bis(dimethylamino)imidazol-2-ylidene)palladium(II)	41
	M40	Al2O3#dotCu(2+)	40
	M41	C33H40ClN3O2Pd	38
	M42	dichloro(1,1'-bis(diphenylphosphanyl)ferrocene)palladium(II)*CH2Cl2	36
	M43	(1,3-bis(2,6-diisopropylphenyl)-3,4,5,6-tetrahydropyrimidin-2-ylidene)Pd(cinnamyl, 3-phenylallyl)Cl	36
	M44	copper(II) iodide	35
ligand	L1	2,2'-bis-(diphenylphosphino)-1,1'-binaphthyl	3014
	L2	tri-tert-butyl phosphine	2137
	L3	4,5-bis(diphenylphos4,5-bis(diphenylphosphino)-9,9-dimethylxanthenephino)-9,9-dimethylxanthene	1995
	L4	N,N`-dimethylethylenediamine	1543
	L5	XPhos	830
	L6	1,10-Phenanthroline	703
	L7	L-proline	620

	L8	1,1'-bis-(diphenylphosphino)ferrocene	653
	L9	johnphos	444
	L10	DavePhos	374
	L11	triphenylphosphine	275
	L12	ruphos	266
	L13	tri tert-butylphosphoniumtetrafluoroborate	265
	L14	tert-butyl XPhos	242
	L15	dicyclohexyl-(2',6'-dimethoxybiphenyl-2-yl)-phosphane	261
	L16	trans-1,2-Diaminocyclohexane	724
	L17	8-quinolinol	206
	L18	CyJohnPhos	192
	L19	trans-N,N'-dimethylcyclohexane-1,2-diamine	535
	L20	ethylenediamine	175
	L21	dimethylaminoacetic acid	167
	L22	dicyclohexyl[3,6-dimethoxy-2?,4?,6?-tris(1-methylethyl)[1,1?-biphenyl]-2-yl]phosphine	165
	L23	2,2,6,6-tetramethylheptane-3,5-dione	163
	L24	1,1'-bi-2-naphthol	162
	L25	bis[2-(diphenylphosphino)phenyl] ether	170
	L26	1-dicyclohexylphosphino-2-di-tert-butylphosphinoethylferrocene	142
	L27	P(i-BuNCH2)3CMe	110
	L28	di-tert-butyl{2?-isopropoxy-[1,1?-binaphthalen]-2-yl}phosphane	108
	L29	di-tert-butyl(2,2-diphenyl-1-methyl-1-cyclopropyl)phosphine	104
	L30	P(i-BuNCH2CH2)3N	98
	L31	N,N-dimethylglycine hydrochoride	96
	L32	N-[2-(di(1-adamantyl)phosphino)phenyl]morpholine	92
	L33	5-(di-tert-butylphosphino)-1?, 3?, 5?-triphenyl-1?H-[1,4?]bipyrazole	91
	L34	2-[2-(dicyclohexylphosphino)-phenyl]-1-methyl-1H-indole	86
	L35	4,4'-di-tert-butyl-2,2'-bipyridine	85

	L36	tris-(o-tolyl)phosphine	77
	L37	2,8,9-tris(2-methylpropyl)-2,5,8,9-tetraaza-1-phosphabicyclo[3.3.3]undecane	75
	L38	cis-N,N'-dimethyl-1,2-diaminocyclohexane	74
	L39	monophosphine 1,2,3,4,5-pentaphenyl-1'-(di-tert-butylphosphino)ferrocene	55
	L40	5-(di(adamantan-1-yl)phosphino)-1?,3?,5?-triphenyl-1?H-1,4?-bipyrazole	55
	L41	t-BuBrettPhos	53
	L42	2-(N,N-dimethylamino)athanol	53
	L43	tricyclohexylphosphine	46
	L44	(E)-3-(dimethylamino)-1-(2-hydroxyphenyl)prop-2-en-1-one	46
	L45	di-tert-butylneopentylphosphonium tetrafluoroborate	38
	L46	2-di-tertbutylphosphino-3,4,5,6-tetramethyl-2',4',6'-triisopropyl-1,1'-biphenyl	37
	L47	N,N,N,N,-tetramethylethylenediamine	26
base	B1	sodium t-butanolate	9103
	B2	potassium carbonate	7129
	B3	caesium carbonate	6957
	B4	potassium phosphate	3274
	B5	potassium tert-butylate	2167
	B6	potassium hydroxide	1420
	B7	triethylamine	500
	B8	lithium hexamethyldisilazane	432
	B9	sodium hydroxide	430
	B10	sodium hydride	228
	B11	sodium carbonate	200
	B12	potassium phosphate monohydrate	130
	B13	sodium hydrogencarbonate	128
solvent	S1	toluene	11970
	S2	1,4-dioxane	5273
	S3	N,N-dimethyl-formamide	4246
	S4	dimethyl sulfoxide	3790
	S5	water	2464

	S6	tetrahydrofuran	1457
	S7	1,2-dimethoxyethane	878
	S8	tert-butyl alcohol	841
	S9	acetonitrile	780
	S10	ethanol	549
	S11	5,5-dimethyl-1,3-cyclohexadiene	497
	S12	isopropyl alcohol	316
	S13	nitrobenzene	315
	S14	1-methyl-pyrrolidin-2-one	292
	S15	hexane	286
	S16	N,N-dimethyl acetamide	281
	S17	1,2-dichloro-benzene	254
	S18	neat (no solvent)	240
	S19	o-xylene	219
	S20	xylene	208
	S21	methanol	180
	S22	ethyl acetate	163
additive	A1	18-crown-6 ether	455
	A2	tetrabutylammomium bromide	372
	A3	8-quinolinol	206
	A4	dimethylaminoacetic acid	167
	A5	1,1'-bi-2-naphthol	162
	A6	water	160
	A7	sodium sulfate	132
	A8	2-(2-methyl-1-oxopropyl)cyclohexanone	121
	A9	phenylboronic acid	120
	A10	1,3-bis[(2,6-diisopropyl)phenyl]imidazolinium chloride	109
	A11	potassium iodide	108
	A12	hydrogenchloride	107
	A13	ethylene glycol	102
	A14	N,N-dimethylglycine hydrochoride	96
	A15	1,3-bis[2,6-diisopropylphenyl]imidazolium chloride	95
	A16	N-ethylmorpholine	93
	A17	tert-butyl alcohol	87

	A18	aluminum oxide	84
	A19	D-glucose	83
	A20	cetyltrimethylammonim bromide	71
	A21	1,3-dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidinone	68
	A22	N',N'-diphenyl-1H-pyrrole-2-carbohydrazide	63
	A23	manganese(II) fluoride	63
	A24	dimethyl sulfoxide	55
	A25	2-(N,N-dimethylamino)athanol	53
	A26	air	48
	A27	iron(III) oxide	48
	A28	(E)-3-(dimethylamino)-1-(2-hydroxyphenyl)prop-2-en-1-one	46
	A29	lithium bromide	44
	A30	6,7-dihydro-5H-quinolin-8-one oxime	43
	A31	CVT-2537	42
	A32	ammonium chloride	42
	A33	1-methyl-pyrrolidin-2-one	42
	A34	tetra(n-butyl)ammonium hydroxide	40
	A35	salicylaldehyde-oxime	39
	A36	potassium fluoride on basic alumina	39
	A37	toluene-4-sulfonic acid	38
	A38	lithium chloride	38
	A39	pipecolic Acid	37
	A40	oxygen	37
	A41	metformin hydrochloride	37
	A42	8-Hydroxyquinoline-N-oxide	37
	A43	1-(5,6,7,8-tetrahydroquinolin-8-yl)ethan-1-one	36
	A44	tetrabutyl ammonium fluoride	36
	A45	N1,N2-bis(thiophen-2-ylmethyl)oxalamide	36
	A46	N-phenyl-2-pyridincarboxamide-1-oxide	35
	A47	N-((1-oxy-pyridin-2-yl)methyl)oxalamic acid	35
	A48	C19H19N5O	35
	A49	manganese(II) chloride tetrahydrate	34
	A50	1-tetralone oxime	32
	A51	N1,N2-bis(2,4,6-trimethoxyphenyl)oxalamide	31

	A52	N-methoxy-1H-pyrrole-2-carboxamide	29
	A53	ammonia	29
	A54	1,2,3-Benzotriazole	29
	A55	dimethylenecyclourethane	28
	A56	isopropylmagnesium chloride	27
	A57	N-(2-cyanophenyl)pyridine-2-carboxamide	27
	A58	C20H18N2O2	27
	A59	2-acetylcyclohexanone	27
	A60	2,6-di-tert-butyl-4-methyl-phenol	26
	A61	2-hydroxy-pyridine N-oxide	26
	A62	TPGS-750-M	25
	A63	N?-phenyl-1H-pyrrole-2-carbohydrazide	25
	A64	lanthanum(III) oxide	25
	A65	ethylmagnesium bromide	25
	A66	ethyl 2-oxocyclohexane carboxylate	25
	A67	1,4-dimethyl-1,2,3,4-tetrahydro-5H-benzo[e][1,4]diazepin-5-one	25
	A68	tetraethoxy orthosilicate	24
	A69	N,N,N',N'-tetramethylguanidine	24
	A70	C20H26N4O4	24
	A71	2-methyl-8-quinolinol	24
	A72	2-carbomethoxy-3-hydroxyquinoxaline-di-N-oxide	24
	A73	1,3-diisopropyl-1H-imidazol-3-ium chloride	24
	A74	MOF-199	24

Table 3.14 *Negishi dataset dictionary.*

Category	bin label	condition name	dataset instances
metal	M1	tetrakis(triphenylphosphine) palladium(0)	1902
	M2	tris-(dibenzylideneacetone)dipalladium(0)	572
	M3	bis-triphenylphosphine-palladium(II) chloride	418
	M4	palladium diacetate	370
	M5	bis(dibenzylideneacetone)-palladium(0)	344

	M6	(1,1'-bis(diphenylphosphino)ferrocene)palladium(II) dichloride	334
	M7	bis(tri-t-butylphosphine)palladium(0)	273
	M8	dichloro(1,1'-bis(diphenylphosphanyl)ferrocene)palladium(II)*CH2Cl2	248
	M9	dichloro{bis[1-(dicyclohexylphosphanyl)piperidine]}palladium(II)	168
	M10	palladium(I) tri-tert-butylphosphine iodide dimer	101
	M11	bis(tricyclohexylphosphine)nickel(II) dichloride	99
	M12	[(C10H13-1,3-(CH2P(C6H11)2)2)Pd(Cl)]	87
	M13	1,3-bis[(diphenylphosphino)propane]dichloronickel(II)	63
	M14	bis(1,5-cyclooctadiene)nickel (0)	56
	M15	nickel dichloride	56
	M16	tris(dibenzylideneacetone)dipalladium(0) chloroform complex	46
	M17	dichlorobis(tri-O-tolylphosphine)palladium	46
	M18	palladium	44
	M19	[1,3-bis(2,6-diisopropylphenyl)imidazol-2-ylidene](3chloro-pyridyl)palladium(II) dichloride	136
	M20	C20H20ClN3Ni	42
	M21	dichloro(1,3-bis(2,6-bis(3-pentyl)phenyl)imidazolin-2-ylidene)(3-chloropyridyl)palladium(II)	39
	M22	bis(triphenylphosphine)nickel(II) chloride	38
	M23	C26H24ClN2NiP*0.1C7H8	35
	M24	cobalt(II) chloride	34
	M25	copper(I) bromide	31
	M26	C40H55Cl5N3Pd	30
	M27	[1,3-bis(2,6-diisoheptylphenyl)-4,5-dichloroimidazol-2-ylidene](3-chloropyridyl)palladium(II) dichloride	29

	M28	dichloro bis(acetonitrile) palladium(II)	29
	M29	palladium(II) trifluoroacetate	27
	M30	1,2-bis(diphenylphosphino)ethane nickel(II) chloride	27
	M31	C27H22Cl2N3NiP	24
	M32	C38H34Br2N4Ni2P2	23
ligand	L1	1,1'-bis-(diphenylphosphino)ferrocene	233
	L2	dicyclohexyl-(2',6'-dimethoxybiphenyl-2-yl)-phosphane	196
	L3	XPhos	187
	L4	triphenylphosphine	161
	L5	trifuran-2-yl-phosphane	128
	L6	monophosphine 1,2,3,4,5-pentaphenyl-1'-(di-tert-butylphosphino)ferrocene	95
	L7	tris-(o-tolyl)phosphine	70
	L8	Ruphos	61
	L9	2?-(dicyclohexylphophanyl)-N2,N2,N6,N6-tetramethyl[1,1?-biphenyl]-2,6-diamine	37
	L10	tripiperidino-phosphine	37
	L11	tri tert-butylphosphoniumtetrafluoroborate	35
	L12	1,2-bis-(dicyclohexylphosphino)ethane	33
	L13	4,5-bis(diphenylphos4,5-bis(diphenylphosphino)-9,9-dimethylxanthenephino)-9,9-dimethylxanthene	31
	L14	N,N,N,N,-tetramethylethylenediamine	24
	L15	[2,2]bipyridinyl	22
	L16	4,4'-di-tert-butyl-2,2'-bipyridine	21
	L17	1,2-Ph2-3,4-bis(2,4,6-(t-Bu)3-phenylphophinidene)cyclobutene	20
	L18	johnphos	20
	L19	tri-tert-butyl phosphine	19
	L20	tricyclohexylphosphine	18
temperature	T1	T<18	101
	T2	18≤T<23	2313
	T3	23≤T<50	643
	T4	50≤T<61	975

	T5	61≤T<80	658
	T6	80≤T<100	673
	T7	100≤T<120	696
	T8	T≥120	479
solvent	S1	tetrahydrofuran	4525
	S2	N,N-dimethyl-formamide	1003
	S3	1-methyl-pyrrolidin-2-one	674
	S4	toluene	541
	S5	1,4-dioxane	335
	S6	N,N-dimethyl acetamide	247
	S7	hexane	219
	S8	diethyl ether	203
	S9	water	122
	S10	1,2-dimethoxyethane	67
additive	A1	lithium chloride	243
	A2	zinc	207
	A3	copper(I) iodide	154
	A4	water	62
	A5	diisobutylaluminium hydride	59
	A6	tetrabutylammomium bromide	52
	A7	ammonium chloride	51
	A8	n-butyllithium	46
	A9	1-Methylpyrrolidine	42
	A10	Li2CoCl4	42
	A11	sodium formate	42
	A12	hydrogenchloride	36
	A13	caesium carbonate	36
	A14	zinc diacetate	32
	A15	potassium carbonate	30
	A16	norborn-2-ene	30
	A17	lithium bromide	28
	A18	1,3-dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidinone	23
	A19	methylzinc chloride	22
	A20	1-methyl-pyrrolidin-2-one	21
	A21	zinc(II) chloride	21

	A22	isoquinoline	20
	A23	sodium carbonate	19
	A24	1-ethyl-2-pyrrolidinone	18
	A25	sodium	16
	A26	1-methyl-1H-imidazole	15
	A27	oxovanadium(V) ethoxydichloride	12
	A28	2-(N,N-dimethylamino)athanol	11
	A29	[bdmim][BF4]	11
	A30	1-butyl-2-(diphenylphosphanyl)-3-methylimidazolium hexafluorophosphate	11

Table 3.15 *PKR dataset dictionary.*

category	bin label	condition name	dataset instances
metal	M1	dicobalt octacarbonyl	614
	M2	di(rhodium)tetracarbonyl dichloride	333
	M3	chloro(1,5-cyclooctadiene)rhodium(I) dimer	140
	M4	[RhCl(CO)dppp]2	92
	M5	cobalt(II) bromide	44
	M6	palladium dichloride	33
	M7	dodecacarbonyl-triangulo-triruthenium	32
	M8	Co2Rh2 nanoparticles immobilized on charcoal	50
	M9	tetracobaltdodecacarbonyl	44
	M10	molybdenum hexacarbonyl	23
	M11	Rh(dppp)2Cl	19
	M12	cobalt nanoparticles on charcoal	36
	M13	methylidynetricobalt nonacarbonyl	25
	M14	bis(triphenylphosphine)(carbonyl)rhodium chloride	11
	M15	PdCl(OHNCCH3C6H4)(C5H5N)	10
	M16	bis(1,5-cyclooctadiene)diiridium(I) dichloride	9
	M17	diiron nonacarbonyl	9
	M18	iron(II) bis(trimethylsilyl)amide	9
ligand	L1	1,1,3,3-tetramethyl-2-thiourea	128
	L2	1,3-bis-(diphenylphosphino)propane	93
	L3	2,2'-bis-(diphenylphosphino)-1,1'-binaphthyl	31

	L4	triphenylphosphine	16
	L5	tri-n-butylphosphine sulfide	15
	L6	(S)-3,5-di-tert-butyl-4-methoxyphenyl-(6,6?-dimethoxybiphenyl-2,2?-diyl)-bis(diphenylphosphine)	12
temperature	T1	T<20	83
	T2	T=20	961
	T3	20<T≤60	299
	T4	60<T≤77	370
	T5	77<T≤94	338
	T6	94<T≤120	395
	T7	T>120	303
solvent	S1	toluene	966
	S2	dichloromethane	601
	S3	tetrahydrofuran	318
	S4	1,2-dichloro-ethane	171
	S5	1,2-dimethoxyethane	145
	S6	acetonitrile	141
	S7	not listed	102
	S8	water	71
	S9	benzene	76
	S10	para-xylene	136
	S11	hexane	43
	S12	dimethyl sulfoxide	39
	S13	1,4-dioxane	33
	S14	dibutyl ether	33
	S15	diethyl ether	22
activator	A1	4-methylmorpholine N-oxide	420
	A2	trimethylamine-N-oxide	212
	A3	dimethyl sulfoxide	137
	A4	cyclohexylamine	68
	A5	n-butyl methyl sulfide	27
	A6	silver trifluoromethanesulfonate	23
	A7	silver tetrafluoroborate	18
	A8	silver hexafluoroantimonate	19
	A9	(4-fluorobenzyl)(methyl)sulfide	14

	A10	dinitrogen monoxide	14
	A11	4-methylmorpholine 4-oxide monohydrate	13
CO (g)	G1	carbon monoxide	1169
	G2	none	1580
additive (other)	O1	4 A molecular sieve	84
	O2	zinc	50
	O3	hydrogen	40
	O4	ethylene glycol	30
	O5	cetyltrimethylammonim bromide	22
	O6	Celite	17
	O7	Triton X(R)-100	37
	O8	acetic anhydride	15
	O9	lithium chloride	15
	O10	water	11
	O11	oxygen	10
	O12	potassium carbonate	8
	O13	triethylsilane	8
pressure	P1	P<760	35
	P2	P=760	2392
	P3	760<P≤7600	169
	P4	P>7600	153

3.6.2 COMPUTATIONAL DETAILS AND HYPERPARAMETERS

3.6.2.1 Gradient-Boosting Machines (GBMs)

Numerical inputs for GBM models[33,45] were constructed by tokenizing SMILES strings[39] for each molecule in a reaction with character-to-number mappings, and calculating chemical descriptor vectors using the Python API of the Mordred calculator.[46] Coded examples for these processing protocols are provided in the associated github repository at the path `models/gbms/parsing_example_suzuki/`. All GBM classifiers were implemented using Microsoft's lightGBM.[48,49] Specific non-default parameter settings are included in Table 3.16.

Table 3.16 *Computational details and general parameters used for GBM models.*

parameter	value	description
train/valid/test	81/9/10	data splitting
max_depth	7	maximum tree depth for base learners
tree_method	gpu_hist	split continuous features into discrete bins
eval_metric	aucpr	evaluation metric

3.6.2.1.1 Binary Relevance Method (BM)

In BM experiments, an independent `lightgbm.LGBMClassifier` was fit for each label bin in a dataset's dictionary using the full input representation.[51]

3.6.2.1.2 Classifier Trellises (CT)

In CT experiments,[53] `lightgbm.LGBMClassifiers` were fit for each label bin in a dataset's dictionary as part of a grid structure in which predictions are made sequentially and are passed to downstream models as additional inputs (see main text for explanation). Mutual information (MI) matrices were constructed for each dataset's label dictionary using sci-kit learn's `sklearn.metrics.mutual_info_score` module.[68] Classifier trellises were then constructed following the algorithm reported by Read et al. (see main text and associated code for details).[37,53] As shown in the example in the main text, each model takes additional input from the bins in directions `north`, `west`, and `northwest` of it. Models on the edges of the trellis take input only from those bins in the available directions (i.e., propagation does not wrap between rows). Here each trellis was initialized using the label `M1`, the most commonly used metal in each dataset. This can be chosen by user preference, expert intuition, or at random. Full MI matrices and trellis structures for all four datasets are provided below.

Figure 3.19 *Optimized classifier trellis for the Suzuki dataset.*

Suzuki	0	1	2	3	4	5	6	7	8	9	10
A	M1	M_NULL	M2	L_NULL	L1	M3	S4	S7	S2	S1	S5
B	M5	M4	B2	B1	B3	L2	S_NULL	S9	S6	L8	B9
C	M6	L3	B4	B_NULL	B5	L4	M13	L22	B6	B15	L6
D	L7	L13	M11	B8	A_NULL	A1	M21	B13	L23	M12	M14
E	B26	M8	B14	M27	A3	A4	S3	B28	L12	A11	L9
F	L14	L5	A9	B11	B22	B7	M10	B18	B16	L21	B10
G	B25	M9	M7	A17	M16	M24	A6	L19	M19	L20	M18
H	M15	B30	B23	B31	L10	A5	B27	B12	M23	S10	B19
I	B35	B21	M17	A14	B20	B17	L15	B24	L11	A8	M20
J	L16	M26	A7	L18	A10	A12	L17	A13	B33	M22	A15
K	M25	B29	A16	B32	B34	M28	S8	A2	None	None	None

Figure 3.20 *Mutual information matrix for the Suzuki dataset.*

Suzuki	0	1	2	3	4	5	6	7	8	9	10
A	0	0.096509	0.022812	0.088898	0.069872	0.00271	0.014156	0.05331	0.052397	0.01702	0.013595
B	0.032864	0.048547	0.024051	0.144962	0.071762	0.016851	0.025646	0.091712	0.012334	0.009231	0.006801
C	0.002359	0.008967	0.031369	0.023736	0.01403	0.007808	0.005451	0.014139	0.003967	0.003043	0.00406
D	0.00682	0.00365	0.002436	0.001583	0.002287	0.066936	0.000948	0.000183	0.000755	0.000702	0.000681
E	0.001079	0.000182	0.000941	0.001686	0.018303	0.017265	0.001083	0.000846	0.000162	0.000657	0.000576
F	4.60E-05	0.000387	0.000201	0.001154	0.000893	0.000775	0.000563	0.000979	0.000263	9.15E-05	0.000275
G	4.44E-05	0.000164	0.000112	0.000673	0.000912	0.000506	0.000619	0.001873	0.000835	0.000212	0.000182
H	7.74E-06	0.001508	0.000412	0.000295	2.57E-05	0.000116	0.000527	3.03E-05	0.000373	5.41E-05	0.00013
I	0.000791	9.31E-05	4.80E-05	1.21E-05	0.000433	2.48E-05	6.91E-05	3.25E-05	3.70E-05	1.00E-05	0.008369
J	1.01E-06	0.002866	5.53E-06	5.72E-06	5.82E-06	2.09E-05	6.06E-06	3.68E-05	6.95E-06	5.40E-06	6.35E-05
K	1.45E-06	0.008351	2.73E-06	3.12E-06	2.78E-06	2.27E-06	1.78E-15	3.55E-15	None	None	None

Figure 3.21 *Optimized classifier trellis for the C–N dataset.*

C–N	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	M1	M_NULL	L_NULL	M2	B1	S1	L1	M3	B3	B2	M4	S13	A7	A_NULL	A1
B	B4	B_NULL	S_NULL	L3	L2	S3	S2	S4	L7	B5	S17	A16	M8	A2	S5
C	L16	L4	S10	S6	B8	M5	L24	B6	L31	S7	L26	S12	A13	B9	A19
D	A23	M26	M14	L10	M36	L28	S15	M7	A70	M39	A73	M11	A52	A74	M16
E	M6	S9	M23	L35	A69	S8	L5	L15	L8	M12	M20	L18	B12	A49	L19
F	A46	A35	L44	B7	A29	L34	A9	L6	A18	S20	M13	A20	A26	L42	M44
G	S11	A67	A11	B11	S16	M33	M10	L29	A40	M15	L23	L9	M24	L25	M21
H	L13	A65	L11	M17	S19	L43	L32	L40	M18	A34	M9	A42	L36	S14	M19
I	L12	S18	M43	B10	A17	A56	L14	A62	A22	L17	A72	A43	L20	S21	A10
J	M29	M28	A60	L22	A6	L41	M30	L21	A33	A53	S22	A32	B13	A12	A8
K	L27	L30	A15	L33	L46	L38	L37	A21	M22	M25	M27	A24	L39	M31	M34
L	M32	M37	A31	A30	M35	M38	A36	M40	A38	M42	L45	A39	A41	M41	A37
M	A48	A44	A45	A47	A50	A51	A54	A55	L47	A57	A59	A61	A64	A66	A58
N	A63	A68	A71	A3	A4	A5	A14	A25	A27	A28	None	None	None	None	None

Figure 3.22 *Mutual information matrix for the C–N dataset.*

C–N	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	0	0.073547	0.120096	0.083775	0.055793	0.114291	0.053129	0.025424	0.018588	0.046073	0.044959	0.021508	0.012579	0.008307	0.027749
B	0.049414	0.055036	0.059134	0.074614	0.06015	0.078562	0.066955	0.036756	0.035659	0.027453	0.021502	0.003319	0.007042	0.023636	0.036117
C	0.012605	0.013552	0.006146	0.004989	0.01415	0.012901	0.009452	0.009538	0.005913	0.005255	0.01089	0.000406	0.007037	0.006807	0.006894
D	0.006856	0.005034	0.004996	0.005457	0.005872	0.007714	0.005609	0.005052	0.00504	0.002448	0.002415	0.001498	0.003485	0.003168	0.008106
E	4.42E-05	0.003866	0.005761	0.010098	0.004365	0.003708	0.013564	0.003323	0.000338	0.002307	0.002772	0.003058	0.000993	0.002919	0.004763
F	0.003656	0.00644	0.004916	0.005547	0.004101	0.005527	0.020156	0.000646	0.002133	0.000985	0.00025	0.00552	0.002737	1.00E-04	0.00187
G	1.25E-05	0.003128	0.002674	0.000967	0.003233	0.004753	8.59E-05	0.013501	0.001922	0.003306	0.00346	0.000136	0.000183	0.000745	8.71E-05
H	0.001328	0.001261	0.001202	0.001433	0.003342	0.000258	0.010713	0.007045	0.000557	0.001991	7.92E-05	0.004235	0.000586	6.64E-05	0.000463
I	4.80E-05	0.002084	0.004955	0.000705	0.005093	0.005229	2.99E-05	0.002696	7.90E-05	3.93E-05	0.00057	0.00039	2.02E-05	7.13E-05	0.004068
J	0.001079	0.002274	0.000352	3.68E-05	6.62E-05	0.004622	0.000946	3.95E-05	0.000296	0.000111	0.000229	0.004983	2.51E-05	7.02E-05	3.75E-05
K	8.14E-06	1.85E-05	1.19E-05	1.94E-05	0.000124	1.44E-05	1.01E-05	1.49E-05	1.35E-05	5.68E-06	9.68E-06	1.00E-05	8.89E-06	1.12E-05	0.004299
L	4.26E-06	9.19E-06	8.02E-06	7.56E-06	5.26E-06	4.38E-06	5.58E-06	5.23E-06	5.00E-06	4.50E-06	4.13E-06	3.94E-06	3.88E-06	3.67E-06	3.43E-06
M	1.33E-06	3.38E-06	3.21E-06	3.12E-06	3.06E-06	2.75E-06	2.62E-06	2.50E-06	2.36E-06	2.31E-06	2.19E-06	2.15E-06	2.04E-06	1.91E-06	1.78E-06
N	6.21E-07	1.56E-06	1.45E-06	3.47E-18	3.47E-18	3.47E-18	3.47E-18	1.78E-15	1.77E-15	0	None	None	None	None	None

Figure 3.23 *Optimized classifier trellis for the Negishi dataset.*

Negishi	0	1	2	3	4	5	6	7	8	9	10
A	M1	L_NULL	M2	L1	A2	A_NULL	A1	A3	M_NULL	O1	O_NULL
B	S3	M9	L3	S1	S2	T8	T2	T4	T3	S_NULL	L17
C	T7	M12	S5	T5	T6	S6	M3	M10	A10	A11	M16
D	M4	L2	M6	M14	L12	M29	L15	S4	M26	M5	L5
E	A16	A13	S10	L4	M18	M7	M8	M27	A19	L6	S8
F	S7	A8	M15	L10	M22	A7	A4	L11	M11	S9	A12
G	L19	A5	A21	M21	A24	L8	M19	A17	A20	A14	A15
H	T1	M13	A25	L20	L7	M17	M20	A9	M23	M24	A22
I	M30	L9	M25	L13	L14	M28	M32	A18	M31	L16	L18
J	A23	A26	A6	A28	A29	A30	A27	T_NULL	None	None	None

Figure 3.24 *Mutual information matrix for the Negishi dataset.*

Negishi	0	1	2	3	4	5	6	7	8	9	10
A	0	0.084283	0.085225	0.044055	0.030646	0.063871	0.072934	0.025056	0.011172	0.082987	0.437132
B	0.030077	0.082105	0.082442	0.039851	0.154657	0.050113	0.0354	0.070173	0.024463	0.036936	0.025029
C	0.062908	0.066378	0.047095	0.028718	0.043761	0.014863	0.025536	0.015061	0.014118	0.05371	0.025252
D	0.003846	0.059757	0.019728	0.015168	0.036481	0.01511	0.010263	0.009652	0.013173	0.00102	0.039501
E	0.014352	0.041992	0.02248	0.013086	0.016298	0.000712	0.00201	0.010624	0.006797	0.023566	0.011486
F	0.000134	0.009139	0.005586	0.029858	0.002824	0.001286	0.016869	0.006107	0.000223	0.000564	0.010772
G	0.005944	0.001132	0.000382	0.006866	0.000311	0.011533	0.000385	0.000847	0.000372	0.009572	0.002582
H	3.18E-05	0.000215	0.012096	0.017011	7.66E-05	0.000149	0.000209	0.042064	8.24E-05	7.84E-05	0.018103
I	8.86E-05	0.000151	9.11E-05	5.23E-05	0.002081	8.62E-05	7.51E-05	6.89E-05	5.97E-05	5.38E-05	4.30E-05
J	1.39E-05	3.35E-05	2.63E-05	1.98E-05	1.71E-05	0.012082	3.54E-15	3.54E-15	None	None	None

Figure 3.25 *Optimized classifier trellis for the PKR dataset.*

PKR	0	1	2	3	4	5	6	7	8	9
A	M1	M_NULL	G1	G_NULL	A_NULL	A1	S2	S1	S3	M6
B	M2	S10	P2	P3	A2	T2	T6	S4	T5	T3
C	O_NULL	O1	P4	T7	M12	T4	L1	L_NULL	L2	M3
D	O2	M5	M13	S14	O8	S5	A4	S9	L5	S_NULL
E	S6	A10	N1	N_NULL	L3	S13	M9	O7	S8	O5
F	O10	T1	O3	A3	P1	O4	S7	M11	O6	M4
G	M8	M17	A8	M14	S11	O11	S12	M10	M7	A7
H	A5	A6	L6	S15	L4	M16	O9	A9	M15	O12
I	O13	M18	A11	T_NULL	P_NULL	None	None	None	None	None

Figure 3.26 *Mutual information matrix for the PKR dataset.*

PKR	0	1	2	3	4	5	6	7	8	9
A	0	0.148624	0.160515	0.68243	0.102405	0.186185	0.131708	0.071628	0.049108	0.023795
B	0.035102	0.053809	0.084487	0.233286	0.135261	0.170539	0.104031	0.05365	0.039864	0.047309
C	0.018397	0.07792	0.133582	0.086451	0.039753	0.070014	0.051156	0.118731	0.084137	0.072301
D	0.041782	0.115446	0.008828	0.028101	0.040167	0.02588	0.086418	0.043118	0.03207	0.0205
E	0.001884	0.015424	0.002866	0.453932	0.013475	0.019953	0.015342	0.024111	0.008457	0.032315
F	0.007421	0.006052	0.012843	0.012669	0.009257	0.007902	0.005033	0.022851	0.023849	0.001589
G	8.69E-05	0.006557	0.001302	0.022536	0.001213	0.002937	0.000731	0.029565	0.000277	0.003329
H	0.000182	0.000289	0.026024	0.00011	0.000167	0.00096	0.000128	0.000159	0.000113	7.75E-05
I	2.94E-05	4.65E-05	2.08E-05	1.77E-15	8.82E-16	None	None	None	None	None

3.6.2.2 Graph Convolutional Networks (GCNs)

Molecular graph calculations and all neural network (NN) architectures tested herein were implemented using the Chainer Chemistry (ChainerChem) library.[54] In all cases, a graph processing network (GPN) was constructed and combined with a dense multi-layer perceptron (MLP), which were trained together as a joint network. All models were trained for 100 epochs on 1 NVIDIA K80 GPU device, unless otherwise specified. Training and test sets were held consistent between models for each reaction dataset. This was done by first splitting each dataset into 90/10 train/test, then splitting the training set into 90/10 train/validation, resulting in a final split of 81/9/10 train/validation/test overall. A dummy predictor that always predicts the most frequent bin in each label category was also created for each dataset as a baseline performance reference.

General parameters and hyperparameter settings are summarized in Table 3.17, which are held constant across all models for all datasets unless otherwise specified in Table 3.18. Every attempt was made to keep shared hyperparameters consistent between model types, and the majority were set to defaults. However, there were certain cases where this resulted in excessive memory requirements and crashes during model training. To adapt to these cases, parameters such as the hidden dimension (`hidden_dim`) were incrementally decreased until training was successful. These cases are noted in Table 3.18.

Model names listed follow those from original references (see main text), names in parentheses refer to those used by Chainer functions. It should be noted that while the predicted vectors contain a single label from each category, it is possible that the ground truth contains more than one or zero. We add a `Null` label to each sub-dictionary to handle the zero case, and with multiple ground truths we treat a category's prediction as accurate

if any are correctly identified. The `Null` case was found commonly in ligand and additive categories, where a `Null` ligand often resulted from use of a pre-ligated metal source. The multi-output scenario arose most frequently in the form of mixed solvent systems.

Table 3.17 *Computational details and general parameters used for GCN models.*

parameter	value	description
loss	sigmoid cross entropy	loss function used for training
optimizer	Adam[56]	model optimization algorithm
train/valid/test	81/9/10	data splitting
batch size	32	batch size used for gradient calculations
epochs	100	number of training iterations
out_dim	128	number of units in the readout
hidden_dim	128	number of units in the hidden layers
n_layers	4	number of convolutional layers
n_atom_types	117	number of allowed atom types
concat_hidden	False	readouts concatenated at each layer

Table 3.18 *Additional architecture-specific hyperparameter settings.*

model	parameter	value	description
NFP	max_degree	6	max degree of atom nodes in graphs
GGNN	weight_tying	True	use weight tying
	num_edge_type	4	edge (bond) types allowed (4 includes single, double, triple, and aromatic)
MPNN	weight_tying	True	use weight tying
	message_func	edgenet	message function
	readout_func	set2set	readout function (pooling)
	hidden_dim	16	default 128 caused excessive memory
	batch size	8	default 32 required excessive memory
	epochs	32 & 5	Suzuki & CN; memory errors found at higher epochs; validation loss had converged

Weave (WeaveNet)	weave_channels	200	weave channel dimensionality
	n_atom	20	number of atoms in input arrays
	n_sub_layer	1	number of layers in each pairing layer
	readout_mode	sum	readout mode (pooling)
	epochs	10	Suzuki only; memory errors found at higher epochs; validation loss had converged
R-GAT (RelGAT)	n_heads	3	number of attention heads
	negative_slope	0.2	LeakyRELU activation negative angle
	dropout_ratio	-1	dropout for normalized attention coefficients
	softmax_mode	across	method for taking softmax over logits
	concat_heads	False	concatenate multi-head attention results
	weight_tying	False	use weight tying
	hidden_dim	12	default 128 caused excessive memory
R-GCN (RelGCN) & AR-GCN	out_channels	128	output feature vector dimensionality
	ch_list	None	channels in update layers
	input_type	int	input vector type
	scale_adj	True	normalize adjacency matrix
RS-GCN (RSGCN)	use_batch_norm	False	apply bach normalization after convolutions
	readout	None	readout mode (None defaults to 'sum')
	dropout_ratio	0.5	dropout function ratio
MLP	out_dim	class_num	custom for number of classes in each dataset
	n_layers	2	number of dense layers
	activation	relu	activation function for hidden layers

3.6.3 ADVERSARIAL CONTROLS

Several control studies were conducted to ensure the validity of chemical feature learning.[65] The main text and results tables above describe the reference model known as the "dummy" predictor. This model simply returns the most frequently occurring top-k labels in each dataset category as its prediction for each task. This gives a baseline accuracy to compare

the actual learning our models achieve beyond simply fitting the dictionary frequency distributions.

3.6.3.1 Binary Relevance GBMs (BM-GBMs)

Additional controls conducted for the BM-GBM models included:

1. Shuffling full inputs, leaving outputs in place (**shuffle inputs**);
2. ablating the SMILES region of the inputs (**Mordred only**);
3. substituting Mordred vectors with random unique numerical vectors (**random Mordred**);
4. shuffling Mordred inputs, leaving outputs in place (**shuffle Mordred**);
5. ablating the descriptor regions of the inputs (**SMILES only**);
6. SMILES vectors with random unique numerical vectors (**random SMILES**); and
7. shuffling SMILES inputs, leaving outputs in place (**shuffle SMILES**)

Each of these control models are compared to the standard BM-GBM model as well as the dummy predictor for all four datasets in Tables 3.19 – 3.22.

The results of these studies have very important implications, as they indicate that SMILES tokens are insufficient chemical representations for structure learning. On its own, that the SMILES-only models perform similarly to the hybrid (full input, **BM-GBM**) models in many cases could simply indicate the reduced dimensionality (300 vs 2304) helps feature learning and/or prevents underfitting. However, since some of the random SMILES control models retain similar accuracy, it must be concluded that little chemical feature learning is taking place in the SMILES models. Rather, SMILES are likely acting as unique "barcodes" for reaction components,[65,69] and models are presumably able to glean enough information from the presence of specific molecules in reactions to establish

some accuracy greater than baseline. This is perhaps unsurprising in datasets such as the Negishi, where the chemical space is limited by dataset size and less-accessible organometallic nucleophiles than in datasets like Suzuki, C–N, and PKR. This results in fewer unique molecules and thus the ability to fit models to their presence or absence.

We strongly recommend caution in reaction learning when using SMILES representations. In line with recent commentary highlighting these control experiments,[65,69] we do not suggest that SMILES modeling is entirely futile, but that in our case models trained on SMILES only will likely be unable to generalize to reactions containing new molecules not in the current dataset. That said, these models could perhaps be used to predict reaction conditions with high accuracy for yet untested substrate pairs whose individual components have been previously reported. With the goal of creating generalizable models, we do not explore SMILES-only models any further.

On the other hand, the full-input and Mordred-only models pass the described adversarial controls in most every case, leading to the conclusion that these input types are capable of representing our chemical space. Replacing the hybrid or Mordred representations with randomized vectors led to a large breakdown of model accuracy (**random inputs** and **random Mordred**), as did shuffling their input-output pairs (**shuffle inputs** and **shuffle Mordred**). To aid in interpreting these results, we compare AERs between the control models ("straw," AER_s) and their featurized reference models ("feature," AER_f).[65] We calculate the difference in AER between these two models and take the ratio of this difference to the reference model's AER to give AER_d, which allows for comparison between model types as below. This translates to measuring the percent

change in AER observed when true chemical features are replaced with random variables, and thus significant negative numbers should be expected.

$$\mathrm{AER}_d = \frac{\mathrm{AER}_s - \mathrm{AER}_f}{\mathrm{AER}_f} \times 100\% \qquad (3.11)$$

Table 3.19 *Modeling results for control models run on the Suzuki dataset.*

Suzuki coupling condition prediction by control models (categorical accuracies)											
category (num)	**top -k**	**naïve (freq)**	**BM-GBM**	**shuffle inputs**	**random inputs**	**Mordred only**	**shuffle Mordred**	**random Mordred**	**SMILES only**	**shuffle SMILES**	**random SMILES**
AER_d	**1**[i]	–	–	-142%	-149%	–	-108%	-122%	–	-105%	-76%
	3	–	–	-112%	-20%	–	-116%	-37%	–	-113%	-11%
metal (28)	1	0.3777	**0.5732**	0.3798	0.5202	0.4749	0.3785	0.4354	0.5459	0.3841	0.5298
	3	0.6744	**0.8516**	0.6718	0.8297	0.7996	0.6748	0.7680	0.8264	0.6747	0.8183
ligand (23)	1	0.8722	0.8390	0.8536	0.7887	0.8466	**0.8724**	0.8352	0.8556	0.8668	0.8211
	3	0.9269	**0.9635**	0.9234	0.9546	0.9533	0.9216	0.9461	0.9606	0.9221	0.9522
base (35)	1	0.3361	**0.4908**	0.3071	0.4529	0.4490	0.3093	0.4145	0.4649	0.3276	0.4497
	3	0.7344	**0.8338**	0.7223	0.8092	0.8138	0.7277	0.7856	0.8078	0.7286	0.7960
solvent (10)	1	0.6377	**0.6729**	0.6463	0.6595	0.6575	0.6459	0.6512	0.6724	0.6462	0.6699
	3	0.8013	**0.8637**	0.7996	0.8624	0.8375	0.8008	0.8306	0.8525	0.8015	0.8620
additive (17)	1	**0.9511**	0.9259	0.9097	0.7589	0.8977	0.9203	0.8905	0.9161	0.9285	0.8979
	3	0.9771	0.9842	0.9740	0.9812	0.9834	0.9738	0.9783	**0.9863**	0.9733	0.9846
[i]Excludes additive											

Table 3.20 *Modeling results for control models run on the C–N dataset.*

C–N coupling condition prediction by control models (categorical accuracies)											
category (num)	**top-k**	**naïve (freq)**	**BM-GBM**	**shuffle inputs**	**random inputs**	**Mordred only**	**shuffle Mordred**	**random Mordred**	**SMILES only**	**shuffle SMILES**	**random SMILES**
AER_d	**1**[i]	–	–	-114%	-56%	–	-136%	-30%	–	-109%	-13%
	3	–	–	-122%	-17%	–	-156%	-30%	–	-121%	-20%
metal (44)	1	0.2452	0.4825	0.2259	0.4758	0.3602	0.2281	0.3114	**0.5217**	0.2281	0.4928
	3	0.6526	0.7928	0.6287	0.7702	0.7078	0.6248	0.6822	**0.8203**	0.6248	0.7791
ligand (47)	1	0.5219	0.5538	0.5139	0.3164	0.5351	0.5067	0.5212	**0.6351**	0.5095	0.5825
	3	0.6647	0.7933	0.6454	0.7616	0.7401	0.6496	0.7181	**0.8407**	0.6379	0.8042
base (13)	1	0.2479	0.5028	0.2167	0.4877	0.3777	0.2162	0.3546	**0.5298**	0.2240	0.5212
	3	0.6400	0.8008	0.6078	0.7877	0.7106	0.6022	0.6944	**0.8109**	0.6103	0.7955
solvent (22)	1	0.3219	0.4582	0.2944	0.4639	0.3593	0.2880	0.3618	0.4983	0.3008	**0.5006**
	3	0.5677	0.7370	0.5404	0.7284	0.6393	0.5331	0.6228	**0.7741**	0.5370	0.7532
additive (74)	1	**0.8904**	0.7669	0.7939	0.6978	0.7660	0.7981	0.7880	0.8061	0.8379	0.7053
	3	0.9156	0.9290	0.9058	0.9212	0.9228	0.9022	0.9203	**0.9370**	0.9033	0.9270
[i]Excludes additive											

Table 3.21 *Modeling results for control models run on the Negishi dataset.*

Negishi coupling condition prediction by control models (categorical accuracies)											
category (num)	**top-k**	**naïve (freq)**	**BM-GBM**	**shuffle inputs**	**random inputs**	**Mordred only**	**shuffle Mordred**	**random Mordred**	**SMILES only**	**shuffle SMILES**	**random SMILES**
AER_d	**1**	–	–	-113%	-24%	–	-103%	-57%	–	-129%	-27%
	3	–	–	-115%	-15%	–	-133%	-38%	–	-115%	-8%
metal (32)	1	0.2887	0.5444	0.2294	0.4927	0.4249	0.2342	0.3328	**0.6284**	0.2068	0.5541
	3	0.5008	0.7771	0.4814	0.7528	0.6737	0.4556	0.6123	**0.8045**	0.4330	0.7900
ligand (20)	1	0.7879	0.8174	0.7738	0.8045	0.7835	0.7528	0.7722	**0.8320**	0.7577	0.8174
	3	0.8549	**0.9548**	0.8304	0.9370	0.9079	0.8352	0.8740	0.9451	0.8449	0.9354
temperature (8)	1	0.3317	**0.6656**	0.2924	0.6107	0.6204	0.2698	0.5331	0.5654	0.2763	0.5396
	3	0.5885	**0.9031**	0.5735	0.8885	0.8643	0.5283	0.8078	0.8498	0.5574	0.8401
solvent (10)	1	0.6938	0.8562	0.6866	0.8433	0.7383	0.8352	0.7076	**0.8724**	0.6737	0.8401
	3	0.8788	0.9321	0.8675	0.9208	0.8934	0.8627	0.8853	**0.9435**	0.8659	0.9370
additive (30)	1	0.8309	**0.8691**	0.8320	0.8417	0.8336	0.8061	0.8304	0.8595	0.8110	0.8340
	3	0.9043	**0.9548**	0.8950	0.9402	0.9241	0.8982	0.9160	**0.9548**	0.8982	0.9499

Table 3.22 *Modeling results for control models run on the PKR dataset.*

PKR coupling condition prediction by control models (categorical accuracies)											
category (num)	**top-k**	**naïve (freq)**	**BM-GBM**	**shuffle inputs**	**random inputs**	**Mordred only**	**shuffle Mordred**	**random Mordred**	**SMILES only**	**shuffle SMILES**	**random SMILES**
AER_d	**1**	–	–	-117%	-57%	–	-149%	-70%	–	-153%	-49%
	3	–	–	-76%	-30%	–	-75%	-45%	–	-73%	-23%
metal (18)	1	0.4302	**0.7901**	0.3969	0.6221	0.7405	0.3359	0.5763	0.6183	0.3321	0.4809
	3	0.7132	**0.9351**	0.7672	0.8588	0.9046	0.7481	0.8130	0.8588	0.7634	0.8359
ligand (6)	1	0.8792	**0.9351**	0.8740	0.9008	0.9160	0.8626	0.9008	0.8969	0.8702	0.8893
	3	0.9019	**0.9962**	0.9733	0.9885	0.9885	0.9695	0.9847	0.9885	0.9618	0.9847
temperature (7)	1	0.2830	**0.5954**	0.2939	0.4580	0.4695	0.2481	0.3740	0.5687	0.2443	0.5115
	3	0.5962	**0.8740**	0.6031	0.7481	0.7863	0.6221	0.6756	0.8359	0.5954	0.7519
solvent (15)	1	0.3321	**0.6183**	0.2863	0.4466	0.5687	0.2366	0.3893	**0.6183**	0.3321	0.5076
	3	0.5925	**0.8779**	0.6412	0.7672	0.7977	0.5992	0.7061	0.8626	0.6374	0.7748
activator (11)	1	0.6906	**0.8244**	0.6412	0.7061	0.7634	0.6374	0.6756	0.8053	0.6565	0.7290
	3	0.8830	0.9466	0.8855	0.9122	0.9351	0.8779	0.8855	**0.9618**	0.8779	0.9198
CO (g) (2)	1	0.7245	**0.8855**	0.5573	0.7405	0.8817	0.5115	0.7061	0.7939	0.5076	0.6870
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
additive (13)	1	**0.9057**	0.9008	0.8893	0.8969	0.8931	0.8702	0.8855	0.8855	0.8626	0.8931
	3	0.9321	**0.9885**	0.9275	0.9771	0.9733	0.9313	0.9542	0.9656	0.9389	0.9618
pressure (4)	1	0.6528	**0.8588**	0.8282	0.8435	**0.8588**	0.8168	0.8244	**0.8588**	0.8015	0.8473
	3	0.9623	0.9771	0.9695	0.9733	**0.9847**	0.9695	0.9733	0.9771	0.9695	0.9771

3.6.3.2 Classifier Trellis GBMs (CT-GBMs)

Additional controls were conducted for the classifier trellis (CT) models where the propagated label predictions were withheld from downstream models. As above, AER_d values were recorded for these straw models relative to their featurized CT reference models. The results indicate that models heavily rely on upstream label information, and perhaps overfit to these input features.

It is interesting to note that the strongest drop in performance on holdout comes from heavily skewed categories (see C–N *ligand* and *additive*, for example). This is perhaps sensible in that the highest frequency bin in these categories is the `NULL` label. As such, it appears that interdependent models for these categories base their predictions

largely on whether or not a reagent is expected to be used in their category at all. This seems to overtake most structural information in the inputs and when propagated predictions are removed, the models break down.

Table 3.23 *Modeling results for CT control models run on the Suzuki dataset.*

Suzuki coupling condition prediction by control models (categorical accuracies)				
category (num)	**top-k**	**naïve (freq)**	**CT-GBM**	**CT-holdout**
AER_d^j	**1**	–	–	-2271%
	3	–	–	-1229%
metal (28)	1	0.3777	**0.5629**	0.3967
	3	0.6744	**0.8475**	0.3981
ligand (23)	1	**0.8722**	0.8408	0.0302
	3	0.9269	**0.9606**	0.0325
base (35)	1	0.3361	**0.4777**	0.3329
	3	0.7344	**0.8250**	0.3350
solvent (10)	1	0.6377	**0.6751**	0.1143
	3	0.8013	**0.8577**	0.3110
additive (17)	1	**0.9511**	0.9196	0.0217
	3	0.9771	**0.9832**	0.0238
[i]Excludes additive				

Table 3.24 *Modeling results for CT control models run on the C–N dataset.*

C–N coupling condition prediction by control models (categorical accuracies)				
category (num)	**top-k**	**naïve (freq)**	**CT-GBM**	**CT-holdout**
AER_d^j	**1**	–	–	-204%
	3	–	–	-327%
metal (44)	1	0.2452	**0.4582**	0.2284
	3	0.6526	**0.7772**	0.2320
ligand (47)	1	0.5219	**0.5710**	0.0869
	3	0.6647	**0.7928**	0.0922
base (13)	1	0.2479	**0.5003**	0.2435
	3	0.6400	**0.7916**	0.6373
solvent (22)	1	0.3219	**0.4524**	0.3173
	3	0.5677	**0.7281**	0.3290
additive (74)	1	**0.8904**	0.7031	0.0136
	3	0.9156	**0.9184**	0.0148
[i]Excludes additive				

Table 3.25 *Modeling results for CT control models run on the Negishi dataset.*

Negishi coupling condition prediction by control models (categorical accuracies)				
category (num)	**top-k**	**naïve (freq)**	**CT-GBM**	**CT-holdout**
AER_d	**1**	–	–	-461%
	3	–	–	-229%
metal (32)	1	0.2887	**0.5218**	0.2827
	3	0.5008	**0.7674**	0.2924
ligand (20)	1	0.7879	**0.7900**	0.0420
	3	0.8549	**0.9321**	0.0468
temperature (8)	1	0.3317	**0.6527**	0.0129
	3	0.5885	**0.8772**	0.4572
solvent (10)	1	0.6938	**0.8514**	0.6947
	3	0.8788	**0.9402**	0.8821
additive (30)	1	0.8309	**0.8401**	0.0452
	3	0.9043	**0.9354**	0.0501

Table 3.26 *Modeling results for CT control models run on the PKR dataset.*

PKR coupling condition prediction by control models (categorical accuracies)				
category (num)	**top-k**	**naïve (freq)**	**CT-GBM**	**CT-holdout**
AER_d[i]	**1**	–	–	-532%
	3	–	–	-495%
metal (18)	1	0.4302	**0.7786**	0.2137
	3	0.7132	**0.9313**	0.2137
ligand (6)	1	0.8792	**0.9237**	0.0496
	3	0.9019	**0.9924**	0.1107
temperature (7)	1	0.2830	**0.5649**	0.0191
	3	0.5962	**0.8321**	0.4580
solvent (15)	1	0.3321	**0.6260**	0.3626
	3	0.5925	**0.8550**	0.6756
activator (11)	1	0.6906	**0.8015**	0.1412
	3	0.8830	**0.9275**	0.2519
CO (g) (2)	1	0.7245	**0.8855**	0.4580
	3	1.0000	1.0000	1.0000
additive (13)	1	**0.9057**	0.8893	0.0267
	3	0.9321	**0.9885**	0.0611
pressure (4)	1	0.6528	**0.8702**	0.0267
	3	0.9623	**0.9847**	0.9389
[i]Excludes additive				

3.6.4 INTERPRETABILITY

3.6.4.1 GBM Feature Importances (FIs)

FIs were calculated and averaged over all classifiers in all four BM-GBM models and the randomized control models (random inputs). The FIs plotted over the full input space are shown below. These show uniform decay in the SMILES region, where the token vectors for the three molecules (two reactants + one product) were padded to length 100 each, and thus the frequency of character presence decays with the vector position.

The FIs from the Mordred vector region were isolated and analyzed for their chemical significance (see main text for discussion). The top 20 FIs from this region are summarized in the tables following the FI plots. Full length FI rankings and FI charts for all four models and controls are included in the code repository.

Figure 3.27 *Relative feature importances for the full vector inputs averaged over the Suzuki BM-GBM classifiers.*

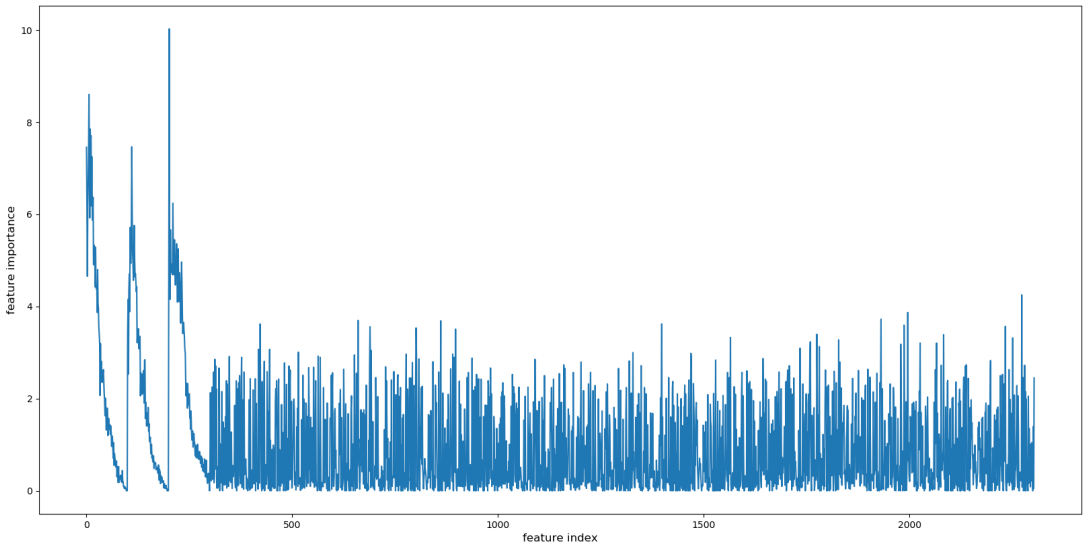


Figure 3.28 *Relative feature importances for random vector inputs (equal length to full input) averaged over the Suzuki BM-GBM classifiers.*

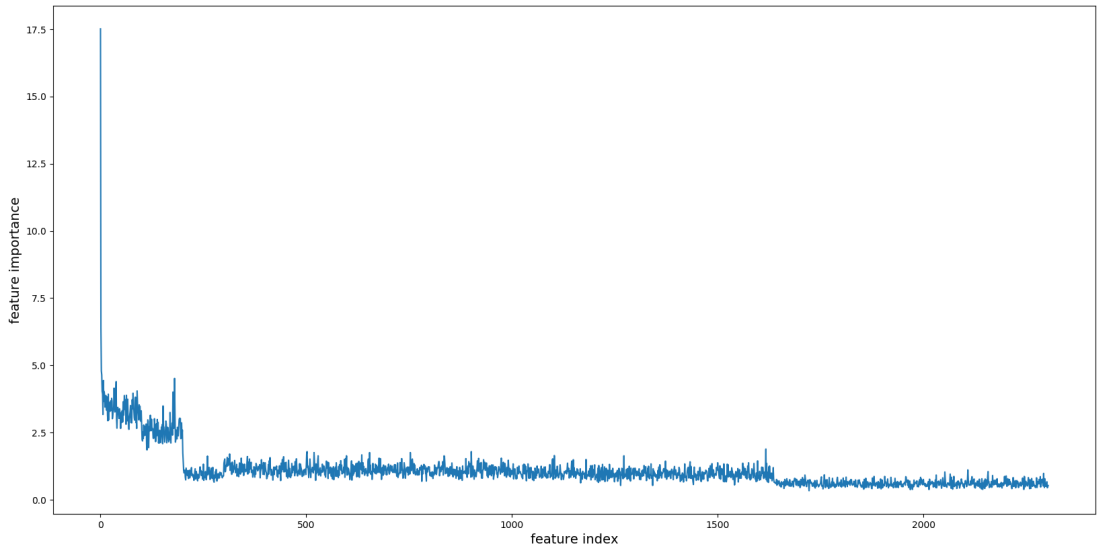


Figure 3.29 *Relative feature importances for Mordred vector inputs averaged over the Suzuki BM-GBM classifiers.*

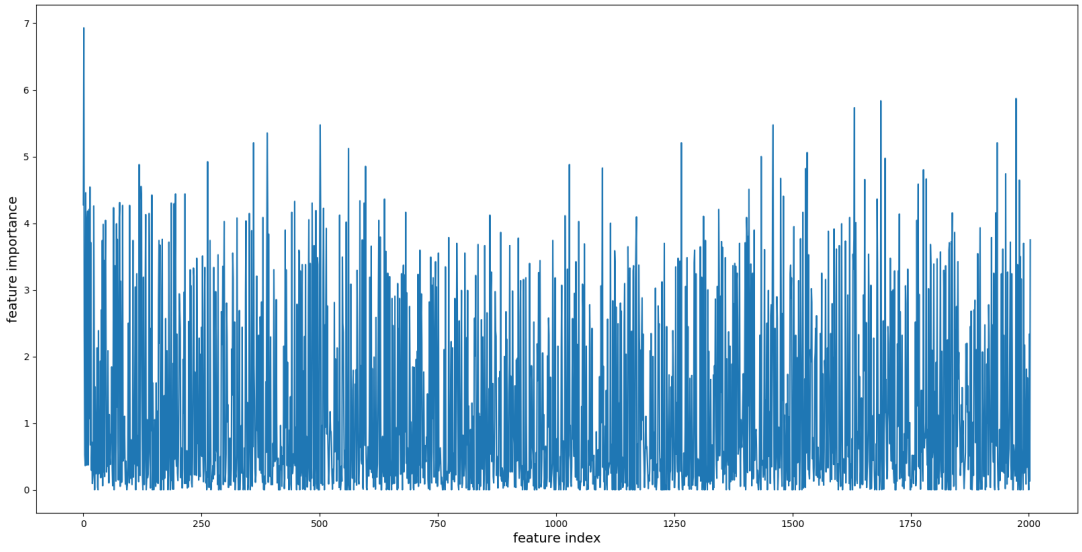


Figure 3.30 *Relative feature importances for random vector inputs (equal length to Mordred) averaged over the Suzuki BM-GBM classifiers.*

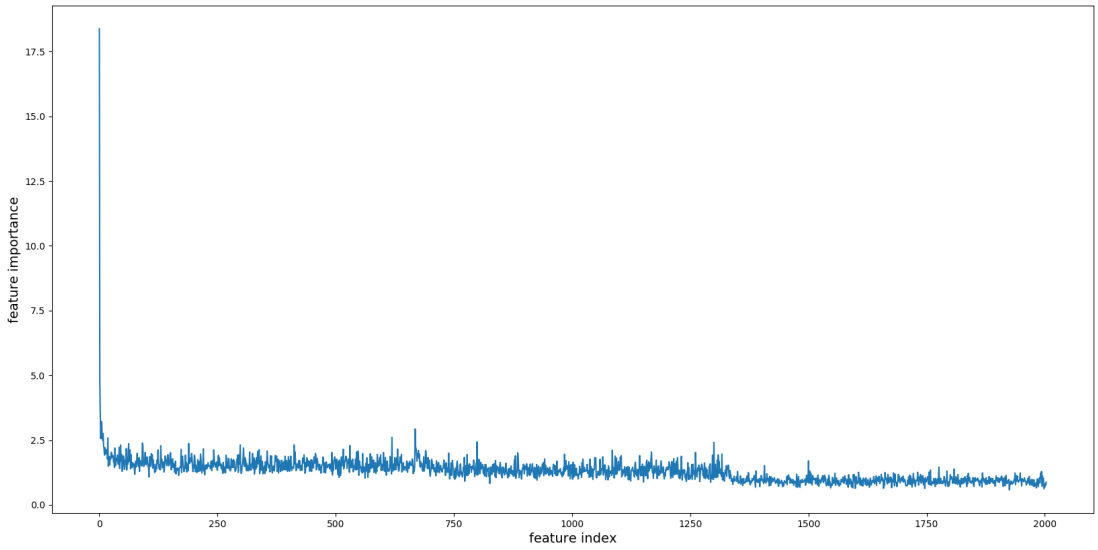


Figure 3.31 *Relative feature importances for the full vector inputs averaged over the Suzuki CT-GBM classifiers.*

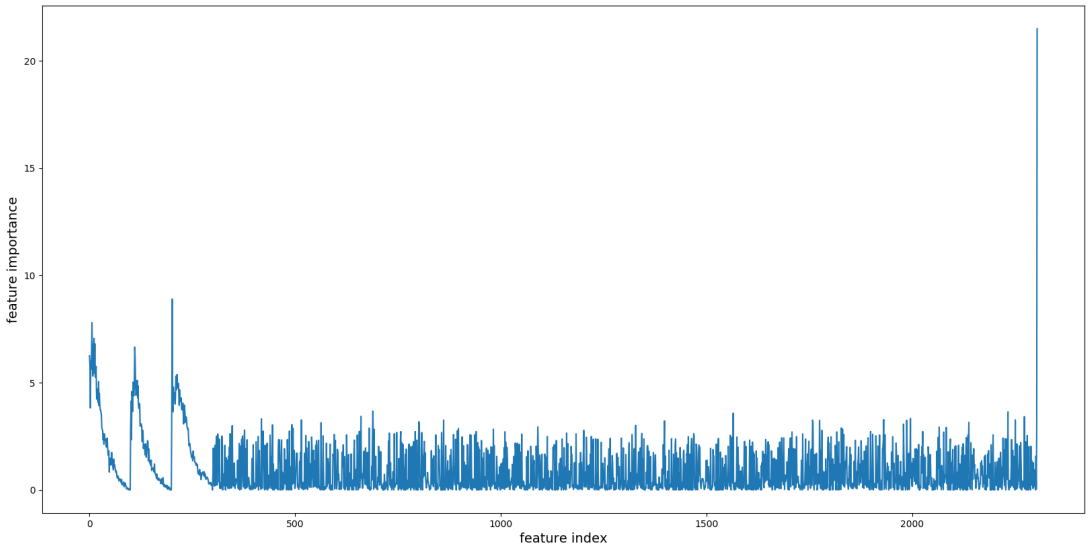


Table 3.27 *Top-20 Mordred descriptor FIs for Suzuki BM-GBMs with chemical explanations.*

rank	descriptor	species	FI score	description
1	JGI6	product	4.2500	6-ordered mean topological charge[a]
2	JGI3	product	3.8707	3-ordered mean topological charge[a]
3	JGI5	product	3.7241	5-ordered mean topological charge[a]
4	JGI3	reactant 1	3.6983	3-ordered mean topological charge[a]
5	AATSC0p	reactant 1	3.6897	averaged and centered Moreau–Broto autocorrelation of lag 0 weighted by polarizability[b]
6	SsOH	reactant 2	3.6207	sum of sOH[c]
7	IC1	reactant 1	3.6207	1-ordered neighborhood information content[d]
8	JGI4	product	3.5948	4-ordered mean topological charge[a]
9	SdssC	product	3.5690	sum of dssC[e]
10	ATSC3m	reactant 1	3.5603	centered Moreau–Broto autocorrelation of lag 3 weighted by mass[b]
11	Estate_VSA6	reactant 1	3.5345	Estate VSA Descriptor 6 ($1.81 \leq x < 2.05$)[f]
12	SdssC	reactant 1	3.5086	sum of dssC[e]
13	JGI9	product	3.3966	9-ordered mean topological charge[a]
14	ATSC4i	product	3.3879	centered Moreau–Broto autocorrelation of lag 4 weighted by ionization potential[b]
15	SdssC	reactant 2	3.3276	sum of dssC[e]
16	SlogP_VSA8	product	3.3190	MOE logP VSA Descriptor 8 ($0.25 \leq x < 0.30$)[g]
17	JGI2	product	3.2759	2-ordered mean topological charge[a]
18	JGI8	product	3.2328	8-ordered mean topological charge[a]
19	PEOE_VSA8	product	3.2069	MOE charge VSA Descriptor 8 ($0.00 \leq x < 0.05$)[h]
20	SsOH	product	3.2069	sum of sOH[c]

[a] *n*-ordered mean topological charge describes sum of atom-pair charge-transfer terms up to edge-distance *n*, averaged over all atoms in a molecule.[61]

[b] Moreau–Broto autocorrelation of lag *n* weighted by property *p* describes the distribution of *p* values over all atom pairs of edge-distance *n*.[62,63]

[c] Sum of electrotopological state of free-alcohol oxygens.[70,71]

[d]Measures graph complexity by summing local symmery over nodes with unique neighborhoods at edge-distance 1.[72]

[e]Sum of electrotopological state of disubstituted sp^2 carbons.[70,71]

[f]Describes the sum of the van der Waals surface area (VSA) with electrotopological state in the range 1.81–2.05.[70,72,73]

[g]Describes the sum of the VSA with SlogP (hybrid atomistic logP) in the range 0.25–0.30.[70,74]

[h]Describes the sum of the VSA with partial charge in the range 0.00–0.05.[70]

Figure 3.32 *Relative feature importances for the full vector inputs averaged over the C–N BM-GBM classifiers.*

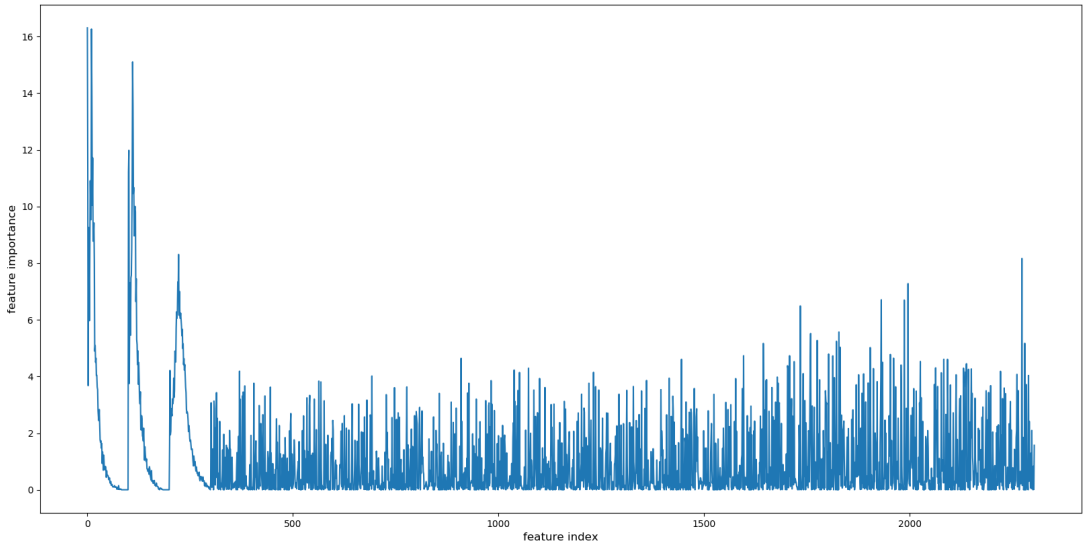


Figure 3.33 *Relative feature importances for random vector inputs (equal length to full input) averaged over the C–N BM-GBM classifiers.*

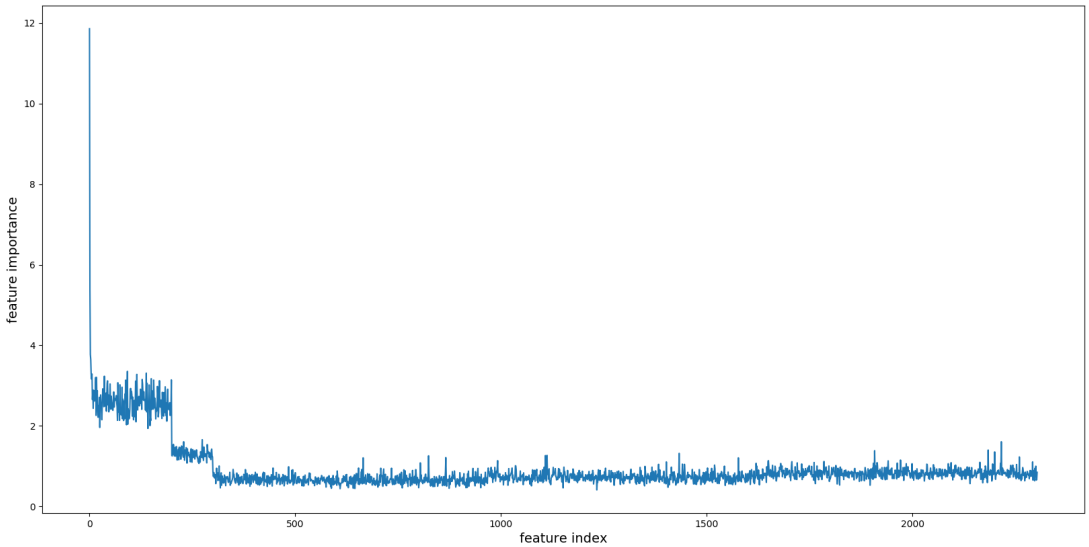


Figure 3.34 *Relative feature importances for Mordred vector inputs averaged over the C–N BM-GBM classifiers.*

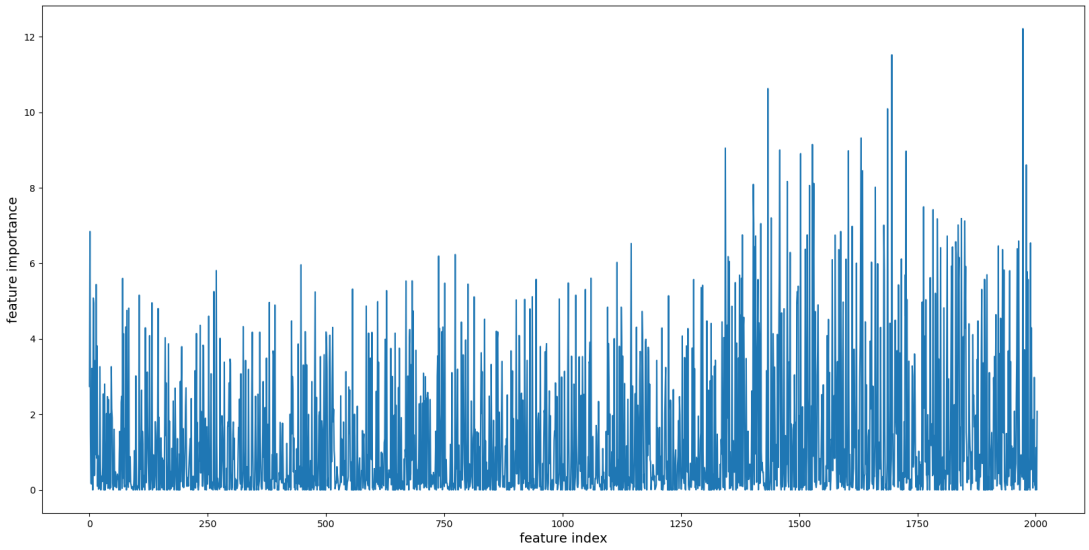


Figure 3.35 *Relative feature importances for random vector inputs (equal length to Mordred) averaged over the C–N BM-GBM classifiers.*

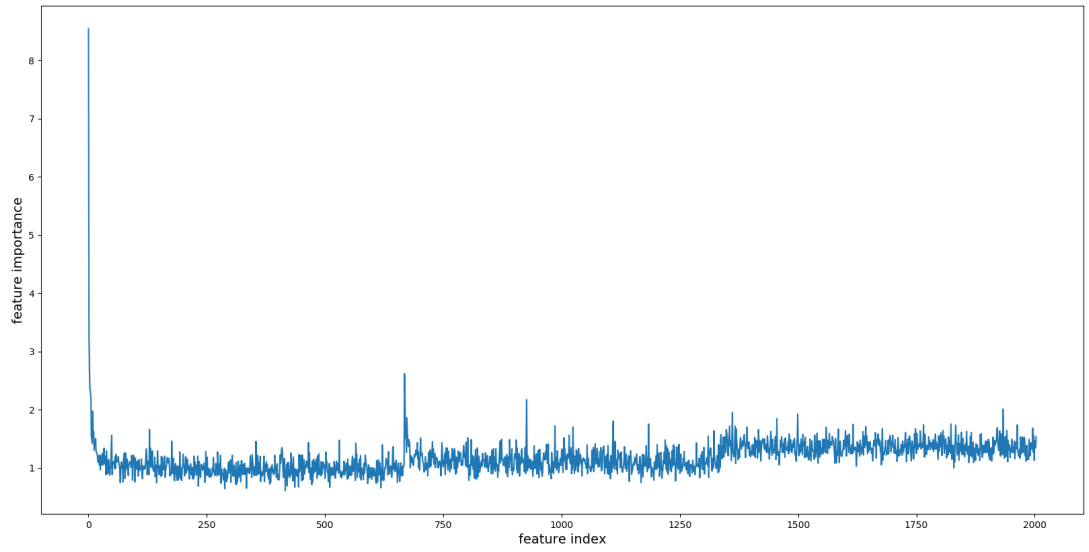


Figure 3.36 *Relative feature importances for the full vector inputs averaged over the C–N CT-GBM classifiers.*

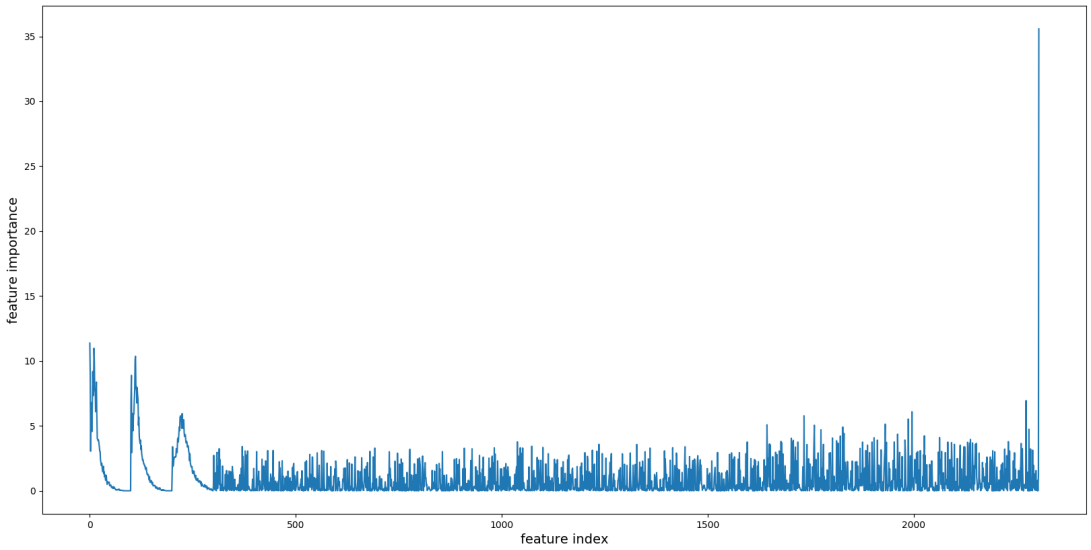


Table 3.28 *Top-20 Mordred descriptor FIs for C–N BM-GBMs with chemical explanations.*

rank	descriptor	species	FI score	description
1	JGI6	product	8.1667	6-ordered mean topological charge[a]
2	JGI3	product	7.2778	3-ordered mean topological charge[a]
3	JGI5	product	6.7071	5-ordered mean topological charge[a]
4	JGI4	product	6.6970	4-ordered mean topological charge[a]
5	JGI7	product	6.4899	7-ordered mean topological charge[a]
6	JGI2	product	5.5707	2-ordered mean topological charge[a]
7	JGI8	product	5.5152	8-ordered mean topological charge[a]
8	JGI9	product	5.2727	9-ordered mean topological charge[a]
9	JGI10	product	5.2424	10-ordered mean topological charge[a]
10	ATSC8d	product	5.1717	centered Moreau–Broto autocorrelation of lag 8 weighted by sigma electrons[b]

11	CIC3	product	5.1667	3-ordered complementary information content[c]
12	ATSC2dv	product	5.0303	centered Moreau–Broto autocorrelation of lag 2 weighted by valence electrons[b]
13	ATSC5i	product	5.0202	centered Moreau–Broto autocorrelation of lag 5 weighted by ionization potential[b]
14	ATSC7i	product	4.7929	centered Moreau–Broto autocorrelation of lag 7 weighted by ionization potential[b]
15	MIC1	product	4.7778	1-ordered modified information content weighted by mass[c]
16	ATSC5p	reactant 2	4.7323	centered Moreau–Broto autocorrelation of lag 5 weighted by polarizability[b]
17	Estate_VSA7	product	4.7323	EState VSA Descriptor 7 ($1.81 \leq x < 2.05$)[d]
18	CIC4	product	4.7273	4-ordered complementary information content[c]
19	ATSC4dv	product	4.6465	centered Moreau–Broto autocorrelation of lag 4 weighted by valence electrons[b]
20	ATSC5se	reactant 1	4.6414	centered Moreau–Broto autocorrelation of lag 5 weighted by Sanderson electronegativity[b]

[a]*n*-ordered mean topological charge describes sum of atom-pair charge-transfer terms up to edge-distance *n*, averaged over all atoms in a molecule.[61]

[b]Moreau–Broto autocorrelation of lag *n* weighted by property *p* describes the distribution of *p* values over all atom pairs of edge-distance *n*.[62,63]

[c]Difference between actual and maximum possible graph complexity as sum of local symmetry over nodes with unique neighborhoods at edge-distance 3.[72]

[d]Describes the sum of the van der Waals surface area (VSA) with electrotopological state in the range 1.81–2.05.[70,72,73]

Figure 3.37 *Relative feature importances for the full vector inputs averaged over the Negishi BM-GBM classifiers.*

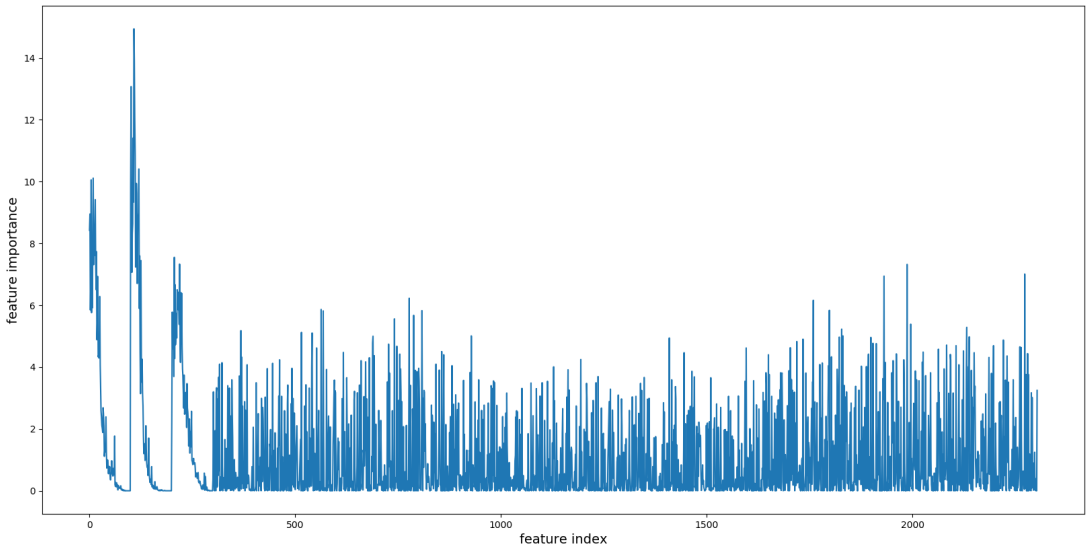


Figure 3.38 *Relative feature importances for random vector inputs (equal length to full input) averaged over the Negishi BM-GBM classifiers.*

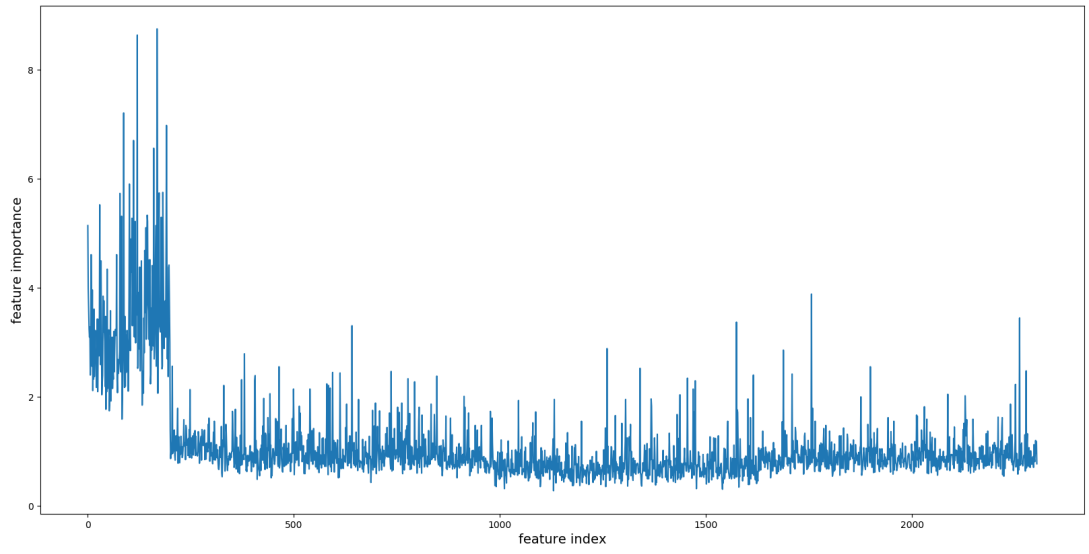


Figure 3.39 *Relative feature importances for Mordred vector inputs averaged over the Negishi BM-GBM classifiers.*

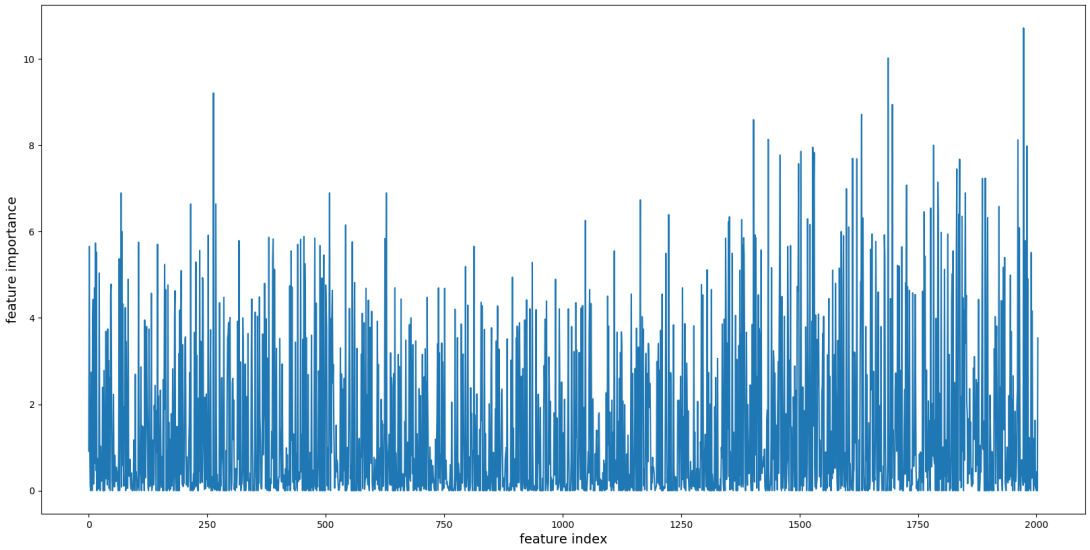


Figure 3.40 *Relative feature importances for random vector inputs (equal length to Mordred) averaged over the Negishi BM-GBM classifiers.*

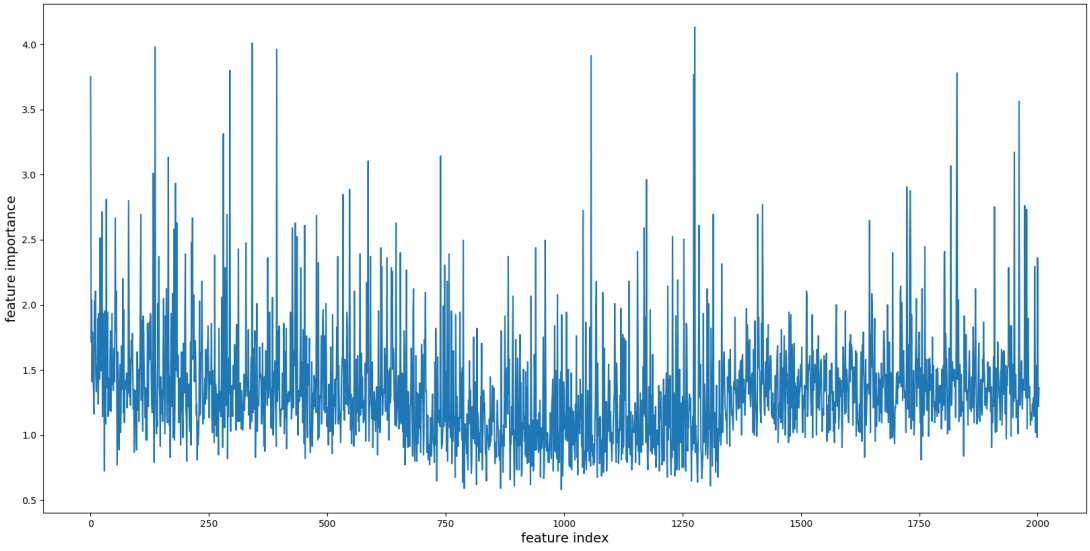


Figure 3.41 *Relative feature importances for the full vector inputs averaged over the Negishi CT-GBM classifiers.*

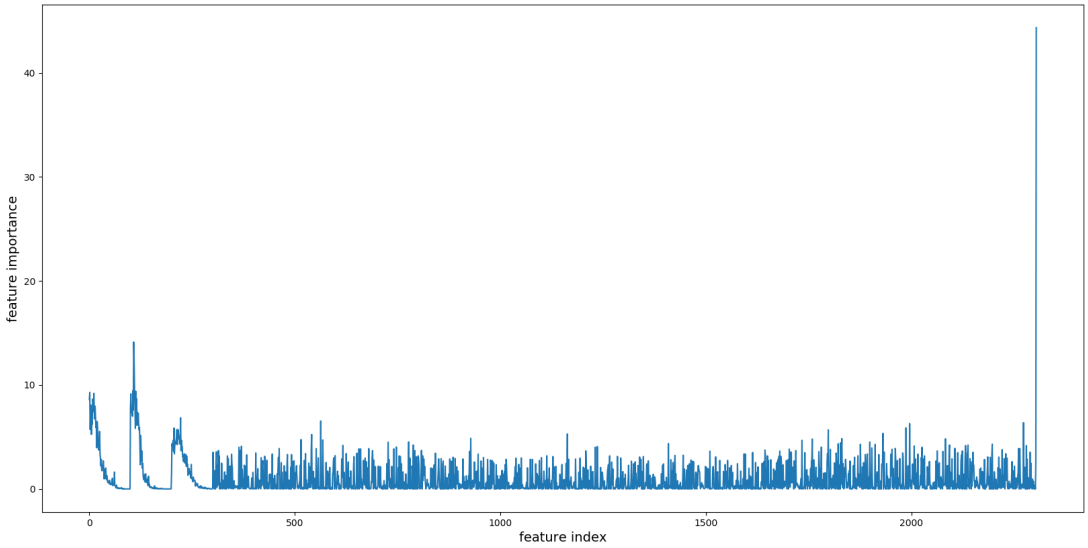


Table 3.29 *Top-20 Mordred descriptor FIs for Negishi BM-GBMs with chemical explanations.*

rank	descriptor	species	FI score	description
1	JGI4	product	7.3238	4-ordered mean topological charge[a]
2	JGI6	product	7.0095	6-ordered mean topological charge[a]
3	JGI5	product	6.9429	5-ordered mean topological charge[a]
4	ATSC4p	reactant 1	6.2286	centered Moreau–Broto autocorrelation of lag 4 weighted by polarizability[b]
5	JGI8	product	6.1619	8-ordered mean topological charge[a]
6	AATSC0i	reactant 1	5.8667	averaged and centered Moreau–Broto autocorrelation of lag 0 weighted by ionization potential[b]
7	SaasC	product	5.8381	sum of assC[c]
8	SsBr	reactant 1	5.8286	sum of sBr[d]
9	ATSC5i	reactant 1	5.8190	centered Moreau–Broto autocorrelation of lag 5 weighted by ionization potential[b]

10	ATSC6se	reactant 1	5.6762	centered Moreau–Broto autocorrelation of lag 6 weighted by Sanderson electronegativity[b]
11	ATSC1se	reactant 1	5.5619	centered Moreau–Broto autocorrelation of lag 1 weighted by Sanderson electronegativity[b]
12	JGI3	product	5.3905	3-ordered mean topological charge[a]
13	ATSC4d	product	5.2857	centered Moreau–Broto autocorrelation of lag 4 weighted by sigma electrons[b]
14	JGI2	product	5.2286	2-ordered mean topological charge[a]
15	ZMIC2	reactant 1	5.1810	2-ordered Z-modified information content weighted by atomic number[e]
16	IC0	reactant 1	5.1238	0-ordered information content[e]
17	bpol	reactant 1	5.1048	bond polarizability[f]
18	ATSC2dv	product	5.0095	centered Moreau–Broto autocorrelation of lag 2 weighted by valence electrons[b]
19	ATSC5p	reactant 1	5.0095	centered Moreau–Broto autocorrelation of lag 5 weighted by polarizability[b]
20	ATSC3m	reactant 1	5.0000	centered Moreau–Broto autocorrelation of lag 3 weighted by mass[b]

[a] *n*-ordered mean topological charge describes sum of atom-pair charge-transfer terms up to edge-distance *n*, averaged over all atoms in a molecule.[61]

[b] Moreau–Broto autocorrelation of lag *n* weighted by property *p* describes the distribution of *p* values over all atom pairs of edge-distance *n*.[62,63]

[c] Sum of electrotopological state of substituted aromatic carbons.[70]

[d] Sum of electrotopological state of organobromides.[70]

[e] Measures graph complexity by summing local symmetry over nodes with unique neighborhoods at edge-distance 2.[72]

[f] Sum of absolute value of polarizability differences between bound atom pairs.[46]

Figure 3.42 *Relative feature importances for the full vector inputs averaged over the PKR BM-GBM classifiers.*

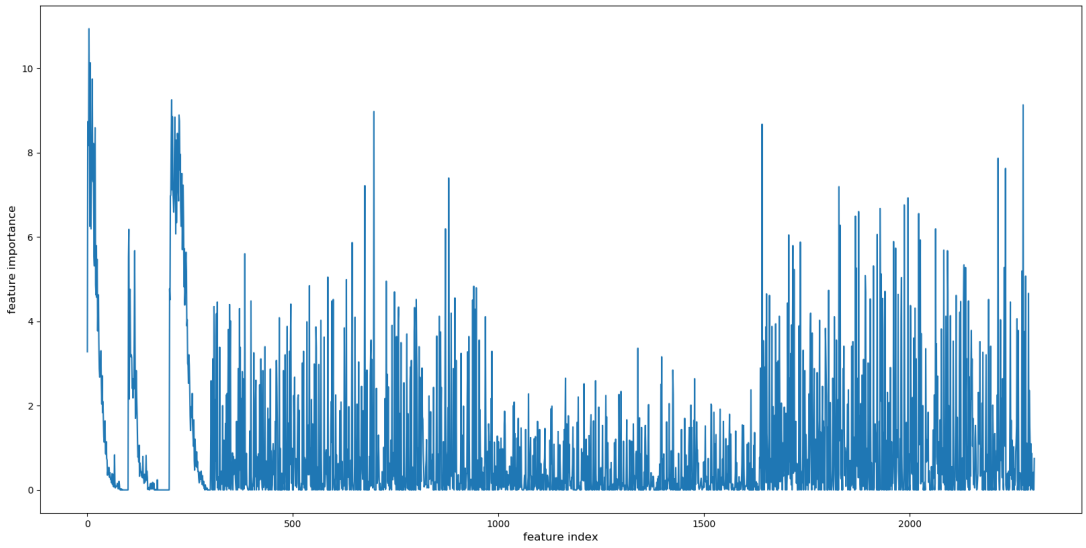


Figure 3.43 *Relative feature importances for random vector inputs (equal length to full input) averaged over the PKR BM-GBM classifiers.*

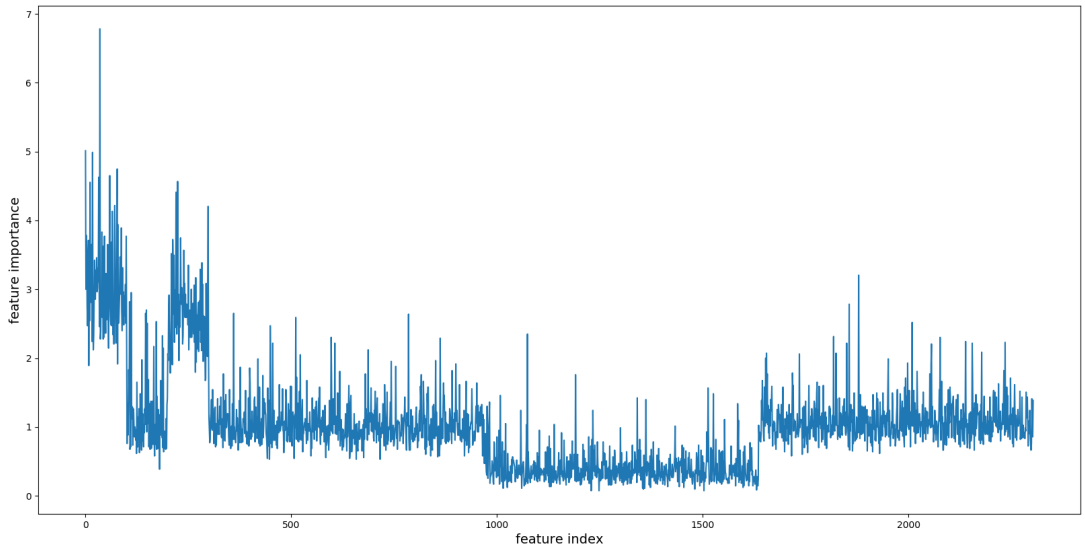


Figure 3.44 *Relative feature importances for Mordred vector inputs averaged over the PKR BM-GBM classifiers.*

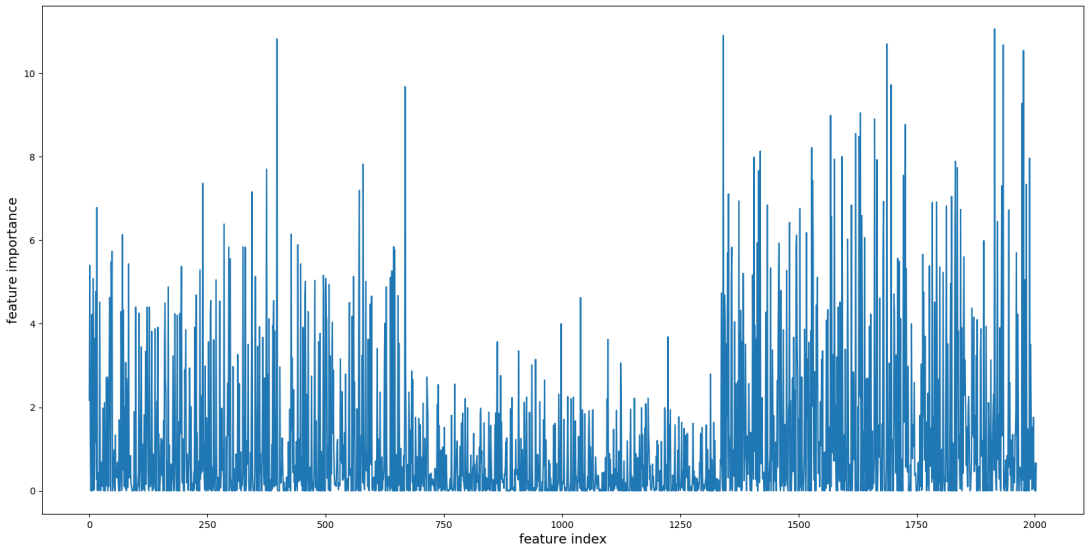


Figure 3.45 *Relative feature importances for random vector inputs (equal length to Mordred) averaged over the PKR BM-GBM classifiers.*

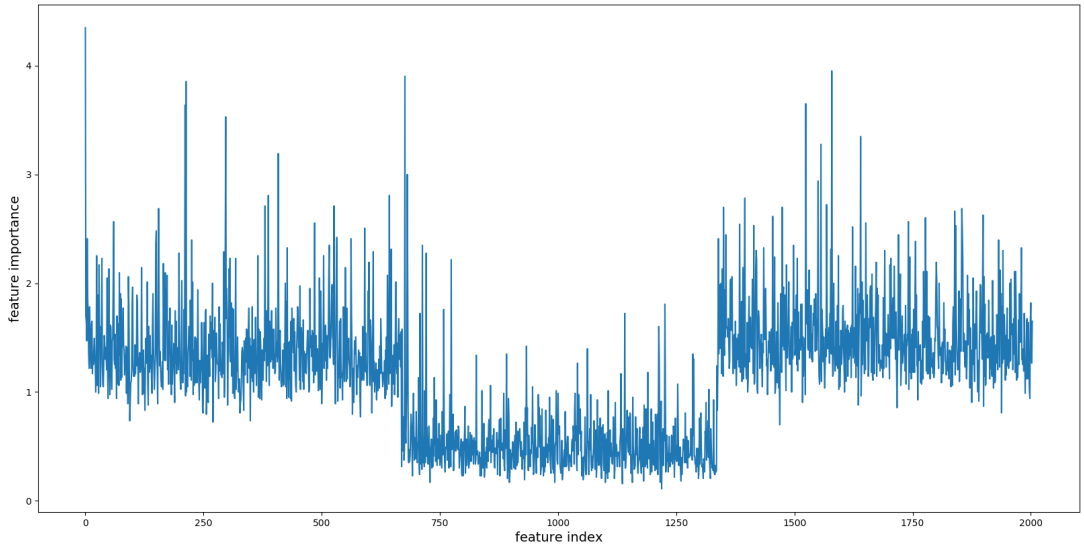


Figure 3.46 *Relative feature importances for the full vector inputs averaged over the PKR CT-GBM classifiers.*

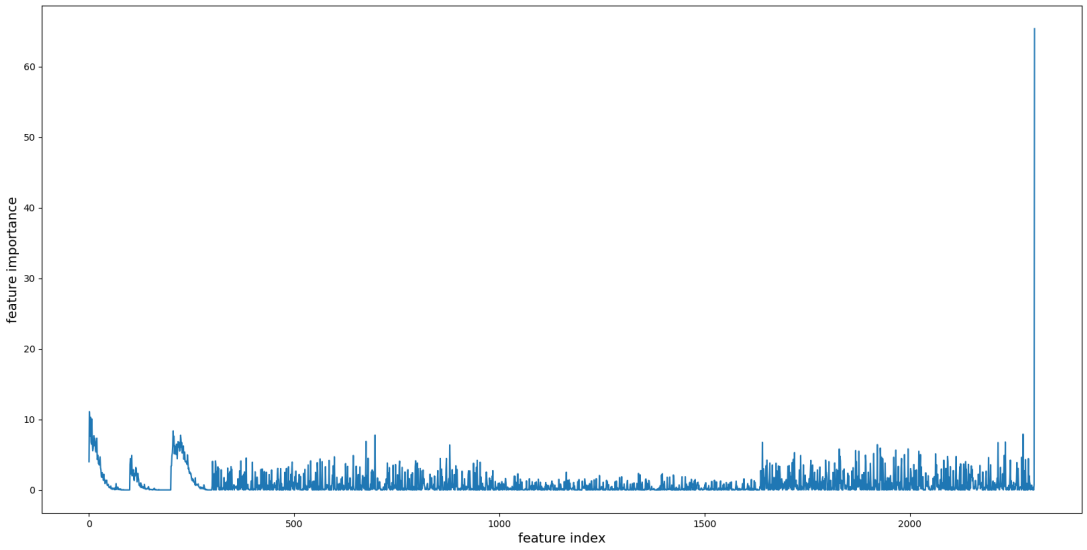


Table 3.30 *Top-20 Mordred descriptor FIs for PKR BM-GBMs with chemical explanations.*

rank	descriptor	species	FI score	description
1	SsssCH	product	9.1325	sum of sssCH[a]
2	SddC	reactant 1	8.9759	sum of ddC[a]
3	EState_VSA3	product	8.6747	EState VSA Descriptor 3 (0.29 ≤ x < 0.72)[b]
4	SdsCH	product	7.8675	sum of dsCH[a]
5	SdssC	product	7.6265	sum of dssC[a]
6	SdsCH	reactant 1	7.3976	sum of dsCH[a]
7	StsC	reactant 1	7.2169	sum of tsC[a]
8	JGI2	product	7.1928	2-ordered mean topological charge[c]
9	JGI3	product	6.9277	3-ordered mean topological charge[c]
10	JGI4	product	6.7590	4-ordered mean topological charge[c]
11	Xch-6dv	product	6.6747	6-ordered Chi chain weighted by valence electrons[d]

12	ATSC3d	product	6.6024	centered Moreau–Broto autocorrelation of lag 3 weighted by sigma electrons[e]
13	Xch-6d	product	6.5542	6-ordered Chi chain weighted by sigma electrons[d]
14	Xch-5dv	product	6.4940	5-ordered Chi chain weighted by valence electrons[d]
15	ATSC2dv	product	6.2771	centered Moreau–Broto autocorrelation of lag 2 weighted by valence electrons[e]
16	SdCH2	reactant 1	6.1928	sum of dCH2[a]
17	ATSC3dv	product	6.1928	centered Moreau–Broto autocorrelation of lag 3 weighted by valence electrons[e]
18	PEOE_VSA7	product	6.0602	MOE charge VSA Descriptor 7 ($-0.05 \leq x < 0.00$)[f]
19	ATSC4v	product	6.0482	centered Moreau–Broto autocorrelation of lag 4 weighted by van der Waals volume[e]
20	PEOE_VSA8	product	5.9277	MOE charge VSA Descriptor 8 ($0.00 \leq x < 0.05$)[f]

[a]Sum of electrotopological state of: sssCH = tertiary aliphatic carbons; ddC = central allenic carbons; dsCH = monosubstituted sp^2 carbons; dssC = disubstituted sp^2 carbons; tsC = internal alkyne carbons; dCH2 = terminal alkene carbons.[70]

[b]Describes the fraction of the van der Waals surface area (VSA) with electrotopological state in the range listed.[70,72,73]

[c]*n*-ordered mean topological charge describes sum of atom-pair charge-transfer terms up to edge-distance *n*, averaged over all atoms in a molecule.[61]

[d]Sum of the products of connectivity degrees of atoms in edge-distance of *n*-order weighted by property *p*.[70]

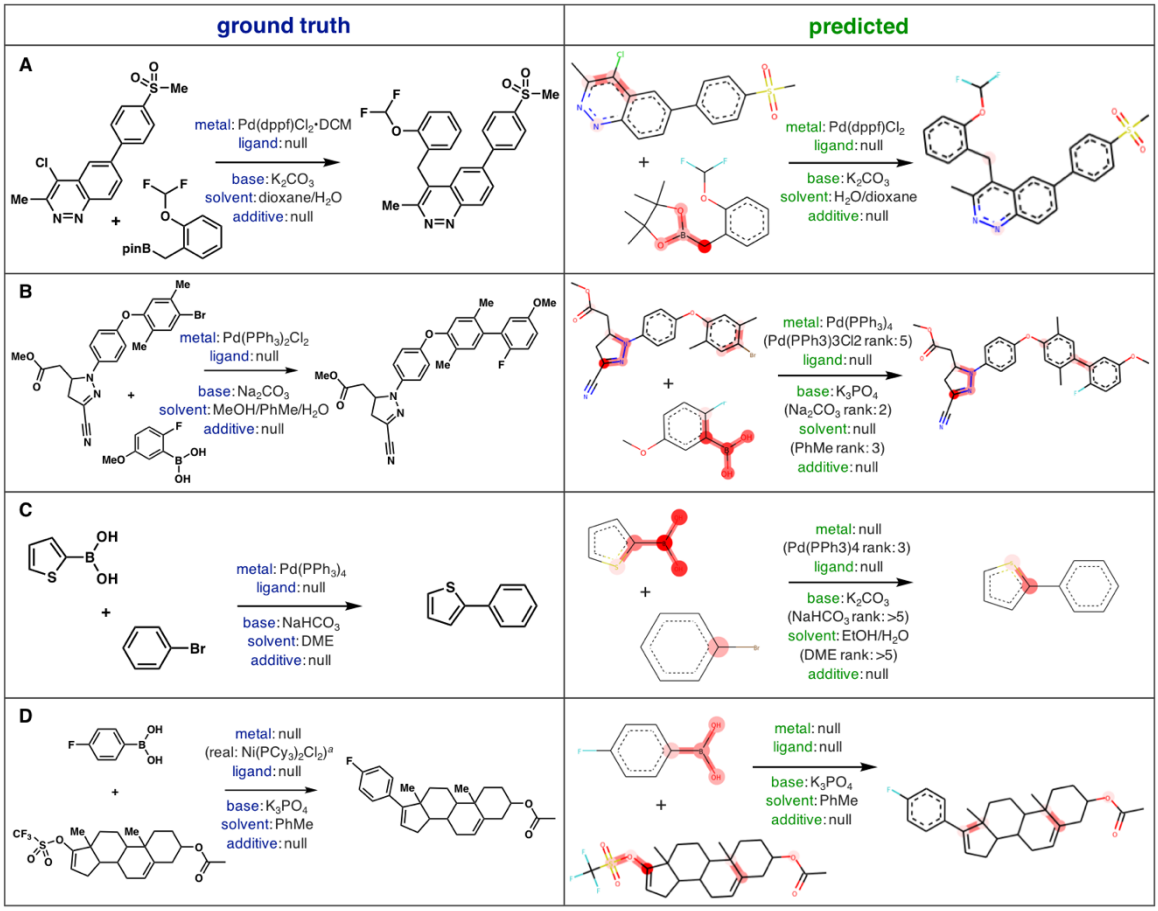
[e]Moreau–Broto autocorrelation of lag *n* weighted by property *p* describes the distribution of *p* values over all atom pairs of edge-distance *n*.[62,63]

[f]Sum of the VSA with partial charge in the range specified.[70]

3.6.4.2 AR-GCN Attention Visualizations

Four random reactions were chosen from each dataset for AR-GCN attention visualization (see main text for explanation).[66] The ground truth category labels and AR-GCN predictions are included in each example.

Figure 3.47 *Attention visualizations for randomly chosen Suzuki couplings.*



[a] Ground truth reagent below threshold dictionary frequency; read as null.

Figure 3.48 *Attention visualizations for randomly chosen C–N couplings.*

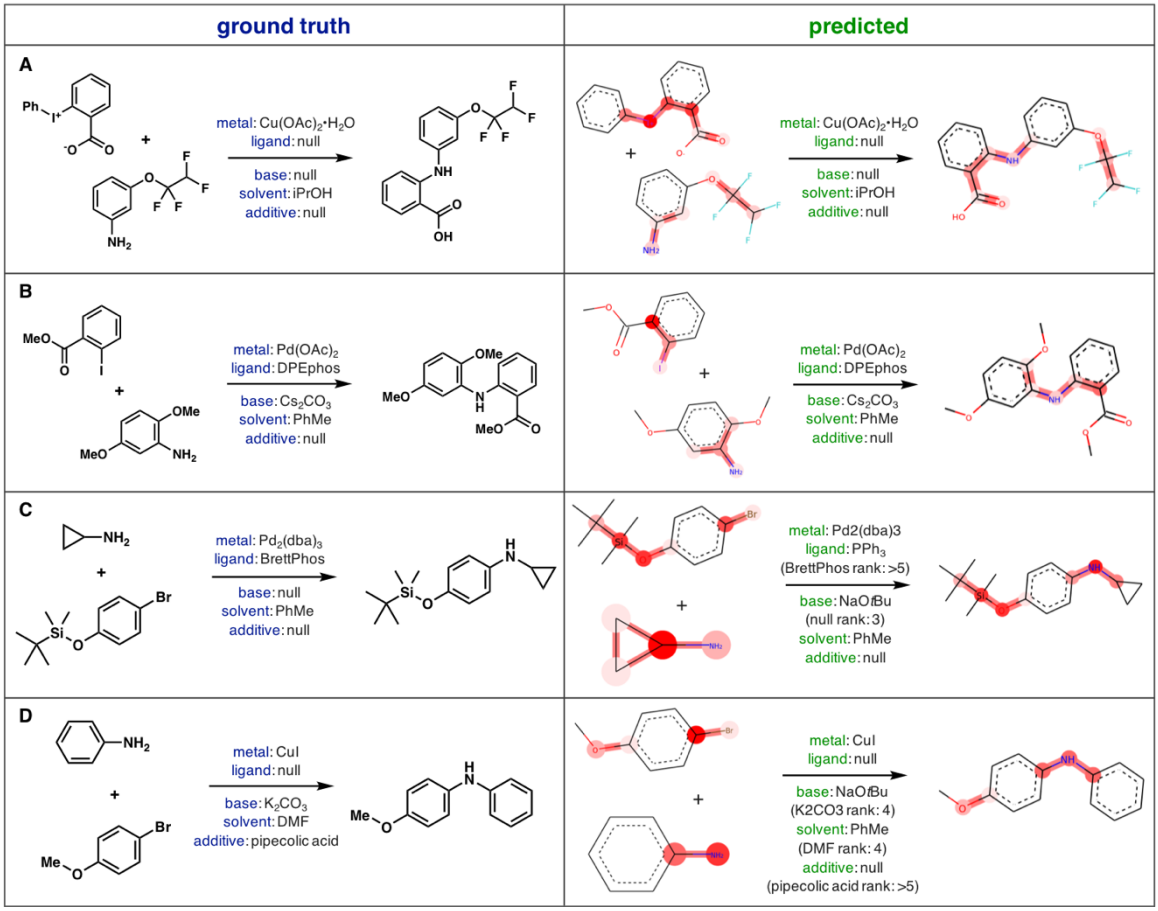
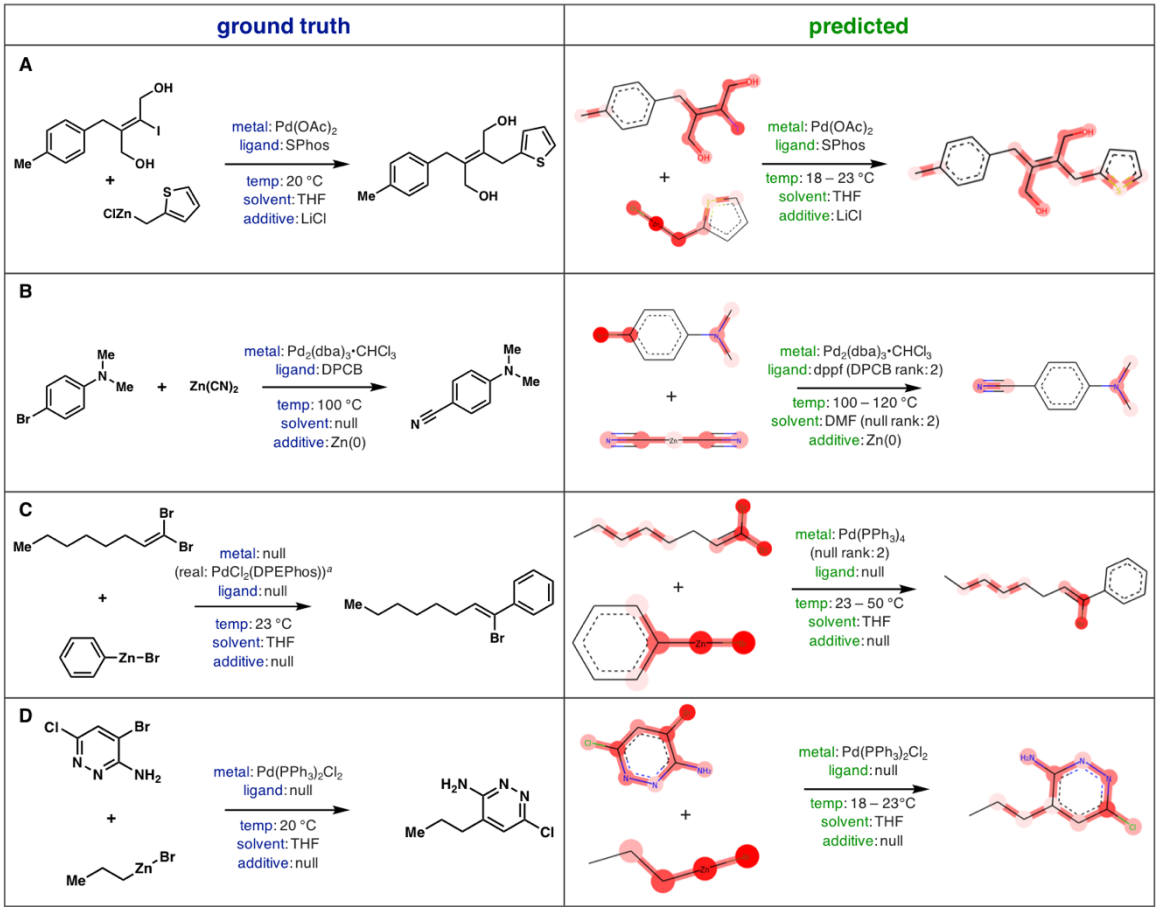
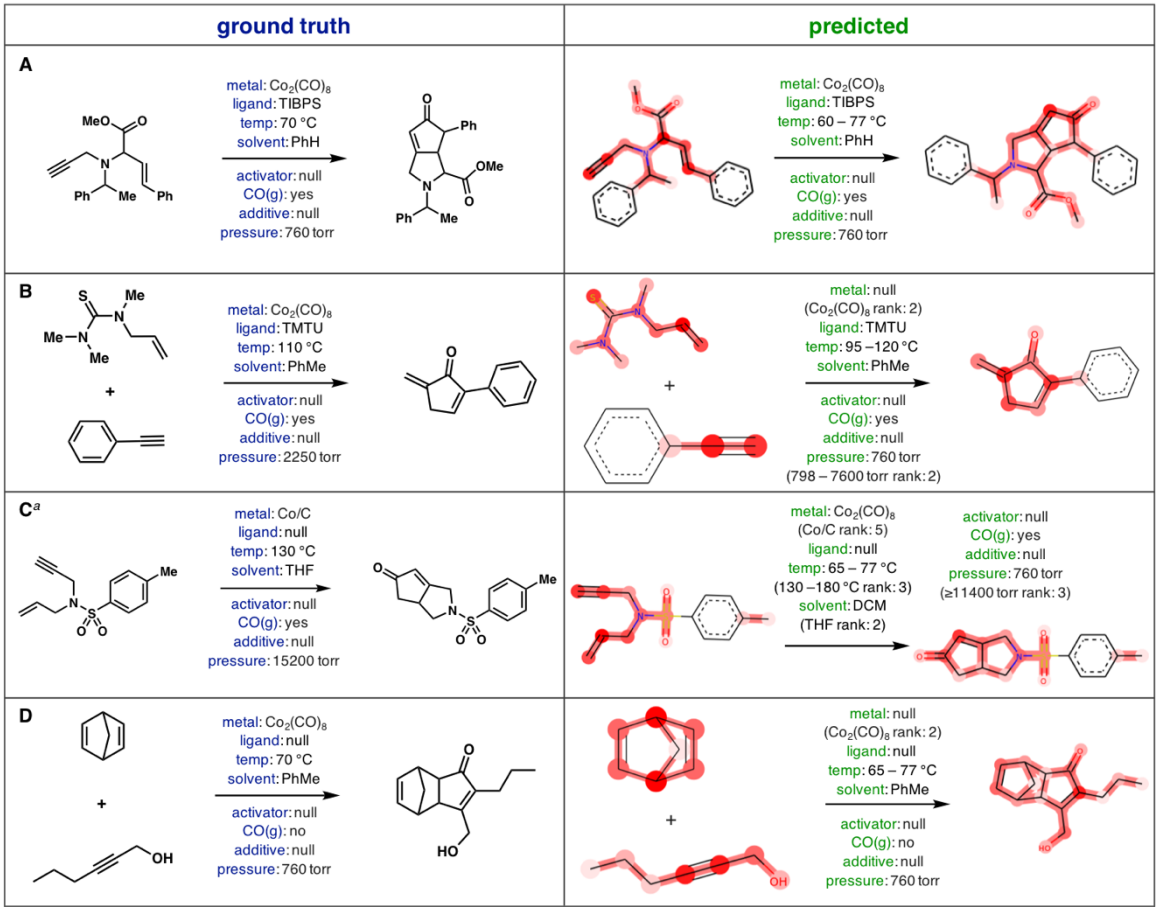


Figure 3.49 *Attention visualizations for randomly chosen Negishi couplings.*



[a] Ground truth reagent below threshold dictionary frequency; read as null.

Figure 3.50 *Attention visualizations for randomly chosen PKR couplings.*



3.6.5 EXPLORATORY DATA ANALYSIS (EDA)

A statistical analysis of the chemical space and reaction dictionary was conducted for each dataset. The analysis included distributions of reaction yields, categorical label frequencies, and the following 16 molecular descriptors of reaction products, calculated with RDKit:[67]

1. MolWt = molecular weight (g/mol)
2. MolLogP = molecular logP (lipophilicity — octanol/water partition coefficient)
3. TPSA = topological polar surface area (Å^2)

4. HeavyAtomCount = number of non-H atoms
5. NumHeteroatoms = number of heteroatoms (non-H or -C)
6. NumValenceElectrons = number of valence electrons
7. NumHAcceptors = number of hydrogen-bond acceptors
8. NumHDonors = number of hydrogen-bond donors
9. NumRotatableBonds = number of rotatable bonds
10. RingCount = number of rings
11. NumAromHeterocycles = number of aromatic heterocycles (aromatic rings with at least one non-C atom)
12. NumAromCarbocycles = number of aromatic carbocycles (aromatic rings made entirely of C atoms)
13. NumSatHeterocycles = number of saturated heterocycles (saturated rings with at least one non-C atom)
14. NumSatCarbocycles = number of saturated carbocycles (saturated rings made entirely of C atoms)
15. FractionCSP3 = fraction of atoms sp^3-hybridized
16. QED = quantitative estimation of drug-likeness

Note: The descriptor names above are not all exactly as written in their respective RDKit functions. For the figures below, the property distributions were truncated at the 1st and 99th percentile in each dataset analysis to avoid sparsity. The full span of the distributions is reflected in the summary tables. For full code, see the EDA jupyter notebooks in the associated GitHub repository.[37]

3.6.5.1 Suzuki Dataset

Table 3.31 *Statistical summary of product molecular property descriptors in the Suzuki dataset.*

property	count	mean	std	min	25%	50%	75%	max
Yield	145413	0.68	0.24	0.00	0.53	0.73	0.88	1.00
MolWt	144972	352.07	146.42	82.11	233.23	331.76	438.44	2339.16
MolLogP	144972	4.70	2.26	-5.97	3.35	4.14	5.52	46.85
TPSA	144972	49.81	36.78	0.00	20.23	43.60	72.45	472.09
HeavyAtomCount	144972	25.32	10.44	6.00	17.00	24.00	31.00	164.00
NumHeteroatoms	144972	4.94	3.42	0.00	2.00	4.00	7.00	86.00
NumValenceElectrons	144972	129.04	54.93	32.00	84.00	120.00	160.00	952.00
NumHAcceptors	144972	3.62	2.59	0.00	2.00	3.00	5.00	39.00
NumHDonors	144972	0.53	0.83	0.00	0.00	0.00	1.00	16.00
NumRotatableBonds	144972	3.97	3.48	0.00	2.00	3.00	5.00	115.00
RingCount	144972	3.47	1.64	0.00	2.00	3.00	4.00	11.00
NumAromHeterocycles	144972	1.03	1.05	0.00	0.00	1.00	2.00	8.00
NumAromCarbocycles	144972	1.98	1.25	0.00	1.00	2.00	2.00	9.00
NumSatHeterocycles	144972	0.17	0.45	0.00	0.00	0.00	0.00	7.00
NumSatCarbocycles	144972	0.10	0.39	0.00	0.00	0.00	0.00	5.00
FractionCSP3	144972	0.19	0.17	0.00	0.07	0.14	0.30	1.00
QED	144972	0.54	0.20	0.01	0.40	0.59	0.70	0.95

Figure 3.51 *Distribution of reaction yields in Suzuki dataset.*

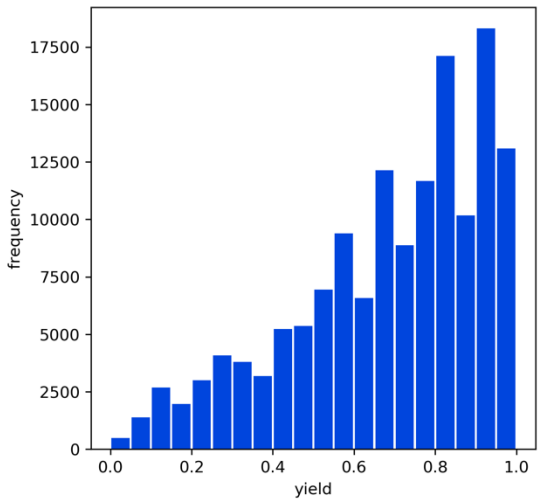


Figure 3.52 *Distribution of each molecular descriptor in Suzuki dataset products.*

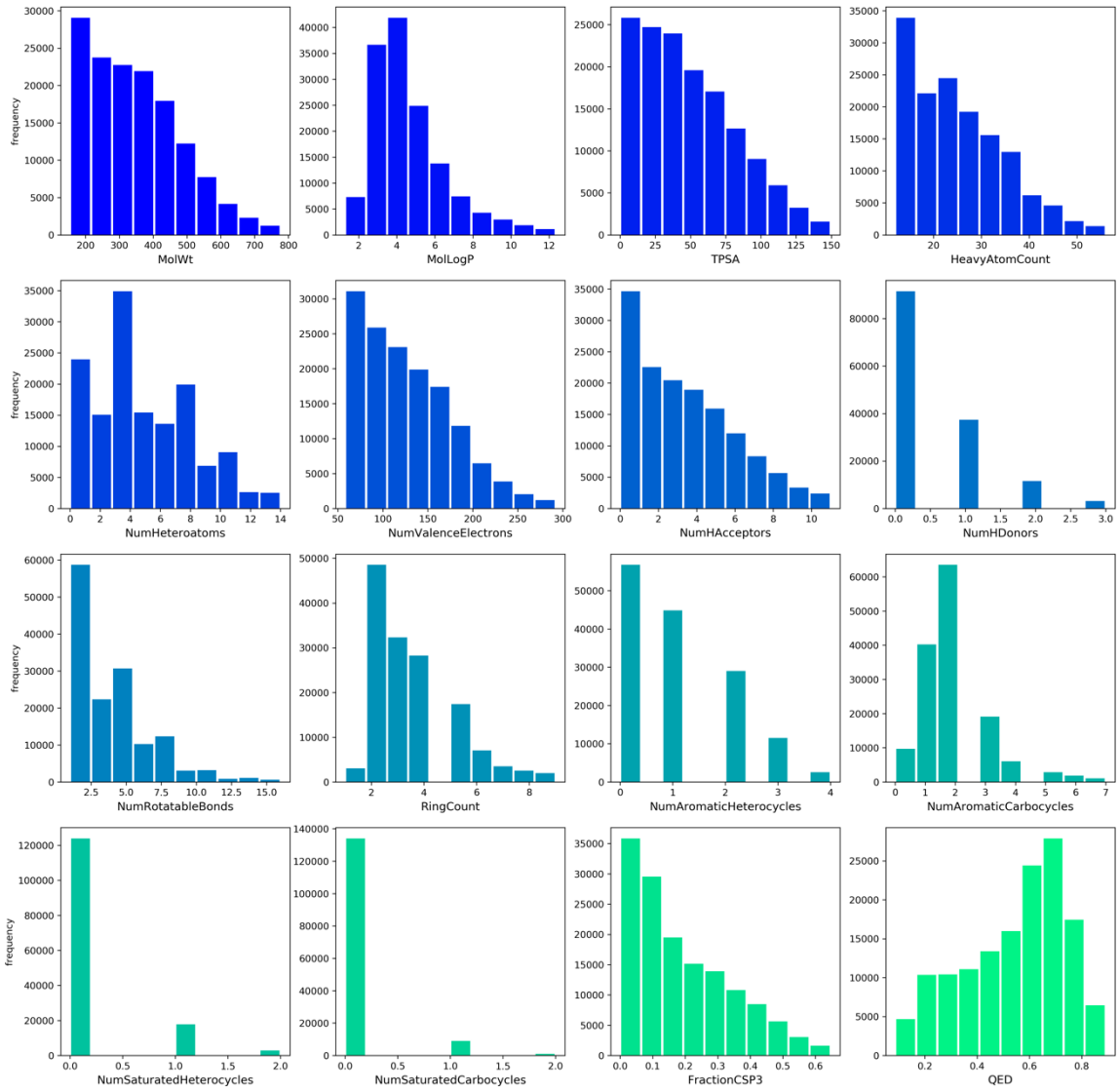
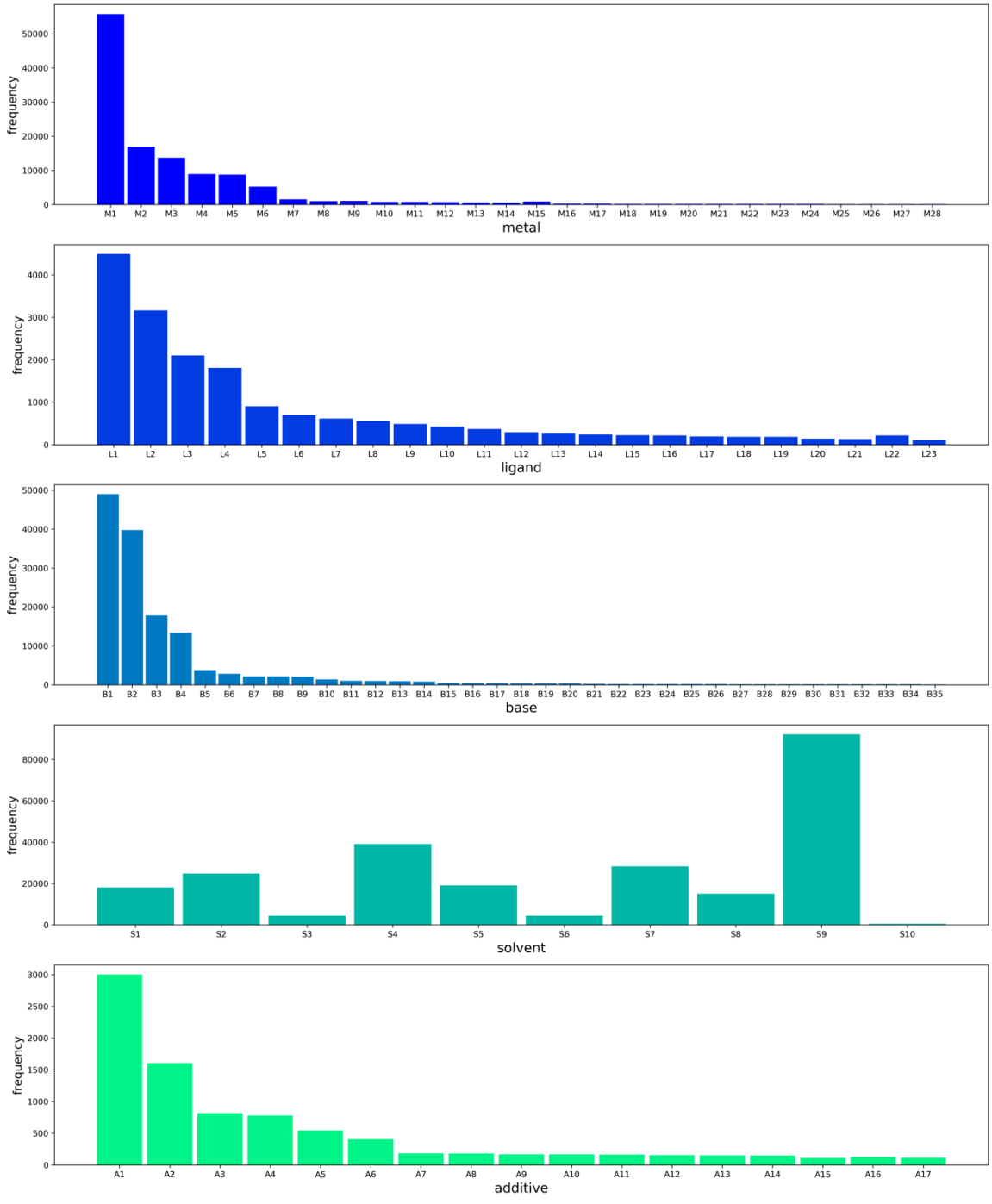


Figure 3.53 *Distribution of dictionary bin frequencies in Suzuki dataset.*



3.6.5.2 C–N Dataset

Table 3.32 *Statistical summary of product molecular property descriptors in the C–N dataset.*

property	count	mean	std	min	25%	50%	75%	max
Yield	36519	0.70	0.23	0.00	0.57	0.75	0.88	1.00
MolWt	36504	310.30	123.81	107.16	216.24	279.33	376.40	2527.86
MolLogP	36504	4.28	2.28	-1.44	2.87	3.76	5.17	30.83
TPSA	36504	41.02	32.61	3.01	12.47	33.20	58.44	355.67
HeavyAtomCount	36504	22.27	8.89	8.00	16.00	20.00	27.00	172.00
NumHeteroatoms	36504	4.24	2.82	1.00	2.00	4.00	6.00	40.00
NumValenceElectrons	36504	114.15	44.88	42.00	82.00	102.00	136.00	960.00
NumHAcceptors	36504	3.13	2.02	1.00	2.00	3.00	4.00	39.00
NumHDonors	36504	0.07	0.77	0.00	0.00	1.00	1.00	8.00
NumRotatableBonds	36504	3.60	2.80	1.00	2.00	3.00	5.00	114.00
RingCount	36504	3.02	1.64	1.00	2.00	3.00	4.00	12.00
NumAromHeterocycles	36504	0.05	0.69	0.00	0.00	0.00	1.00	7.00
NumAromCarbocycles	36504	2.09	1.29	1.00	1.00	2.00	2.00	9.00
NumSatHeterocycles	36504	0.29	0.53	0.00	0.00	0.00	1.00	4.00
NumSatCarbocycles	36504	0.07	0.36	0.00	0.00	0.00	0.00	5.00
FractionCSP3	36504	0.23	0.19	0.00	0.07	0.18	0.40	0.92
QED	36504	0.60	0.20	0.02	0.48	0.65	0.76	0.95

Figure 3.54 *Distribution of reaction yields in C–N dataset.*

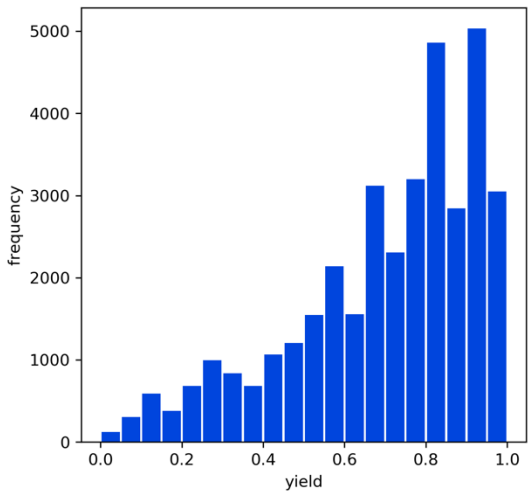


Figure 3.55 *Distribution of each molecular descriptor in C–N dataset products.*

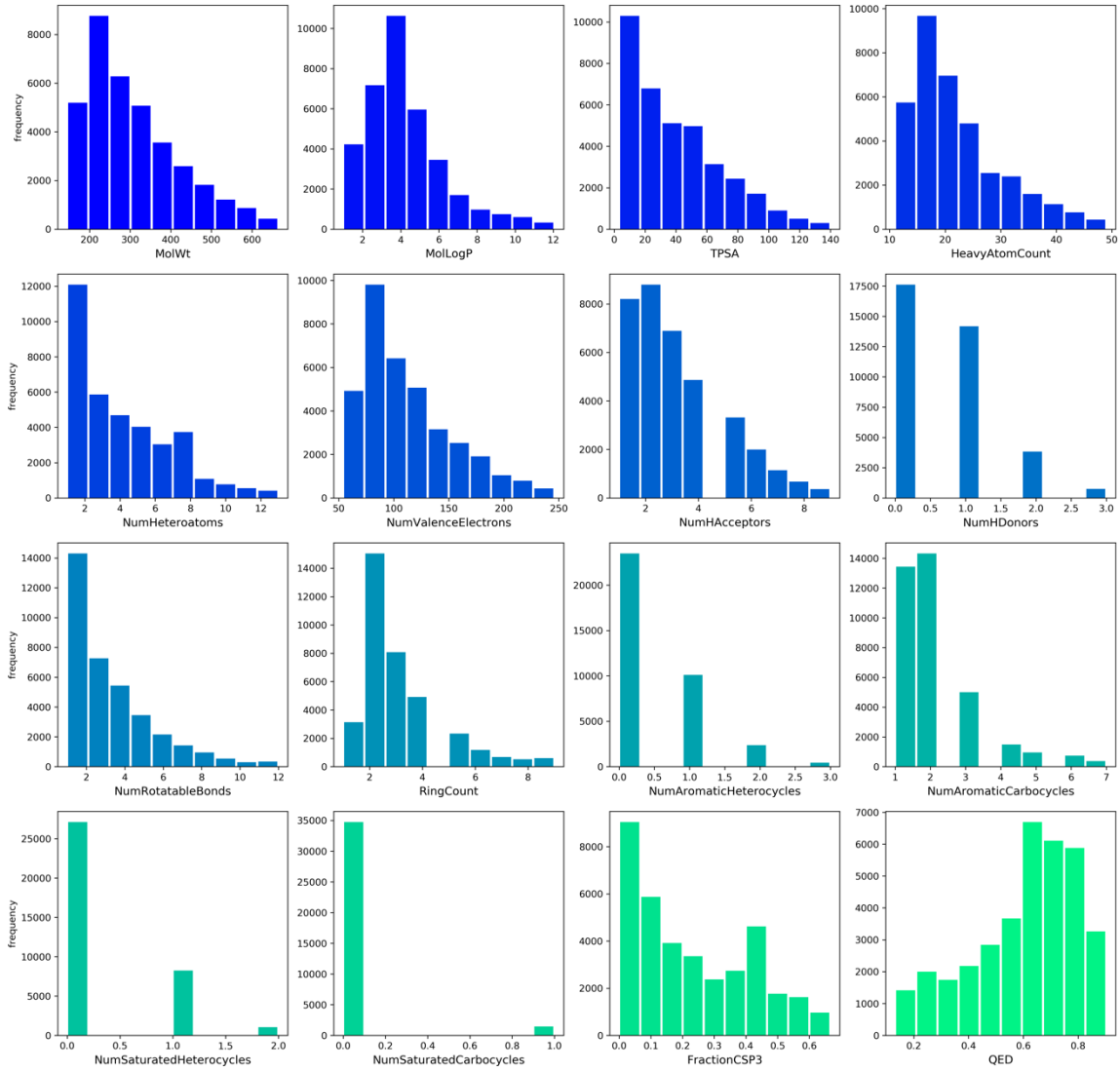
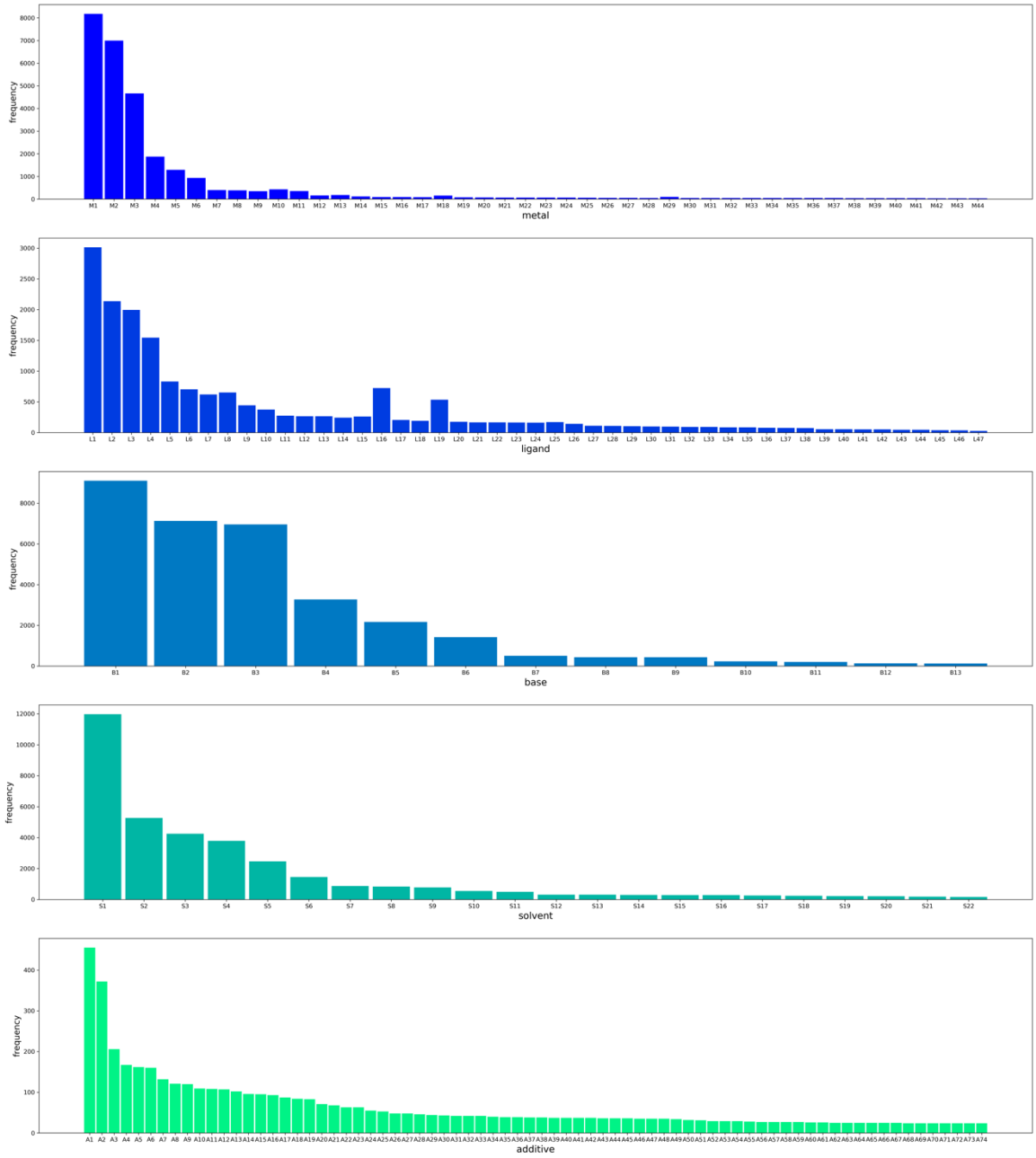


Figure 3.56 *Distribution of dictionary bin frequencies in C–N dataset.*



3.6.5.3 Negishi Dataset

Table 3.33 *Statistical summary of product molecular property descriptors in the Negishi dataset.*

property	count	mean	std	min	25%	50%	75%	max
Yield	6391	0.71	0.22	0.01	0.06	0.76	0.89	1.00
MolWt	6383	288.30	111.96	82.10	209.25	263.24	339.37	1209.10
MolLogP	6383	3.89	1.68	-1.69	2.82	3.64	4.65	17.40
TPSA	6383	39.46	28.72	0.00	18.46	35.53	55.14	213.58
HeavyAtomCount	6383	20.40	7.86	6.00	15.00	19.00	24.00	80.00
NumHeteroatoms	6383	3.92	2.68	0.00	2.00	3.00	5.00	28.00
NumValenceElectrons	6383	106.51	42.46	32.00	76.00	98.00	126.00	456.00
NumHAcceptors	6383	2.87	1.99	0.00	1.00	3.00	4.00	17.00
NumHDonors	6383	0.25	0.53	0.00	0.00	0.00	0.00	6.00
NumRotatableBonds	6383	3.60	2.82	0.00	2.00	3.00	5.00	31.00
RingCount	6383	2.34	1.24	0.00	2.00	2.00	3.00	9.00
NumAromHeterocycles	6383	0.68	0.82	0.00	0.00	0.00	1.00	5.00
NumAromCarbocycles	6383	1.22	0.92	0.00	1.00	1.00	2.00	7.00
NumSatHeterocycles	6383	0.13	0.39	0.00	0.00	0.00	0.00	4.00
NumSatCarbocycles	6383	0.15	0.49	0.00	0.00	0.00	0.00	4.00
FractionCSP3	6383	0.30	0.21	0.00	0.12	0.27	0.45	0.94
QED	6383	0.62	0.17	0.04	0.53	0.65	0.74	0.95

Figure 3.57 *Distribution of reaction yields in Negishi dataset.*

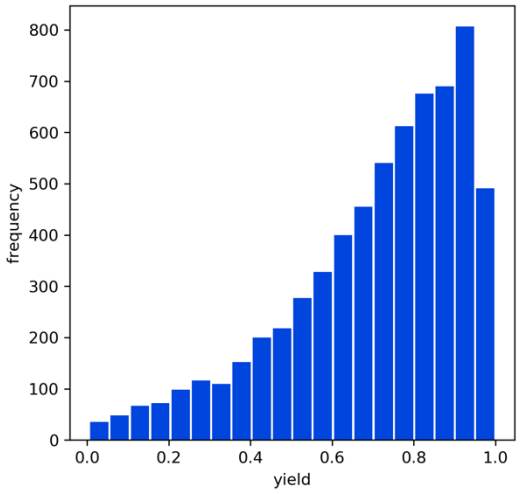


Figure 3.58 *Distribution of each molecular descriptor in Negishi dataset products.*

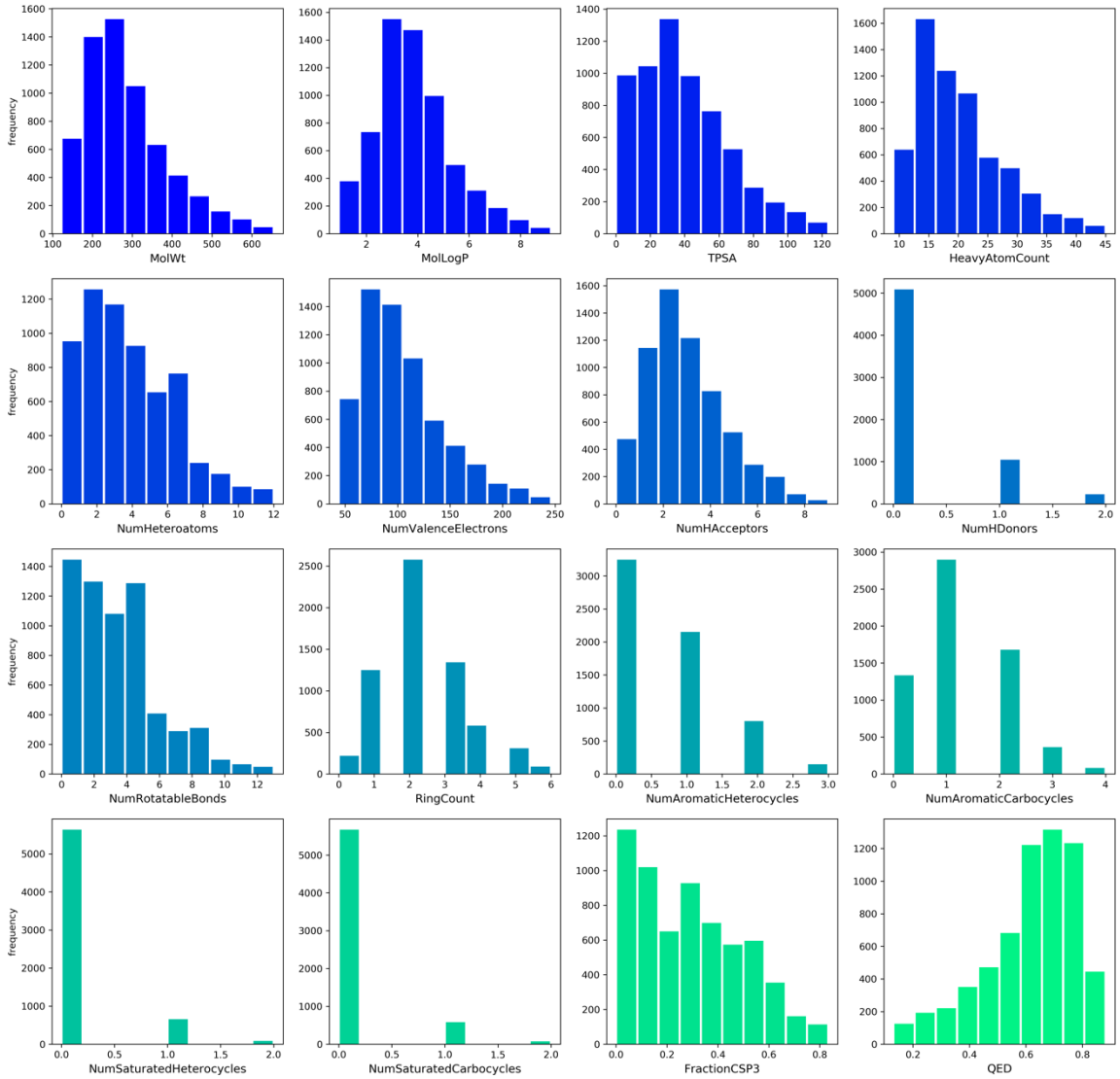
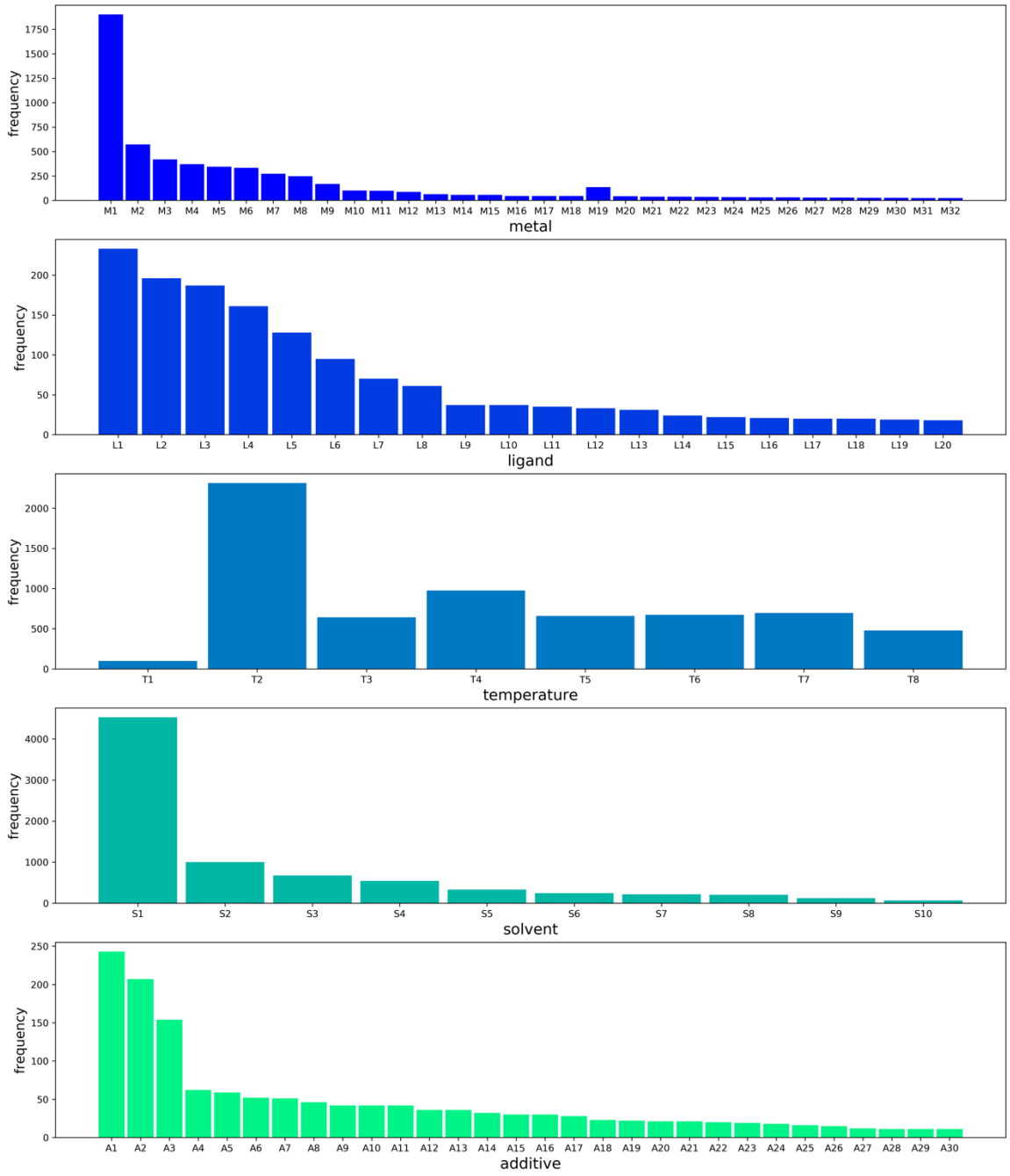


Figure 3.59 *Distribution of dictionary bin frequencies in Negishi dataset.*



3.6.5.4 PKR Dataset

Table 3.34 *Statistical summary of product molecular property descriptors in the PKR dataset.*

property	count	mean	std	min	25%	50%	75%	max
Yield	2749	0.66	0.22	0.03	0.51	0.69	0.84	1.00
MolWt	2746	312.95	94.15	122.17	240.30	303.39	364.40	905.24
MolLogP	2746	3.40	1.71	-1.18	2.10	3.13	4.37	11.22
TPSA	2746	45.98	23.24	0.00	26.30	44.76	63.68	151.35
HeavyAtomCount	2746	22.33	6.51	9.00	17.00	22.00	26.00	64.00
NumHeteroatoms	2746	4.00	1.99	0.00	2.00	4.00	5.00	13.00
NumValenceElectrons	2746	117.86	34.34	48.00	92.00	114.00	136.00	340.00
NumHAcceptors	2746	3.15	1.64	0.00	2.00	3.00	4.00	13.00
NumHDonors	2746	0.11	0.35	0.00	0.00	0.00	0.00	3.00
NumRotatableBonds	2746	3.04	2.36	0.00	1.00	3.00	4.00	24.00
RingCount	2746	3.26	1.09	1.00	3.00	3.00	4.00	8.00
NumAromHeterocycles	2746	0.05	0.22	0.00	0.00	0.00	0.00	2.00
NumAromCarbocycles	2746	0.10	0.98	0.00	0.00	1.00	1.00	5.00
NumSatHeterocycles	2746	0.43	0.64	0.00	0.00	0.00	1.00	4.00
NumSatCarbocycles	2746	0.54	0.71	0.00	0.00	0.00	1.00	4.00
FractionCSP3	2746	0.47	0.20	0.00	0.31	0.45	0.64	0.93
QED	2746	0.64	0.15	0.05	0.56	0.66	0.75	0.92

Figure 3.60 *Distribution of reaction yields in PKR dataset.*

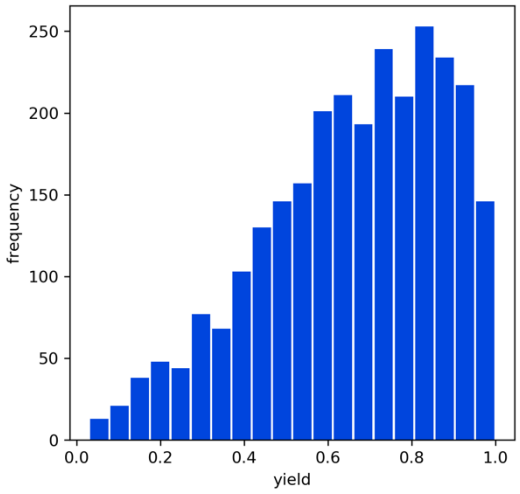


Figure 3.61 *Distribution of each molecular descriptor in PKR dataset products.*

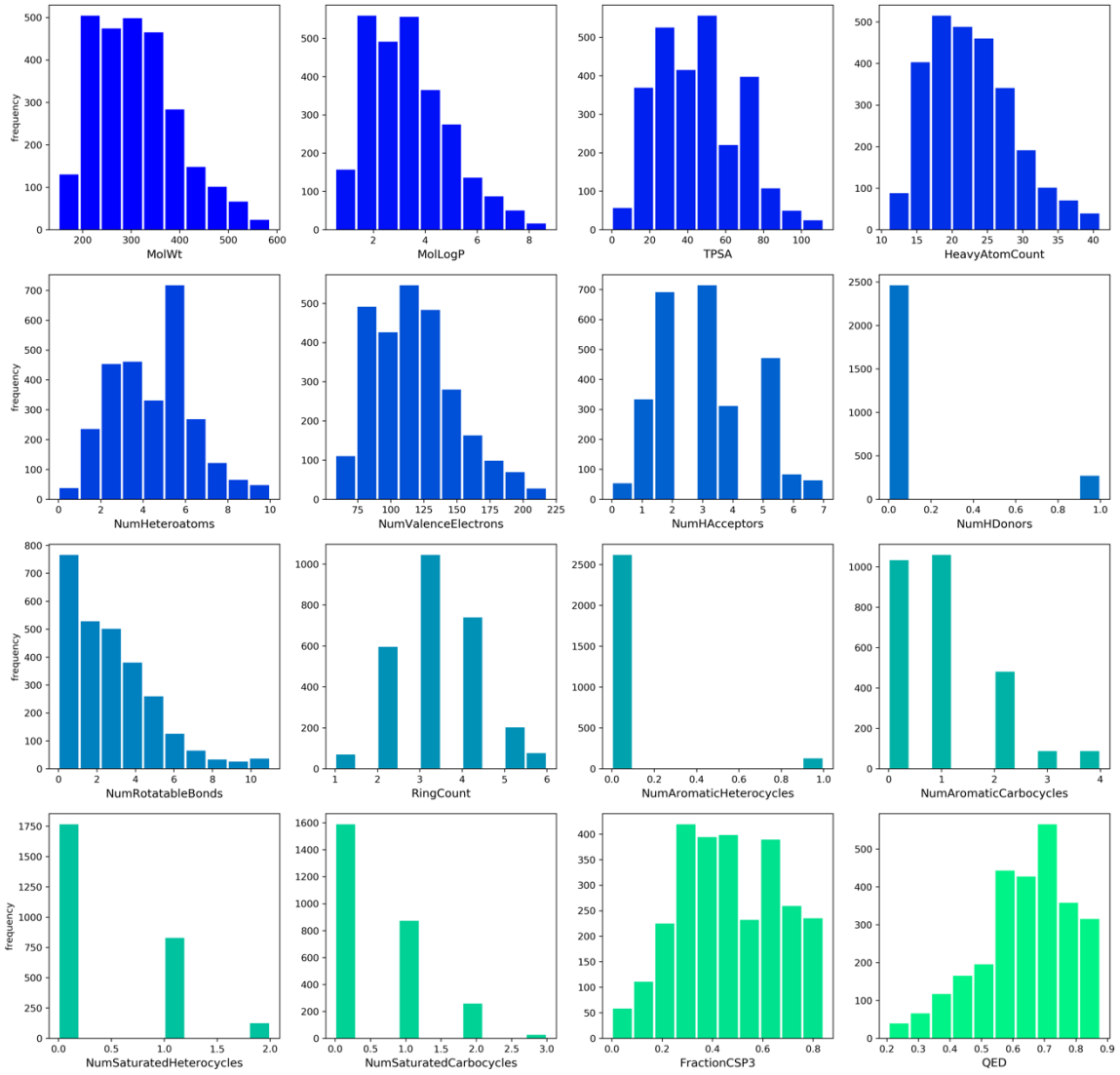
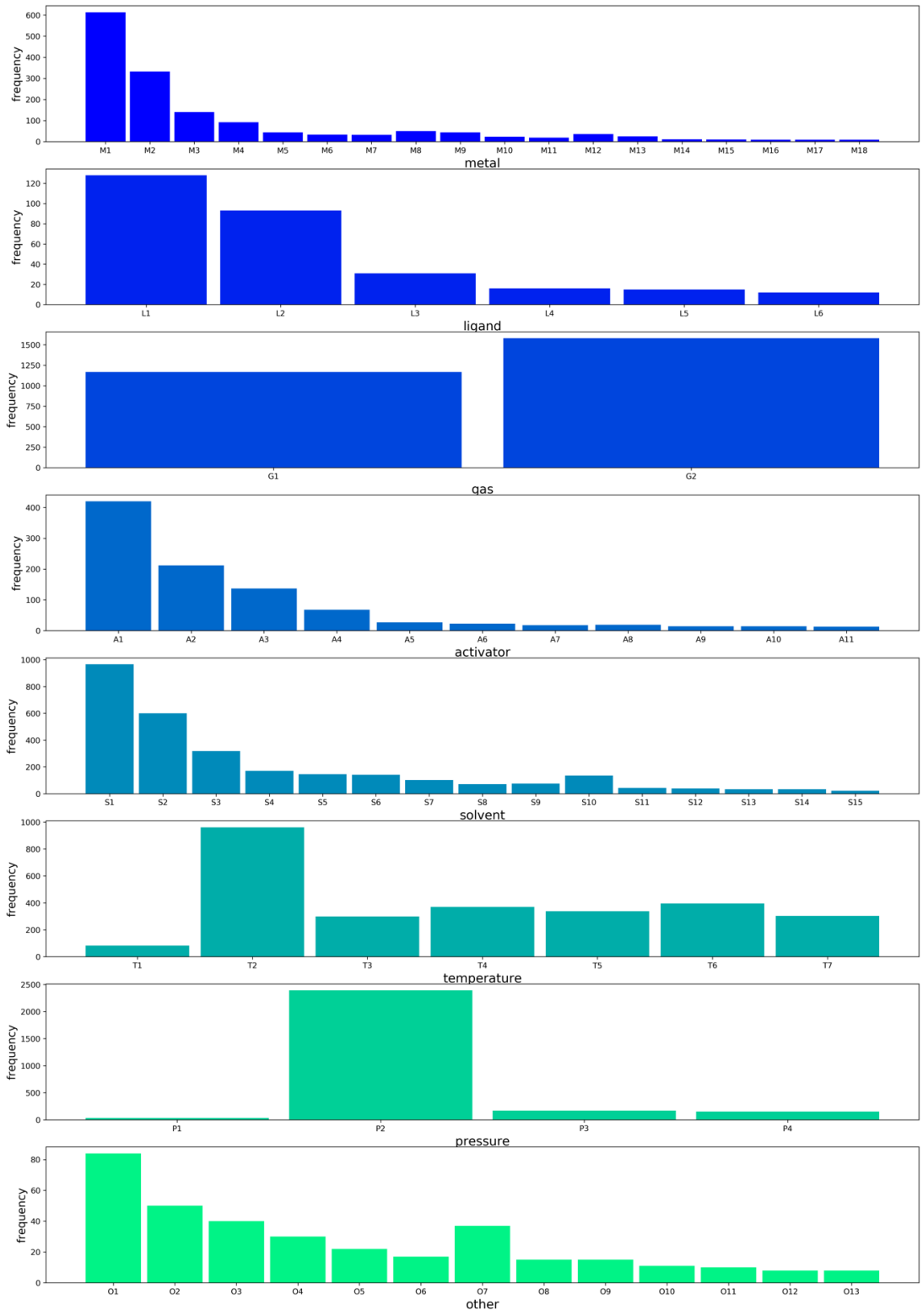


Figure 3.62 *Distribution of dictionary bin frequencies in PKR dataset.*



3.7 REFERENCES

(1) Dreher, S. D. Catalysis in Medicinal Chemistry. *React. Chem. Eng.* **2019**, *4*, 1530–1535.

(2) Blakemore, D. C.; Castro, L.; Churcher, I.; Rees, D. C.; Thomas, A. W.; Wilson, D. M.; Wood, A. Organic Synthesis Provides Opportunities to Transform Drug Discovery. *Nat. Chem.* **2018**, *10*, 383–394.

(3) Mahatthananchai, J.; Dumas, A. M.; Bode, J. W. Catalytic Selective Synthesis. *Angew. Chem., Int. Ed.* **2012**, *51*, 10954–10990.

(4) Reid, J. P.; Sigman, M. S. Comparing Quantitative Prediction Methods for the Discovery of Small-Molecule Chiral Catalysts. *Nat. Rev. Chem.* **2018**, *2*, 290–305.

(5) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559*, 547–555.

(6) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: A Benchmark for Molecular Machine Learning. *Chem. Sci.* **2018**, *9*, 513–530.

(7) Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; Palmer, A.; Settels, V.; Jaakkola, T.; Jensen, K.; Barzilay, R. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **2019**, *59*, 3370–3388.

(8) Withnall, M.; Lindelöf, E.; Engkvist, O.; Chen, H. Building Attention and Edge Message Passing Neural Networks for Bioactivity and Physical–Chemical Property Prediction. *J. Cheminf.* **2020**, *12*, 1.

(9) Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackerman, Z.; Tran, V. M.; Chiappino-Pepe, A.; Badran, A. H.; Andrews, I. W.; Chory, E. J.; Church, G. M.; Brown, E. D.; Jaakkola, T. S.; Barzilay, R.; Collins, J. J. A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020**, *180*, 688–702.

(10) Blaschke, T.; Olivecrona, M.; Engkvist, O.; Bajorath, J.; Chen, H. Application of Generative Autoencoder in De Novo Molecular Design. *Mol. Inf.* **2018**, *37*, 1700123.

(11) Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W. Deep Learning for Molecular Design—A Review of the State of the Art. *Mol. Syst. Des. Eng.* **2019**, *4*, 828–849.

(12) Prykhodko, O.; Johansson, S. V.; Kotsias, P.-C.; Arús-Pous, J.; Bjerrum, E. J.; Engkvist, O.; Chen, H. A De Novo Molecular Generation Method Using Latent Vector Based Generative Adversarial Network. *J. Cheminf.* **2019**, *11*, 74.

(13) Moret, M.; Friedrich, L.; Grisoni, F.; Merk, D.; Schneider, G. Generating Customized Compound Libraries for Drug Discovery with Machine Intelligence. *ChemRxiv* **2019**, *Preprint*.

(14) Panteleev, J.; Gao, H.; Jia, L. Recent Applications of Machine Learning in Medicinal Chemistry. *Bioorg. Med. Chem. Lett.* **2018**, *28*, 2807–2815.

(15) Skoraczyński, G.; Dittwald, P.; Miasojedow, B.; Szymkúc, S.; Gajewska, E. P.; Grzybowski, B. A.; Gambin, A. Predicting the Outcomes of Organic Reactions via Machine Learning: Are Current Descriptors Sufficient? *Sci. Rep.* **2017**, *7*.

(16) Coley, C. W.; Barzilay, R.; Jaakkola, T. S.; Green, W. H.; Jensen, K. F. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **2017**, *3*, 434–443.

(17) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186–190.

(18) Nielsen, M. K.; Ahneman, D. T.; Riera, O.; Doyle, A. G. Deoxyfluorination with Sulfonyl Fluorides: Navigating Reaction Space with Machine Learning. *J. Am. Chem. Soc.* **2018**, *140*, 5004–5008.

(19) Simón-Vidal, L.; García-Calvo, O.; Oteo, U.; Arrasate, S.; Lete, E.; Sotomayor, N.; Gonzaáez-Díaz, H. Perturbation-Theory and Machine Learning (PTML) Model for High-Throughput Screening of Parham Reactions: Experimental and Theoretical Studies. *J. Chem. Inf. Model.* **2018**, *58*, 1384–1396.

(20) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an Organic Synthesis Robot with Machine Learning to Search for New Reactivity. *Nature* **2018**, *559*, 377–381.

(21) Hughes, T. B.; Miller, G. P.; Swamidass, S. J. Modeling Epoxidation of Drug-like Molecules with a Deep Machine Learning Network. *ACS Cent. Sci.* **2015**, *1*, 168–180.

(22) Peng, Q.; Duarte, F.; Paton, R. S. Computing Organic Stereoselectivity – From Concepts to Quantitative Calculations and Predictions. *Chem. Soc. Rev.* **2016**, *45*, 6093–6107.

(23) Banerjee, S.; Sreenithya, A.; Sunoj, R. B. Machine Learning for Predicting Product Distributions in Catalytic Regioselective Reactions. *Phys. Chem. Chem. Phys.* **2018**, *20*, 18311–18318.

(24) Beker, W.; Gajewska, E. P.; Badowski, T.; Grzybowski, B. A. Prediction of Major Regio-, Site-, and Diastereoisomers in Diels–Alder Reactions by Using Machine-Learning: The Importance of Physically Meaningful Descriptors. *Angew. Chem., Int. Ed.* **2019**, *58*, 4515–4519.

(25) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363*, eaau5631.

(26) Coley, C. W.; Green, W. H.; Jensen, K. F. Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **2018**, *51*, 1281–1289.

(27) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature* **2018**, *555*, 604–610.

(28) Coley, C. W.; Green, W. H.; Jensen, K. F. RDChiral: An RDKit Wrapper for Handling Stereochemistry in Retrosynthetic Template Extraction and Application. *J. Chem. Inf. Model.* **2019**, *59*, 2529–2537.

(29) Badowski, T.; Gajewska, E. P.; Molga, K.; Grzybowski, B. A. Synergy Between Expert and Machine-Learning Approaches Allows for Improved Retrosynthetic Planning. *Angew. Chem., Int. Ed.* **2020**, *59*, 725–730.

(30) Nicolaou, C. A.; Watson, I. A.; LeMasters, M.; Masquelin, T.; Wang, J. Context Aware Data-Driven Retrosynthetic Analysis. *J. Chem. Inf. Model.* **2020**, *60*, 2728–2738.

(31) Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **2018**, *4*, 1465–1476.

(32) Nakayama, Y.; Maser, M. R.; Okita, T.; Dubrovskiy, A. V.; Campbell, T. L.; Reisman, S. E. Total Synthesis of Ritterazine B. *J. Am. Chem. Soc.* **2021**, *143*, 4187–4192.

(33) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Statist.* **2001**, *29*, 1189–1232.

(34) (a) Fitzner, M.; Wuitschik, G.; Adam, J.-M.; Schindler, T.; Reymond, J.-L. What Can Reaction Databases Teach Us About Buchwald–Hartwig Cross-Couplings? *Chem. Sci.* **2020**, *11*, 13085–13093. (b) Huerta, F.; Hallinder, S.; Minidis, A. Machine Learning to Reduce Reaction Optimization Lead Time – Proof of Concept with Suzuki, Negishi and Buchwald–Hartwig Cross-Coupling Reactions. *ChemRxiv* **2020**, *Preprint*.

(35) (a) Miyaura, N.; Suzuki, A. Palladium-Catalyzed Cross-Coupling Reactions of Organoboron Compounds. *Chem. Rev.* **1995**, *95*, 2457–2483. (b) Bariwal, J.; Van der Eycken, E. CN Bond Forming Cross-Coupling Reactions: An Overview. *Chem. Soc. Rev.* **2013**, *42*, 9283. (c) Negishi, E.; King, A. O.; Okukado, N. Selective Carbon-Carbon Bond Formation via Transition Metal Catalysis. 3. A Highly Selective Synthesis of Unsymmetrical Biaryls and Diarylmethanes by the Nickel- or Palladium-Catalyzed Reaction of Aryl- and Benzylzinc Derivatives with Aryl Halides. *J. Org. Chem.* **1977**, *42*, 1821–1823. (d) Khand, I. U.; Knox, G. R.; Pauson, P. L.; Watts, W. E.; Foreman, M. I. Organocobalt Complexes. Part II. Reaction of Acetylenehexacarbonyldicobbalt Complexes, $(R^1C_2R^2)Co_2(CO)_6$, with Norbornene and its Derivatives. *J. Chem. Soc. Perkin Trans. 1* **1973**, *0*, 977–981.

(36) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular Graph Convolutions: Moving Beyond Fingerprints. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 595–608.

(37) Full data-processing and modeling code is available in the accompanying repository. https://github.com/slryou41/reaction-gcnn.

(38) Reaxys. https://new.reaxys.com/ (accessed on May 13, 2019).

(39) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28*, 31–36.

(40) Labels were trimmed at 95% cumulative density during both the initial enumeration and within each category.

(41) Given their relative frequency and to maintain consistent formatting, intramolecular couplings were dropped from the Suzuki, CN, and Negishi datasets, but were retained for the PKR dataset.

(42) Sanchez-Lengeling, B.; Wei, J. N.; Lee, B. K.; Gerkin, R. C.; Aspuru-Guzik, A.; Wiltschko, A. B. Machine Learning for Scent: Learning Generalizable Perceptual Representations of Small Molecules. *arXiv:1910.10685* **2019**, *Preprint*.

(43) Cui, Y.; Jia, M.; Lin, T.-Y.; Song, Y.; Belongie, S. Class-Balanced Loss Based on Effective Number of Samples. *arXiv:1901.05555* **2019**, *Preprint*.

(44) Wu, X.-Z.; Zhou, Z.-H. A Unified View of Multi-Label Performance Measures. *arXiv:1609.00288* **2017**, *Preprint*.

(45) Natekin, A.; Knoll, A. Gradient Boosting Machines, a Tutorial. *Front. Neurorobot.* **2013**, *7*, 21.

(46) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *J. Cheminf.* **2018**, *10*, 4.

(47) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.

(48) Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y.; Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems 30*; Curran Associates, Inc., **2017**, 3146–3154.

(49) Gradient boosting was implemented using Microsoft's LightGBM. https://github.com/microsoft/LightGBM.

(50) Li, J.; Eastgate, M. D. Making Better Decisions During Synthetic Route Design: Leveraging Prediction to Achieve Greenness-by-Design. *React. Chem. Eng.* **2019**, *4*, 1595–1607.

(51) Zhang, M.-L.; Zhou, Z.-H. A Review on Multi-Label Learning Algorithms. *IEEE Trans. Knowl. Data. Eng.* **2014**, *26*, 1819–1837.

(52) Read, J.; Pfahringer, B.; Holmes, G.; Frank, E. Classifier Chains for Multilabel Classification. Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg, **2009**; pp 254–269.

(53) Read, J.; Martino, L.; Olmos, P.; Luengo, D. Scalable Multi-Output Label Prediction: From Classifier Chains to Classifier Trellises. *Pattern Recognition* **2015**, *48*, 2096–2109.

(54) All NN models were implemented using the Chainer Chemistry (ChainerChem) deep learning library. https://github.com/chainer/chainer-chemistry. See: Tokui, S.; Okuta, R.; Akiba, T.; Niitani, Y.; Ogawa, T.; Saito, S.; Suzuki, S.; Uenishi, K.; Vogel, B.; Vincent, H. Y. Chainer: A Deep Learning Framework for Accelerating the Research Cycle. *arXiv:1908.00213* **2019**, *Preprint*.

(55) (a) Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. *arXiv:1509.09292* **2015**, *Preprint*. (b) Li, Y.; Tarlow, D.; Brockschmidt, M.; Zemel, R. Gated Graph Sequence Neural Networks. *arXiv:1511.05493* **2017**, *Preprint*. (c) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *arXiv:1704.01212* **2017**, *Preprint*. (d) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv:1710.10903* **2018**, *Preprint*. (e) Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; Berg, R. v. d.; Titov, I.; Welling, M. Modeling Relational Data with Graph Convolutional Networks. *arXiv:1703.06103* **2017**, *Preprint*. (f) Kipf, T. N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv:1609.02907* **2017**, *Preprint*.

(56) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980* **2017**, *Preprint*.

(57) Busbridge, D.; Sherburn, D.; Cavallo, P.; Hammerla, N. Y. Relational Graph Attention Networks. *arXiv:1904.05811* **2019**, *Preprint*.

(58) We found it necessary to reduce the hidden dimension of R-GATs to avoid excessive memory requirements relative to other GCNs, and thus do not make a direct comparison of their performance.

(59) Newman, M. E. J. Power Laws, Pareto Distributions and Zipf's Law. *Contemp. Phys.* **2005**, *46*, 323–351.

(60) (a) Zhou, K.; Dong, Y.; Lee, W. S.; Hooi, B.; Xu, H.; Feng, J. Effective Training Strategies for Deep Graph Neural Networks. *arXiv:2006.07107* **2020**, *Preprint*. (b) Jiang, D.; Wu, Z.; Hsieh, C.-Y.; Chen, G.; Liao, B.; Wang, Z.; Shen, C.; Cao, D.; Wu, J.; Hou, T. Could Graph Neural Networks Learn Better Molecular Representation for Drug Discovery? A Comparison Study of Descriptor-Based and Graph-Based Models. *J. Cheminform.* **2021**, *13*, 12.

(61) Galvez, J.; Garcia, R.; Salabert, M. T.; Soler, R. Charge Indexes. New Topological Descriptors. *J. Chem. Inf. Model.* **1994**, *34*, 520–525.

(62) Moreau, G.; Broto, P. The Autocorrelation of a Topological Structure: A New Molecular Descriptor. *New J. Chem.* **1980**, *4*, 359–360.

(63) Hollas, B. An Analysis of the Autocorrelation Descriptor for Molecules. *J. Math. Chem.* **2003**, *33*, 91–101.

(64) Detailed adversarial control studies for all GBM models are included in the Experimental Section.

(65) Chuang, K. V.; Keiser, M. J. Adversarial Controls for Scientific Machine Learning. *ACS Chem. Biol.* **2018**, *13*, 2819–2821.

(66) Sanchez-Lengeling, B.; Wei, J.; Lee, B.; Reif, E.; Wang, P. Y.; Qian, W.; McCloskey, K.; Colwell, L.; Wiltschko, A. Evaluating Attribution for Graph Neural Networks. *NeurIPS 34*, **2020**, 5898–5910.

(67) Landrum, G. A. RDKit: Open-Source Cheminformatics Software. https://github.com/rdkit/rdkit, (accessed Nov 20, 2018).

(68) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(69) Chuang, K. V.; Keiser, M. J. Comment on "Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning." *Science* **2018**, *362*, eaat8603.

(70) Hall, L. H.; Mohney, B.; Kier, L. B. The Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs. *J. Chem. Inf. Model.* **1991**, *31*, 76–82.

(71) Hall, L. H.; Kier, L. B. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *J. Chem. Inf. Model.* **1995**, *35*, 1039–1045.

(72) Todeschini, R.; Consonni, V. Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing / Volume II: Appendices, References, 1st ed.; Methods and Principles in Medicinal Chemistry; Wiley, **2009**; *Vol. 41*.

(73) Labute, P. A Widely Applicable Set of Descriptors. *J. Mol. Graphics Modell.* **2000**, *18*, 464–477.

(74) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873.

Appendix 3

Robotic High-Throughput Experimentation and Active Learning for Automated Reaction Optimization[1]

A3.1 INTRODUCTION

Section 3.2 details challenges faced in predicting experimental reaction yields from literature datasets (See Figures 3.4 and 3.5).[1] We note the difficulty in obtaining machine learning (ML) models that reliably predict low yields when datasets are heavily biased toward high-yielding reactions. Though we expect that oversampling and/or data augmentation techniques could be useful for accurately modeling such datasets,[2] we sought to address the issue of obtaining biased data at its root. Despite the cultural influences that result in such biased reporting, we recognize that the reality of synthetic chemistry is less fantastical; the majority of tested reactions "fail," giving zero to low yields of desired products (~20% or less).[1] As such, we initiated an effort to obtain representative, non-biased datasets of modern organic reactions from which to train practically useful yield predictors.

---

[1] The research discussed in this appendix was completed in collaboration with former graduate student Dr. Caitlin R. Lacker and current graduate student Emily P. Chen of the Reisman Lab, postdoctoral scholar Dr. Uriah Israel of the Yue and Van Valen Groups, graduate student Jennifer J. Sun of the Perona and Yue Groups, undergraduate Megan Santhumayor of the Reisman Lab, and Elena Zhong of the Yue Group. We warmly acknowledge Dr. Scott Virgil for engineering and experimental assistance. Please refer to the PhD Thesis of Dr. Caitlin R. Lacker for experimental procedures relevant to this appendix.

Robotic high-throughput experimentation (HTE) has modernized reaction screening in the pharmaceutical industry, enabling hundreds to thousands of experiments to be conducted simultaneously.[3-6] The technology has yet to see widespread use in academia due to its economic burden, though a handful of reports have appeared in recent years.[5,7,8] Critical to robotic HTE workflows is the ability to perform equally high-throughput analysis to quantify reaction outcomes. The most useful data from HTE screens is that with precise, calibrated performance measurements such as peak-integrated HPLC reaction yield and/or ee (enantiomeric excess) relative to an internal standard, as opposed to raw, uncalibrated UV peak areas. The Caltech Center for Catalysis and Chemical Synthesis (3CS) is endowed with two robotic chemical handling systems, one of which—a Freeslate® Core Module N_2 Drybox System—resides in an air-free, N_2-filled glovebox and can be used for screening of air-sensitive reactions such as those catalyzed by transition metals.

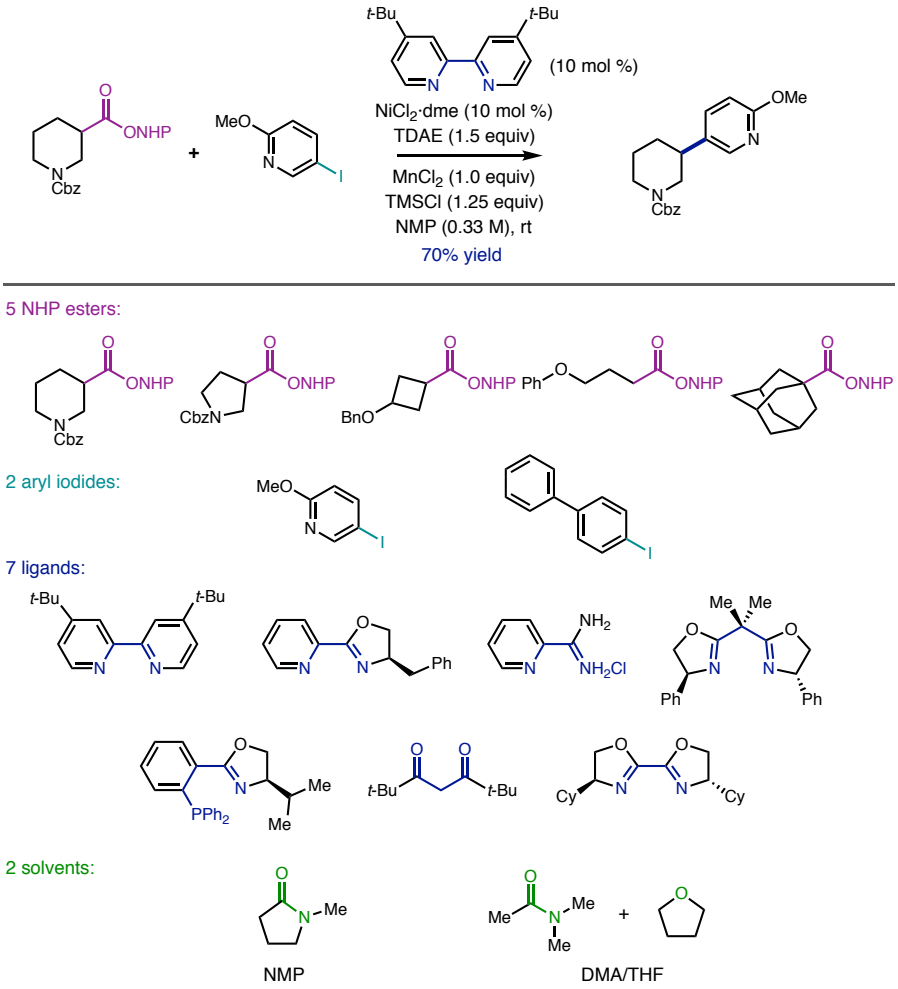
A3.2 INITIAL STUDIES

A reaction of interest in the Reisman Lab is the Ni-catalyzed reductive sp^3–sp^2 cross-coupling of redox-active esters and aryl halides (**Figure A3.1**). It was recently found in our group that *N*-hydroxyphthalimide (NHP) esters could serve as competent alkyl electrophiles for the desired coupling.[9] In batch studies, a 3-piperidyl NHP ester underwent coupling with 2-methoxy-5-iodo-pyridine in 70% yield using a Ni(II) catalyst, a di-*t*Bu-bipyridine ligand, the organic reductant tetrakis(dimethylamino)ethylene (TDAE), $MnCl_2$ and trimethylsilyl chloride (TMSCl) additives, and N-methylpyrrolidone (NMP) as solvent at room temperature (**Figure A3.1**, top).

While exciting, this initial hit gave racemic product by nature of the achiral bipyridine ligand, though we ultimately desired an enantioselective coupling. Our group and others have

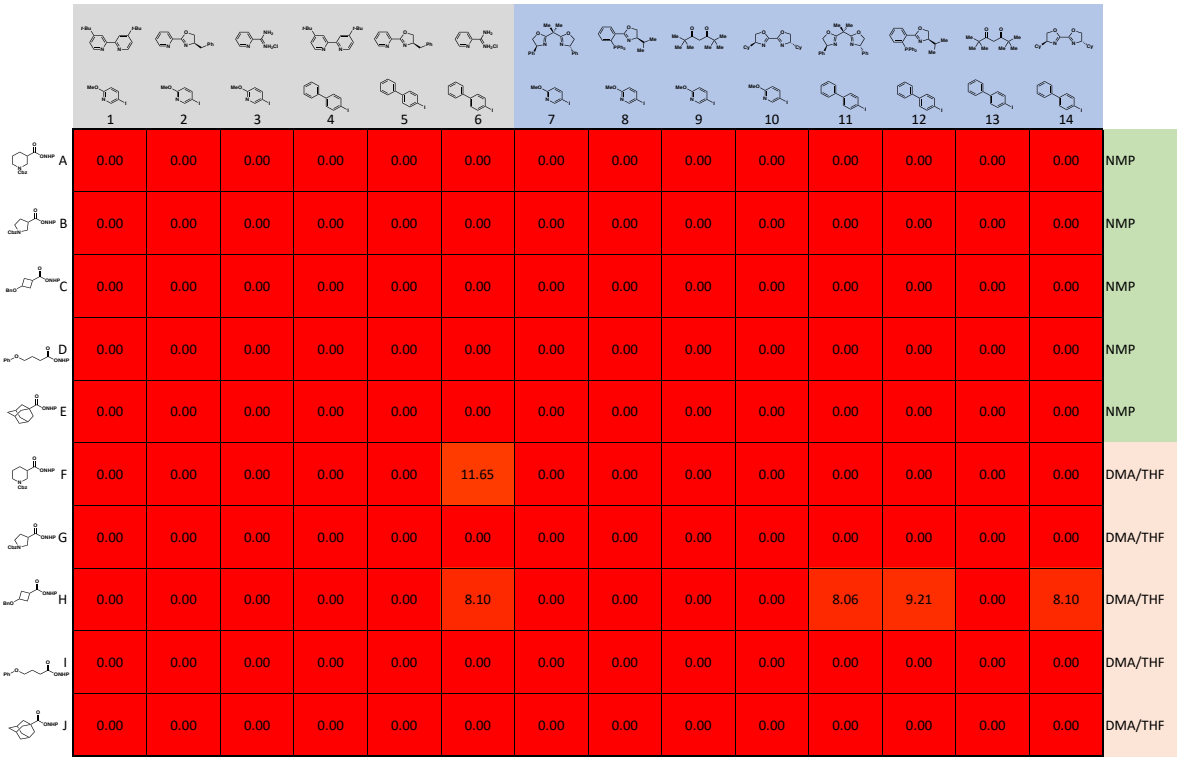
previously noted the challenge of rendering racemic couplings asymmetric due to the relatively weak reactivity of chiral ligands compared to, e.g., bipyridines.[10] We thus sought to evaluate the performance landscape of the transformation across the chemical space of substrates, ligands, and solvents (**Figure A3.1**, bottom), in order to better understand the structure-reactivity relationships (SRRs) of each and in the hopes of establishing reactivity with chiral ligands. To this end, we devised a HTE screen of five NHP esters, two aryl iodides, seven bidentate ligands, and two solvents, giving a combinatorial total of 140 reactions.

Figure A3.1 *Ni-catalyzed reductive cross-coupling of NHP esters and aryl iodides. Initial hit in batch optimization (top); HTE chemical space screen components (bottom).*



Initial screening efforts utilizing the Freeslate® robotic system in the 3CS were met with several engineering challenges. For example, the NMP solvent is particularly viscous, and we thus observed substantial lag times for liquid dispensing that resulted in non-uniform reaction compositions. We were able to circumvent such problems through trial and error, e.g., by increasing the wait time between liquid withdrawal and needle raising. Unfortunately, this resulted in significant time delays between rounds of reagent dispensing over the 140 reactions. As a result, the first full-scale HTE attempt was met with broad failure, with the vast majority of reactions giving 0% yield and all giving < 15% (**Table A3.1**).

Table A3.1 *Initial failed full-scale HTE screen.*



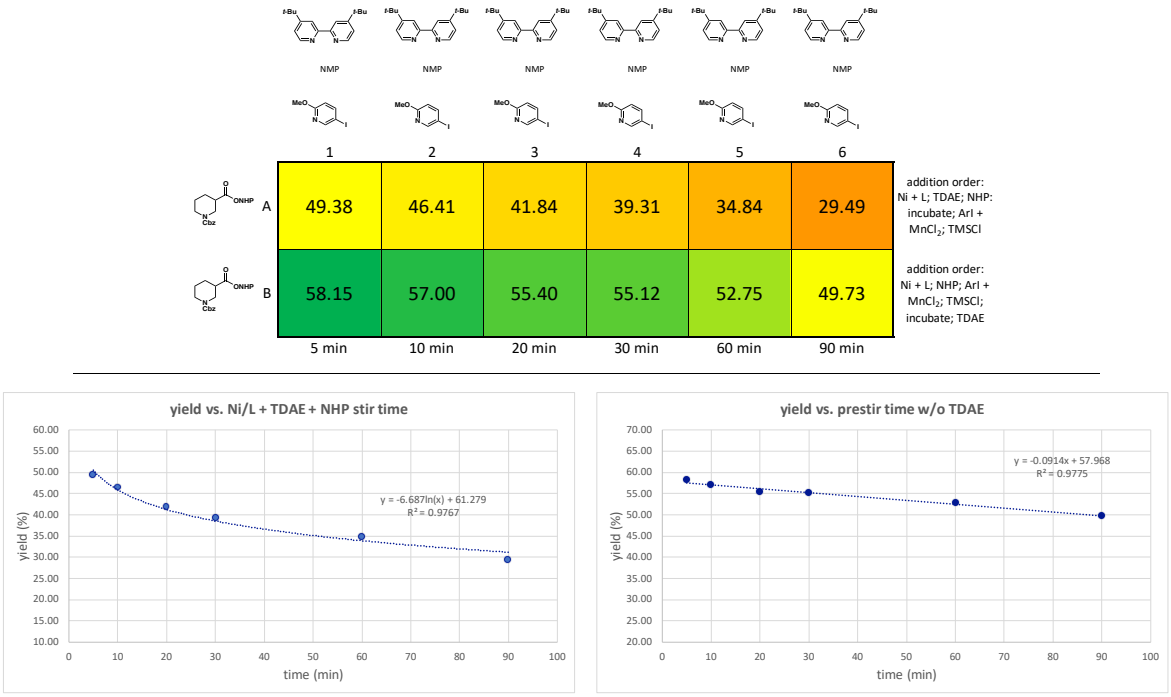
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NMP
B	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NMP
C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NMP
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NMP
E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NMP
F	0.00	0.00	0.00	0.00	0.00	11.65	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	DMA/THF
G	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	DMA/THF
H	0.00	0.00	0.00	0.00	0.00	8.10	0.00	0.00	0.00	0.00	8.06	9.21	0.00	8.10	DMA/THF
I	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	DMA/THF
J	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	DMA/THF

A3.3 TROUBLESHOOTING AND HTE OPTIMIZATION

We hypothesized that prolonged exposure of the NHP esters and Ni catalysts to TDAE was leading to deleterious reactivity before the productive reaction could be initiated by

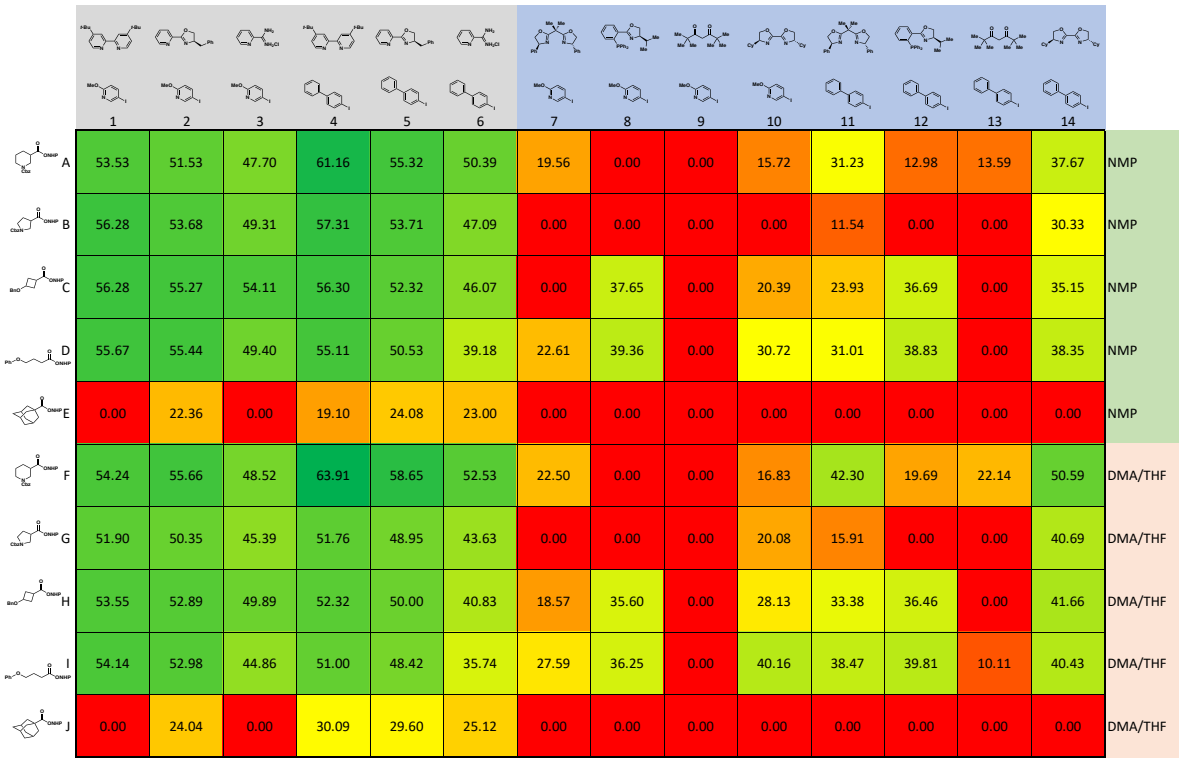
addition of TMSCl. To test this hypothesis, we devised two parallel time-course studies of six reactions each. In the first (*screen A*), the Ni/L, TDAE, and NHP ester/internal standard were added sequentially and allowed to incubate for periods of 5, 10, 20, 30, 60, or 90 minutes before addition of the aryl iodide/$MnCl_2$ and TMSCl. In the second (*screen B*), the Ni/L, NHP ester/internal standard, aryl iodide/MnCl2, and TMSCl were added sequentially and allowed to incubate for the same intervals before addition of TDAE. As expected, a marked drop in yield was observed in screen A over the increasing TDAE incubation time (**Figure A3.2**, top, row A), from 49% to 29%. On plotting, the decay appeared roughly exponential (**Figure A3.2**, bottom-left), though replicate experiments would add confidence to the relationship. Conversely, in screen B, only a minor drop in yield was observed (**Figure A3.2**, top, row B), from 58% to 50%. This decay was roughly linear with a slope of –0.1% yield/min (**Figure A3.2**, bottom-right).

Figure A3.2 *Time course study of TDAE and TMSCl incubation. Tabulated results (top); plotted yield decay for two addition order setups (bottom).*



Puzzlingly, even in screen A, the reaction yield did not drop to 0% in any case, despite incubation times of up to 90 min. Though the full cause of the failed results in Table A3.1 is not completely understood, we elected to divide the second attempt into two sub-screens of 70 reactions by solvent in hopes that incubation times and reproducibility could be optimized. Gratifyingly, we observed far superior results with the revised addition order (as in *screen B* above) and split setup, obtaining a clean spread of reaction yields across the chemical space (**Table A3.2**).

Table A3.2 *Results of full-scale HTE screen.*



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
A	53.53	51.53	47.70	61.16	55.32	50.39	19.56	0.00	0.00	15.72	31.23	12.98	13.59	37.67	NMP
B	56.28	53.68	49.31	57.31	53.71	47.09	0.00	0.00	0.00	0.00	11.54	0.00	0.00	30.33	NMP
C	56.28	55.27	54.11	56.30	52.32	46.07	0.00	37.65	0.00	20.39	23.93	36.69	0.00	35.15	NMP
D	55.67	55.44	49.40	55.11	50.53	39.18	22.61	39.36	0.00	30.72	31.01	38.83	0.00	38.35	NMP
E	0.00	22.36	0.00	19.10	24.08	23.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NMP
F	54.24	55.66	48.52	63.91	58.65	52.53	22.50	0.00	0.00	16.83	42.30	19.69	22.14	50.59	DMA/THF
G	51.90	50.35	45.39	51.76	48.95	43.63	0.00	0.00	0.00	20.08	15.91	0.00	0.00	40.69	DMA/THF
H	53.55	52.89	49.89	52.32	50.00	40.83	18.57	35.60	0.00	28.13	33.38	36.46	0.00	41.66	DMA/THF
I	54.14	52.98	44.86	51.00	48.42	35.74	27.59	36.25	0.00	40.16	38.47	39.81	10.11	40.43	DMA/THF
J	0.00	24.04	0.00	30.09	29.60	25.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	DMA/THF

Unsurprisingly (*vide supra*), pyridine-containing ligands performed stronger than non-pyridine-containing ligands (cf Table A3.2 columns 1–6 and columns 7–14). However, notable reactivity was observed with bis(oxazoline) (box, columns 7 & 11), phosphino-oxazoline (phox, columns 8 &12), and bioxazoline (biox, columns 10 & 14) ligands. The most reactive chiral ligand was the pyridyl-oxazoline (pyox, columns 2 & 5), which gave nearly identical yields to the standard bipyridine (columns 1 & 4). In fact, the pyox ligand effected successful coupling of the adamantyl NHP ester and the iodo-pyridine (E2 & J2)—the most challenging of all substrate combinations, for which no other ligand gave observable yield. Altogether, the observed reactivity with chiral ligands is promising, as it provides an avenue to render the transformation asymmetric (see Introduction Section A3.1). Studies to this end are ongoing.

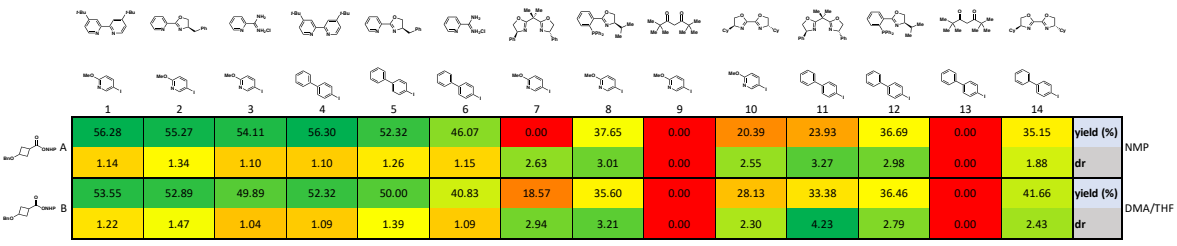
In general, promising reactivity was observed across the NHP esters, barring the 3° adamantyl substrate, as well as for both aryl iodides, though the non-heteroaryl iodide

performed slightly better. There seemed to be no major difference in results between the two solvent systems, though DMA/THF did give slightly higher yields (2–3%) in several cases. Altogether, it is encouraging to see a spread of reaction yields across the chemical space that could potentially lead to robust ML models (see Introduction Section A3.1).

A3.4 FOLLOW-UP CYCLOBUTANE STUDIES

From the HTE results above, we identified the cyclobutyl-NHP ester as an intriguing substrate for follow-up studies. Substituted cyclobutanes are of significant value to pharmaceutical chemistry, though methods for their diastereoselective synthesis are scarce. Upon NMR analysis of the arylated cyclobutane products (**Table A3.2**, rows C & H), we observed promising diastereomeric ratios (dr's) favoring the *cis*-1,3-disubstituted cyclobutanes (**Table A3.3**). Unfortunately, as is often encountered in selective synthesis, the ligands providing the highest dr's were those that also gave the lowest yields. The highest dr observed was 4.23:1 from the coupling of iodo-biphenyl using the bis(oxazoline) ligand in DMA/THF (cell B11). The same reaction proceeded in a modest 33% yield.

Table A3.3 *Yield and dr results for cyclobutyl-NHP ester couplings.*



	1	2	3	4	5	6	7	8	9	10	11	12	13	14		
A	56.28	55.27	54.11	56.30	52.32	46.07	0.00	37.65	0.00	20.39	23.93	36.69	0.00	35.15	yield (%)	NMP
	1.14	1.34	1.10	1.10	1.26	1.15	2.63	3.01	0.00	2.55	3.27	2.98	0.00	1.88	dr	
B	53.55	52.89	49.89	52.32	50.00	40.83	18.57	35.60	0.00	28.13	33.38	36.46	0.00	41.66	yield (%)	DMA/THF
	1.22	1.47	1.04	1.09	1.39	1.09	2.94	3.21	0.00	2.30	4.23	2.79	0.00	2.43	dr	

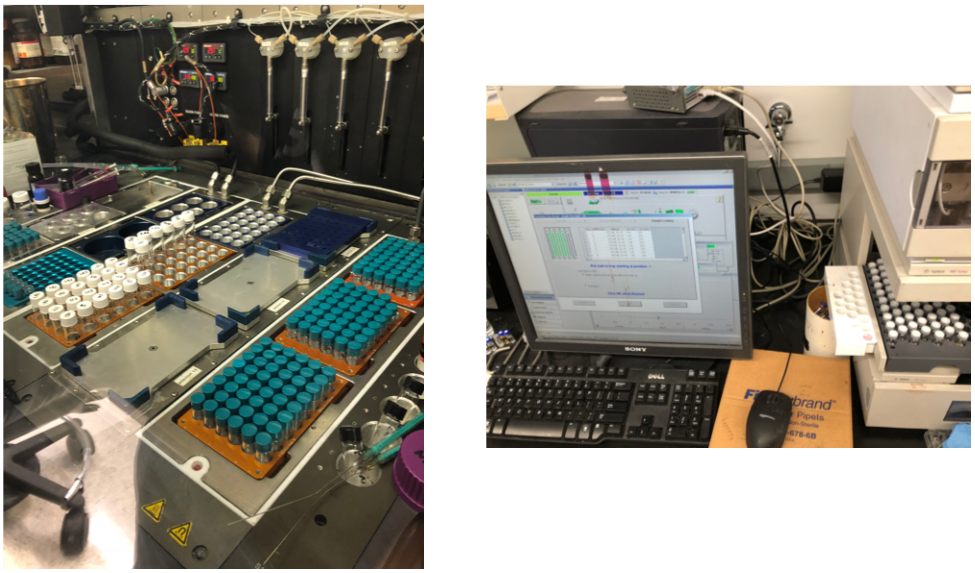
Additional efforts to increase the yield and dr of these couplings by traditional methods have been fruitful (not shown). Yields as high as 80% and dr's of up to 15:1 have been observed, using a box ligand modified from the one tested here. Additionally, a preliminary substrate scope has been demonstrated, with a variety of substituted cyclobutyl-NHP esters

and (hetero)aryl halides undergoing successful coupling. These results demonstrate the utility of HTE chemical space screening in initial reaction development efforts, where reactivity trends arise that may have otherwise been missed in batch optimization. Such data can serve as jumping-off points for further investigation into novel reactivity. The results of these combined studies will be reported in due course.

A3.5 CONCLUSION AND FUTURE OUTLOOK

In summary, we have devised a high-throughput robotic workflow for the screening of Ni-catalyzed, reductive Csp^3–Csp^2 couplings between NHP esters and aryl iodides. After initial setbacks, engineering controls and optimized addition orders allowed the acquisition of hundreds of quantitative reaction results on the order of 1–2 work days (**Figure A3.3**). With our approach, we successfully established structure-reactivity relationships across a variety of carefully chosen coupling partners and catalyst systems. We uncovered reactivity with chiral ligands in the box, biox, pyox, and phox families simultaneously, demonstrating the value of HTE. We expect the developed protocols to be of significant practical utility for group efforts in reaction development going forward.

Figure A3.3 *Photographs of robotic HTE setup. 140-reaction screen with stock solution and reaction plates (left); high-throughput LCMS analysis (right).*



It is worth noting that, on HTE scale, reaction yields have topped out around 60% (see Table A3.2). We have thus become interested in using this preliminary dataset as a launching point for reaction optimization. To this end, we have initiated an effort utilizing active machine learning (AL),[11] wherein a model trained on our current data would predict yields for a virtual library of potential next experiments to add to the dataset. We plan to conduct AL studies under a variety of algorithmic configurations, testing explorative, exploitative, and mixed selection strategies.[11,12] Further, we have begun the development of pre-trained models built on large corpus datasets with self-supervision for modeling of the types of small reaction datasets described herein.[2a,13] The results of these studies will be reported in due course.

A3.6 REFERENCES

(1) (a) Fitzner, M.; Wuitschik, G.; Koller, R. J.; Adam, J.-M.; Schindler, T.; Reymond, J.-L. What Can Reaction Databases Teach Us about Buchwald–Hartwig Cross-Couplings? *Chem. Sci.* **2020**, *11*, 13085–13093. (b) Schwaller, P.; Vaucher, A. C.; Laino, T.; Reymond, J.-L. Prediction of Chemical Reaction Yields using Deep Learning. *Mach. Learn.: Sci. Technol.* **2021**, *2*, 015016.

(2) For an example of augmenting reaction datasets with conjured examples having guaranteed outcomes, see: (a) Fortunato, M. E.; Coley, C. W.; Barnes, B. C.; Jensen, K. F. Data Augmentation and Pretraining for Template-Based Retrosynthetic Prediction in Computer-Aided Synthesis Planning. *J. Chem. Inf. Model.* **2020**, *60*, 3398–3407. For an example of oversampling minority class examples in molecular activity prediction, see: (b) McCloskey, K.; Sigel, E. A.; Kearnes, S.; Xue, L.; Tian, X.; Moccia, D.; Gikunju, D.; Bazzaz, S.; Chan, B.; Clark, M. A.; Cuozzo, J. W.; Guié, M.-A.; Guilinger, J. P.; Huguet, C.; Hupp, C. D.; Keefe, A. D.; Mulhern, C. J.; Zhang, Y.; Riley, P. Machine Learning on DNA-Encoded Libraries: A New Paradigm for Hit Finding. *J. Med. Chem.* **2020**, *63*, 8857–8866.

(3) Perera, D.; Tucker, J. W.; Brahmbhatt, S.; Helal, C. J.; Chong, A.; Farrell, W.; Richardson, P.; Sach, N. W. A Platform for Automated Nanomole-Scale Reaction Screening and Micromole-Scale Synthesis in Flow. *Science* **2018**, *359*, 429-434.

(4) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186-190.

(5) Gesmundo, N. J.; Sauvagnat, B.; Curran, P. J.; Richards, M. P.; Andrews, C. L.; Dandliker, P. J.; Cernak, T. Nanoscale Synthesis and Affinity Ranking. *Nature* **2018**, *557*, 228-232.

(6) Benz, M.; Asperger, A.; Hamester, M.; Welle, A.; Heissler, S.; Levkin, P. A. A Combined High-Throughput and High-Content Platform for Unified On-Chip Synthesis, Characterization and Biological Screening. *Nature Commun.* **2020**, *11*, 5391.

(7) Mahjour, B.; Shen, Y.; Cernak, T. Ultrahigh-Throughput Experimentation for Information-Rich Chemical Synthesis. *Acc. Chem. Res.* **2021**, *54*, 2337–2346.

(8) Potgieter, K.; Aimon, A.; Smit, E.; Delft, F.; Meijboom, R. Robotic Catalysis: A High-Throughput Method for Miniature Screening of Mesoporous Metal Oxides. *Chem. Methods* **2021**, *1*, 192–200.

(9) *Unpublished results*.

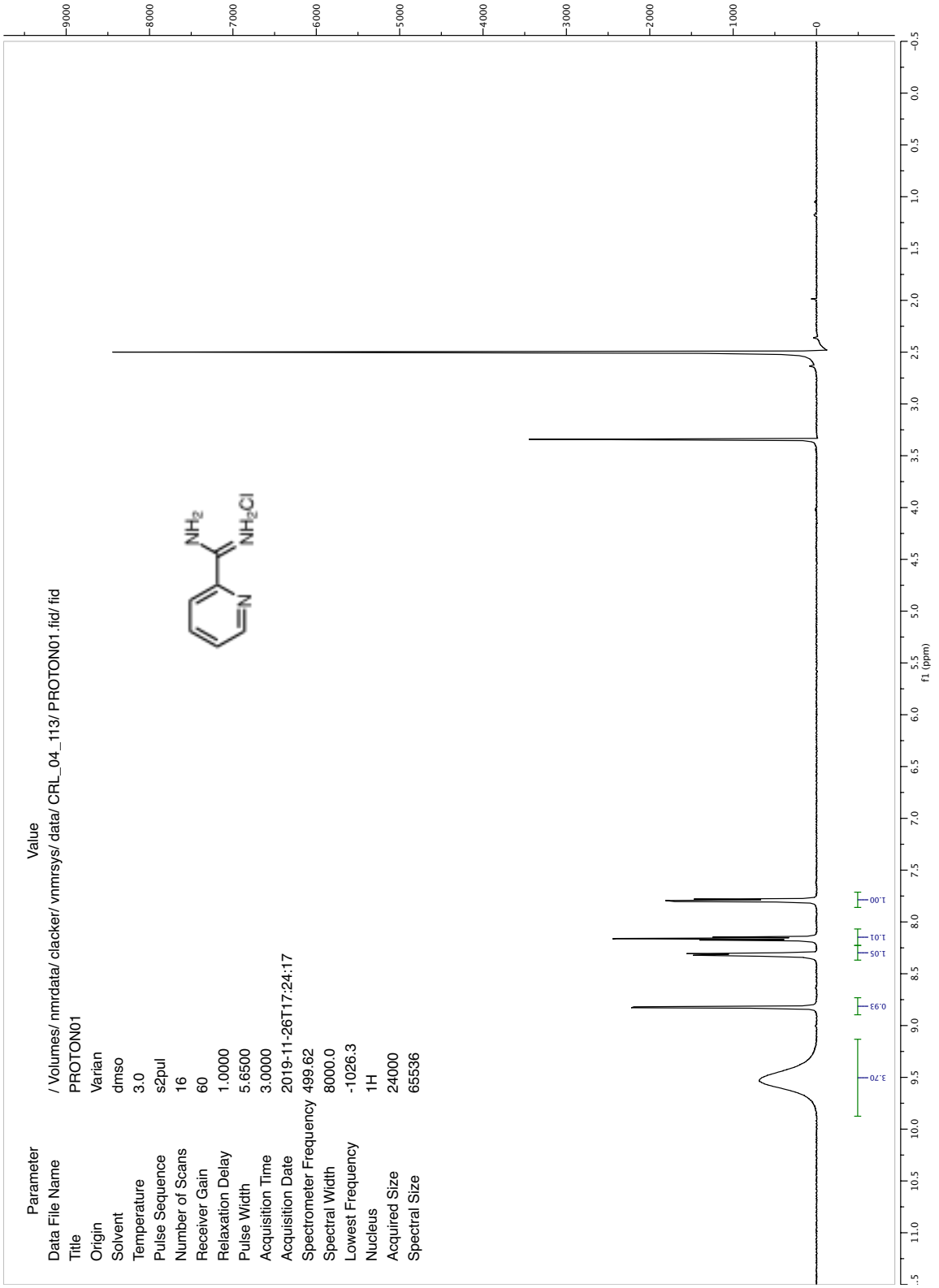
(10) Poremba, K. E.; Dibrell, S. E.; Reisman, S. E. Nickel-Catalyzed Enantioselective Reductive Cross-Coupling Reactions. *ACS Catal.* **2020**, *10*, 8237–8246.

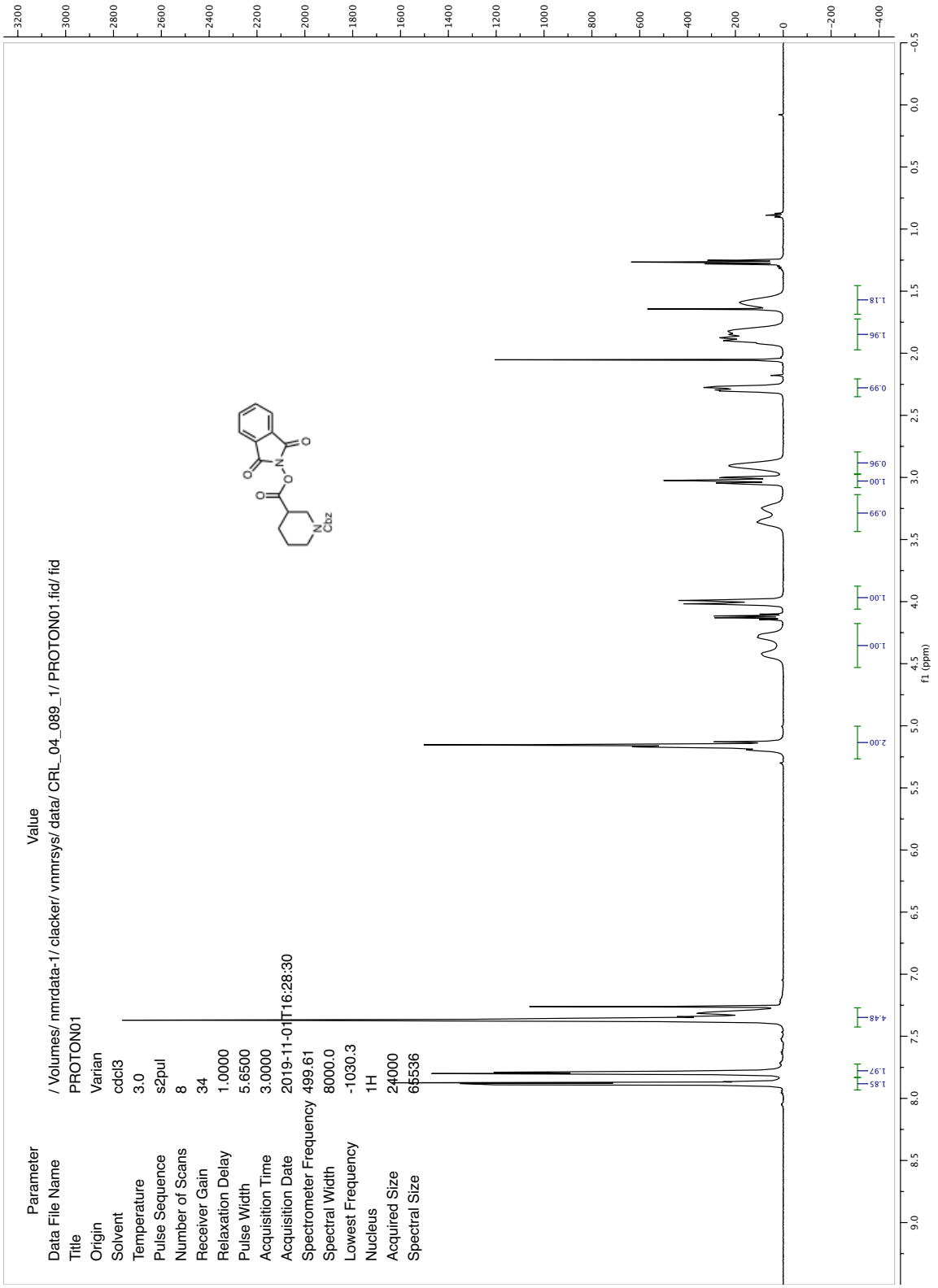
(11) (a) Settles, B. Active Learning *VI*. **2012** (Morgan & Claypool Publishers), p1–100. (b) Reker, D.; Schneider, G. Active-Learning Strategies in Computer-Assisted Drug Discovery. *Drug Discovery Today* **2015**, *20*, 458–465. (c) Ren, P.; Ziao, Y.; Chang, X.; Huang, P.-Y.; Li, Z.; Chen, X.; Wang, X. A Survey of Deep Active Learning. *arXiv:2009.00236* **2020**, *Preprint*.

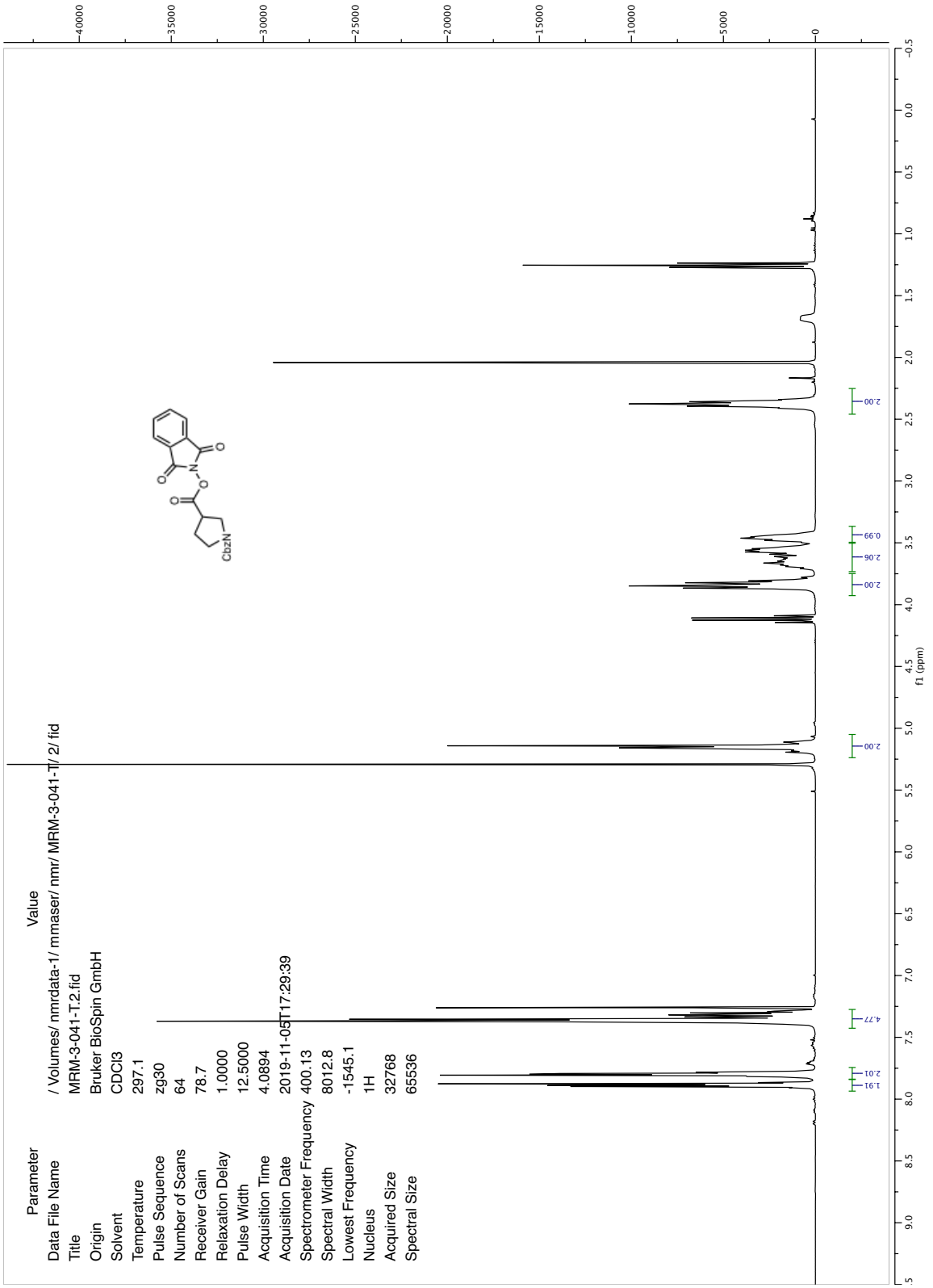
(12) (a) Baram, Y.; El-Yaniv, R.; Luz, K. Online Choice of Active Learning Algorithms. *Journal of Machine Learning Research* **2004**, *5*, 255–291. (b) Zuluaga, M.; Krause, A.; Sergent, G.; Puschel, M. Active Learning for Multi-Objective Optimization. *PMLR* 2013, 28, 462–470. (c) Reker, D. Practical Considerations for Active Machine Learning in Drug Discovery. *Drug Discovery Today: Technologies* **2019**, *32*, 73–79. (d) Eyke, N. S.; Green, W. H.; Jensen, K. F. Iterative Experimental Design Based on Active Machine Learning Reduces the Experimental Burden Associated with Reaction Screening. *React. Chem. Eng.* **2020**, *5*, 1963–1972.

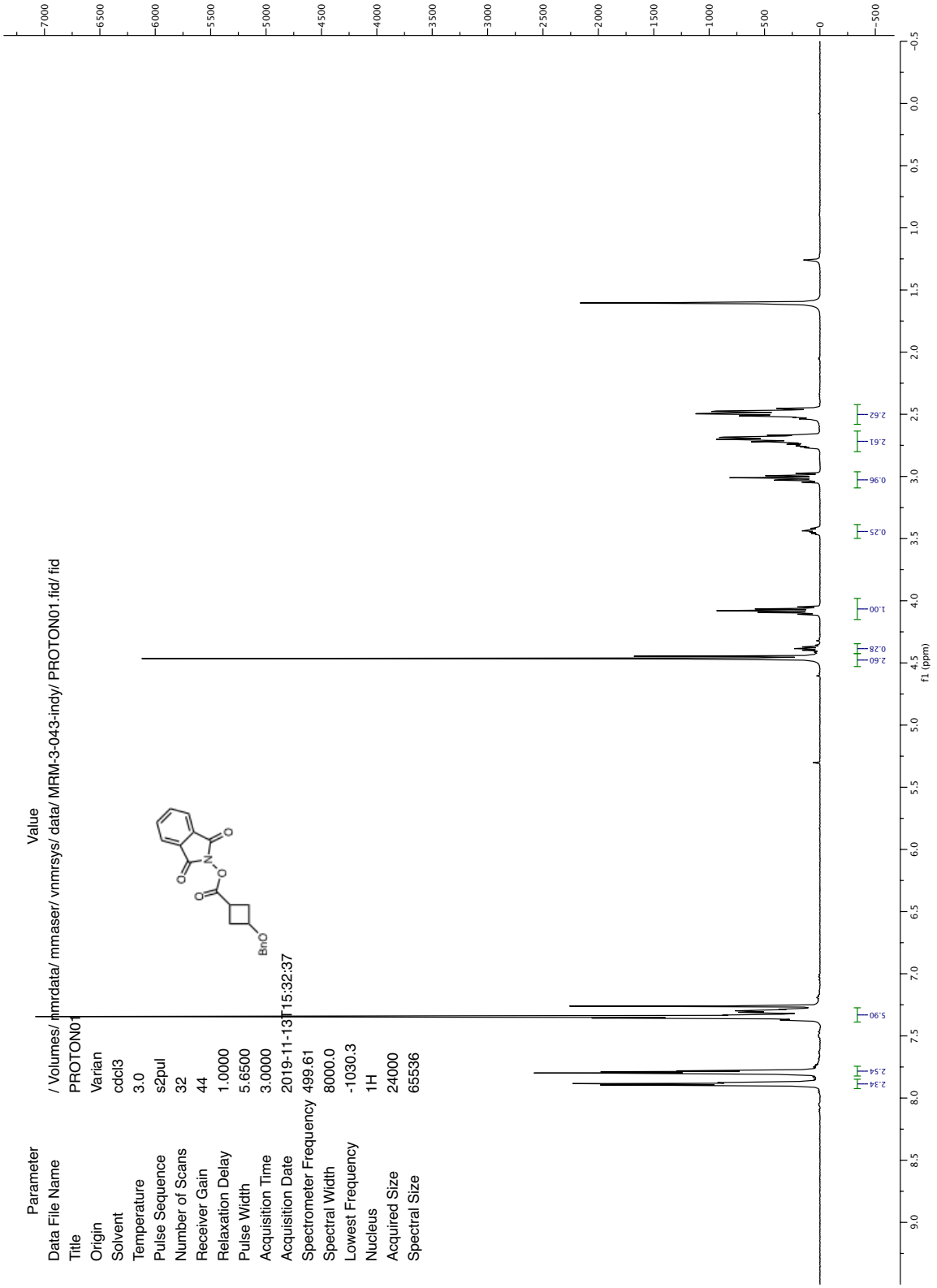
(13) (a) Hu, W.; Liu, B.; Gomes, J.; Zitnik, M.; Liang, P.; Pande, V.; Leskovec, J. Strategies for Pre-training Graph Neural Networks. *arXiv:1905.12265* **2020**, *Preprint*. (b) Reed, C. J.; Yue, X.; Nrusimha, A.; Ebrahimi, S.; Vijaykumar, V.; Mao, R.; Li, B.; Zhang, S.; Guillory, D.; Metzger, S.; Keutzer, K.; Darrell, T. Self-Supervised Pretraining Improves Self-Supervised Pretraining. *arXiv:2103.12718* **2021**, *Preprint*. (c) Li, X.; Fourches, D. Inductive Transfer Learning for Molecular Activity Prediction: Next-Gen QSAR Models with MolPMoFiT. *J. Cheminform.* **2020**, *12*, 27.

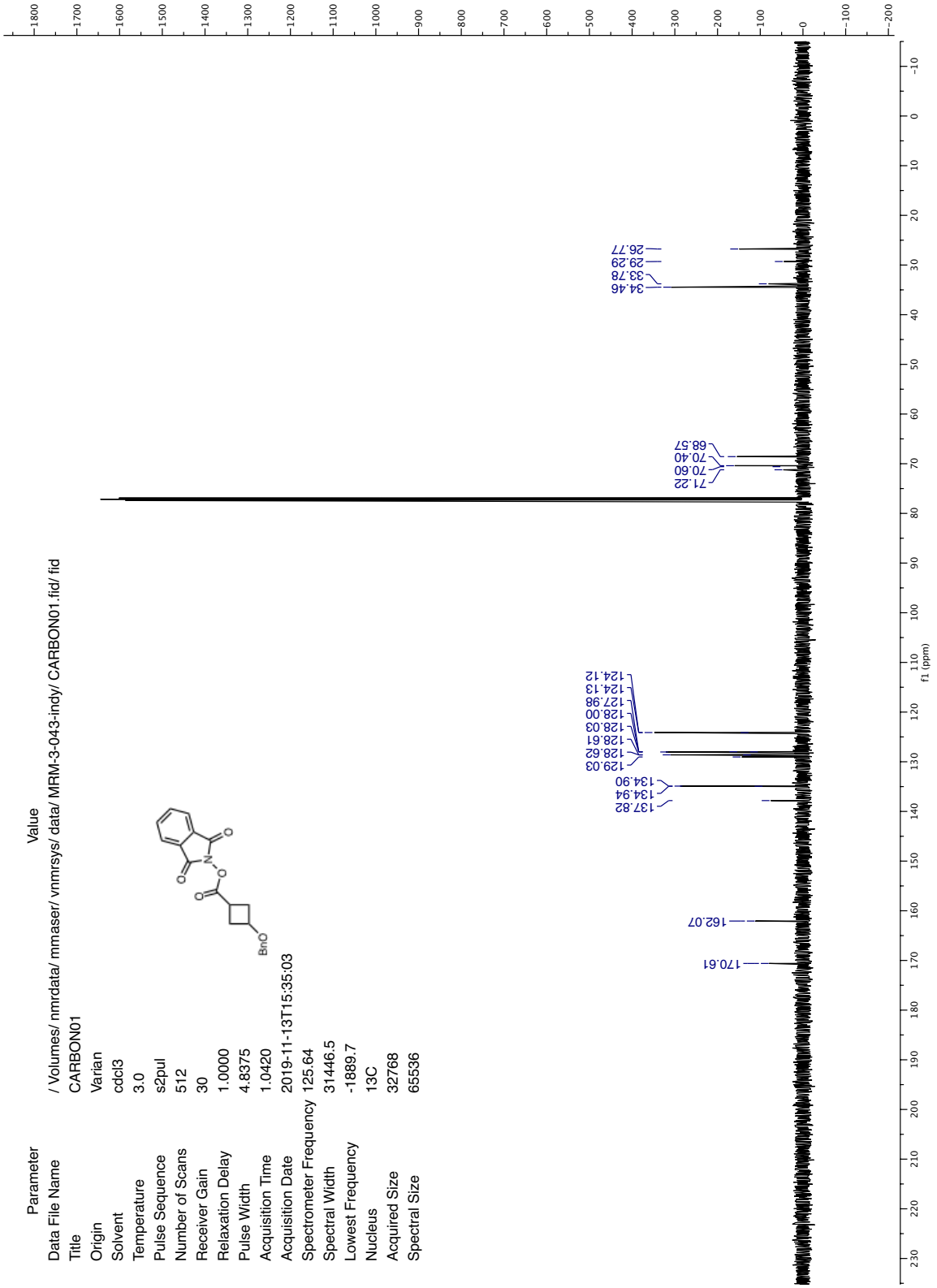
A3.7 NMR SPECTRA

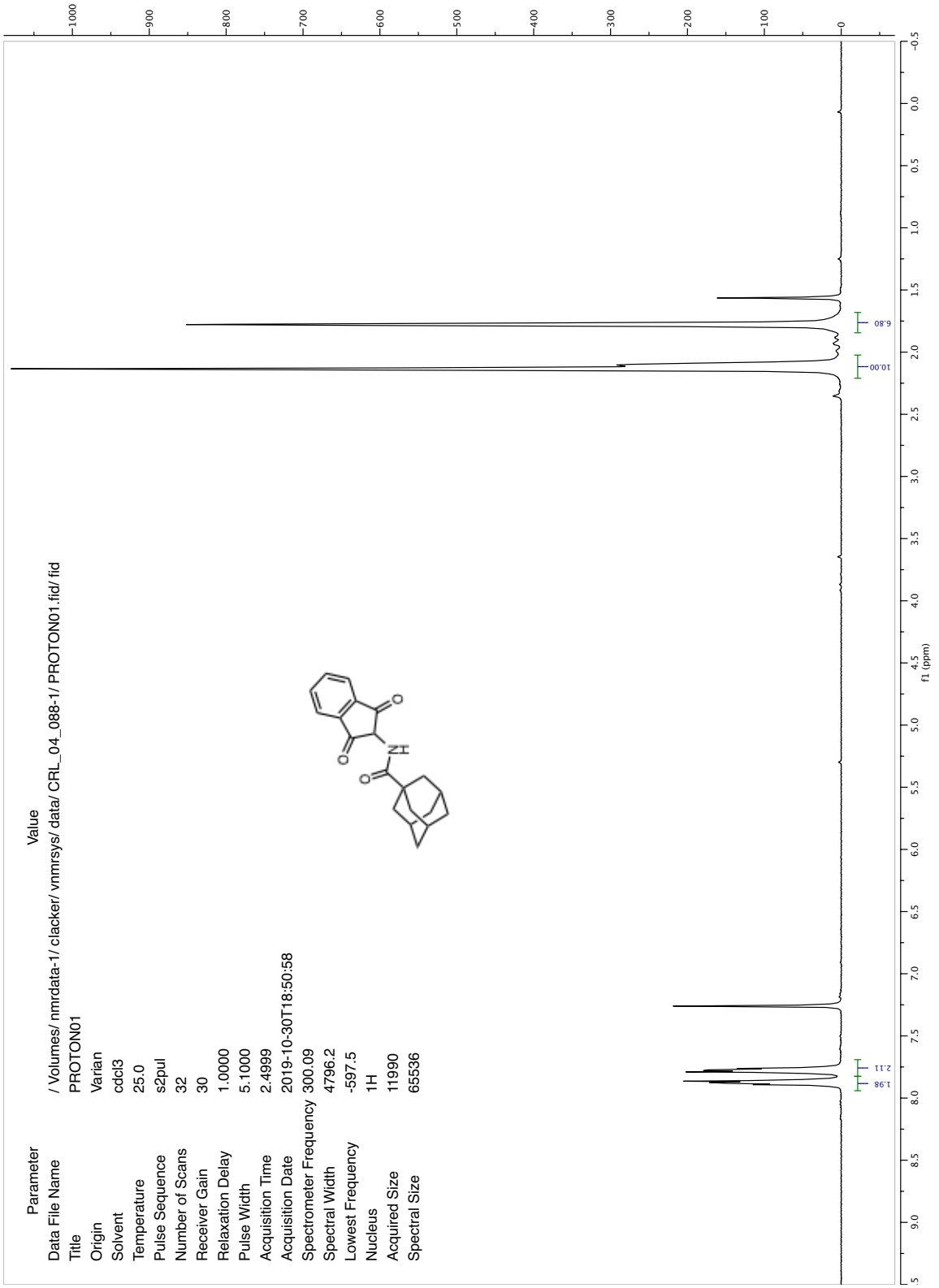

Parameter
Value
Data File Name / Volumes/ nmrdata/ clacker/ vnmrsys/ data/ CRL_04_113/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent dmso
Temperature 3.0
Pulse Sequence s2pul
Number of Scans 16
Receiver Gain 60
Relaxation Delay 1.0000
Pulse Width 5.6500
Acquisition Time 3.0000
Acquisition Date 2019-11-26T17:24:17
Spectrometer Frequency 499.62
Spectral Width 8000.0
Lowest Frequency -1026.3
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
NH2
NH2Cl
N
3.70
0.93
1.05
1.01
1.00
f1 (ppm)

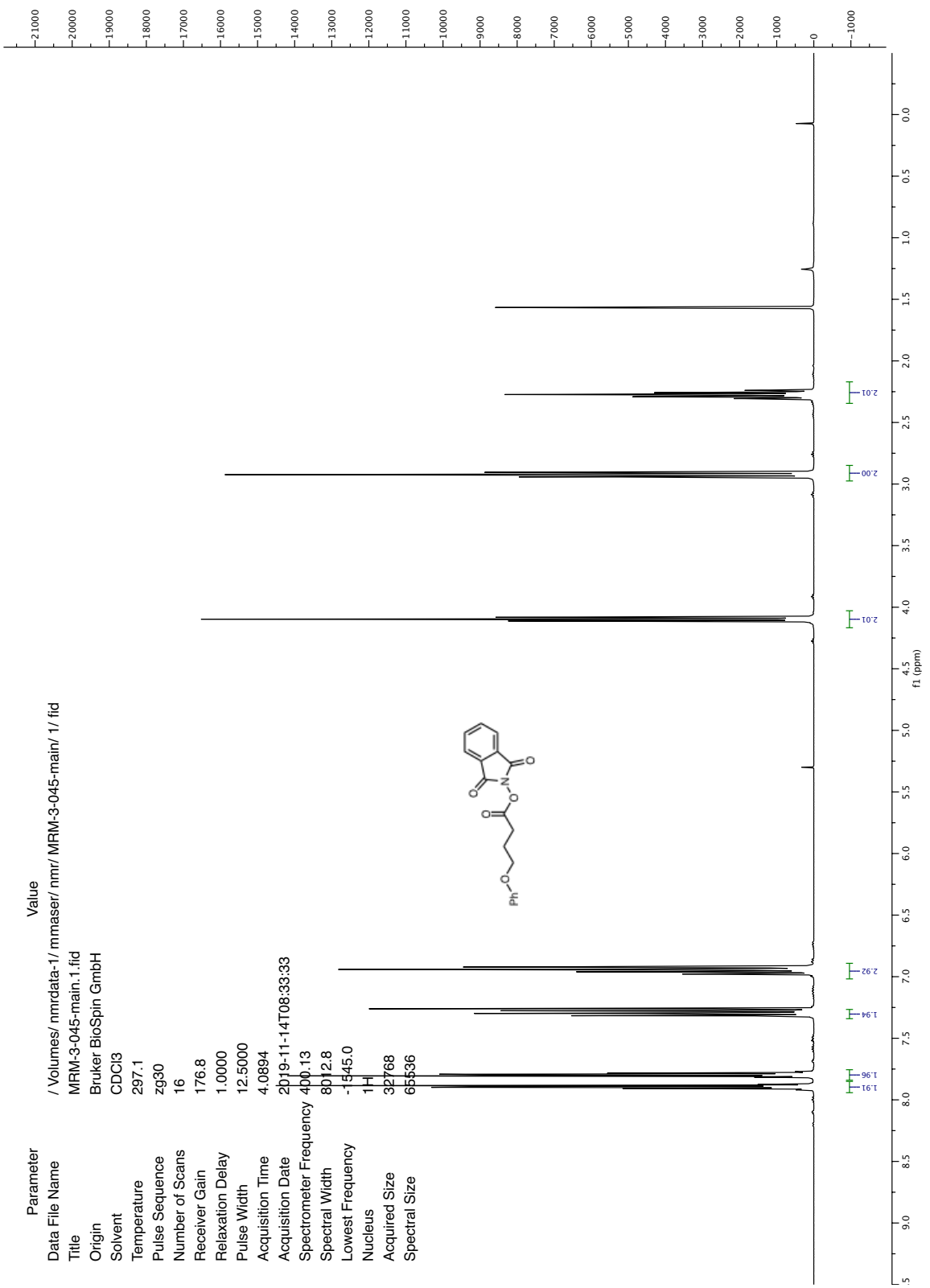

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ clacker/ vnmrsys/ data/ CRL_04_089_1/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 3.0
Pulse Sequence s2pul
Number of Scans 8
Receiver Gain 34
Relaxation Delay 1.0000
Pulse Width 5.6500
Acquisition Time 3.0000
Acquisition Date 2019-11-01T16:28:30
Spectrometer Frequency 499.61
Spectral Width 8000.0
Lowest Frequency -1030.3
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
Cbz
f1 (ppm)

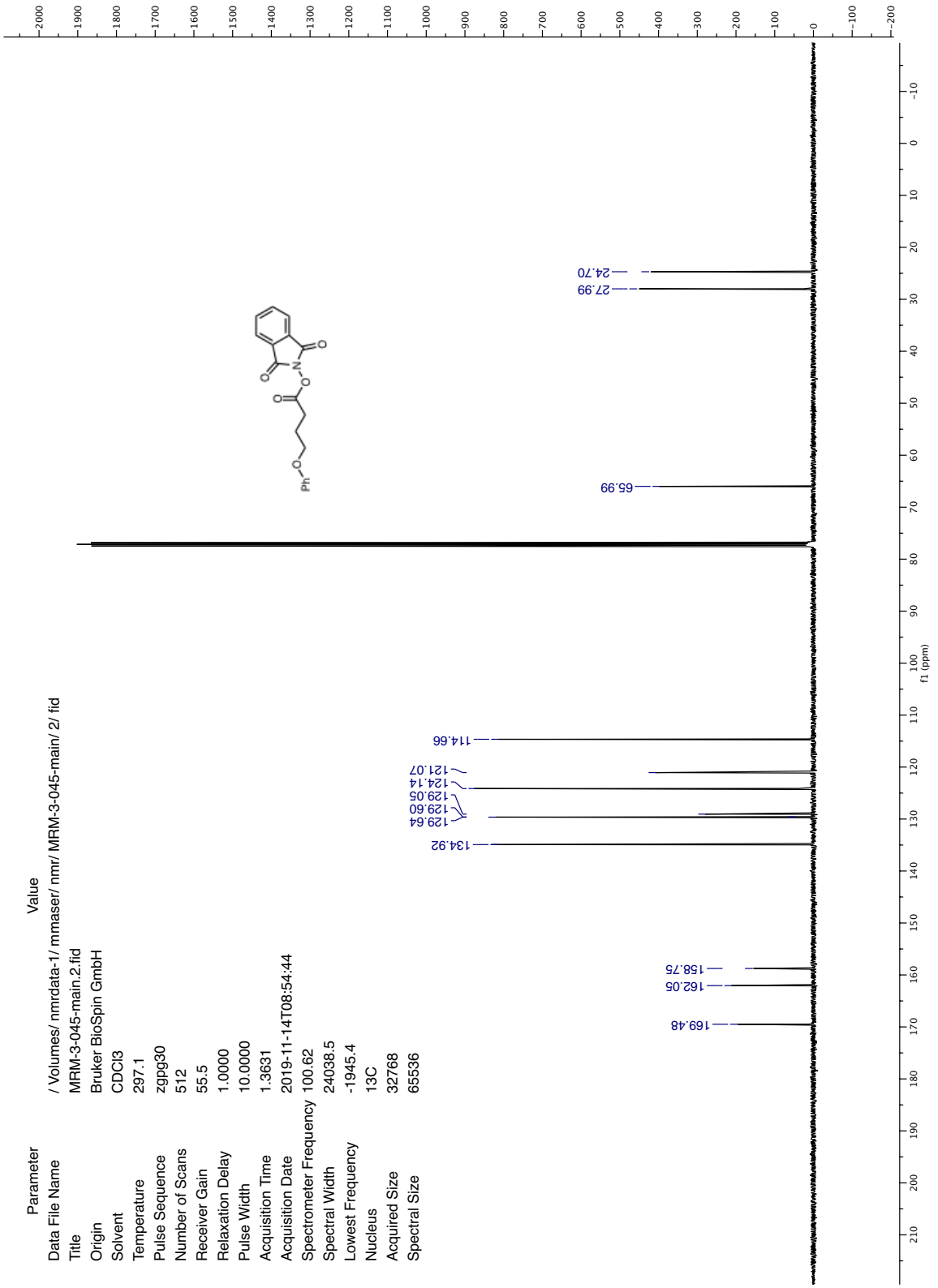

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-041-T/ 2/ fid
Title MRM-3-041-T.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 64
Receiver Gain 78.7
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-05T17:29:39
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.1
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
CbzN
2.00
0.99
2.06
2.00
2.00
4.77
1.91
2.01
f1 (ppm)

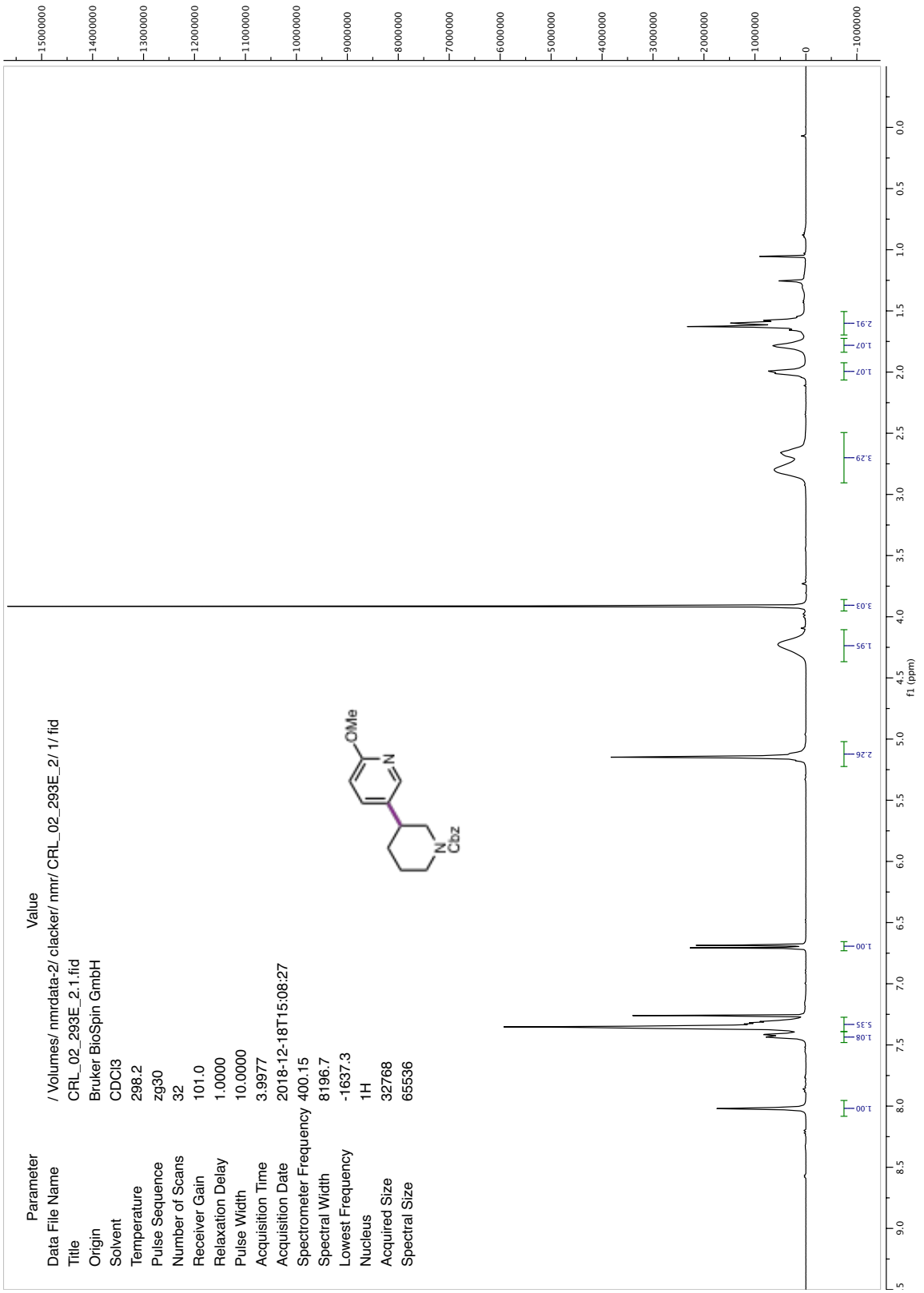

Parameter
Value
Data File Name / Volumes/ nmrdata/ mmaser/ vnmrsys/ data/ MRM-3-043-indy/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 3.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 44
Relaxation Delay 1.0000
Pulse Width 5.6500
Acquisition Time 3.0000
Acquisition Date 2019-11-13T15:32:37
Spectrometer Frequency 499.61
Spectral Width 8000.0
Lowest Frequency -1030.3
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
BnO
2.62
2.61
0.96
0.25
1.00
0.28
2.60
5.90
2.54
2.34
f1 (ppm)

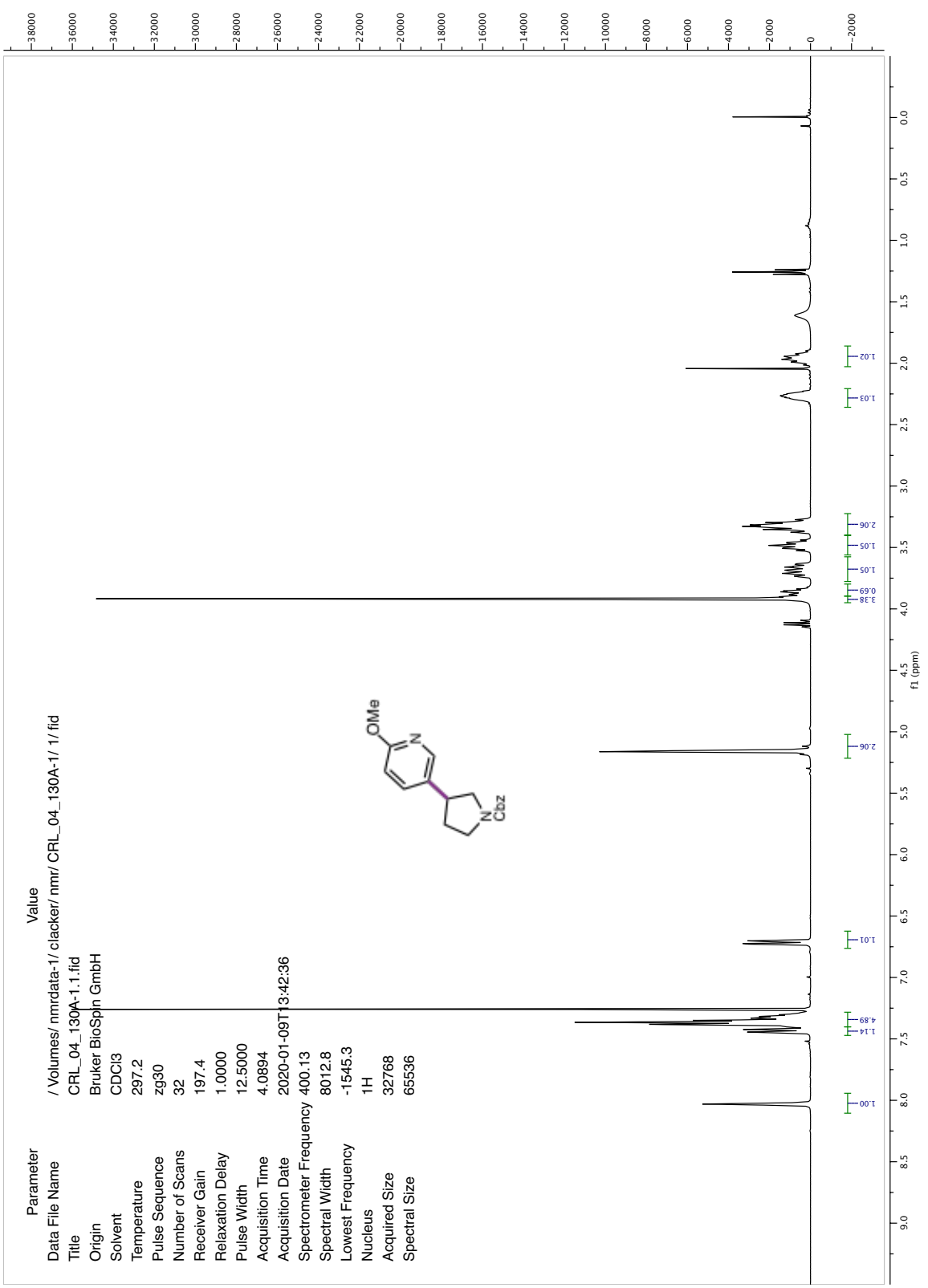

Parameter Value
Data File Name / Volumes/ nmrdata/ mmaser/ vnmrsys/ data/ MRM-3-043-indy/ CARBON01.fid/ fid
Title CARBON01
Origin Varian
Solvent cdcl3
Temperature 3.0
Pulse Sequence s2pul
Number of Scans 512
Receiver Gain 30
Relaxation Delay 1.0000
Pulse Width 4.8375
Acquisition Time 1.0420
Acquisition Date 2019-11-13T15:35:03
Spectrometer Frequency 125.64
Spectral Width 31446.5
Lowest Frequency -1889.7
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
BnO
26.77
29.29
33.78
34.46
68.57
70.40
70.60
71.22
124.12
124.13
127.98
128.00
128.03
128.61
128.62
129.03
134.90
134.94
137.82
162.07
170.61
f1 (ppm)

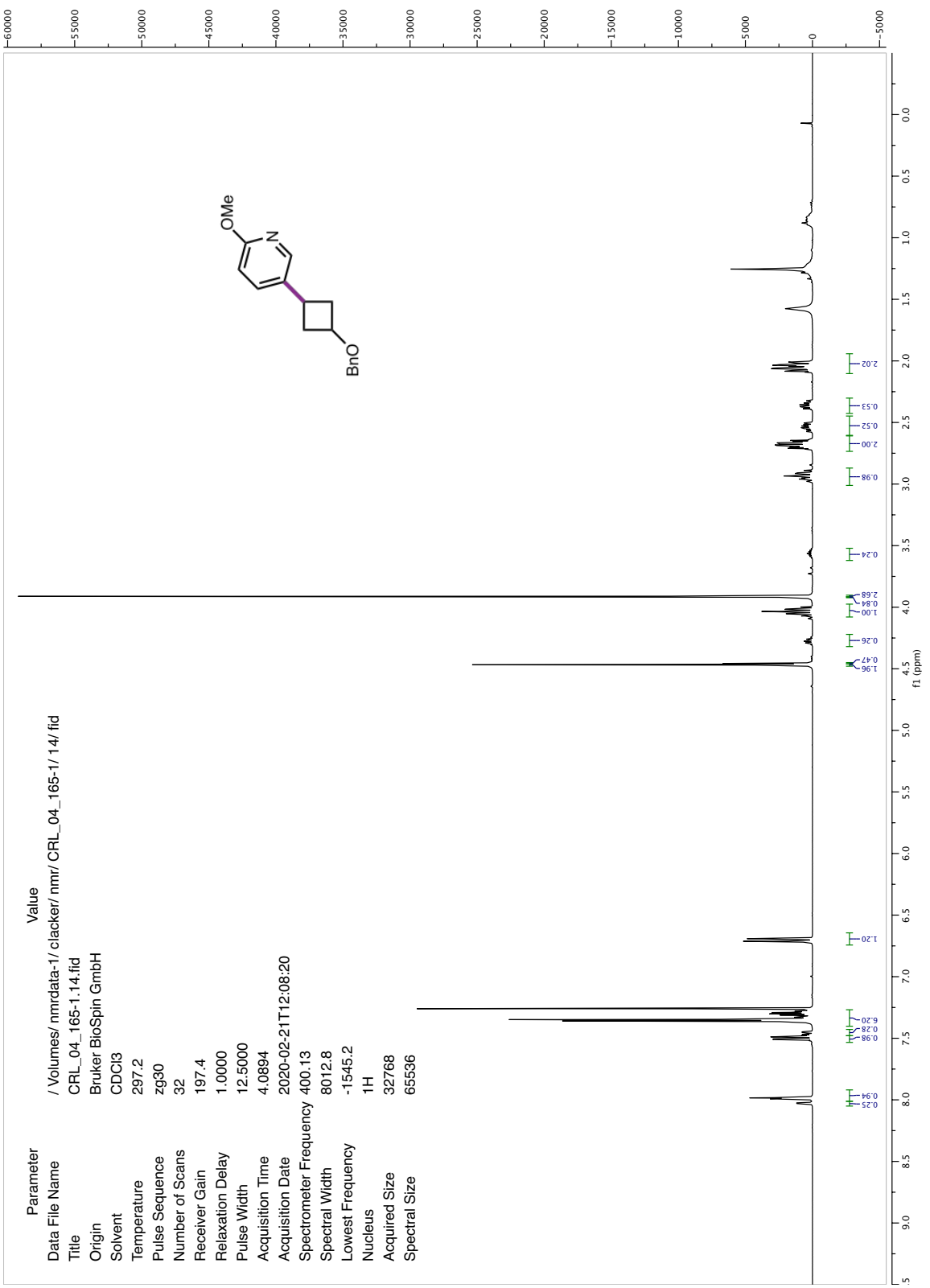

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ clacker/ vnmrsys/ data/ CRL_04_088-1/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 25.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 30
Relaxation Delay 1.0000
Pulse Width 5.1000
Acquisition Time 2.4999
Acquisition Date 2019-10-30T18:50:58
Spectrometer Frequency 300.09
Spectral Width 4796.2
Lowest Frequency -597.5
Nucleus 1H
Acquired Size 11990
Spectral Size 65536
6.80
10.00
2.11
1.98
f1 (ppm)


Parameter
Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-045-main/ 1/ fid
Title MRM-3-045-main.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 176.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-14T08:33:33
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.0
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Ph
2.01
2.00
2.01
2.92
1.94
1.96
1.91
f1 (ppm)

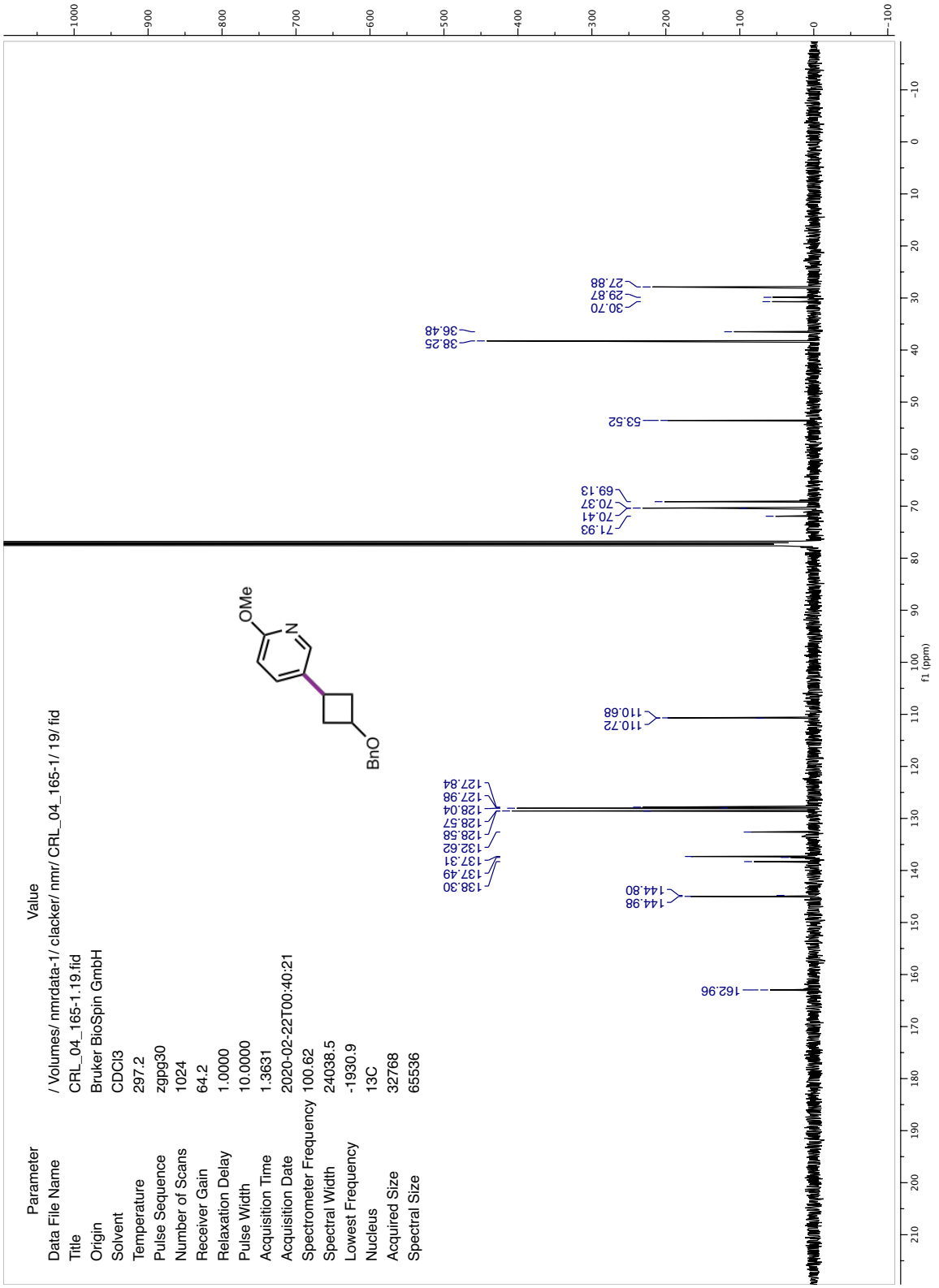

Parameter Value
Data File Name / Volumes/ nmrdata-1/ mmaser/ nmr/ MRM-3-045-main/ 2/ fid
Title MRM-3-045-main.2.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zgpg30
Number of Scans 512
Receiver Gain 55.5
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2019-11-14T08:54:44
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1945.4
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
24.70
27.99
65.99
114.66
121.07
124.14
129.05
129.60
129.64
134.92
158.75
162.05
169.48
Ph
f1 (ppm)

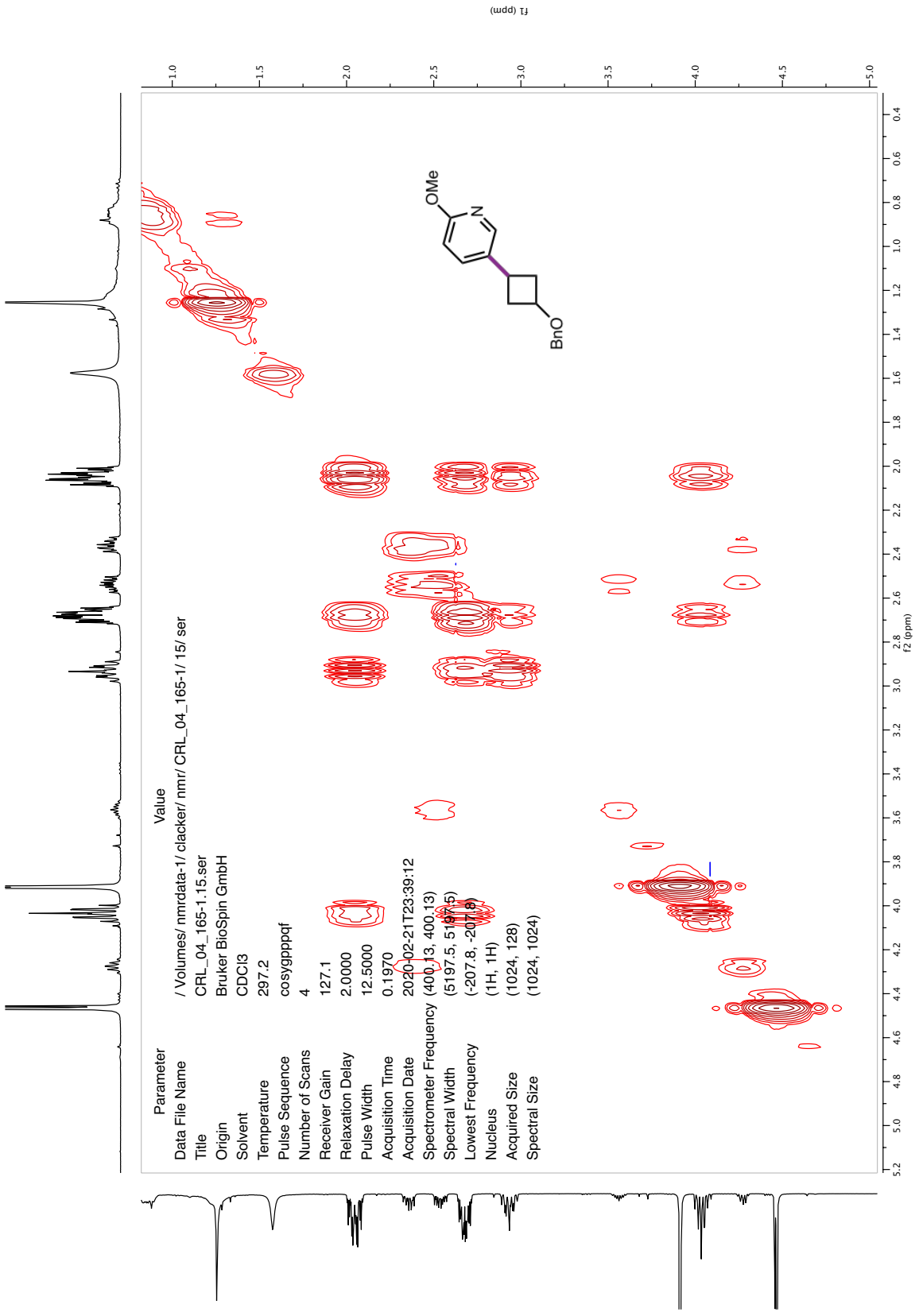

Parameter Value
Data File Name / Volumes/ nmrdata-2/ clacker/ nmr/ CRL_02_293E_2/ 1/ fid
Title CRL_02_293E_2.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 298.2
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 101.0
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 3.9977
Acquisition Date 2018-12-18T15:08:27
Spectrometer Frequency 400.15
Spectral Width 8196.7
Lowest Frequency -1637.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OMe
N
N
Cbz
1.00
1.08
5.35
1.00
2.26
1.95
3.03
3.29
1.07
1.07
2.91
f1 (ppm)

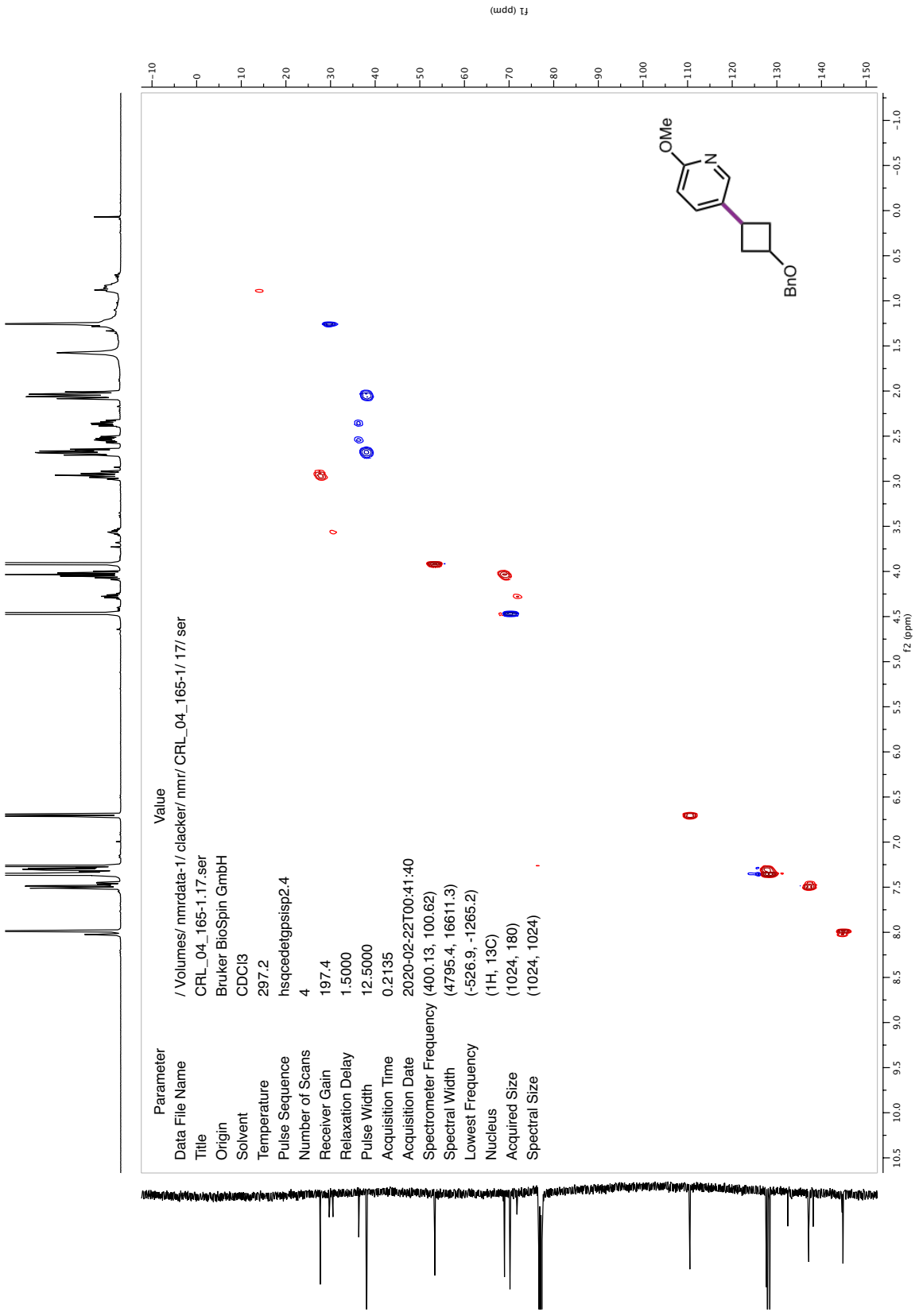

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ clacker/ nmr/ CRL_04_130A-1/ 1/ fid
Title CRL_04_130A-1.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-01-09T13:42:36
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OMe
N
N
Cbz
f1 (ppm)

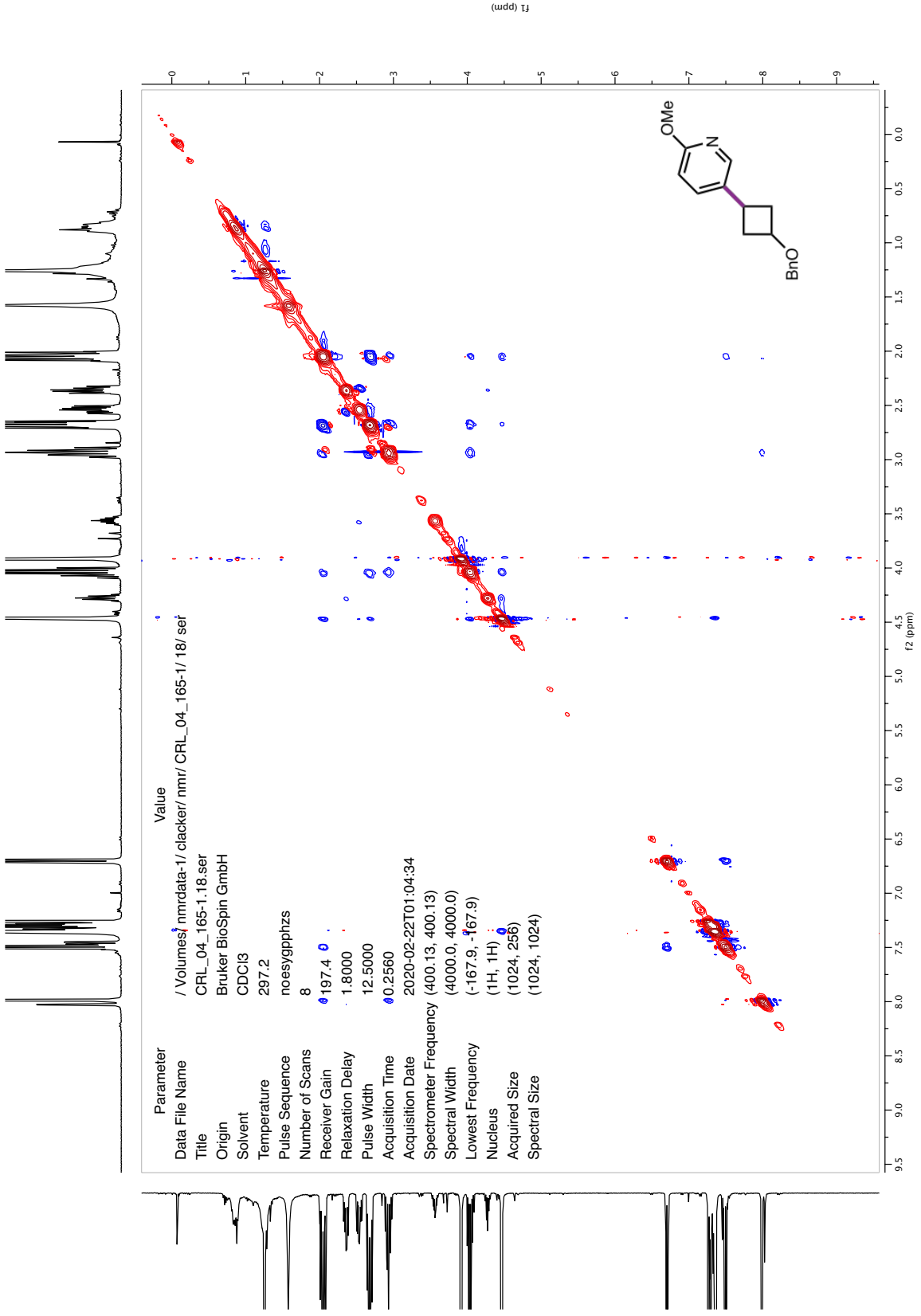


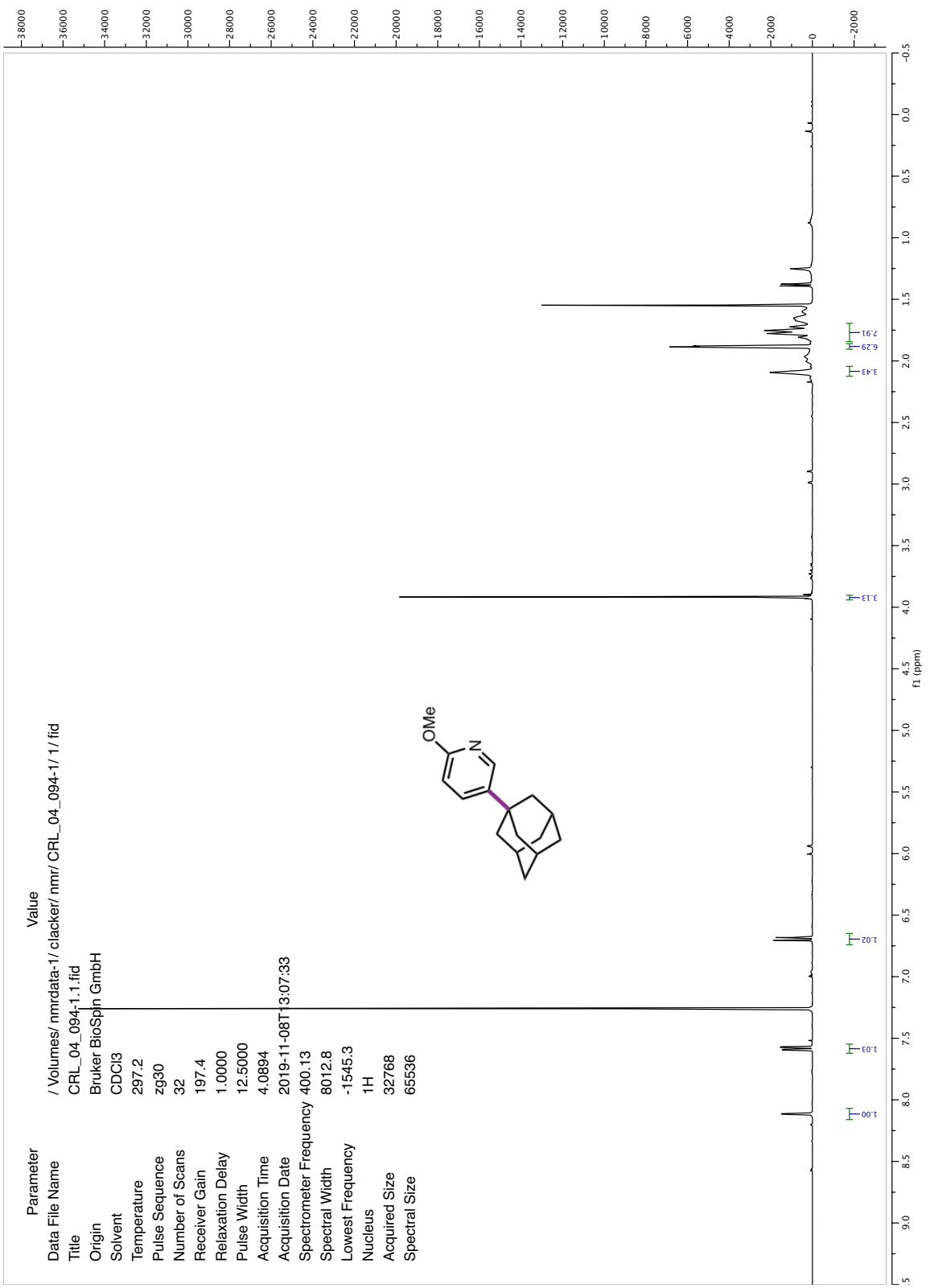
Parameter	Value
Data File Name	/ Volumes/ nmrdata-1/ clacker/ nmr/ CRL_04_165-1/ 14/ fid
Title	CRL_04_165-1.14.fid
Origin	Bruker BioSpin GmbH
Solvent	CDCl3
Temperature	297.2
Pulse Sequence	zg30
Number of Scans	32
Receiver Gain	197.4
Relaxation Delay	1.0000
Pulse Width	12.5000
Acquisition Time	4.0894
Acquisition Date	2020-02-21T12:08:20
Spectrometer Frequency	400.13
Spectral Width	8012.8
Lowest Frequency	-1545.2
Nucleus	1H
Acquired Size	32768
Spectral Size	65536

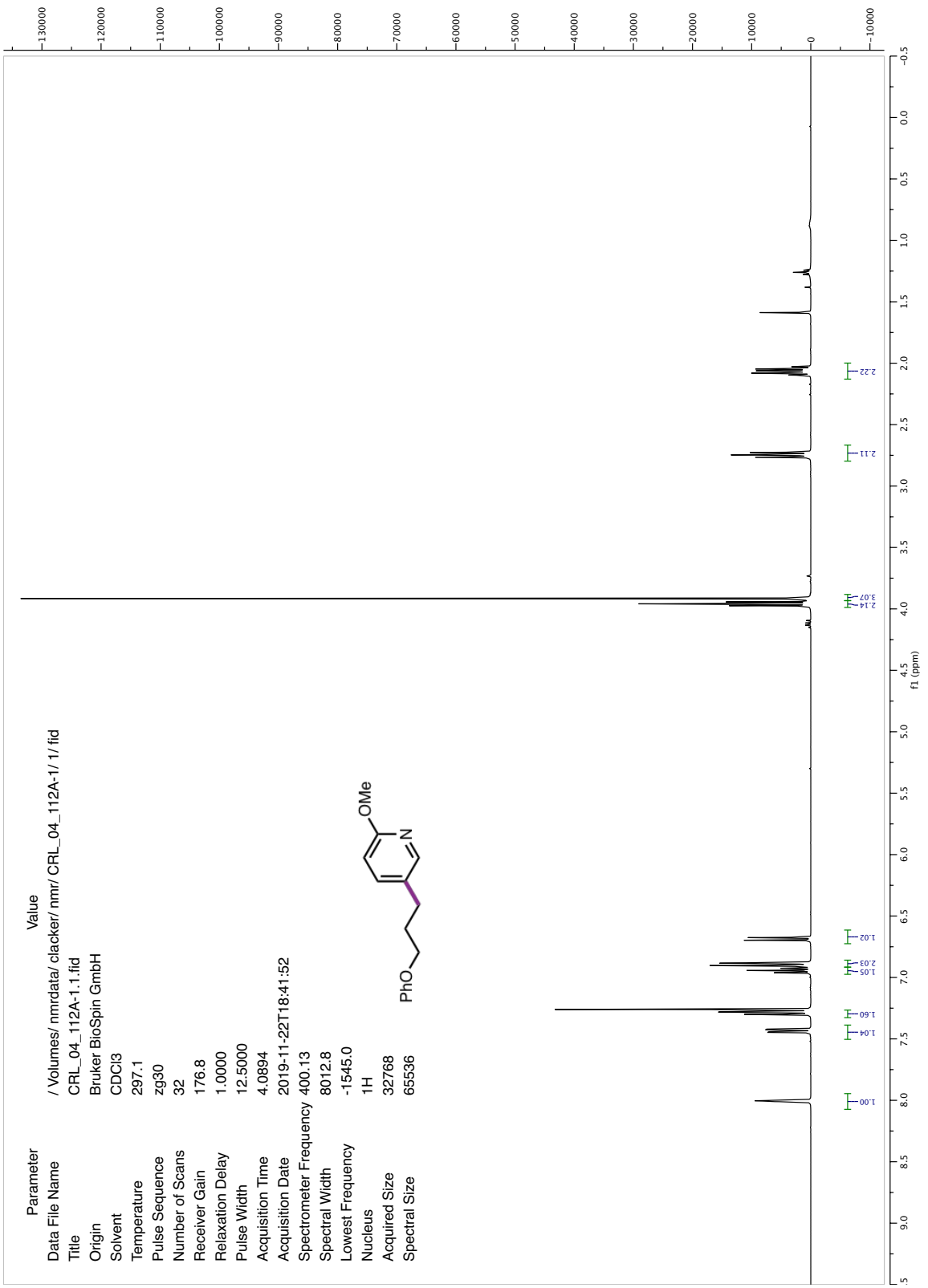

Parameter Value
Data File Name / Volumes/ nmrdata-1/ clacker/ nmr/ CRL_04_165-1/ 19/ fid
Title CRL_04_165-1.19.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zgpg30
Number of Scans 1024
Receiver Gain 64.2
Relaxation Delay 1.0000
Pulse Width 10.0000
Acquisition Time 1.3631
Acquisition Date 2020-02-22T00:40:21
Spectrometer Frequency 100.62
Spectral Width 24038.5
Lowest Frequency -1930.9
Nucleus 13C
Acquired Size 32768
Spectral Size 65536
OMe
N
BnO
162.96
144.98
144.80
138.30
137.49
137.31
132.62
128.58
128.57
128.04
127.98
127.84
110.72
110.68
71.93
70.41
70.37
69.13
53.52
38.25
36.48
30.70
29.87
27.88
f1 (ppm)

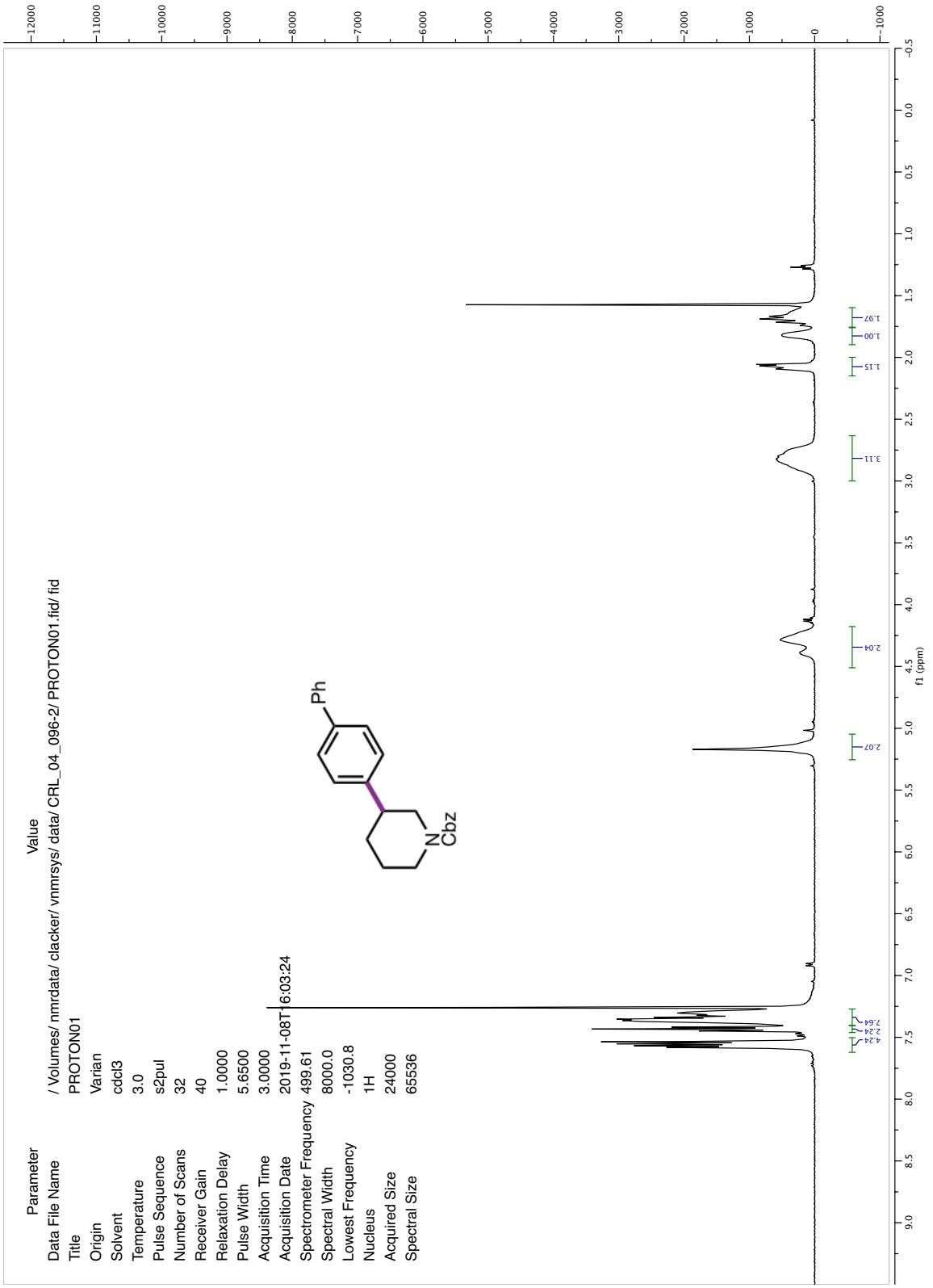

f1 (ppm)
f2 (ppm)
OMe
N
BnO
Parameter
Value
Data File Name
/ Volumes/ nmrdata-1/ clacker/ nmr/ CRL_04_165-1/ 15/ ser
Title
CRL_04_165-1.15.ser
Origin
Bruker BioSpin GmbH
Solvent
CDCl3
Temperature
297.2
Pulse Sequence
cosygpppqf
Number of Scans
4
Receiver Gain
127.1
Relaxation Delay
2.0000
Pulse Width
12.5000
Acquisition Time
0.1970
Acquisition Date
2020-02-21T23:39:12
Spectrometer Frequency
(400.13, 400.13)
Spectral Width
(5197.5, 5197.5)
Lowest Frequency
(-207.8, -207.8)
Nucleus
(1H, 1H)
Acquired Size
(1024, 128)
Spectral Size
(1024, 1024)

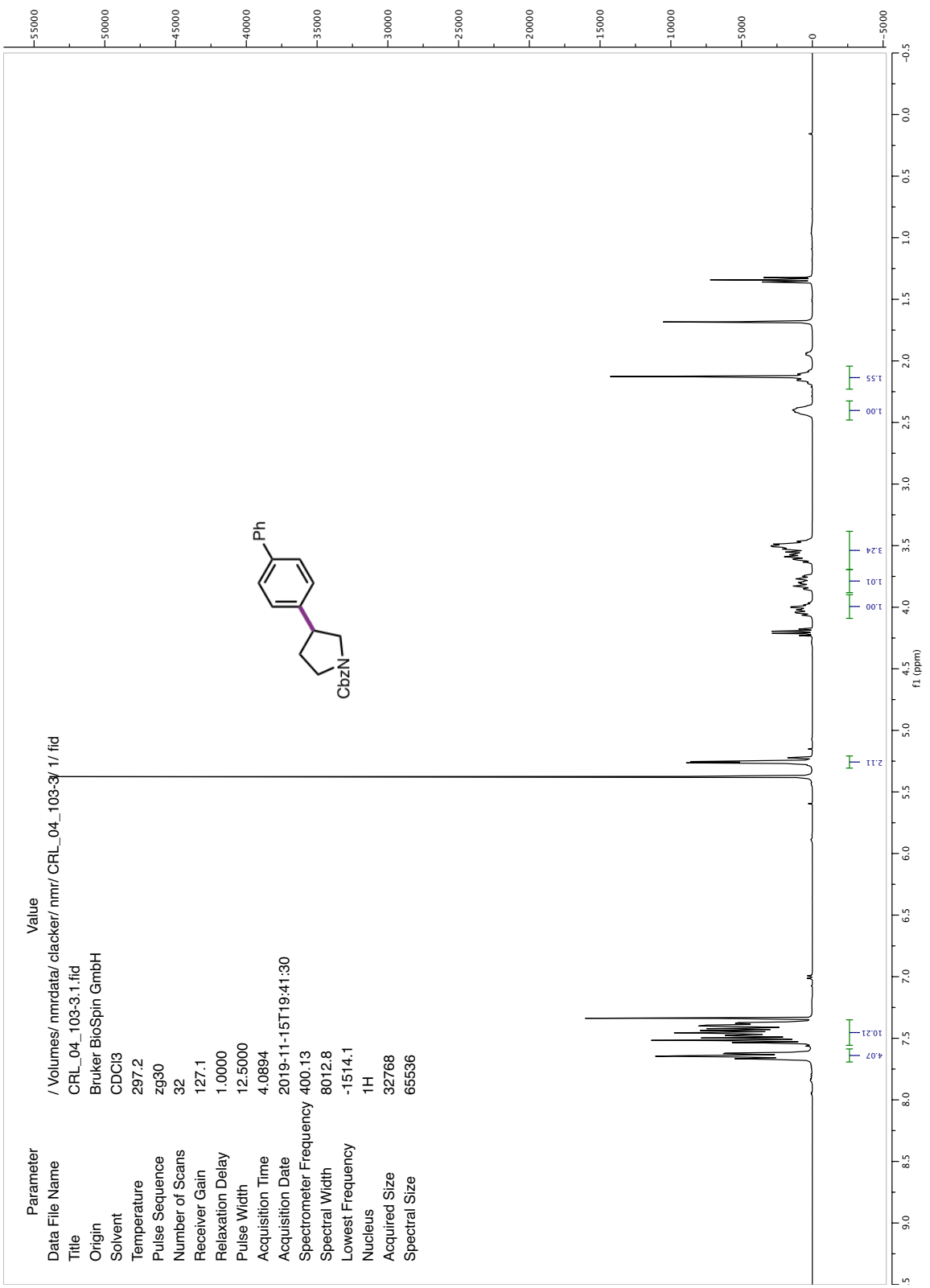

f1 (ppm)
f2 (ppm)
OMe
N
BnO
Parameter Value
Data File Name / Volumes/ nmrdata-1/ clacker/ nmr/ CRL_04_165-1/ 17/ ser
Title CRL_04_165-1.17.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence hsqcedetgpsisp2.4
Number of Scans 4
Receiver Gain 197.4
Relaxation Delay 1.5000
Pulse Width 12.5000
Acquisition Time 0.2135
Acquisition Date 2020-02-22T00:41:40
Spectrometer Frequency (400.13, 100.62)
Spectral Width (4795.4, 16611.3)
Lowest Frequency (-526.9, -1265.2)
Nucleus (1H, 13C)
Acquired Size (1024, 180)
Spectral Size (1024, 1024)

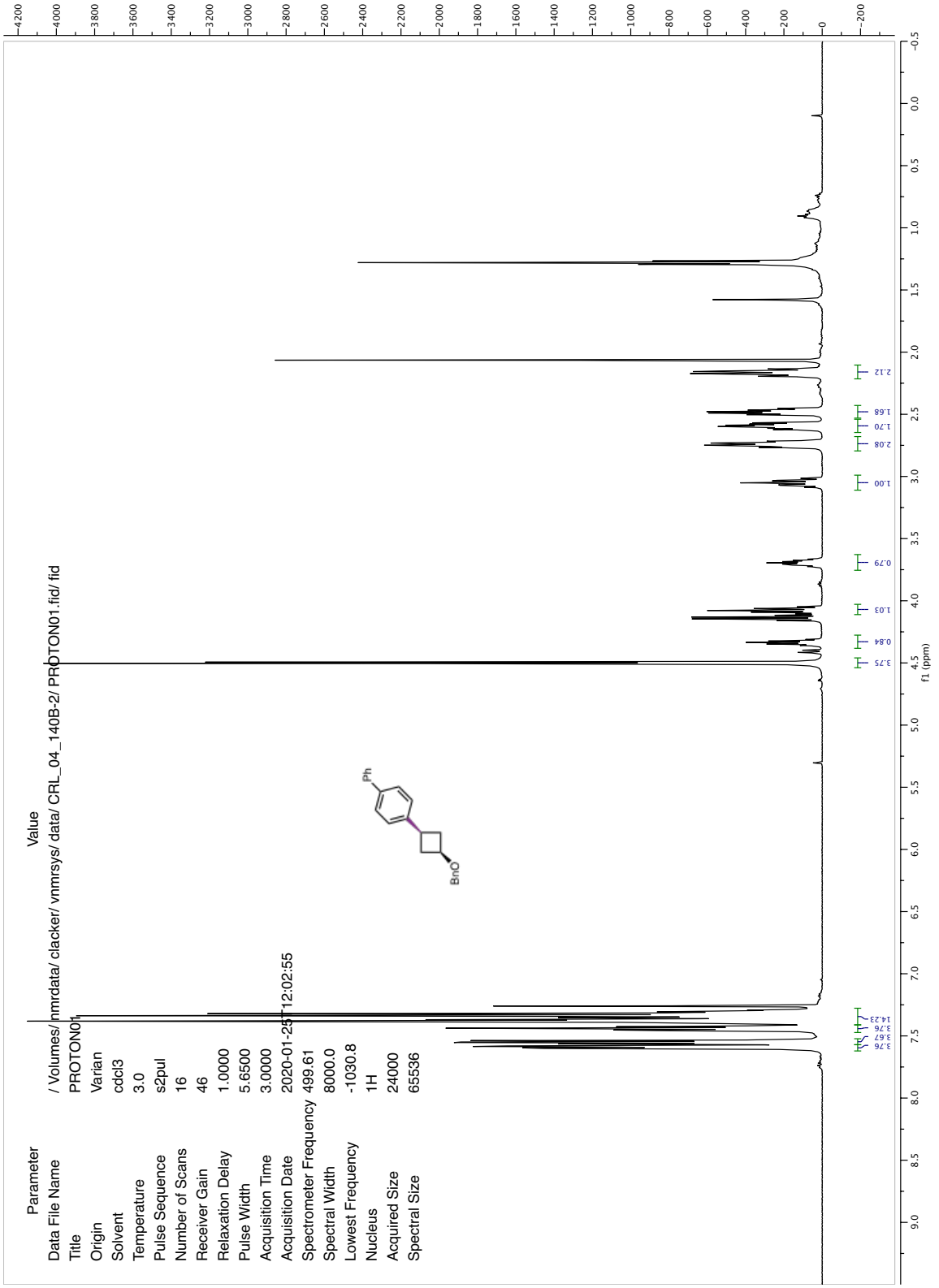

f1 (ppm)
f2 (ppm)
OMe
N
BnO
Parameter Value
Data File Name / Volumes/ nmrdata-1/ clacker/ nmr/ CRL_04_165-1/ 18/ ser
Title CRL_04_165-1.18.ser
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence noesygpphzs
Number of Scans 8
Receiver Gain 197.4
Relaxation Delay 1.8000
Pulse Width 12.5000
Acquisition Time 0.2560
Acquisition Date 2020-02-22T01:04:34
Spectrometer Frequency (400.13, 400.13)
Spectral Width (4000.0, 4000.0)
Lowest Frequency (-167.9, -167.9)
Nucleus (1H, 1H)
Acquired Size (1024, 256)
Spectral Size (1024, 1024)

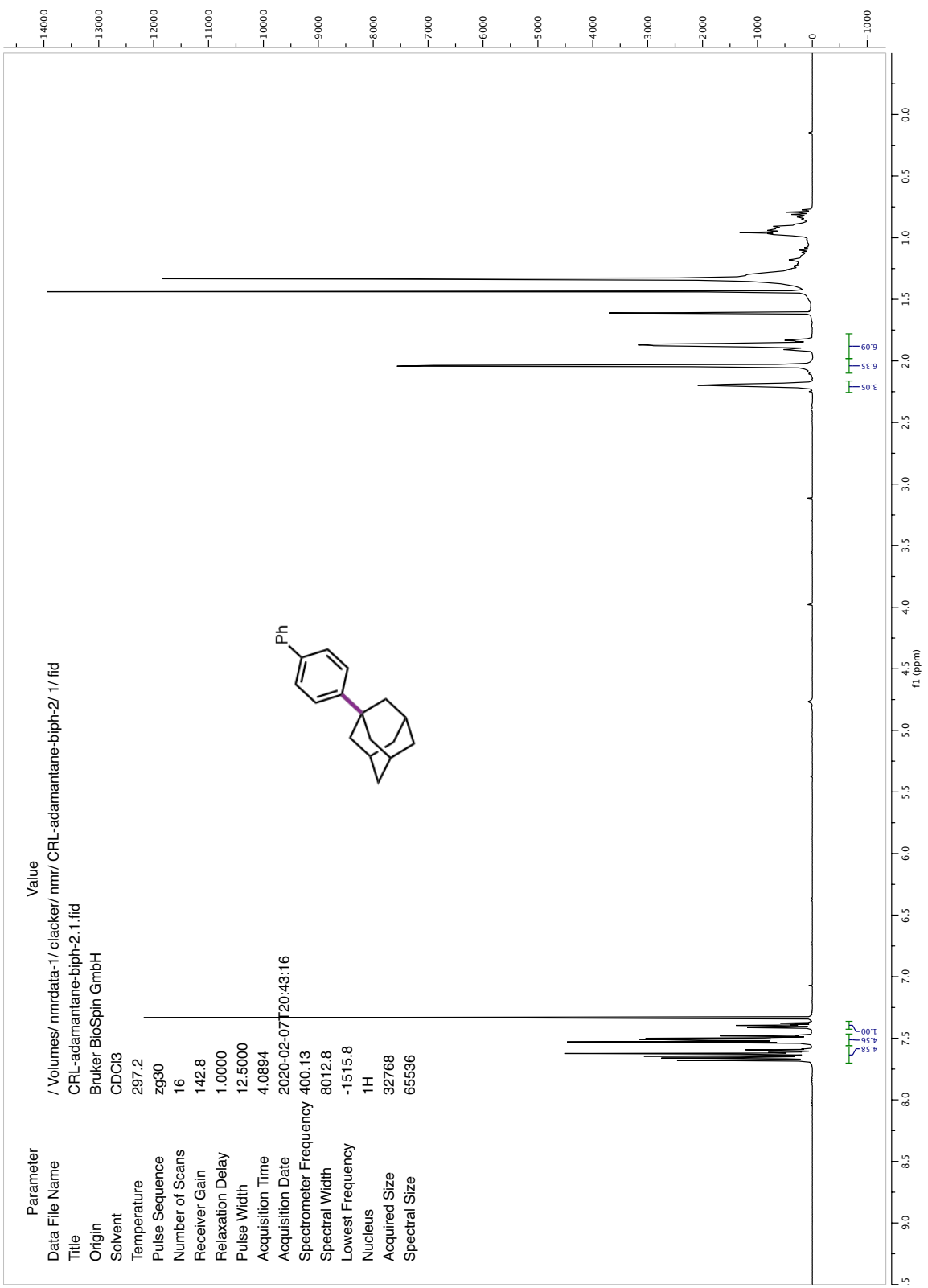

Parameter
Value
Data File Name / Volumes/ nmrdata-1/ clacker/ nmr/ CRL_04_094-1/ 1/ fid
Title CRL_04_094-1.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 197.4
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-08T13:07:33
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.3
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OMe
N
1.00
1.03
1.02
3.13
3.43
6.29
7.91
f1 (ppm)

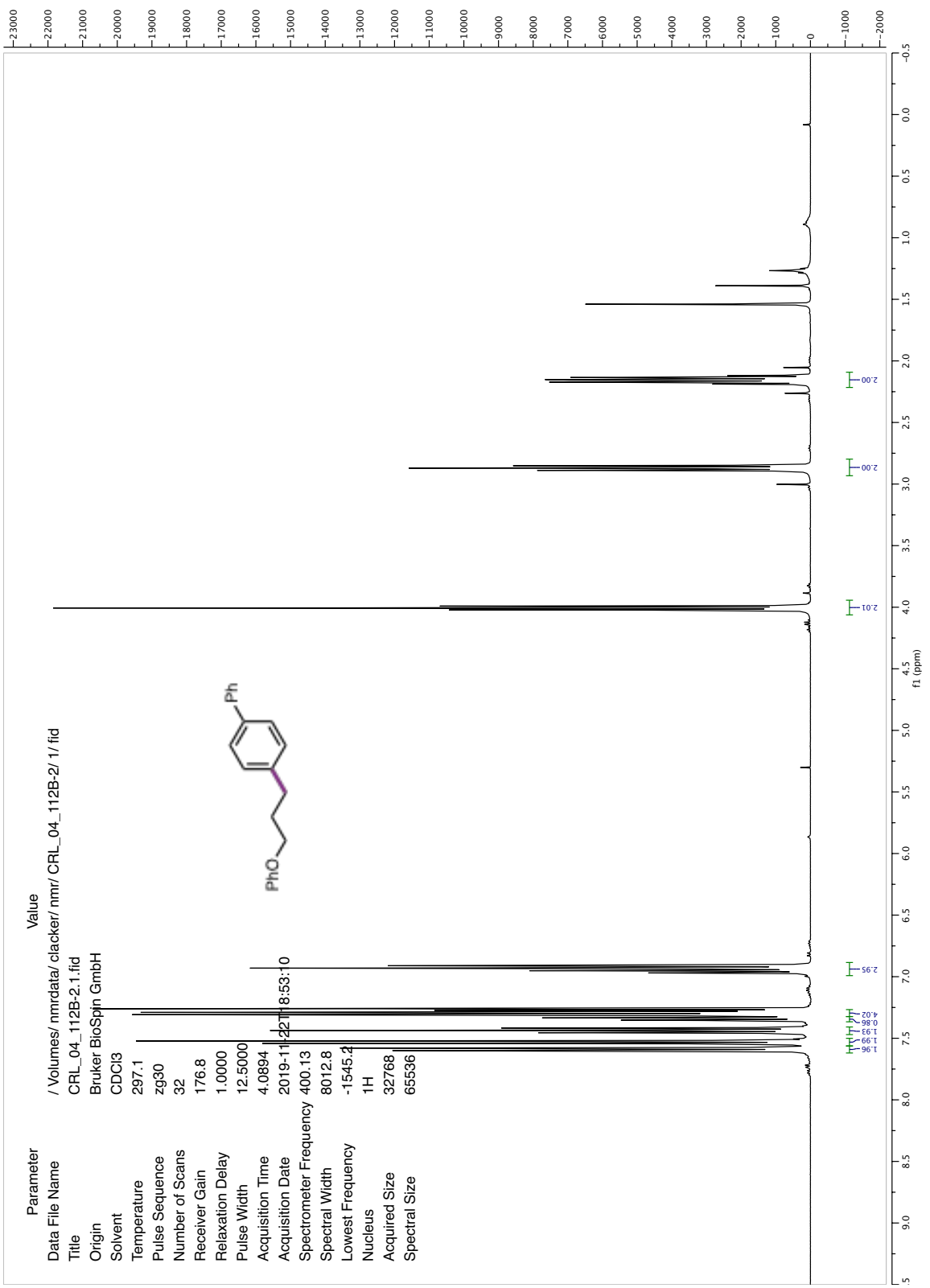

Parameter
Value
Data File Name / Volumes/ nmrdata/ clacker/ nmr/ CRL_04_112A-1/ 1/ fid
Title CRL_04_112A-1.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 176.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-22T18:41:52
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.0
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
OMe
N
PhO
2.22
2.11
3.07
2.14
1.02
1.05
2.03
1.60
1.04
1.00
f1 (ppm)


Parameter
Value
Data File Name / Volumes/ nmrdata/ clacker/ vnmrsys/ data/ CRL_04_096-2/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 3.0
Pulse Sequence s2pul
Number of Scans 32
Receiver Gain 40
Relaxation Delay 1.0000
Pulse Width 5.6500
Acquisition Time 3.0000
Acquisition Date 2019-11-08T16:03:24
Spectrometer Frequency 499.61
Spectral Width 8000.0
Lowest Frequency -1030.8
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
Ph
N
Cbz
1.97
1.00
1.15
3.11
2.04
2.07
7.64
2.24
4.24
f1 (ppm)


Parameter
Value
Data File Name / Volumes/ nmrdata/ clacker/ nmr/ CRL_04_103-3/ 1/ fid
Title CRL_04_103-3.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 127.1
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-15T19:41:30
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1514.1
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Ph
CbzN
1.55
1.00
3.24
1.01
1.00
2.11
10.21
4.07
f1 (ppm)


Parameter
Value
Data File Name / Volumes/ nmrdata/ clacker/ vnmrsys/ data/ CRL_04_140B-2/ PROTON01.fid/ fid
Title PROTON01
Origin Varian
Solvent cdcl3
Temperature 3.0
Pulse Sequence s2pul
Number of Scans 16
Receiver Gain 46
Relaxation Delay 1.0000
Pulse Width 5.6500
Acquisition Time 3.0000
Acquisition Date 2020-01-25T12:02:55
Spectrometer Frequency 499.61
Spectral Width 8000.0
Lowest Frequency -1030.8
Nucleus 1H
Acquired Size 24000
Spectral Size 65536
Ph
BnO
f1 (ppm)


Parameter
Value
Data File Name / Volumes/ nmrdata-1/ clacker/ nmr/ CRL-adamantane-biph-2/ 1/ fid
Title CRL-adamantane-biph-2.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.2
Pulse Sequence zg30
Number of Scans 16
Receiver Gain 142.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2020-02-07T20:43:16
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1515.8
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Ph
6.09
6.35
3.05
4.58
4.56
1.00
f1 (ppm)


Parameter
Value
Data File Name / Volumes/ nmrdata/ clacker/ nmr/ CRL_04_112B-2/ 1/ fid
Title CRL_04_112B-2.1.fid
Origin Bruker BioSpin GmbH
Solvent CDCl3
Temperature 297.1
Pulse Sequence zg30
Number of Scans 32
Receiver Gain 176.8
Relaxation Delay 1.0000
Pulse Width 12.5000
Acquisition Time 4.0894
Acquisition Date 2019-11-22T18:53:10
Spectrometer Frequency 400.13
Spectral Width 8012.8
Lowest Frequency -1545.2
Nucleus 1H
Acquired Size 32768
Spectral Size 65536
Ph
PhO
2.00
2.00
2.01
2.95
1.96
1.99
0.86
4.02
f1 (ppm)

Appendix 4

The Open Reaction Database[1]

A4.1 BACKGROUND

Section 3.2 details challenges faced in predicting experimental reaction yields from literature datasets and Appendix 3 details our attempts to address the challenges by generation of in-house high-throughput datasets. Another direction for addressing these challenges is at the source of literature data. Our previous studies utilized reaction data stored in commercial databases such as Reaxys®. This data is often fraught with recording mistakes (such as reagent mis-spellings) and missing critical information (such as reaction temperatures or yields). Our previous work thus required major efforts in data curation to refine initial datasets and retain only reliable records. Then still, significant pre-processing protocols were devised in order to

[1] Portions of this appendix were adapted from the following communications: Kearnes, S. M.; Maser, M. R.; Wleklinski, M.; Kast, A.; Doyle, A. G.; Dreher, S. D.; Hawkins, J. M.; Jensen, K. F.; Coley, C. W. The Open Reaction Database. *J. Am. Chem. Soc.* **2021**, *143*, 18820. Copyright © 2021 American Chemical Society. The research discussed in this appendix was completed in collaboration with Dr. Steven M. Kearnes of Relay Therapeutics, Dr. Michael Wleklinski of Merck & Co., Dr. Anton Kast of Google Research, Prof. Abigail G. Doyle of UCLA, Dr. Spencer D. Dreher of Merck & Co., Dr. Joel M. Hawkins of Pfizer Inc., and Profs. Klavs F. Jensen and Connor W. Coley of MIT. We warmly acknowledge all of the respondents to our original use case survey in 2019, the full advisory board, and all of the direct contributors to `ord-data`. We also thank Nathan Kim for helping with infrastructure development, Brian Lee for helping with feedback and beta testing, and Devin Sandberg and Zan Armstrong for design advice. Cloud computing resources and legal support were provided by Google. Additional GitHub data storage and bandwidth were provided through the GitHub Education program.

format such reaction data into digital representations usable for ML studies. Finally, due to copyright issues, our resulting datasets and trained models could not be made public for the community to use. Resulting from this paradigm is the inability to share research and tools with the broader community as well as a general lack of reproducibility in such studies.

Seeking to overcome these limitations, we joined an effort initiated at Google Research and MIT to construct The Open Reaction Database (ORD). The ORD provides a flexible but standardized digital format for defining and housing recorded reaction data. The result is a highly useful data curation and serving system that can be directly integrated with python ML workflows such as ours and is freely available to the world, thus overcoming the pitfalls of commercial data systems discussed above. What follows is a culminative description of the effort to design and construct the ORD software, to which I contributed functional code, datasets, stress testing, and chemical/ML expertise. That said, I wish to make clear that the bulk of the design was first completed by Dr. Steven Kearnes and Prof. Connor Coley, to whom I am grateful for the opportunity to learn and collaborate in building this tool for the community.

A4.2 INTRODUCTION

The opportunity to learn complex patterns of chemical reactivity from organic reaction data is increasingly clear. Data-driven machine-learning models have been designed and applied to planning synthetic pathways, recommending reaction conditions for each putative transformation, and even predicting what the major products of a yet-untested reaction might be, e.g., for impurity prediction. These and other tasks in “predictive synthesis” promise to streamline chemical development at stages where the synthetic route development occurs and eventually enable automated multistep synthesis.

Chemical reaction data support these modeling efforts by capturing the details of how an experiment was performed and its outcome. There are few curated databases that record organic reactions and even fewer that are generally available to researchers, even commercially. Only one prominent data set—a set of reactions extracted from the USPTO—is open access.[1]

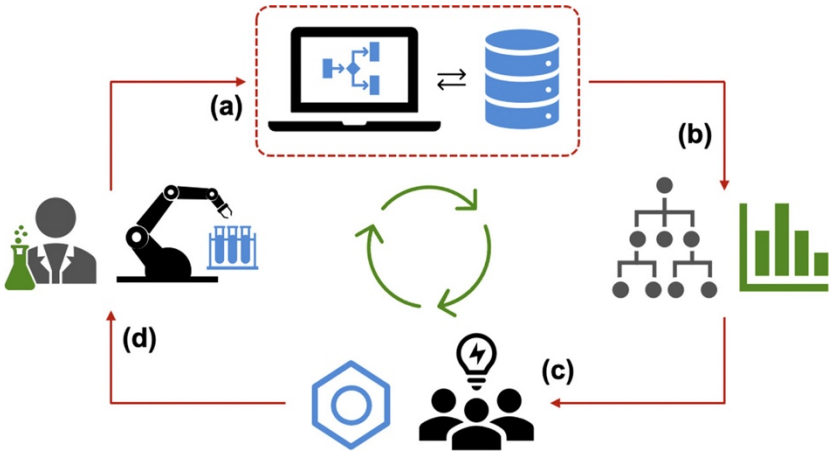
Beyond their role in supporting data-driven models, tabulated reaction data have become an indispensable tool in the workflow of nearly every chemist, if only for information retrieval. Few practicing chemists approach the synthesis of a new compound without performing a database or literature search (e.g., through SciFinder or Reaxys) and referring to procedural details in a precedent article or patent describing a relevant synthesis. While these databases tabulate and make searchable many important aspects of a chemical reaction such as the structures of reactants, agents, and products, full procedural details are left as unstructured text in the original document.

The Open Reaction Database[29] is an initiative to support machine learning and related efforts in reaction prediction, chemical synthesis planning, and experimental design (**Figure A4.1**). Our goals are to provide a structured data format for chemical reaction data, to make data freely and publicly available, to encourage sharing of precompetitive proprietary data, and to provide an interface for browsing/downloading that data. One key use case is the standardization and sharing of high-throughput reaction screening data. Underlying this effort are the FAIR principles for scientific data management.[2]

We have defined a reaction schema that provides a thorough coverage of experimental details that we know are important for reproducibility. Importantly, we capture the most essential information in a structured format, rather than an unstructured text format as is

currently used in publications (e.g., as Supporting Information). Analytical data, both raw and processed, can be directly coupled to experimental outcomes. Reaction data can be recorded programmatically (e.g., using Python) or using an interactive web editor that is more approachable for those without coding experience. A templating mechanism for data set enumeration allows one reaction entry to be rapidly expanded into hundreds or thousands, as in the case of a high-throughput screen.

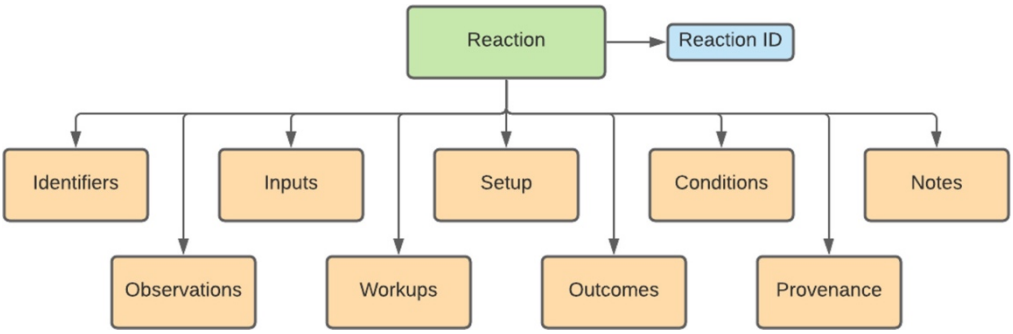
Figure A4.1 *Computer-aided chemical discovery cycle: (a) the Open Reaction Database; (b) machine learning and cheminformatics; (C) human or automated interpretation and material design; (d) manual or robotic chemical synthesis.*



A4.3 SCHEMA

The ORD schema contains structured and unstructured (free text) fields for documenting chemical reactions. It is intended to be descriptive, rather than prescriptive; each `Reaction` record in the database should describe what was actually done in the lab, and not an idealized protocol or instruction set for, e.g., automated synthesis hardware (although this may also be included). The schema is implemented using Protocol Buffers,[30] which have distinct usability advantages over other common technologies such as XML.

Figure A4.2 *Overview of the Open Reaction Database schema. Each* `Reaction` *record contains sections for reaction identifiers, inputs, outcomes, etc.*

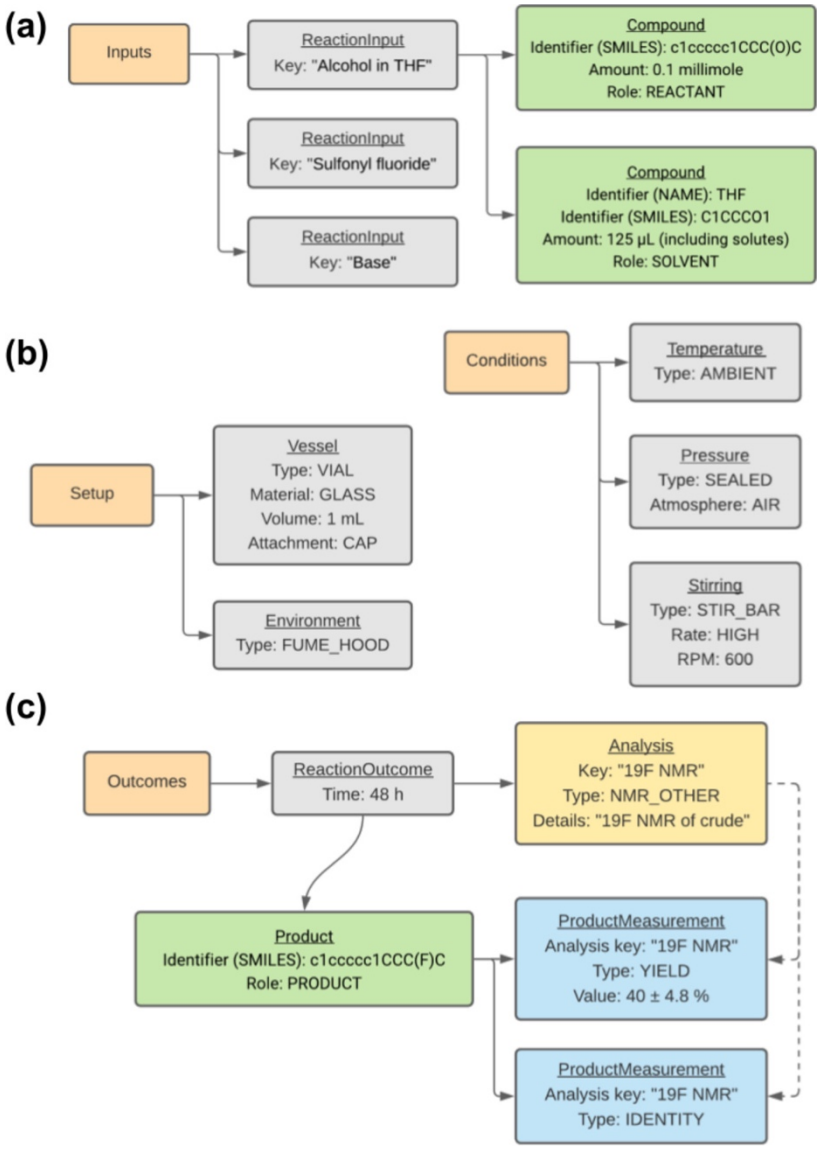


At a high level, the schema is divided into nine sections: reaction identifiers, inputs, setup, conditions, notes, observations, workups, outcomes (products and analytics), and provenance (**Figure A4.2**). Conceptually, each section can be thought of as a first-class object in the schema. Each of these sections contains data fields and child objects (with their own data fields) for describing the reaction. For example, each `ReactionInput` contains one or more components (usually `Compound` objects), each with their own identifiers(s), amount, and reaction role (**Figure A4.3a**). The fields of each schema object are structured to constrain their types or values, such as only allowing positive numeric values for amounts or limiting the units to a set of predefined constants. Many schema objects also include an unstructured `details` field for providing additional information that is not captured by other fields. The set of structured fields is largely driven by (a) a relatively small set of common types, such as for reaction roles, and (b) downstream use cases for machine learning, where structured types can be used as categorical features.

The schema is designed to support arbitrary levels of detail depending on the available information. For reactions taken from the patent literature, it may only be possible to describe the inputs and outputs at the level of their identifiers and amounts. For reactions submitted by

the original experimenter, it is possible to use structured and unstructured fields to include every detail required for reproducibility (including and beyond those in Figure A4.3b). To enforce a base level of consistency between records in the database, we use validation functions written in Python to require the presence of certain fields and check for reasonable values. For instance, each reaction must have at least one input, and every input compound requires an amount. Warnings are issued for ambiguous values, such as percentages entered as fractions. These validations are performed automatically in the interactive web editor (see below) and during the data set submission process.

Figure A4.3 *Example schema representations for a deoxyfluorination reaction.*[3] *(a) Inputs. The "Alcohol in THF" input is expanded to show its solute and solvent components. (b) Setup and conditions. The setup includes a description of the reaction vessel and there are separate objects for temperature, pressure, and stirring conditions. (c) Outcomes. Each outcome has an associated reaction time and one or more analyses that are linked to product-specific measurements of product identity, yield, selectivity, etc. There are several dozen additional fields not shown in this figure.*



A4.4 INTERFACES

Although submissions can be generated programmatically with Python, we are sensitive to the reality that experimentalists are not always comfortable with programming or using the command line. Accordingly, we have built web interfaces for creating submissions and searching the database. We also include instructions for web-only submissions in the online documentation.

The ORD Interactive Editor[31] is a tool for generating submissions to the database. Users can create data sets and use a responsive web form to fill in structured and unstructured data fields. Additionally, the editor can be used to enumerate a factorial data set based on a reaction template and a CSV or Excel spreadsheet with a row for each reaction. We have also recorded tutorial videos demonstrating the use of the editor, available on the ORD YouTube channel. Submissions to the ORD are made as pull requests to the `ord-data` GitHub repository[32] or directly by email.[33]

The ORD Search Interface[34] is used to search for reactions in the database. Reactions can be queried by SMILES/SMARTS substructure patterns for input or output molecules, by reaction ID, by publication DOI, etc. Each reaction has a summary page that displays the full record along with a graphical schematic. We have also written a simple Python API to enable downstream users to easily access data programmatically. The client library supports all of the search functionality in the web-based search interface, as well as the ability to download full data sets or lists of reactions.

A4.5 DATA SET EXAMPLES

The ORD has been designed to accommodate information about organic reactions spanning many distinct types, each requiring the definition of different metadata to ensure

reproducibility. Here, we briefly mention several common categories we anticipate and discuss the data that might be associated with each; examples in the ORD are given in Table A4.1.

The simplest but perhaps broadest category of reactions to capture are those performed through "traditional" benchtop operations. A data set of this type typically contains a small number of reactions defined by their inputs (**Figure A4.3a**), conditions/setup (**Figure A4.3b**), and product outcomes at a single time point (**Figure A4.3c**). Reaction inputs should always contain information about the quantity of each component and any additional metadata that might be relevant, such as the source of the compound. Conditions should at least include temperature and stirring, and the setup should specify what kind of vessel (e.g., flask, vial) was used. The level of detail associated with each reaction is a function of what information is available; for retrospective data entries, such as importing the open-source reaction data set parsed from the USPTO literature, many procedural details will be unspecified.

Table A4.1 *Example Data Sets Currently Available in the ORD.*

category	description	ref	size	example
single-step batch	deoxyfluorination reaction screening as a function of substrate, base, and fluoride source (entries in Figure 1)	3	80	link
single-step batch	microwave synthesis of a small library using the Biginelli multicomponent condensation reaction	4	48	link
kinetic profiling	online monitoring of a Suzuki coupling reaction by HPLC	5	7	link
high throughput	subset of "chemistry informer" screen of copper-catalyzed Buchwald–Hartwig aminations (entries 11–15 in Figure 4)	6	90	link
high throughput	C–N cross-coupling reaction yields varying aryl halide, additive, Pd catalyst, and base identities	7	4312	link
high throughput	Suzuki coupling reaction performance as a function of aryl halide, boronic acid, ligand, base, and solvent performed under pseudoflow conditions	8	5760	link
high throughput	C–N cross-coupling reaction performance of 3-bromopyridine with various nucleophiles, varying precatalysts and bases (entries in Experiment 2)	9	1536	link
high throughput	combinatorial nanochemistry screen of a complex aryl halide library using dual-metal photoredox C–N coupling (entries in Figure 6)	10	1728	link

photochemistry	substrate scope tables regarding coupling of α-carboxyl sp^3 carbons with aryl halides	11	24	link
photochemistry	Ir-catalyzed debromination conversions as a function of photocatalyst ligands	12	1152	link
electrochemistry	electroreductive coupling of alkenyl and benzyl halides via nickel catalysis (entries in Figures 2 and 3)	13	27	link
flow chemistry	sulfonamide library synthesis in flow	14	39	link
enzymatic	multistep biocatalytic cascade for the manufacture of islatravir	15	3	link
multistep	copper-catalyzed enantioselective hydroamination of alkenes	16	3	link
literature extracted	reactions extracted by text-mining United States published patents; imported from CML documents	1	1771017	link

In most cases, analytical data will at least include the yield of the desired product and how it was calculated (isolated yield, NMR yield, enantiomeric excess by chiral SFC, etc.). Ideally, more detailed analytical data will be included, such as the usual slate of analyses when a new compound is reported (i.e., ^{1}H NMR, ^{13}C NMR, HRMS, IR). Reactions can accommodate arbitrary analytical data as uploaded files or URLs to externally hosted data. Kinetic profiling experiments where multiple samples or aliquots are taken from a single vessel over time are described by one reaction with several distinct outcomes, each with their own analytical and product information; if different reactors are used for each time point, these are described as distinct reactions.

Data sets corresponding to high-throughput experiments will typically follow the case where reactions are identical other than a small number of variable fields such as substrate identity, catalyst identity, reagent identity, temperature, and product yield. Depending on the analytical workflow, the recorded yields from HTE may be uncalibrated: i.e., correspond only to relative UV peak areas in an LC or integrated TIC/EIC traces from LCMS. Specifying how yields were calculated or estimated is essential for understanding the fidelity of the data and

how it should be treated in downstream tasks. It is easiest to define data sets of this type through Python code or through the data set enumeration feature.

Additional types of reactions make use of the many optional fields in the schema. Photochemical reactions require additional specification of the illumination conditions, such as the light source and its distance to the vessel. Electrochemical reactions require details about the cell configuration: whether it is divided or undivided and the materials of the cathode and anode. Reactions employing flow chemistry include information about the type of pump and delivery rate used for all stock solutions, in addition to details about the type of flow reactor used. Reactions employing enzymatic catalysts take advantage of the ability to define compounds by their primary amino acid sequence or UnitProt/PDB identifiers.

Multistep reactions in the ORD are split into separate entries within one data set; their treatment depends on whether intermediate products are isolated. Reaction sequences with incomplete intermediate isolations (e.g., one-pot reactions or telescoped reaction sequences) are linked by considering the product of one step as a "crude" input of the subsequent one. Otherwise, an isolated product from an in-house reaction is treated no differently from a purchased starting material, except for a cross-reference to the reaction ID defining how the compound was prepared.

A4.6 DOWNSTREAM USE CASES

We anticipate that one of the major applications of the ORD will be for the generation of structured data sets for machine learning (ML). Reaction modeling with ML has become increasingly common, with reports frequently citing the importance of high-quality data for success. Examples of the types of data-driven reactivity studies that could benefit from the ORD include synthesis planning,[17,18] reaction product prediction,[19] reaction yield

prediction,[3,7,20] reaction condition prediction,[21,22] selective catalyst design,[23,24] and reaction optimization,[25] among others. These predictive chemistry tools illustrate the value of curating reaction data, yet we believe they are just scratching the surface in terms of both their utility and the complexity of tasks that they can help address.

The schema was designed in part around these use cases and provides descriptive, easily accessible fields for reaction featurization (see above). As a result, data sets require minimal processing before model training and can be quickly integrated into Python ML workflows. We have made an example data processing and ML pipeline available as a Jupyter Notebook in the ORD schema repository.[35] This example constructs a yield predictor from a Suzuki−Miyaura coupling data set[8] and reproduces regression results from the literature.[20]

It is our hope that this type of ML workflow will aid in advancing the understanding of reaction performance (e.g., yield, selectivity, etc.) and thus accelerate discovery. We expect that another valuable use case will be the construction of retrosynthesis models. These tools may enable more quantitative evaluations of new reactions given these additional details (e.g., concentrations, orders of addition, vendor information) and eventually allow training directly on raw analytical data instead of processed analytical data. Data captured by the ORD schema is richer than that in previous database efforts and should facilitate training of the next generation of predictive chemistry tools.

A4.7 DISCUSSION

Commitment to Open Access. As the name implies, the Open Reaction Database is designed for open access and community contributions. All reaction data in the database is available under a CC BY-SA license.[36] The various software tools and code, such as the schema definitions and the interactive web editor, are available under an Apache license.[37]

Both the data and code are hosted on GitHub under the Open Reaction Database organization (https://github.com/open-reaction-database).

Data Quality and Validation. There are several mechanisms for ensuring, or at least encouraging, that data are of high quality. At a basic level, Python scripts are used to validate each reaction within a data set, as mentioned above. This requires that all reactions contain a minimal level of information and, for example, that quantities are associated with units. Beyond these checks, the schema itself was designed to capture important metadata such as how yields were quantified, so that subsequent analyses can distinguish between true isolated yields and estimates from LCMS peak areas (where the latter is more common in high-throughput workflows).

Additionally, all submissions are reviewed manually for completeness and correctness by a reviewer not involved in the preparation of the submission. During the review process, the reviewer will require the details listed above unless the submitter does not have access to more information about the experiment (e.g., more than what was published); when a previously published data set is reviewed, the publication and submission should be compared directly. If a data set is not associated with a journal or patent article, then the reviewer may only check for internal consistency and completeness and must trust the submitter to provide the correct structures and technical details. As the rate of ORD submissions grows, we anticipate recruiting additional volunteer reviewers and adjusting the review process as needed.

Information about the submitter and experimenter (who might not be the same person) is captured as part of the provenance metadata for each entry, optionally including email addresses. This will allow data consumers to contact either party if more information is needed

for their downstream use case. If data quality issues or discrepancies are identified, we will request that the original submitter help revise or review any proposed changes.

Evolution of the Schema. The ORD schema is a "living document" that will change and adapt to the needs of its users and of the community. For example, future contributions from electrochemists or advances in downstream machine learning may require the addition of new structured fields to better represent the data. In other cases, structured fields that are hardly ever used may be deprecated and removed in favor of unstructured descriptions. We will use versioning and migration processes to ensure that backward-incompatible changes (if any) are deployed without disruption to existing workflows pinned to earlier versions of the data and code.

A4.8 OUTLOOK

By providing a structured schema, submission mechanism, and search/retrieval tools, the ORD reduces several technical barriers to data sharing and model building. For instance, adoption of the ORD schema will allow the Supporting Information in publications to be standardized and readable by anyone without manual parsing or back and forth with the authors. Importantly, the use of a well-defined schema will make the omission of procedural details needed for reproducibility more apparent, as well as enable easier comparisons between reactions (both within and across publications). We plan to support translation software to convert between the ORD format and other emerging data standards[26–28] as well as electronic lab notebook formats. To be clear: we believe that PDFs without accompanying structured data should no longer clear the bar for publication.

However, we are still sensitive to the many social and cultural barriers surrounding publishing that persist in the field. There is not a culture of data sharing beyond what is required

for peer-reviewed publication. In particular, the unwillingness to share "unsuccessful" or "failed" reactions with low yields or selectivities—whether due to intrinsic reactivity, procedural details, or human errors—presents an overly optimistic view of chemistry where most reactions succeed, impeding the development of models to explain when they may not. Successful and unsuccessful reactions are not differentiated by the ORD schema beyond product measurements; this categorization is somewhat arbitrary and depends on the downstream use case.

Though the ORD is still young, initial feedback and adoption have been promising. The success of this effort will depend on buy-in from data generators, contributors, and consumers and a shared recognition of its value. We encourage data generators to explore the schema and infrastructure that we have built for capturing their experimental data and invite discussion from the broader community on how to incorporate these structured data formats throughout the life cycle of reaction data—from benchtop to laptop.

A4.9 REFERENCES

(1) Lowe, D. Chemical reactions from US patents (1976–September 2016). https://figshare.com/articles/dataset/Chemical_reactions_from_US_patents_1976-Sep2016_/5104873. DOI: 10.6084/m9.fig-share.5104873.v1. (accessed 5/19/2021).

(2) Wilkinson, M. D. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018.

(3) Nielsen, M. K.; Ahneman, D. T.; Riera, O.; Doyle, A. G. Deoxyfluorination with Sulfonyl Fluorides: Navigating Reaction Space with Machine Learning. *J. Am. Chem. Soc.* **2018**, *140*, 5004–5008.

(4) Stadler, A.; Kappe, C. O. Automated Library Generation Using Sequential Microwave-Assisted Chemistry. Application toward the Biginelli Multicomponent Condensation. *J. Comb. Chem.* **2001**, *3*, 624–630.

(5) Christensen, M.; Adedeji, F.; Grosser, S.; Zawatzky, K.; Ji, Y.; Liu, J.; Jurica, J. A.; Naber, J. R.; Hein, J. E. Development of an Automated Kinetic Profiling System with Online HPLC for Reaction Optimization. *React. Chem. Eng.* **2019**, *4*, 1555–1558.

(6) Kutchukian, P. S.; Dropinski, J. F.; Dykstra, K. D.; Li, B.; DiRocco, D. A.; Streckfuss, E. C.; Campeau, L.-C.; Cernak, T.; Vachal, P.; Davies, I. W.; Krska, S. W.; Dreher, S. D. Chemistry Informer Libraries: A Chemoinformatics Enabled Approach to Evaluate and Advance Synthetic Methods. *Chem. Sci.* **2016**, *7*, 2604– 2613.

(7) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-coupling Using Machine Learning. *Science* **2018**, *360*, 186–190.

(8) Perera, D.; Tucker, J. W.; Brahmbhatt, S.; Helal, C. J.; Chong, A.; Farrell, W.; Richardson, P.; Sach, N. W. A Platform for Automated Nanomole-scale Reaction Screening and Micromole-scale Synthesis in Flow. *Science* **2018**, *359*, 429–434.

(9) Buitrago Santanilla, A.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.-C.; Schneeweis, J.; Berritt, S.; Shi, Z.-C.; Nantermet, P.; Liu, Y.; Helmy, R.; Welch, C. J.; Vachal, P.; Davies, I. W.; Cernak, T. Dreher, S. D. Nanomole-scale High-throughput Chemistry for the Synthesis of Complex Molecules. *Science* **2015**, *347*, 49–53.

(10) Dreher, S. D.; Krska, S. W. Chemistry Informer Libraries: Conception, Early Experience, and Role in the Future of Cheminformatics. *Acc. Chem. Res.* **2021**, *54*, 1586–1596.

(11) Zuo, Z.; Ahneman, D. T.; Chu, L.; Terrett, J. A.; Doyle, A. G.; MacMillan, D. W. C. Merging Photoredox with Nickel Catalysis: Coupling of α-carboxyl sp^3-carbons with Aryl Halides. *Science* **2014**, *345*, 437–440.

(12) Mdluli, V.; Diluzio, S.; Lewis, J.; Kowalewski, J. F.; Connell, T. U.; Yaron, D.; Kowalewski, T.; Bernhard, S. High-throughput Synthesis and Screening of Iridium(III) Photocatalysts for the Fast and Chemoselective Dehalogenation of Aryl Bromides. *ACS Catal.* **2020**, *10*, 6977–6987.

(13) DeLano, T. J.; Reisman, S. E. Enantioselective Electroreductive Coupling of Alkenyl and Benzyl Halides via Nickel Catalysis. *ACS Catal.* **2019**, *9*, 6751–6754.

(14) Gioiello, A.; Rosatelli, E.; Teofrasti, M.; Filipponi, P.; Pellicciari, R. Building a Sulfonamide Library by Eco-Friendly Flow Synthesis. *ACS Comb. Sci.* **2013**, *15*, 235–239.

(15) Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T.; Halsey, H. M.; Hughes, G. J.; Jo, J.; Joyce, L. A.; Kolev, J. N.; Liang, J.; Maloney, K. M.; Mann, B. F.; Marshall, N. M.; McLaughlin, M.; Moore, J. C.; Murphy, G. S.; Nawrat, C. C.; Nazor, J.; Novick, S.; Patel, N. R.; Rodrigues-Granillo, A.; Robaire, S. A.; Sherer, E. C.; Truppo, M. D.; Whittaker, A. M.; Verma, D.; Xiao, L.; Xu, Y.; Yang, H. Design of an in vitro Biocatalytic Cascade for the Manufacture of Islatravir. *Science* **2019**, *366*, 1255–1259.

(16) Liu, R. Y.; Buchwald, S. L. Copper-Catalyzed Enantioselective Hydroamination of Alkenes. *Org. Synth.* **2018**, *95*, 80–96.

(17) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature* **2018**, *555*, 604–610.

(18) Coley, C. W.; Green, W. H.; Jensen, K. F. Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **2018**, *51*, 1281–1289.

(19) Coley, C. W.; Barzilay, R.; Jaakkola, T. S.; Green, W. H.; Jensen, K. F. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **2017**, *3*, 434–443.

(20) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an Organic Synthesis Robot with Machine Learning to Search for New Reactivity. *Nature* **2018**, *559*, 377–381.

(21) Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **2018**, *4*, 1465–1476.

(22) Maser, M. R.; Cui, A. Y.; Ryou, S.; DeLano, T. J.; Yue, Y.; Reisman, S. E. Multilabel Classification Models for the Prediction of Cross-Coupling Reaction Conditions. *J. Chem. Inf. Model.* **2021**, *61*, 156–166.

(23) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363*, No. eaau5631.

(24) Zhao, S.; Gensch, T.; Murray, B.; Niemeyer, Z. L.; Sigman, M. S.; Biscoe, M. R. Enantiodivergent Pd-catalyzed C–C Bond Formation Enabled through Ligand Parameterization. *Science* **2018**, *362*, 670–674.

(25) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature* **2021**, *590*, 89–96.

(26) Pistoia Alliance, D. Unified Data Model. https://www.pistoiaalliance.org/projects/current-projects/unified-data-model/, Accessed June 1, 2021.

(27) Mehr, S. H. M.; Craven, M.; Leonov, A. I.; Keenan, G.; Cronin, L. A Universal System for Digitization and Automatic Execution of the Chemical Synthesis Literature. *Science* **2020**, *370*, 101–108.

(28) Tremouilhac, P.; Lin, C.-L.; Huang, P.-C.; Huang, Y.-C.; Nguyen, A.; Jung, N.; Bach, F.; Ulrich, R.; Neumair, B.; Streit, A.; Bräse, S. The Repository Chemotion: Infrastructure for Sustainable Research in Chemistry. *Angew. Chem., Int. Ed.* **2020**, *59*, 22771–22778.

(29) https://open-reaction-database.org/.

(30) https://developers.google.com/protocol-buffers.

(31) https://editor.open-reaction-database.org/.

(32) https://github.com/open-reaction-database/ord-data.

(33) submissions@open-reaction-database.org.

(34) https://client.open-reaction-database.org/.

(35) https://github.com/open-reaction-database/ord-schema/blob/main/examples.

(36) https://creativecommons.org/licenses/by-sa/4.0/.

(37) https://www.apache.org/licenses/LICENSE-2.0.

Chapter 4

3D Computer Vision Architectures for Molecular Shape Recognition[1]

4.1 INTRODUCTION

Obtaining quantum mechanical (QM) properties of molecules with high precision is computationally expensive and scales exponentially with system size.[1] Semi-empirical QM methods offer faster alternatives to density functional theory (DFT) at the expense of accuracy.[2] The combination of low-precision QM modeling with machine learning (ML), so called delta-learning,[3-6] has demonstrated success in obtaining chemically accurate results with minimal additional computational cost. This approach, however, is not amenable to high-throughput experimentation (HTE), e.g., for the in-silico screening of a virtual compound library in drug discovery campaigns.[1]

The main bottleneck for HTE delta-learning still lies in the optimization of 3D geometry prior to property prediction. On the other hand, structure optimization using simple force field methods[7] can be performed in ultra-high-throughput, though with low accuracy in downstream QM calculations. Therefore, the direct prediction of high-precision QM properties from force-field-optimized geometries represents a valuable goal for molecular ML.

---

[1] Portions of this chapter were adapted from the following communication: Maser, M. R.; Reisman, S. E. 3D Computer Vison Models Predict DFT-Level HOMO-LUMO Gap Energies from Force-Field Optimized Molecular Geometries. *ChemRxiv*, **2021**, *Preprint*. DOI: 10.33774/chemrxiv-2021-11r61. Copyright © 2021 The Authors. The research discussed in this chapter was completed in collaboration with current undergraduate student Megan Santhumayor and current graduate student Emily P. Chen of the Reisman Lab. Extensions of the work herein are currently being investigated in collaboration with the same students, as well as postdoctoral scholar Dr. Uriah Israel of the Yue and Van Valen Groups, graduate student Jennifer J. Sun of the Perona and Yue Groups, and graduate student Elena Zhong of the Yue Group.

Computer vision (CV) approaches to understanding molecular geometry have seen little development despite their enormous impact in other 3D-shape-learning domains.[8-14] Space-filled volumetric representations match the chemical intuition of molecules as physical objects, though only a handful of reports have utilized voxels[15-18] or 3D shape descriptors derived from voxels[19] as input features for ML. Even so, studies thus far are limited by use of a single core scaffold for shape differentiation,[19] low input resolution at the atomic scale,[17] and the prediction of trivially computed properties.[15]

We herein report on our initial studies to advance these methods, developing 3D convolutional neural network (CNN) architectures for molecular shape recognition. Our methods are demonstrated by the prediction of QM energy values at near-DFT accuracy using rapidly computed force field geometries[7] rendered as voxel grids. We take advantage of modern graphics processing capabilities for ultra-HT voxelization,[18] allowing us to interrogate the effects of atomic radial decay in our input volumes. We find that models based on inception networks[20] provide high accuracy and generalization, and we discuss the chemical relevance of such architectures.

We further find that multi-view augmentation of input data by 10-fold rotation significantly improves accuracy in many cases.[8] We provide a simple workable example for data processing and model construction as an interactive python notebook (link), and will make the full code base as well as trained models freely available in a follow-up full report. The models are readily usable by researchers in chemistry, and the voxelization and learning functions provide a modeling challenge for the CV community.

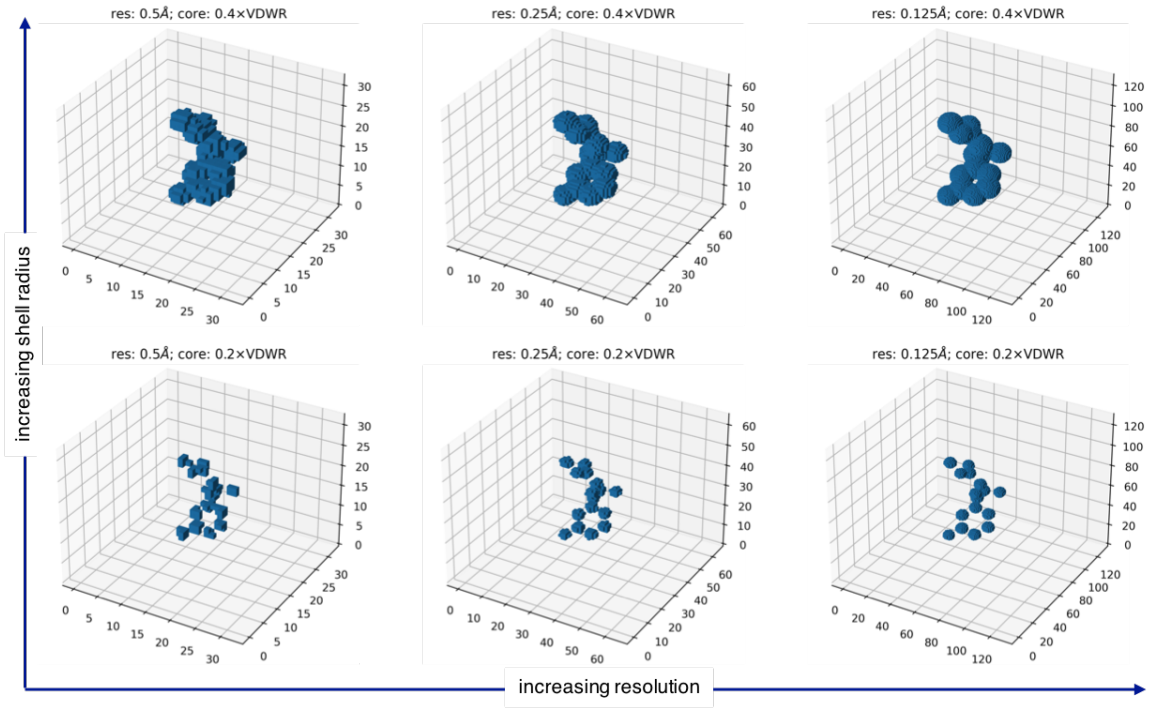
4.2 APPROACH AND METHODS

4.2.1 DATA

The PubChemQC PM6 (PCQCPM6) dataset[1] was recently published and includes over 200 million molecules with geometries optimized at the semi-empirical PM6 theory level.[2] A 2.5 million structure subset was further optimized by high-precision DFT at the B3LYP/6-31G* theory level. The energy difference between a compound's highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), so called the HOMO-LUMO gap, is a valuable property often predicted by DFT calculation. The HOMO-LUMO gap is commonly used as an indicator of a substrate's reactivity, particularly in organic electronics applications.[1,21] HOMO-LUMO gap energies were published for the entirety of the PCQCPM6 dataset, calculated at the theory level of each compound's optimization. Given their high accuracy and cost, the B3LYP energy results will serve as our regression targets. These data are rather Gaussian in nature but contain a few significant outliers. Those outside of 5 standard deviations from the mean were removed before modeling (see the original report for more data information[1]).

We prepared space-filled volumes (voxel arrays) for 1.8 million molecules from the B3LYP subset with molecular weights up to 500 g/mol containing the atoms C, H, N, O, P, S, F, and Cl. Geometries were optimized in high throughput using the Merck Molecular Force Field (MMFF)[7] as implemented in the RDKit cheminformatics package.[22] We toggled several settings in the voxelization procedure, including the voxel value decay from atom centers as a function of their van der Waals radius (VDWR, **Figure 4.1**, see Experimental Section 4.7).

Figure 4.1 *Example renderings of voxelized PCQCPM6 ID: 66098987 (HOMO-LUMO gap 5.76 eV) at various resolutions and decay settings.*



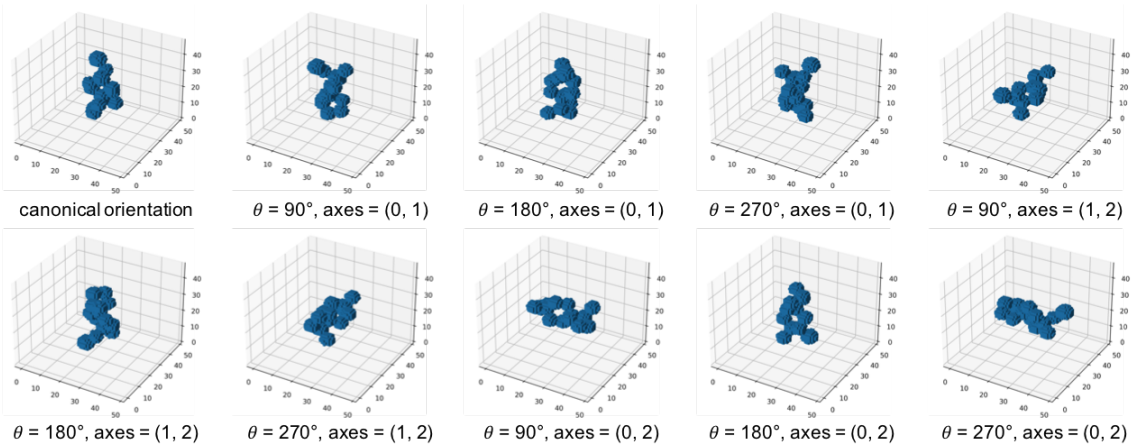
We additionally tested multiple input resolutions, from 32^3 to 128^3, holding the overall grid size constant at 16Å in each dimension (**Figure 4.1**). We ultimately found higher resolutions offered only minor improvements and for computational speed and data storage settled on 32^3 grids (i.e., 0.5Å spacing), maximum voxel values within 0.3 × the VDWR, stepwise decay in shells of 0.15 × the VWDR, and a total of 4 shells. For the sake of direct comparison between theory-level inputs, we aligned the MMFF structures to the PM6-optimized geometries before modeling (see Section 4.4).

For architecture assessment, we used a randomly selected subset of 100k molecules. This was divided into training, validation, and test sets with an 80/10/10 split. For the ultimate models, this set was expanded to 1M structures, trained and evaluated with the same 80/10/10 split.

4.2.2 AUGMENTATION

Inspired by performance improvements obtained in the original VoxNet report,[8] we experimented with rotational data augmentation. This holds analogy to multi-view 3D object recognition,[23] with the extension that our "views" remain in 3D space instead of 2D. To achieve this, we performed 9 right-angle rotations[24] of each structure (90°, 180°, and 270° about each of the principal axes), yielding 10 total orientations with identical target values when added to the original view (**Figure 4.2**). To avoid excessive data storage requirements, we implemented a generator function that performs this augmentation on the fly as examples are passed into the networks (see associated code for details). When using augmentation, mini-batch sizes were multiplied by the number of augmented samples (10 here) such that batches contained only complete instances of individual molecules (i.e., rotated examples were not spread over multiple batches).

Figure 4.2 *Visualization of 10-fold rotational augmentation.*



4.2.3 DEEP LEARNING MODELS

We implemented four published CNN architectures with known success in 3D object recognition[8,25] and shape-based design.[17] Their composition is described below, and

graphical depictions can be found in the Experimental Section. Unless specified otherwise, all models were implemented using the TensorFlow deep learning platform[26] and were trained on four NVIDIA® Tesla T4 GPUs for 20 epochs using the Adam optimizer[27] with a learning rate that decayed linearly from $1e^{-3}$ to $1e^{-4}$.[8]

VoxNet. Reported in 2015 by Maturana and Scherer,[8] VoxNet has become a standard baseline architecture for 3D shape learning tasks. It consists of two convolutional layers with kernels, strides, and channels of size (5, 2, 32) and (3, 1, 32), respectively. These are each treated with LeakyReLu activation ($\alpha = 0.1$) and dropout ($p = 0.2$). The outputs of the CNN undergo [2×2×2] max-pooling and flattening before two ReLU-activated dense layers of output dimensions 128 and 1, the latter providing the predicted HOMO-LUMO gap. In total, VoxNet contains just 0.91M learnable parameters (see Experimental Section 4.7).

FusionNet. An alternative strategy to direct 3D convolutions, FusionNet[25] takes 2D slices of voxel grids and uses standard image-processing CNN machinery for learning. An interesting consequence of this approach is the anisotropic expansion in the channel dimension during propagation. This translates to stretching a molecule along one axis as it passes through the network. In the report, "**V-CNN I**" achieves this by a ReLu-activated CNN layer containing 64 channels with 2D kernels and stride of size (3, 1). 2D max-pooling is followed by two more of the same CNN layers, dropout ($p = 0.5$), and flattening. A dense layer of size 2048 again precedes the output layer. V-CNN I contains 3.4M parameters.

In the same report, "**V-CNN II**" combines the 2D slice approach with inception modules,[20] or multiple kernels of different sizes whose outputs are concatenated and treated

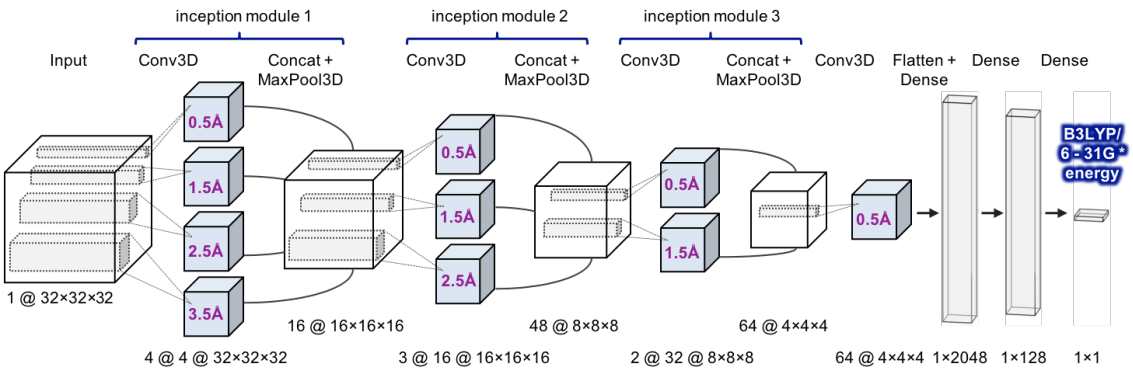
together by a nonlinearity in each layer. Specifically, 2D kernels of size 1, 3, and 5 pixels with stride 1 each produce 20 channels (a contraction relative to the size 32 inputs). The three outputs are concatenated to 60 channels before ReLU activation and dropout (p = 0.2). Another 2-pronged inception module with kernels of size 1 and 3 combine to give 64 channels, again before ReLU and dropout (p = 0.3). A final 32 channel convolution with a [3×3] kernel, activation, and dropout (p = 0.5) is followed by the same dense module as V-CNN I. The largest of all tested architectures, V-CNN II contains over 67M parameters.

ShapeAutoencoder. Skalic et al. reported an autoencoder architecture for generating embeddings from 3D molecular shapes.[17] These embeddings are utilized for generative sampling by decoding into SMILES string representations.[28] The encoder architecture is essentially a higher capacity VoxNet, with five feed-forward 3D CNN layers reaching up to 512 channels. Each layer utilizes [3×3×3] kernels and alternates between strides 1 and 2 for contraction with ReLU activation. After flattening, the embedding is made by a dense layer of size 512; we append a final dense layer to output our scalar predictions. We found the exact implementation of their architecture to give rather unstable loss curves and severe overfitting. We thus added dropout (p = 0.2) after the second and fourth CNN layers and first dense layer, which led to much smoother fitting and holdout performance (see Section 4.3 and Experimental Section 4.7). We refer to the resulting architecture as "**ShapeEncoder-d**" going forward. ShapeEncoder-d contains over 15M parameters.

QMCVNet (ours). To derive the benefits of each architecture above, we merged the inception modules of V-CNN II with the deep 3D network of ShapeEncoder-d. For the task of quantum mechanical computer vision, we call our resulting model QMCVNet,

depicted in Figure 4.3. Given that our voxel grids are spaced at 0.5Å, the size 1, 3, and 5 kernels in V-CNN II correspond to 0.5, 1.5, and 2.5Å in each dimension. As typical carbon-carbon and carbon-heteroatom bond lengths are between 1–1.5Å, these kernels convolve over roughly 1–2 bonds, or 2–3 atom centers at most. With the hope of covering entire functional groups (roughly 3–4 atoms), we added a third inception module with kernels of size 1, 3, 5, and 7 to the beginning of the network. After each block, the multi-scale outputs are concatenated and treated with ReLU activation and dropout (p = 0.2) before 3D max-pooling of size and stride 2. The output of a final 3D convolution is flattened and passed through two dense layers of size 2048 and 128 before the HOMO-LUMO gap is predicted. In total, QMCVNet contains 8.7M parameters.

Figure 4.3 *QMCVNet architecture. Uppermost labels represent shorthand layer names. Nethermost labels represent tensor shapes at each layer. Å labels represent the kernel size on one edge used to obtain each intermediate representation in the inception modules.*



4.3 RESULTS AND ANALYSIS

4.3.1 ARCHITECTURE SCREEN – 100k EXAMPLES

To assess their performance, we tested each architecture for predicting HOMO-LUMO gaps at the B3LYP level from single PM6 and MMFF geometries and orientations. Results from the 100k compound set (see Section 4.2.1) are included in Figure 4.4 and Table 4.1. We present model performances as their mean absolute error (MAE) in electron Volts (eV) on the 10k example test set, as this is the metric used in the PCQCPM6 report for comparing the PM6 method against B3LYP itself.

In these experiments, we found that higher capacity models such as ShapeEncoder-d and QMCVNet performed best overall. Smaller and simpler architectures such as VoxNet and V-CNN I tended to underfit the data, yielding mostly mean-centered predictions and unstable loss curves. The best models gave MAEs as low as 0.539 eV for PM6 geometries (ShapeEncoder-d), and 0.549 eV for MMFF (QMCVNet). These MAEs corresponded to mean absolute percentage errors (MAPEs) of around 10.5%, meaning that, on average, predictions were within 10-11% of the true HOMO-LUMO gap values. Correlation coefficients (R^2) between predictions and ground truths were modest in most cases, topping out around 0.580 (QMCVNet + PM6). Note that the PM6 baseline achieves MAE, MAPE, and R^2 values of 0.400 eV, 7.57%, and 0.783, respectively.

Figure 4.4 *B3LYP HOMO-LUMO gap prediction performance for all architectures on the 100k dataset. Box-and-whisker plots aggregate results over 5 random folds of training data.*

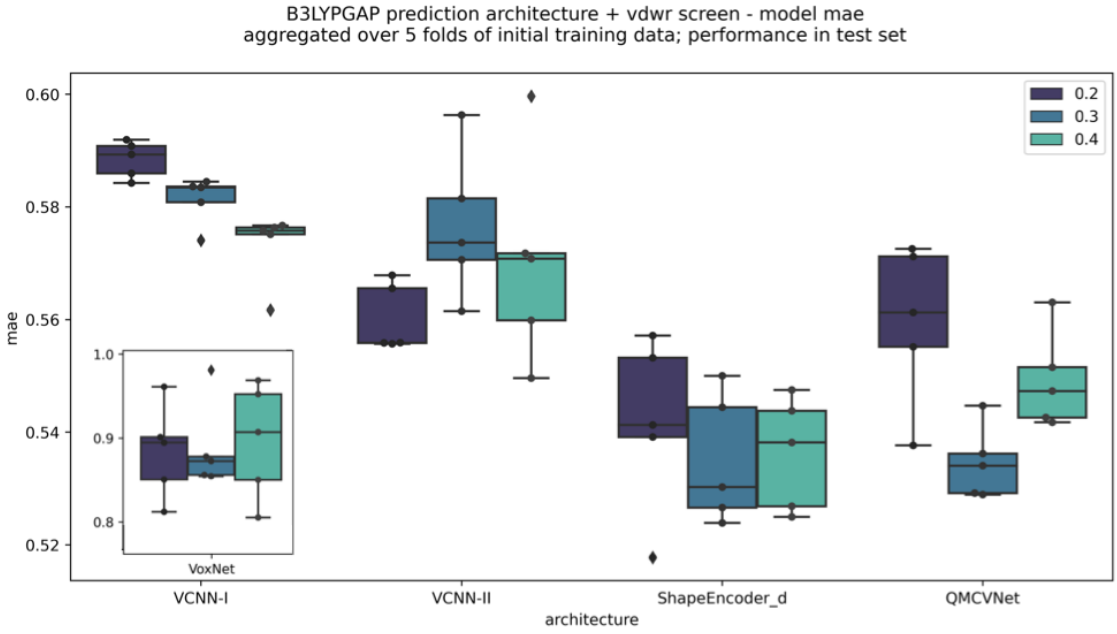


Table 4.1 *B3LYP HOMO-LUMO gap prediction performance for all architectures on the 100k dataset.*

Model screen on 100K geometry inputs predicting B3LYP/6-31G*				
Model (params)	**input theory**	**test MAE (eV, ↓)**	**test MAPE (%, ↓)**	**R^2 (↑)**
naïve (mean)	–	0.851	16.00	0.000
PM6	**–**	**0.400**	**7.566**	**0.783**
VoxNet (916,705)	PM6	0.980	16.84	-0.119
	MMFF	1.256	21.70	-0.719
	MMFF + aug	1.298	22.21	-0.833
V-CNN I (3,373,249)	PM6	0.587	11.22	0.504
	MMFF	0.630	11.89	0.426
	MMFF + aug	0.555	10.53	0.553
V-CNN II (67,173,149)	PM6	0.577	11.18	0.525
	MMFF	0.610	11.38	0.467
	MMFF + aug	0.595	11.26	0.493
ShapeEncoder-d* (15,126,497)	**PM6**	**0.539**	**10.39**	**0.578**
	MMFF	0.554	10.52	0.562
	MMFF + aug	**0.502**	**9.626**	**0.643**
QMCVNet (ours) (8,741,505)	PM6	0.543	10.31	0.580
	MMFF	**0.549**	**10.55**	**0.566**
	MMFF + aug	0.530	10.03	0.607
*w/ dropout				

In attempt to increase the performance and generality of our models, we tested the effects of 10-fold rotational augmentation with each model type using the MMFF geometries (see Section 4.2.2). As in ref. 8 we pass each rotated structure through a given network individually and take the average of their outputs as the ultimate prediction. Here, we found most models improved over the single-view approach, reducing their MAE by about 5-10%, relatively. ShapeEncoder-d achieved a MAE of 0.502, a MAPE of 9.626, and an R^2 of 0.643.

4.3.2 BEST ARCHITECTURE SCALEUP – 1M EXAMPLES

With the results above, we tested the best performing architectures—ShapeEncoder-d and QMCVNet—using a dataset of 1M examples, encoded and split as in the previous experiments. Results are included in Table 4.2, which showed that both models improved substantially with the added data, particularly with PM6 geometries. ShapeEncoder-d achieved a MAE, MAPE, and R^2 of 0.418, 8.051, and 0.732 for PM6 and 0.484, 9.196, and 0.664 for MMFF, respectively. This was particularly exciting as it approaches the accuracy obtained by the true semi-empirical PM6 method, learning from data alone.

Table 4.2 *B3LYP HOMO-LUMO gap prediction performance on the 1M dataset.*

Model scaleup on 1M geometry inputs predicting B3LYP/6-31G*				
Model (params)	**input theory**	**test MAE (eV, ↓)**	**test MAPE (%, ↓)**	**R^2 (↑)**
naïve (mean)	–	0.848	15.86	0.000
PM6	**–**	**0.400**	**7.557**	**0.782**
ShapeEncoder-d* (15,126,497)	**PM6**	**0.418**	**8.051**	**0.732**
	MMFF	0.484	9.196	0.664
	MMFF (100 epochs)	**0.455**	**8.873**	**0.687**
QMCVNet (ours) (8,741,505)	PM6	0.479	9.482	0.654
	MMFF	0.526	10.38	0.583
	MMFF (100 epochs)	0.484	9.590	0.645
*w/ dropout				

If allowed to continue training to 100 epochs, models were able to fit even MMFF geometries down to MAE of 0.455 and MAPE of 8.873 (ShapeEncoder-d). Further, in this experiment, energy evaluations on the 100k compound test set were completed in just 2.5 min, i.e., at a rate of roughly 1.5 ms per prediction. This extrapolates to evaluating the

entire 220M PCQCPM6 dataset in under 4 days, compared to the reported 1.5 years for the full calculations.[1]

4.3.3 BASELINE METHODS

To benchmark the performance of our methods against existing approaches to molecular property prediction, we implemented a variety of canonical ML setups for this task. These included training random forests (RFs), gradient-boosting machines (GBMs), and deep NNs (DNNs) on cheminformatic descriptor vectors and Morgan fingerprints (1024-bit, radius 3).[29,30] Additionally, we implemented a standard message-passing NN (MPNN)[31] to test the efficacy of graph convolutional networks (GCNs).[32,33] The results of each model setup with varying amounts of data are included in Figures 4.5 and Table 4.3.

Figure 4.5 *B3LYP HOMO-LUMO gap prediction performance for baseline model setups including RFs, GBMs, and DNNs with descriptors and fingerprints (top) and MPNNs (bottom) on varying dataset sizes. Box-and-whisker plots aggregate results over 5 random folds of training data.*

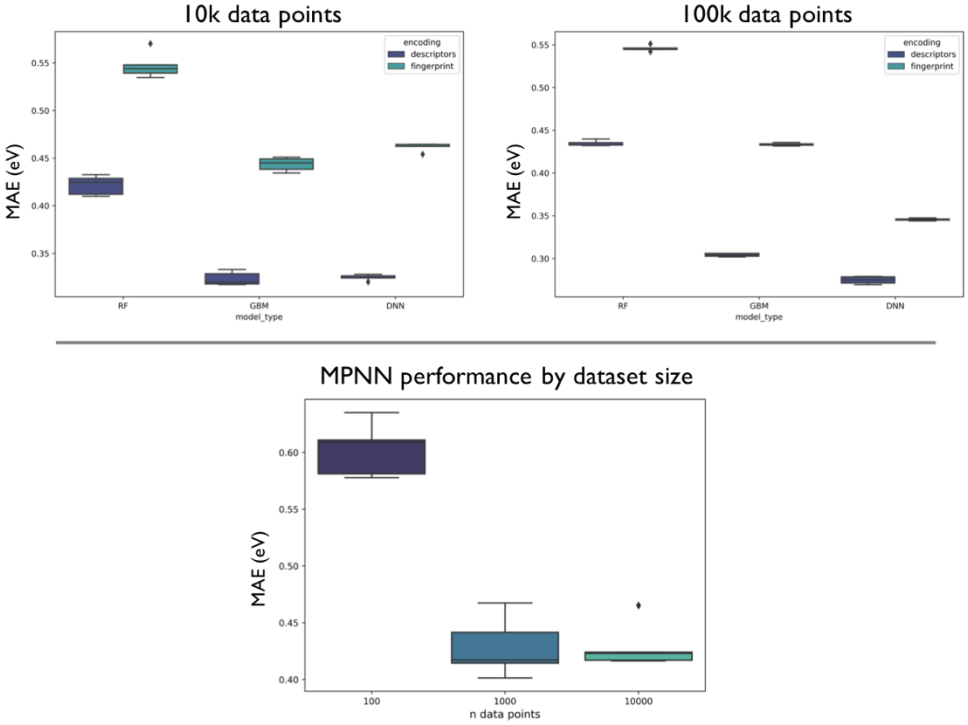


Table 4.3 *B3LYP HOMO-LUMO gap prediction performance for all baseline models on each dataset.*

Baseline model performance predicting B3LYP/6-31G*				
model type	**encoding type**	**n data**	**test MAE (eV, ↓)**	**R^2 (↑)**
naïve (mean)	–	1e5	0.851	0.000
PM6	–	1e5	0.400	0.783
random forest	descriptors	1e4	0.422	0.739
	–	1e5	0.435	0.723
	fingerprints	1e4	0.547	0.582
	–	1e5	0.546	0.577
gradient-boosting machines	**descriptors**	**1e4**	**0.323**	0.839
	–	1e5	0.304	0.857
	fingerprints	1e4	0.444	0.711
	–	1e5	0.434	0.724
deep neural networks	**descriptors**	1e4	0.324	**0.843**
	–	**1e5**	**0.275**	**0.885**
	fingerprints	1e4	0.462	0.697
	–	1e5	0.346	0.820
graph convolutional networks (MPNN)	**graphs**	1e2	0.603	0.322
	–	**1e3**	**0.428**	**0.690**
	–	1e4	0.429	0.679

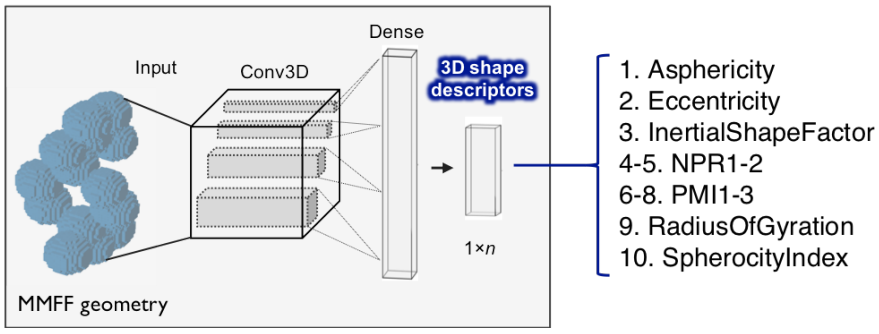
As can be seen, classical ML models can outperform 3D CNNs and indeed even PM6 computation by significant margins, particularly in the low-data regime (10k–100k examples). GBMs and DNNs achieve MAEs as low as 0.323 eV along with R^2 of around 0.84 with as few as 10k examples when trained with descriptor vector encodings, substantially improving over the PM6 method. Further, with 100k examples, DNNs returned a MAE of 0.275 eV and R^2 of 0.885, the best metrics obtained to date. GNNs performed slightly worse than expected, giving MAE and R^2 of 0.429 and 0.679 with 10k examples, respectively. That said, only the standard MPNN architecture has been tested

thus far and much development on GNN architectures has been reported in recent years (see those tested in Chapter 3, Sections 3.3.3.2 and 3.4).

4.4 DEMONSTRATION OF 3D SHAPE LEARNING

Given that the large, complex, heavily parameterized 3D CNN models were outperformed by simpler methods for this particular task, we sought a better understanding of the chemical features being learned by the volumetric models. In particular, we sought a demonstration that VNN models were in fact learning 3D shape features of molecules, with the hypothesis that this would forecast their utility for more inherently 3D tasks (see Introduction). As such, we computed 3D descriptors for the MMFF-optimized geometries in the 100k datasets and re-configured our best volumetric models as multi-task architectures for their prediction (**Figure 4.6**).

Figure 4.6 *Multi-task VNN setup (convolutional module abbreviated).*



We selected a total of 10 3D shape descriptors from the Descriptors3D module in the RDKit.[22] These included computed features such as asphericity, principal moments of inertia, and radius of gyration, among others.[34] Figure 4.7 shows the predicted vs ground truth values for each of the 10 geometric descriptors using the ShapeEncoder-d architecture. Figure 4.8 shows the same predictions using the DNN+descriptor model—our best performing baseline configuration. It can be clearly seen that VNN models are

much more capable of learning such 3D shape features from voxel data than classical models trained on 1D features. We thus hypothesize that our 3D architectures may be useful for modeling 3D chemical phenomena such as predicting reaction enantioselectivity[35] or drug-target binding affinity,[18] two very challenging ML tasks. We note that it could be reasonably expected that 1D models struggle with this task in that they have no information on the particular conformer geometry used for feature calculation and should only have the ability to infer rough geometric features from underlying connectivity, if anything.

Figure 4.7 *Prediction of 3D shape features by ShapeEncoder-d architecture trained on 100k voxelized examples with MMFF optimization.*

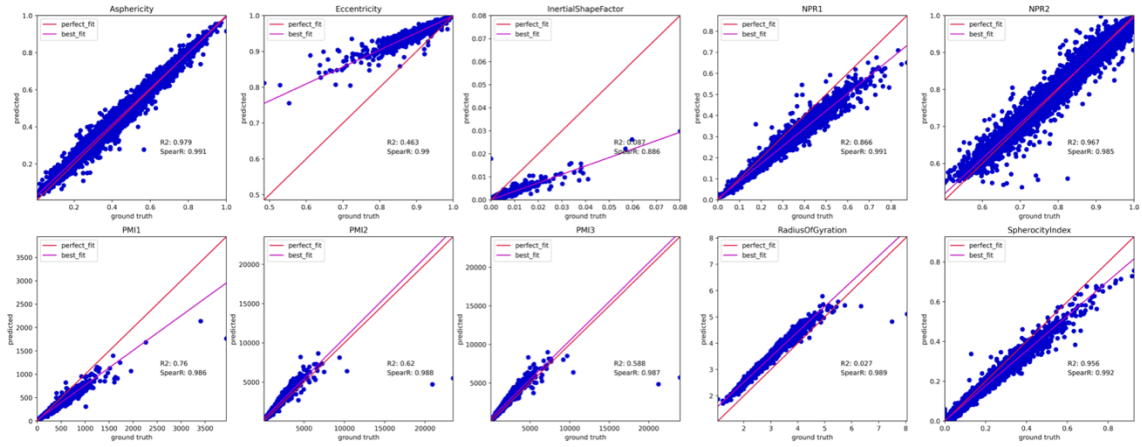
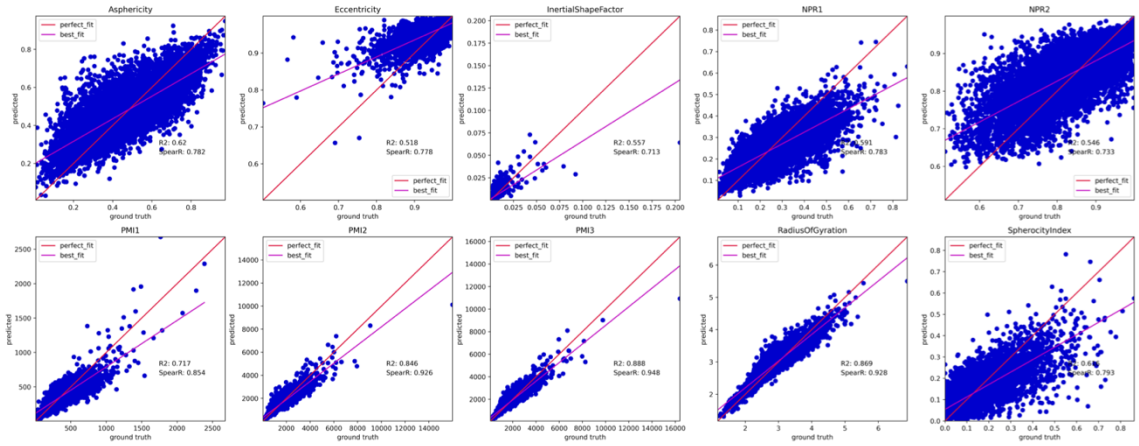
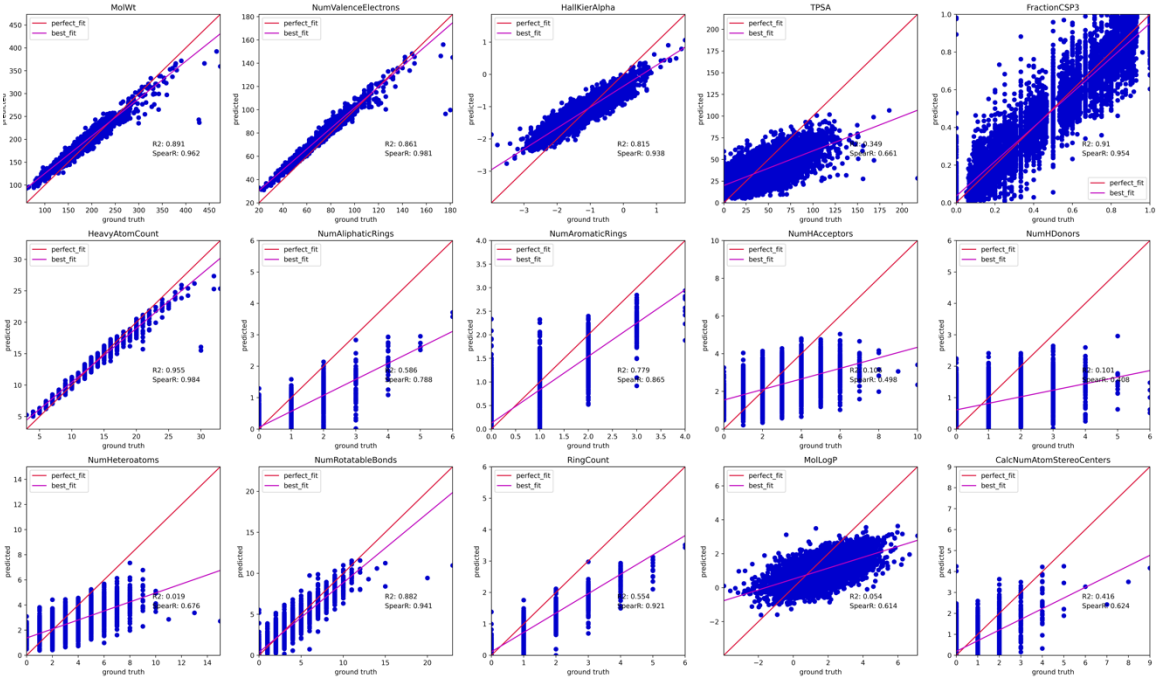


Figure 4.8 *Prediction of 3D shape features by DNN model trained on 100k examples with 1D and 2D cheminformatic descriptors.*



In contrast to the findings above, VNN models struggled to accurately reproduce 1D and 2D molecular properties such as topological polar surface area (TPSA), lipophilicity (MolLogP), and the number of hydrogen-bond donors and acceptors (NumHDonors and NumHAcceptors, respectively). Figure 4.9 shows the prediction performance for this experiment, which essentially amounts to predicting the cheminformatics descriptor vectors that are used as inputs to the 1D models above.

Figure 4.9 *Prediction of 1D & 2D molecular properties by ShapeEncoder-d architecture trained on 100k voxelized examples with MMFF optimization.*



From these results, we infer that 3D models are not able to glean the underlying molecular structure and connectivity from voxel inputs, instead learning largely from their volumetric shapes. We thus hypothesize that the merging of learned representations from our 3D CNNs with other model types, e.g., GNNs, could provide a global representation learning system well-suited for complex modeling tasks like those mentioned above. Such "multimodal representation" networks have been described recently in the context of combined audio-visual input learning.[36] We leave these studies for future works.

4.5 DISCUSSION AND OUTLOOK

There is much room for development based on the work herein. We see several avenues for advancements regarding data, modeling, and applications, each of which are discussed below.

Data. In this work, we focused only on single conformers of each example compound, as it significantly reduces data-storage requirements. However, for this and many other 3D learning tasks, evaluating multiple conformers may be critical. Chemical phenomena such as ligand-target binding,[18] enantioselective catalysis,[35] and structure optimization itself[5] often involve the sampling of many conformers, both (re)active and not. As such, we intend to evaluate the use of conformer ensembles[37] along with our current augmentation techniques to improve on our performance.

Given the Gaussian nature of the B3LYP HL-gap data (see Experimental Section), it was challenging to overcome mean-biased predictions, as is the case in many property prediction tasks. Due to the size of our networks and encodings, we elected to retain all data for the sake of effective modeling. However, it is possible that stratified binning and/or over-sampling of data points at the extrema could prove beneficial for overall error reduction.

Modeling. Perhaps the most pressing avenue for advancing on our current approach is the direct application of equivariant and/or invariant kernel methods. Recent developments have been made in rotationally invariant 3D CNNs,[24] which have been applied successfully for medical imaging[38-40] and texture analysis.[41] Reports thus far have achieved such invariance through the use of rotated convolutional filters,[40] spherical harmonics,[38] and orientational/angular pooling.[24] These methods are directly applicable to our domain, and represent an attractive approach to molecular shape modeling that could avoid the need for rotational augmentation.

Another interesting approach to improve performance is the use of self- or unsupervised pretraining.[42] In fact, our current models using HOMO-LUMO gap

supervision could be used directly as pre-trained models for transfer learning with other tasks. We expect, however, that pre-training with an auxiliary task that allows use of the full PCQCPM6 dataset may lead to more transferable models, particularly for tasks in low-data regimes. We intend to explore this avenue first, with regression of the various QM properties contained in the PCQCPM6 examples,[1] or even larger descriptor vectors[43] as "self-supervision" tasks.[42] Work toward this end has been initiated in collaboration with the Yue Group at Caltech.

Lastly, given the results observed herein, much simpler architecture and hyperparameter adjustments may give some performance boosts. In particular, our QMCVNet is just $1/2$ and $1/8$ the size of ShapeEncoder-d and V-CNN II, respectively, but still performs nearly as well or better in most cases. As such, simply expanding the inception modules or dense layers may prove fruitful, as could neural architecture search.[44] We leave these studies for future work.

Applications. Ultimately, it is our goal to develop 3D CNNs for the understanding of organic reactivity. At the fundamental level, reactions proceed through 3D interactions, and we therefore expect that explicitly modeling the 3D shape of reaction components could lead to better performance for reactivity tasks. These include but are not limited to reaction yield,[45,46] selectivity,[47,48] and condition prediction,[49] and even retrosynthesis planning.[50]

It should be noted that these strategic adjustments may prove necessary to advance VCNN performance beyond the simpler ML methods tested, which proved better suited for the task studied herein. We expect that volumetric encodings may be unreasonably large for this task (particularly in the low-data regime) and for problems where electronic

structure may outweigh steric shape. However, with advancements in high-throughput experimentation[51-53] pre-training methods,[42] and graphics processing and data storage,[18] we expect that 3D learning may become achievable for inherently 3D tasks, and hope that our current work will inform future studies.

4.6 CONCLUSION

Herein, we presented results from 3D deep learning experiments with voxel representations for predicting high-theory-level quantum mechanical properties from rapidly computed force-field geometries. We investigated several state-of-the-art 3D CNN models, and developed our own architecture—QMCVNet—combining concepts from inception networks[20] and 3D shape recognition.[8,25] We found that high capacity models tended to perform best, giving accuracies near semi-empirical quantum methods and with dramatic increase in speed. We presented several opportunities for further improvement on our results, which we are currently investigating. It is our hope that our developed methods and the availability of the PubChemQC PM6 dataset[1] will lead to community engagement to advance research at the intersection of computer vision and chemistry.

4.7 EXPERIMENTAL SECTION

4.7.1 COMPUTATIONAL DETAILS AND HYPERPARAMETERS

Unless specified otherwise, all models were implemented using the TensorFlow deep learning platform[26] and were trained on four NVIDIA® Tesla T4 GPUs. In all experiments, model weights were saved and loaded from the training epoch with the minimum validation set loss, and these models were used for generating predictions and evaluating test set performance. Additional details are contained in the table below.

Table 4.4 *Computational details and general parameters used for all models.*

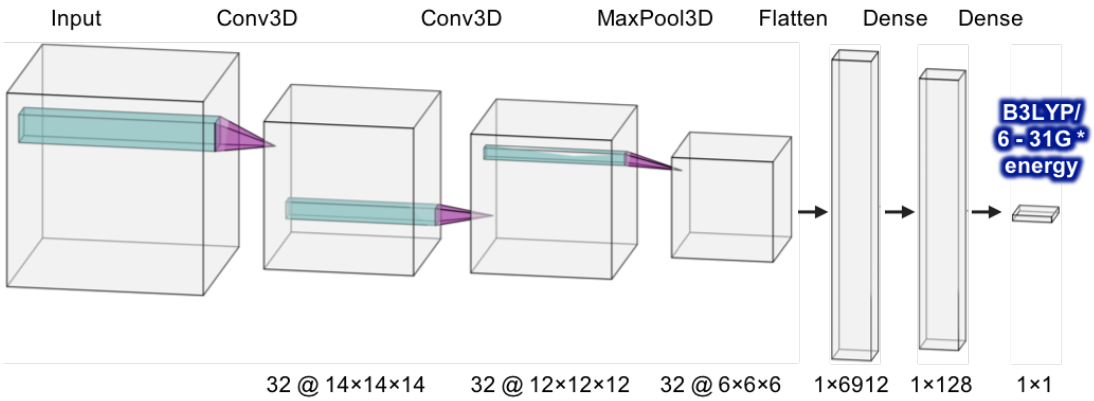
parameter	value	description
loss	mean absolute error (MAE, eV)	loss function used for training
optimizer	Adam	model optimization algorithm[27]
learning rate	$1e^{-3} \rightarrow 1e^{-4}$	linearly decaying learning rate[8]
train/valid/test	80/10/10	data splitting
batch size	32	mini-batch size (per device) used for gradient calculations
epochs	20	number of training epochs (for architecture tests)

4.7.2 MODEL ARCHITECTURES AND IMPLEMENTATIONS

Included below are visual depictions of each VNN architecture tested, as well as model summary outputs and code for their implementation in tensorflow.

4.7.2.1 VoxNet

Figure 4.10 *VoxNet architecture. Uppermost labels represent shorthand layer names. Nethermost labels represent tensor shapes at each layer.*



VoxNet can be implemented as follows:

```
import tensorflow as tf

class VoxNet(tf.keras.Model):
  """
  VoxNet from Maturana and Scherer.
  http://ieeexplore.ieee.org/document/7353481/

  """
  def __init__(self):
    super().__init__()
    self.gauss = tf.keras.initializers.RandomNormal(mean=0, stddev=0.01)
    self.henorm = tf.keras.initializers.HeNormal()

    self.conv1 = tf.keras.layers.Conv3D(
        32, 5, strides=2, kernel_initializer=self.henorm,
        input_shape=(32, 32, 32, 1), data_format='channels_last',
        activation=tf.keras.layers.LeakyReLU(alpha=0.1))
    self.dropout1 = tf.keras.layers.Dropout(0.2)
    self.conv2 = tf.keras.layers.Conv3D(
        32, 3, strides=1, kernel_initializer=self.henorm,
        data_format='channels_last',
        activation=tf.keras.layers.LeakyReLU(alpha=0.1))
    self.dropout2 = tf.keras.layers.Dropout(0.2)
    self.pool1 = tf.keras.layers.MaxPool3D(pool_size=(2, 2, 2))
    self.flatten1 = tf.keras.layers.Flatten()
    self.dense1 = tf.keras.layers.Dense(
        128, kernel_initializer=self.gauss, activation=tf.keras.layers.ReLU())
    self.dropout3 = tf.keras.layers.Dropout(0.2)
    self.dense2 = tf.keras.layers.Dense(
        1, kernel_initializer=self.gauss, activation=tf.keras.layers.ReLU())

  def call(self, inputs):
    x = self.conv1(inputs)
    x = self.dropout1(x)
    x = self.conv2(x)
```

```
        x = self.dropout2(x)
        x = self.pool1(x)
        x = self.flatten1(x)
        x = self.dense1(x)
        x = self.dropout3(x)
        return self.dense2(x)
```

---

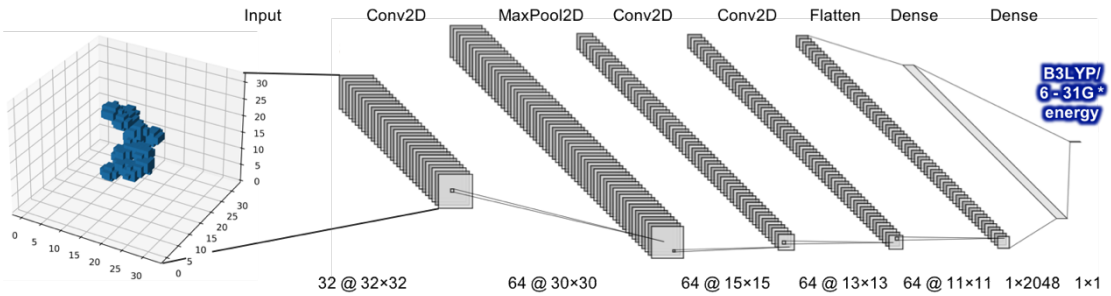
and when instantiated (`model = VoxNet()`) gives the `model.summary` below:

Figure 4.11 *VoxNet architecture summary (from ref. 8).*

```
Model: "VoxNet"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         [(None, 32, 32, 32, 1)]   0
_________________________________________________________________
conv3d (Conv3D)              (None, 14, 14, 14, 32)    4032
_________________________________________________________________
dropout (Dropout)            (None, 14, 14, 14, 32)    0
_________________________________________________________________
conv3d_1 (Conv3D)            (None, 12, 12, 12, 32)    27680
_________________________________________________________________
dropout_1 (Dropout)          (None, 12, 12, 12, 32)    0
_________________________________________________________________
max_pooling3d (MaxPooling3D) (None, 6, 6, 6, 32)       0
_________________________________________________________________
flatten (Flatten)            (None, 6912)              0
_________________________________________________________________
dense (Dense)                (None, 128)               884864
_________________________________________________________________
dropout_2 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 129
=================================================================
Total params: 916,705
Trainable params: 916,705
Non-trainable params: 0
_________________________________________________________________
```

4.7.2.2 V-CNN I

Figure 4.12 *V-CNN I architecture. Uppermost labels represent shorthand layer names. Nethermost labels represent tensor shapes at each layer.*



V-CNN I can be implemented as follows:

```
import tensorflow as tf

class VCNNI(tf.keras.Model):
  """
  V-CNN I from Hegde and Zadeh.
  https://arxiv.org/abs/1607.05695
  """
  def __init__(self):
    super().__init__()

    self.reshape1 = tf.keras.layers.Reshape(
        (32, 32, 32), input_shape=(32, 32, 32, 1))
    self.conv1 = tf.keras.layers.Conv2D(
        64, 3, strides=1, data_format='channels_first',
        activation=tf.keras.layers.ReLU())
    self.pool1 = tf.keras.layers.MaxPool2D(
        pool_size=(2, 2), data_format='channels_first')
    self.conv2 = tf.keras.layers.Conv2D(
        64, 3, strides=1, data_format='channels_first',
        activation=tf.keras.layers.ReLU())
    self.conv3 = tf.keras.layers.Conv2D(
        64, 3, strides=1, data_format='channels_first')
    self.pool2 = tf.keras.layers.MaxPool2D(
        pool_size=(2, 2), data_format='channels_first')
    self.dropout1 = tf.keras.layers.Dropout(0.5)
    self.flatten1 = tf.keras.layers.Flatten()
    self.dense1 = tf.keras.layers.Dense(2048)
    self.dense2 = tf.keras.layers.Dense(1)

  def call(self, inputs):
    x = self.reshape1(inputs)
    x = self.conv1(x)
    x = self.pool1(x)
    x = self.conv2(x)
    x = self.conv3(x)
    x = self.pool2(x)
    x = self.dropout1(x)
    x = self.flatten1(x)
    x = self.dense1(x)
    return self.dense2(x)
```

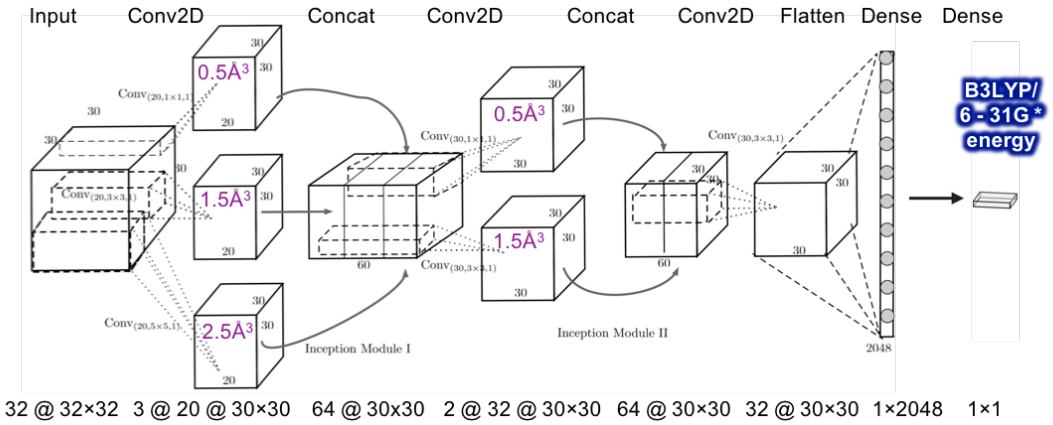
and when instantiated (`model = VCNNI()`) gives the `model.summary` below:

Figure 4.13 *V-CNN I architecture summary (from ref. 25).*

```
Model: "VCNNI"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         [(None, 32, 32, 32, 1)]   0
_________________________________________________________________
reshape (Reshape)            (None, 32, 32, 32)        0
_________________________________________________________________
conv2d (Conv2D)              (None, 64, 30, 30)        18496
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 64, 15, 15)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 64, 13, 13)        36928
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 64, 11, 11)        36928
_________________________________________________________________
max_pooling2d_1 (MaxPooling2D) (None, 64, 5, 5)        0
_________________________________________________________________
dropout (Dropout)            (None, 64, 5, 5)          0
_________________________________________________________________
flatten (Flatten)            (None, 1600)              0
_________________________________________________________________
dense (Dense)                (None, 2048)              3278848
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 2049
=================================================================
Total params: 3,373,249
Trainable params: 3,373,249
Non-trainable params: 0
_________________________________________________________________
```

4.7.2.3 V-CNN II

Figure 4.14 *V-CNN II architecture. Uppermost labels represent shorthand layer names. Nethermost labels represent tensor shapes at each layer.*



V-CNN II can be implemented as follows:

```
import tensorflow as tf

class VCNNII(tf.keras.Model):
```

```
"""
V-CNN II from Hegde and Zadeh.
https://arxiv.org/abs/1607.05695
"""
def __init__(self):
  super().__init__()
  self.reshape1 = tf.keras.layers.Reshape(
      32, 32, 32), input_shape=(32, 32, 32, 1))
  self.conv1 = tf.keras.layers.Conv2D(
      20, 1, strides=1, padding='same', data_format='channels_first')
  self.conv2 = tf.keras.layers.Conv2D(
      20, 3, strides=1, padding='same', data_format='channels_first')
  self.conv3 = tf.keras.layers.Conv2D(
      20, 5, strides=1, padding='same', data_format='channels_first')
  self.concat1 = tf.keras.layers.Concatenate(axis=1)
  self.act1 = tf.keras.layers.ReLU()
  self.dropout1 = tf.keras.layers.Dropout(0.2)
  self.conv4 = tf.keras.layers.Conv2D(
      32, 1, strides=1, padding='same', data_format='channels_first')
  self.conv5 = tf.keras.layers.Conv2D(
      32, 3, strides=1, padding='same', data_format='channels_first')
  self.concat2 = tf.keras.layers.Concatenate(axis=1)
  self.act2 = tf.keras.layers.ReLU()
  self.dropout2 = tf.keras.layers.Dropout(0.3)
  self.conv6 = tf.keras.layers.Conv2D(
      32, 3, strides=1, padding='same', data_format='channels_first')
  self.act3 = tf.keras.layers.ReLU()
  self.dropout3 = tf.keras.layers.Dropout(0.5)
  self.flatten1 = tf.keras.layers.Flatten()
  self.dense1 = tf.keras.layers.Dense(2048)
  self.dense2 = tf.keras.layers.Dense(1)

def call(self, inputs):
  x = self.reshape1(inputs)
  x1 = self.conv1(x)
  x3 = self.conv2(x)
  x5 = self.conv3(x)
  x = self.concat1([x1, x3, x5])
  x = self.act1(x)
  x = self.dropout1(x)
  x1 = self.conv4(x)
  x3 = self.conv5(x)
  x = self.concat2([x1, x3])
  x = self.act2(x)
  x = self.dropout2(x)
  x = self.conv6(x)
  x = self.act3(x)
  x = self.dropout3(x)
  x = self.flatten1(x)
  x = self.dense1(x)
  return self.dense2(x)
```

---

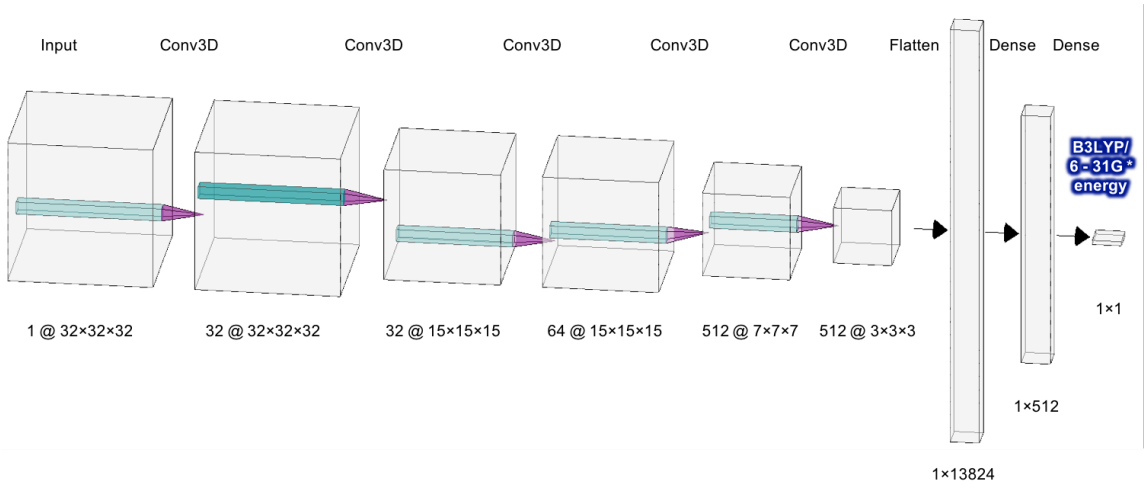
and when instantiated (`model = VCNNII()`) gives the `model.summary` below:

Figure 4.15 *V-CNN II architecture summary (from ref. 25).*

```
Model: "VCNNII"
_________________________________________________________________________________________________
Layer (type)                 Output Shape              Param #       Connected to
=================================================================================================
input_2 (InputLayer)         [(None, 32, 32, 32, 1)]   0
_________________________________________________________________________________________________
reshape (Reshape)            (None, 32, 32, 32)        0             input_2[0][0]
_________________________________________________________________________________________________
conv2d (Conv2D)              (None, 20, 32, 32)        660           reshape[1][0]
_________________________________________________________________________________________________
conv2d_1 (Conv2D)            (None, 20, 32, 32)        5780          reshape[1][0]
_________________________________________________________________________________________________
conv2d_2 (Conv2D)            (None, 20, 32, 32)        16020         reshape[1][0]
_________________________________________________________________________________________________
concatenate (Concatenate)    (None, 60, 32, 32)        0             conv2d[1][0]
                                                                     conv2d_1[1][0]
                                                                     conv2d_2[1][0]
_________________________________________________________________________________________________
re_lu (ReLU)                 (None, 60, 32, 32)        0             concatenate[1][0]
_________________________________________________________________________________________________
dropout (Dropout)            (None, 60, 32, 32)        0             re_lu[1][0]
_________________________________________________________________________________________________
conv2d_3 (Conv2D)            (None, 32, 32, 32)        1952          dropout[1][0]
_________________________________________________________________________________________________
conv2d_4 (Conv2D)            (None, 32, 32, 32)        17312         dropout[1][0]
_________________________________________________________________________________________________
concatenate_1 (Concatenate)  (None, 64, 32, 32)        0             conv2d_3[1][0]
                                                                     conv2d_4[1][0]
_________________________________________________________________________________________________
re_lu_1 (ReLU)               (None, 64, 32, 32)        0             concatenate_1[1][0]
_________________________________________________________________________________________________
dropout_1 (Dropout)          (None, 64, 32, 32)        0             re_lu_1[1][0]
_________________________________________________________________________________________________
conv2d_5 (Conv2D)            (None, 32, 32, 32)        18464         dropout_1[1][0]
_________________________________________________________________________________________________
re_lu_2 (ReLU)               (None, 32, 32, 32)        0             conv2d_5[1][0]
_________________________________________________________________________________________________
dropout_2 (Dropout)          (None, 32, 32, 32)        0             re_lu_2[1][0]
_________________________________________________________________________________________________
flatten (Flatten)            (None, 32768)             0             dropout_2[1][0]
_________________________________________________________________________________________________
dense (Dense)                (None, 2048)              67110912      flatten[1][0]
_________________________________________________________________________________________________
dense_1 (Dense)              (None, 1)                 2049          dense[1][0]
=================================================================================================
Total params: 67,173,149
Trainable params: 67,173,149
Non-trainable params: 0
_________________________________________________________________________________________________
```

4.7.2.4 ShapeEncoder-d

Figure 4.16 *ShapeEncoder-d architecture. Uppermost labels represent shorthand layer names. Nethermost labels represent tensor shapes at each layer.*



ShapeEncoder-d can be implemented as follows:

```
import tensorflow as tf

class ShapeEncoder_d(tf.keras.Model):
  """
  Encoder network from Skalic et al.
  https://pubs.acs.org/doi/10.1021/acs.jcim.8b00706
  Added dropout to prevent overfitting.
  """
  def __init__(self):
    super().__init__()

    self.conv1 = tf.keras.layers.Conv3D(
        32, 3, strides=1, padding='same', input_shape=(32, 32, 32, 1),
        activation='relu', data_format='channels_last')
    self.conv2 = tf.keras.layers.Conv3D(
        32, 3, strides=2, activation='relu', data_format='channels_last')
    self.dropout1 = tf.keras.layers.Dropout(0.2)
    self.conv3 = tf.keras.layers.Conv3D(
        64, 3, strides=1, padding='same', activation='relu',
        data_format='channels_last')
    self.conv4 = tf.keras.layers.Conv3D(
        512, 3, strides=2, activation='relu', data_format='channels_last')
    self.dropout2 = tf.keras.layers.Dropout(0.2)
    self.conv5 = tf.keras.layers.Conv3D(
        512, 3, strides=2, activation='relu', data_format='channels_last')
    self.flatten1 = tf.keras.layers.Flatten()
    self.dense1 = tf.keras.layers.Dense(512, activation='relu')
    self.dropout3 = tf.keras.layers.Dropout(0.2)
    self.dense2 = tf.keras.layers.Dense(1, activation='relu')

  def call(self, inputs):
    x = self.conv1(inputs)
```

```
        x = self.conv2(x)
        x = self.dropout1(x)
        x = self.conv3(x)
        x = self.conv4(x)
        x = self.dropout2(x)
        x = self.conv5(x)
        x = self.flatten1(x)
        x = self.dense1(x)
        x = self.dropout3(x)
        return self.dense2(x)
```

---

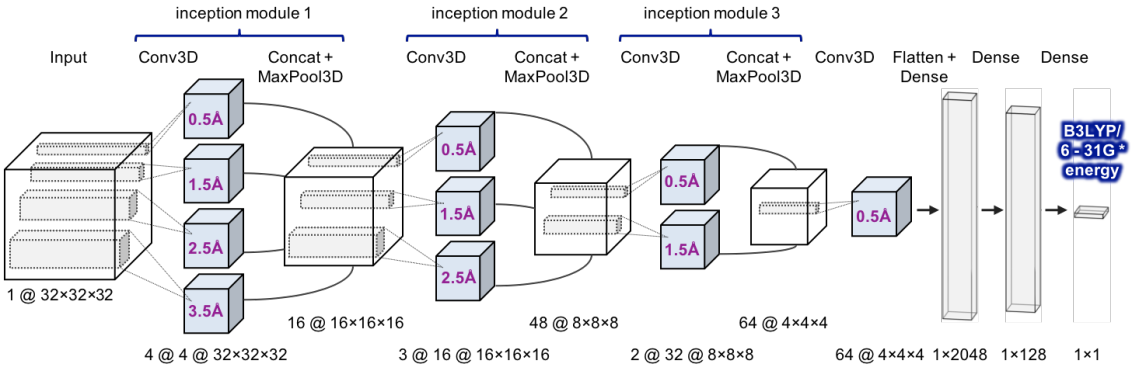
and when instantiated (`model = ShapeEncoder_d()`) gives the `model.summary` below:

Figure 4.17 *ShapeEncoder-d architecture summary (from ref. 17) with the addition of dropout.*

```
Model: "ShapeEncoder_d"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         [(None, 32, 32, 32, 1)]   0
_________________________________________________________________
conv3d (Conv3D)              (None, 32, 32, 32, 32)    896
_________________________________________________________________
conv3d_1 (Conv3D)            (None, 15, 15, 15, 32)    27680
_________________________________________________________________
dropout (Dropout)            (None, 15, 15, 15, 32)    0
_________________________________________________________________
conv3d_2 (Conv3D)            (None, 15, 15, 15, 64)    55360
_________________________________________________________________
conv3d_3 (Conv3D)            (None, 7, 7, 7, 512)      885248
_________________________________________________________________
dropout_1 (Dropout)          (None, 7, 7, 7, 512)      0
_________________________________________________________________
conv3d_4 (Conv3D)            (None, 3, 3, 3, 512)      7078400
_________________________________________________________________
flatten (Flatten)            (None, 13824)             0
_________________________________________________________________
dense (Dense)                (None, 512)               7078400
_________________________________________________________________
dropout_2 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 513
=================================================================
Total params: 15,126,497
Trainable params: 15,126,497
Non-trainable params: 0
_________________________________________________________________
```

4.7.2.5 QMCVNet

Figure 4.18 *QMCVNet architecture. Uppermost labels represent shorthand layer names. Nethermost labels represent tensor shapes at each layer. Å labels represent the kernel size on one edge used to obtain each intermediate representation in the inception modules.*



QMCVNet can be implemented as follows (note that this has been modularized compared to the architectures above to support pre-training, where learned representations can be transferred to downstream tasks by the `get_embeddings()` method):

```
import tensorflow as tf

class QMCVNet(tf.keras.Model):
  """
  QMCVNet from Maser and Reisman (this work).

  Channels, targets, and embedding dimension made flexible.
  Embedding and inference modules sliced out for pretraining/transfer learning.
  """
  def __init__(self, channel, n_targets, embed_dim):
    super().__init__()

    # First inception block
    self.conv1 = tf.keras.layers.Conv3D(
        channel, 1, strides=1, padding='same', data_format='channels_last',
        input_shape=(32, 32, 32, 1))
    self.conv2 = tf.keras.layers.Conv3D(
        channel, 3, strides=1, padding='same', data_format='channels_last')
    self.conv3 = tf.keras.layers.Conv3D(
        channel, 5, strides=1, padding='same', data_format='channels_last')
    self.conv4 = tf.keras.layers.Conv3D(
        channel, 7, strides=1, padding='same', data_format='channels_last')
    self.concat1 = tf.keras.layers.Concatenate(axis=-1)
    self.act1 = tf.keras.layers.ReLU()
    self.dropout1 = tf.keras.layers.Dropout(0.2)
    self.pool1 = tf.keras.layers.MaxPool3D(
        pool_size=(2, 2, 2), strides=2, data_format='channels_last')
```

```
        # Second inception block
        self.conv5 = tf.keras.layers.Conv3D(
            channel*4, 1, strides=1, padding='same', data_format='channels_last')
        self.conv6 = tf.keras.layers.Conv3D(
            channel*4, 3, strides=1, padding='same', data_format='channels_last')
        self.conv7 = tf.keras.layers.Conv3D(
            channel*4, 5, strides=1, padding='same', data_format='channels_last')
        self.concat2 = tf.keras.layers.Concatenate(axis=-1)
        self.act2 = tf.keras.layers.ReLU()
        self.dropout2 = tf.keras.layers.Dropout(0.2)
        self.pool2 = tf.keras.layers.MaxPool3D(
            pool_size=(2, 2, 2), strides=2, data_format='channels_last')

        # Third inception block
        self.conv8 = tf.keras.layers.Conv3D(
            channel*8, 1, strides=1, padding='same', data_format='channels_last')
        self.conv9 = tf.keras.layers.Conv3D(
            channel*8, 3, strides=1, padding='same', data_format='channels_last')
        self.concat3 = tf.keras.layers.Concatenate(axis=-1)
        self.act3 = tf.keras.layers.ReLU()
        self.dropout3 = tf.keras.layers.Dropout(0.2)
        self.pool3 = tf.keras.layers.MaxPool3D(
            pool_size=(2, 2, 2), strides=2, data_format='channels_last')

        # Embedding
        self.conv10 = tf.keras.layers.Conv3D(
            channel*16, 1, strides=1, padding='same', data_format='channels_last')
        self.act4 = tf.keras.layers.ReLU()
        self.dropout4 = tf.keras.layers.Dropout(0.2)
        self.flatten1 = tf.keras.layers.Flatten()
        self.dense1 = tf.keras.layers.Dense(embed_dim, activation='relu')

        # Inference
        self.dense2 = tf.keras.layers.Dense(embed_dim*8, activation='relu')
        self.dense3 = tf.keras.layers.Dense(128, activation='relu')
        self.dense4 = tf.keras.layers.Dense(n_targets)

    def get_embeddings(self, inputs):
        x = inputs
        x1 = self.conv1(x)
        x3 = self.conv2(x)
        x5 = self.conv3(x)
        x7 = self.conv4(x)
        x = self.concat1([x1, x3, x5, x7])
        x = self.act1(x)
        x = self.dropout1(x)
        x = self.pool1(x)
        x1 = self.conv5(x)
        x3 = self.conv6(x)
        x5 = self.conv7(x)
        x = self.concat2([x1, x3, x5])
        x = self.act2(x)
        x = self.dropout2(x)
        x = self.pool2(x)
        x1 = self.conv8(x)
        x3 = self.conv9(x)
        x = self.concat3([x1, x3])
        x = self.act3(x)
        x = self.dropout3(x)
        x = self.pool3(x)
        x = self.conv10(x)
        x = self.act4(x)
```

```
    x = self.dropout4(x)
    x = self.flatten1(x)
    return self.dense1(x)

  def run_inference(self, embeddings):
    x = self.dense2(embeddings)
    x = self.dense3(x)
    return self.dense4(x)

  def call(self, inputs):
    x = self.get_embeddings(inputs)
    return self.run_inference(x)
```

---

and when instantiated (`model = QMCVNet()`) gives the `model.summary` below:

Figure 4.19 *QMCVNet architecture summary (from this work).*

```
Model: "QMCVNet"
_________________________________________________________________________________________________
Layer (type)                 Output Shape              Param #       Connected to
=================================================================================================
input_2 (InputLayer)         [(None, 32, 32, 32, 1)]   0
_________________________________________________________________________________________________
conv3d (Conv3D)              (None, 32, 32, 32, 4)     8             input_2[0][0]
_________________________________________________________________________________________________
conv3d_1 (Conv3D)            (None, 32, 32, 32, 4)     112           input_2[0][0]
_________________________________________________________________________________________________
conv3d_2 (Conv3D)            (None, 32, 32, 32, 4)     504           input_2[0][0]
_________________________________________________________________________________________________
conv3d_3 (Conv3D)            (None, 32, 32, 32, 4)     1376          input_2[0][0]
_________________________________________________________________________________________________
concatenate (Concatenate)    (None, 32, 32, 32, 16)    0             conv3d[1][0]
                                                                     conv3d_1[1][0]
                                                                     conv3d_2[1][0]
                                                                     conv3d_3[1][0]
_________________________________________________________________________________________________
re_lu (ReLU)                 (None, 32, 32, 32, 16)    0             concatenate[1][0]
_________________________________________________________________________________________________
dropout (Dropout)            (None, 32, 32, 32, 16)    0             re_lu[1][0]
_________________________________________________________________________________________________
max_pooling3d (MaxPooling3D) (None, 16, 16, 16, 16)    0             dropout[1][0]
_________________________________________________________________________________________________
conv3d_4 (Conv3D)            (None, 16, 16, 16, 16)    272           max_pooling3d[1][0]
_________________________________________________________________________________________________
conv3d_5 (Conv3D)            (None, 16, 16, 16, 16)    6928          max_pooling3d[1][0]
_________________________________________________________________________________________________
conv3d_6 (Conv3D)            (None, 16, 16, 16, 16)    32016         max_pooling3d[1][0]
_________________________________________________________________________________________________
concatenate_1 (Concatenate)  (None, 16, 16, 16, 48)    0             conv3d_4[1][0]
                                                                     conv3d_5[1][0]
                                                                     conv3d_6[1][0]
_________________________________________________________________________________________________
re_lu_1 (ReLU)               (None, 16, 16, 16, 48)    0             concatenate_1[1][0]
_________________________________________________________________________________________________
dropout_1 (Dropout)          (None, 16, 16, 16, 48)    0             re_lu_1[1][0]
_________________________________________________________________________________________________
max_pooling3d_1 (MaxPooling3D) (None, 8, 8, 8, 48)     0             dropout_1[1][0]
_________________________________________________________________________________________________
conv3d_7 (Conv3D)            (None, 8, 8, 8, 32)       1568          max_pooling3d_1[1][0]
_________________________________________________________________________________________________
conv3d_8 (Conv3D)            (None, 8, 8, 8, 32)       41504         max_pooling3d_1[1][0]
_________________________________________________________________________________________________
concatenate_2 (Concatenate)  (None, 8, 8, 8, 64)       0             conv3d_7[1][0]
                                                                     conv3d_8[1][0]
_________________________________________________________________________________________________
re_lu_2 (ReLU)               (None, 8, 8, 8, 64)       0             concatenate_2[1][0]
_________________________________________________________________________________________________
dropout_2 (Dropout)          (None, 8, 8, 8, 64)       0             re_lu_2[1][0]
_________________________________________________________________________________________________
max_pooling3d_2 (MaxPooling3D) (None, 4, 4, 4, 64)     0             dropout_2[1][0]
_________________________________________________________________________________________________
conv3d_9 (Conv3D)            (None, 4, 4, 4, 64)       4160          max_pooling3d_2[1][0]
_________________________________________________________________________________________________
re_lu_3 (ReLU)               (None, 4, 4, 4, 64)       0             conv3d_9[1][0]
_________________________________________________________________________________________________
dropout_3 (Dropout)          (None, 4, 4, 4, 64)       0             re_lu_3[1][0]
_________________________________________________________________________________________________
flatten (Flatten)            (None, 4096)              0             dropout_3[1][0]
_________________________________________________________________________________________________
dense (Dense)                (None, 2048)              8390656       flatten[1][0]
_________________________________________________________________________________________________
dense_1 (Dense)              (None, 128)               262272        dense[1][0]
_________________________________________________________________________________________________
dense_2 (Dense)              (None, 1)                 129           dense_1[1][0]
=================================================================================================
Total params: 8,741,505
Trainable params: 8,741,505
Non-trainable params: 0
_________________________________________________________________________________________________
```

4.7.3 VOXELIZATION PARAMETER SCREEN

To evaluate the effects of voxel value decay on the learnability of our 3D representations, we devised a screen of the decay parameter settings. The RDKit package[22]

implements voxel encoding in terms of "shells," the first being a sphere about each atom center and additional shells layering from the previous outer edge (see main text and associated code for function details). The core shell radius can be set as a fraction of the van der Waals radius (VDWR) of the atom in question, which gives an implicit mechanism of encoding atom type. The width of the additional shells can again be set as a function of the atom's VDWR, and the maximum number of layers can also be specified. In the voxelized encoding, the core shell takes the maximum grid value, with stepwise linear decay at each of the outer shells.

Using the 100k dataset (see main text), we experimented with three core shell sizes ("core_VDWR_multiplier") and three additional shell widths ("shell_VDWR_multiplier"), all to a maximum of 4 total layers using MMFF-optimized geometries with all five architectures. Mean absolute error ("test_MAE"), mean absolute percentage error ("test_MAPE"), and correlation coefficients between predictions and ground truth HOMO-LUMO gaps ("test_R2") were evaluated for a heldout test set (10%). The results of this screen along with example voxel renderings are shown below.

Table 4.5 *Atom-centered voxel value decay parameter screen results.*

entry	core_VDWR_ multiplier	shell_VDWR_ multiplier	model	test_MAE	test_MAPE	test_R2
1	0.2	0.100	QMCVNet	0.5426	10.50	0.5738
2	0.3	0.150	QMCVNet	0.5487	10.60	0.5668
3	0.4	0.200	ShapeEncoder-d	0.5508	10.62	0.5597
4	0.2	0.200	QMCVNet	0.5523	10.55	0.5666
5	0.3	0.150	ShapeEncoder-d	0.5524	10.68	0.5571
6	0.4	0.100	QMCVNet	0.5529	10.58	0.5619
7	0.3	0.075	QMCVNet	0.5530	10.72	0.5587
8	0.2	0.050	QMCVNet	0.5585	10.56	0.5588
9	0.3	0.300	ShapeEncoder-d	0.5594	10.60	0.5530

10	0.2	0.100	ShapeEncoder-d	0.5624	10.84	0.5446
11	0.2	0.200	ShapeEncoder-d	0.5626	10.82	0.5436
12	0.3	0.075	ShapeEncoder-d	0.5639	11.00	0.5376
13	0.4	0.400	ShapeEncoder-d	0.5666	10.78	0.5431
14	0.3	0.300	QMCVNet	0.5706	10.63	0.5562
15	0.4	0.200	QMCVNet	0.5711	10.68	0.5470
16	0.2	0.050	ShapeEncoder-d	0.5783	11.17	0.5109
17	0.2	0.200	V-CNN II	0.5966	11.51	0.4893
18	0.2	0.100	V-CNN II	0.5993	11.27	0.4913
19	0.4	0.400	QMCVNet	0.6042	10.98	0.4954
20	0.4	0.200	V-CNN I	0.6088	11.40	0.4672
21	0.3	0.075	V-CNN II	0.6107	11.37	0.4673
22	0.3	0.300	V-CNN II	0.6129	11.33	0.4565
23	0.3	0.150	V-CNN II	0.6139	11.31	0.4595
24	0.4	0.100	V-CNN II	0.6171	11.34	0.4536
25	0.3	0.300	V-CNN I	0.6199	11.69	0.4405
26	0.4	0.400	V-CNN I	0.6281	11.80	0.4332
27	0.4	0.100	V-CNN I	0.6327	11.93	0.4238
28	0.3	0.150	V-CNN I	0.6352	11.95	0.4194
29	0.2	0.200	V-CNN I	0.6385	11.99	0.4085
30	0.2	0.050	V-CNN II	0.6446	11.85	0.4004
31	0.4	0.400	V-CNN II	0.6455	11.82	0.3900
32	0.3	0.075	V-CNN I	0.6459	12.09	0.4049
33	0.4	0.200	V-CNN II	0.6543	11.70	0.3949
34	0.2	0.050	V-CNN I	0.6655	12.45	0.3614
35	0.2	0.100	V-CNN I	0.6705	12.52	0.3593
36	0.4	0.100	ShapeEncoder-d	0.8451	15.53	-0.0128
37	0.4	0.200	VoxNet	1.1737	20.14	-0.5251
38	0.4	0.400	VoxNet	1.2083	20.68	-0.6242
39	0.4	0.100	VoxNet	1.2085	20.77	-0.6218
40	0.3	0.300	VoxNet	1.2560	21.36	-0.7705
41	0.3	0.150	VoxNet	1.2769	21.59	-0.8773
42	0.2	0.200	VoxNet	1.3554	23.45	-0.9615
43	0.3	0.075	VoxNet	1.4965	25.84	-1.3379
44	0.2	0.100	VoxNet	1.5350	26.60	-1.4337
45	0.2	0.050	VoxNet	1.5698	27.18	-1.5517

Figure 4.20 *Example compounds from the 100k dataset encoded with core VDWR multiplier = 0.4 and shell width multiplier = 0.2.*

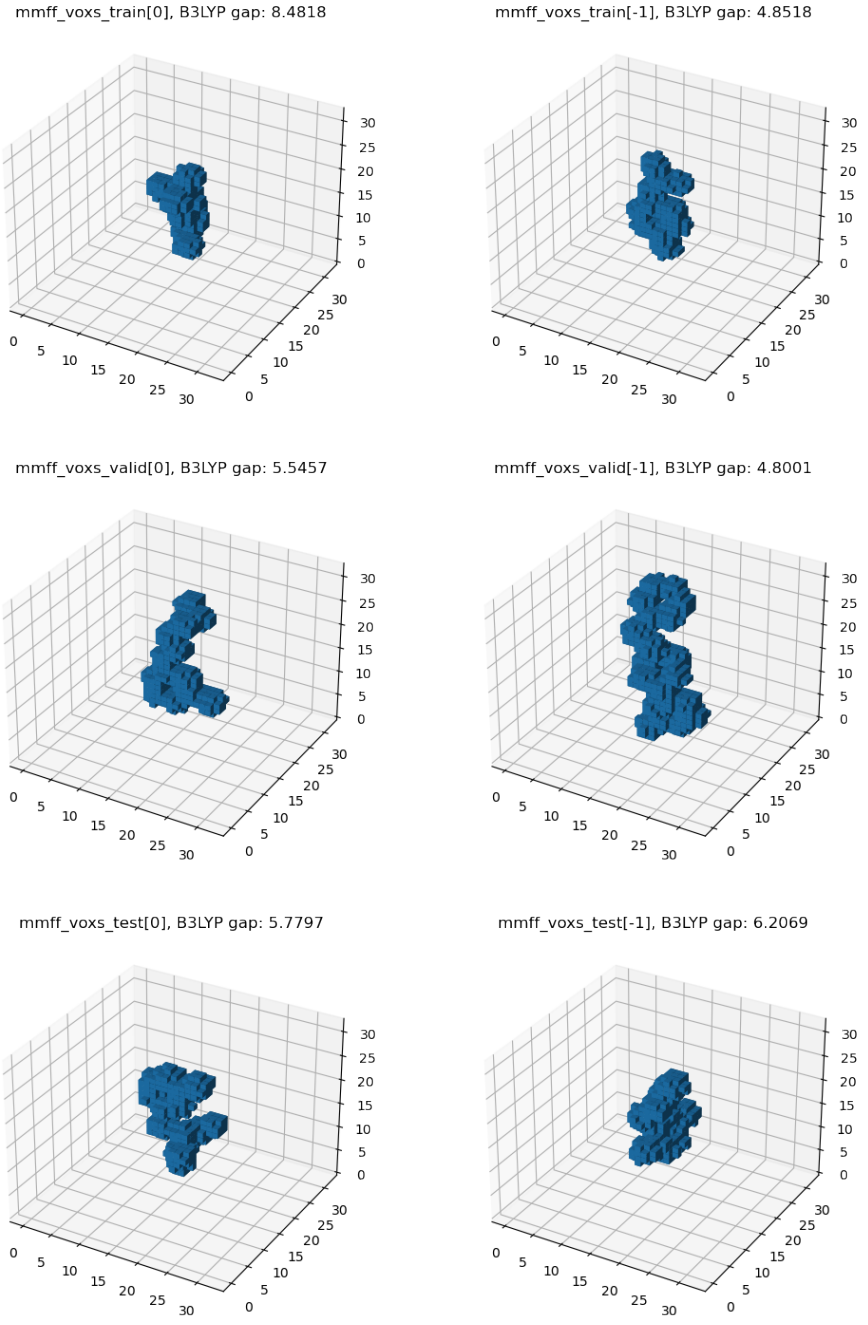


Figure 4.21 *Example compounds from the 100k dataset encoded with core VDWR multiplier = 0.3 and shell width multiplier = 0.15.*

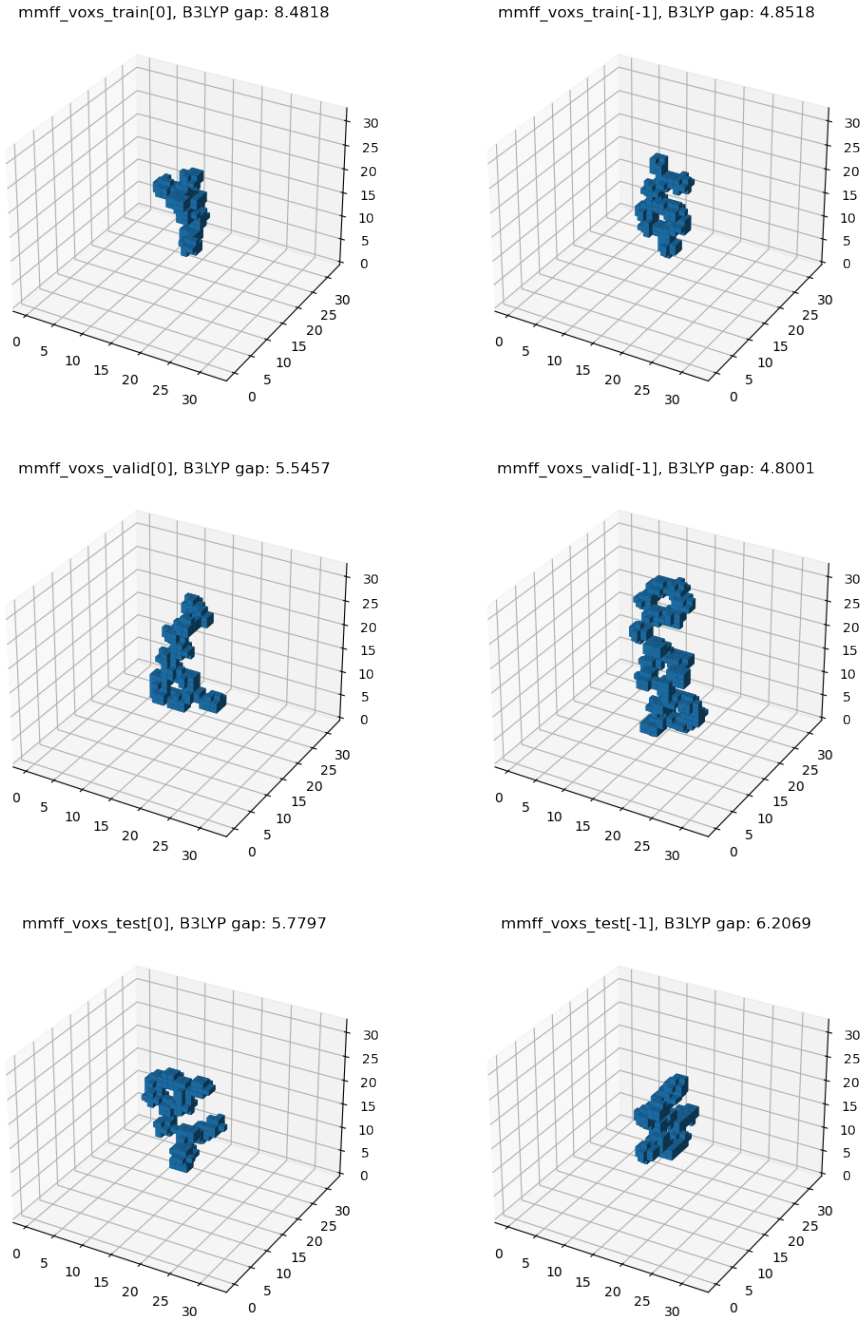
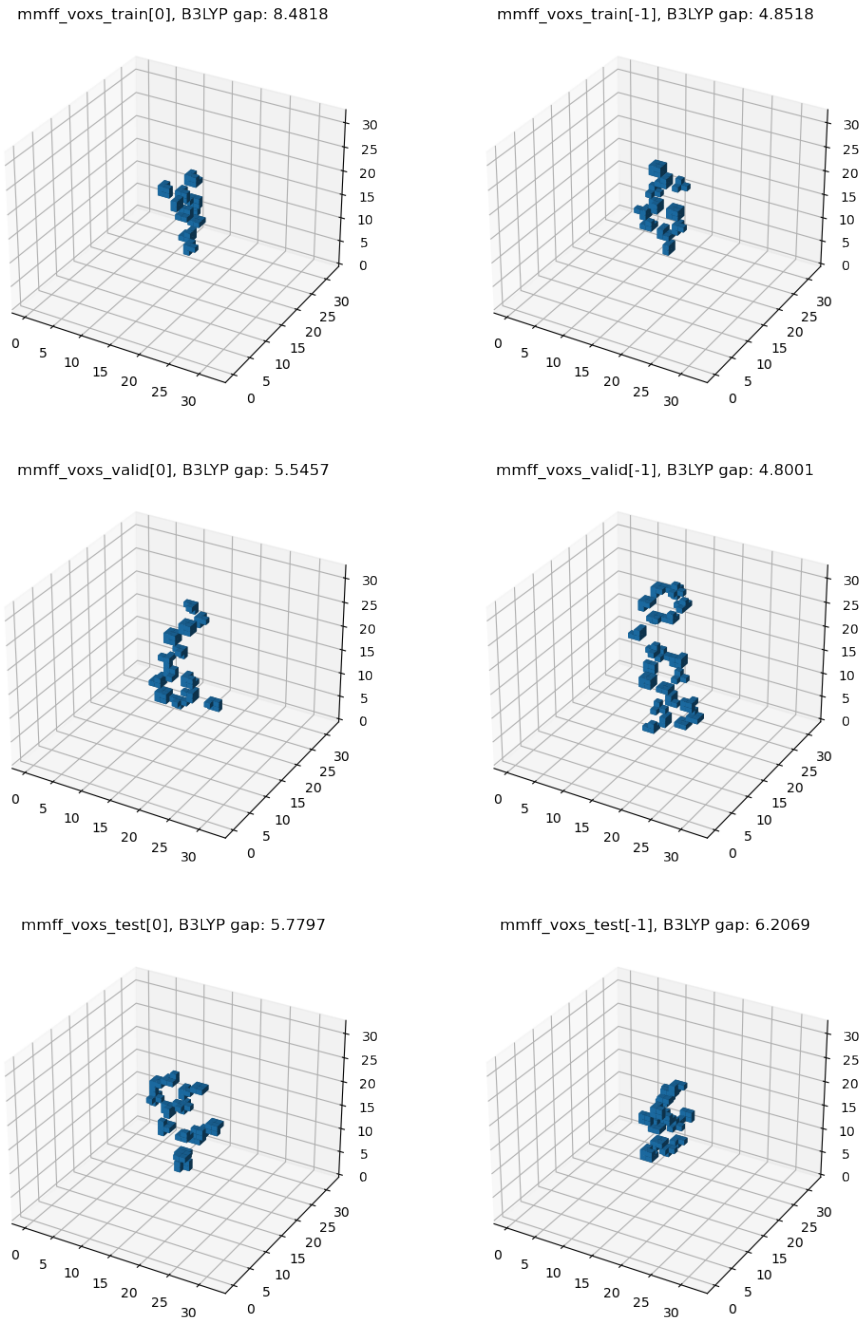


Figure 4.22 *Example compounds from the 100k dataset encoded with core VDWR multiplier = 0.2 and shell width multiplier = 0.1.*



These voxel visualizations can be rendered individually using the following code:

```
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import numpy as np
```

```
# Visualize an example voxel array
fig, ax = plt.subplots(figsize=(8, 8), subplot_kw={'projection': '3d'})
vox = np.random.randint(0, high=4, size=(32, 32, 32, 1))  # Default vox shape
ax.voxels(vox[:, :, :, 0])
plt.show()
```

4.7.4 MODELING RESULTS: SCALEUP AND EXTENDED TRAINING

The best performing architectures from the 100k dataset experiments (see Section 4.7.3 and main text) were retrained with the 1M dataset. The MMFF models were additionally allowed to fully converge over 100 epochs. Their tabulated performance metrics can be found in the main text, and their predictions are shown below along with the baseline PM6 semi-empirical method.

Figure 4.23 *Predicted vs. ground truth B3LYP HOMO-LUMO gaps for QMCVNet (top) and ShapeEncoder-d (bottom) trained on the 1M MMFF dataset for 100 epochs.*

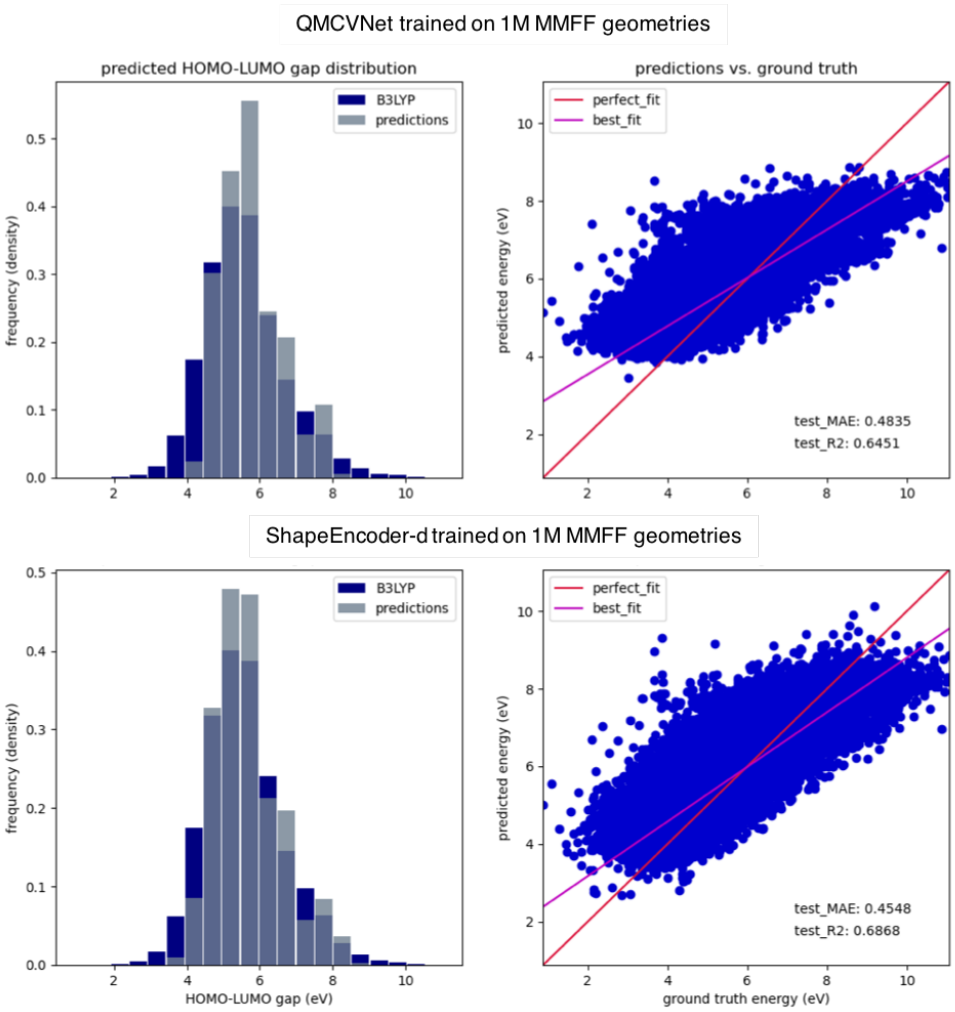
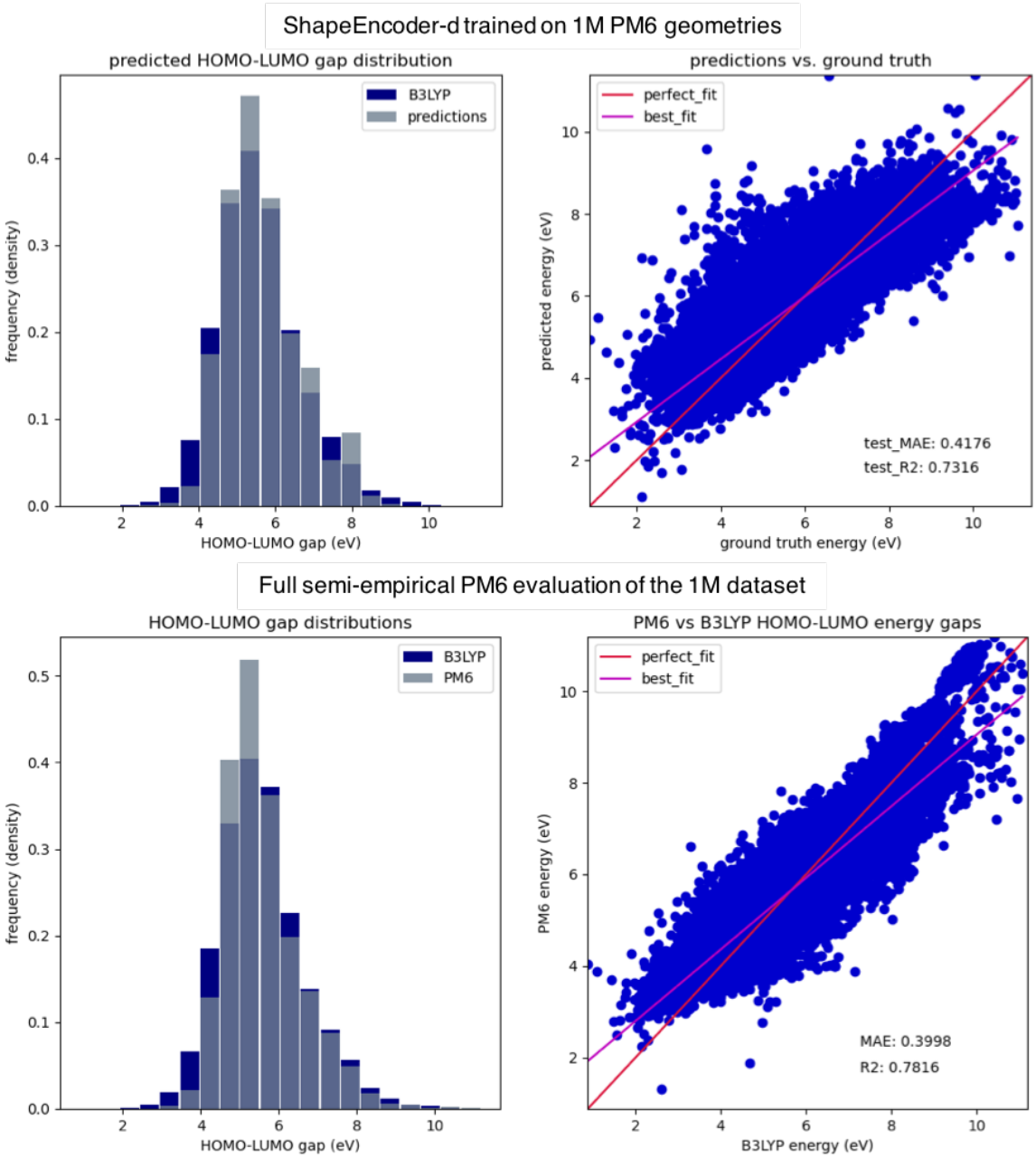


Figure 4.24 *Predicted vs. ground truth B3LYP HOMO-LUMO gaps for the overall best model (ShapeEncoder-d + PM6, 20 epochs) vs. the PM6 method.*



4.7.5 ADDITIONAL SAMPLE CODE

Voxelization protocols were implemented in Python utilizing the RDKit cheminformatics package.[22] A sample function for converting a molecule's SMILES string representation to a voxelized array is included below:

```
from typing import Tuple
import numpy as np

from rdkit import Chem
from rdkit.Chem import AllChem, rdDistGeom, rdForceFieldHelpers
from rdkit.Geometry import rdGeometry
from rdkit.DataStructs import cDataStructs

def smiles_to_voxels(
    smiles: str, box_dim: Tuple[int, int, int] = (16, 16, 16),
    spacing: float = 0.5, vdw_scale: float = 0.3, shell_scale: float = 0.15,
    n_shells: int = 4, num_confs: int = 5, rms_thresh: float = 0.1,
    reshape: bool = True, return_mol: bool = False) -> np.ndarray:
  """Creates voxel grid encodings of molecular volumes.

  Args:
    smiles: SMILES string of the molecule to voxelize.
    box_dim: Shape of voxel cube in Angstroms.
    spacing: Size of each voxel edge in Angstroms.
    vdw_scale: Fraction of van der Waals radius to include in the core shell.
    shell_scale: Fraction of VDWR to include in outer shells.
    n_shells: The total number of shells to encode.
    num_confs: The number of conformers to screen for the embedding.
    rms_thresh: The threshold similarity to retain conformers.
    reshape: Whether to add a batch dimension to the voxel array.
    return_mol: Whether to return the mol object with the voxel array.

  Returns:
    np array containing the voxelized encoding and (optionally) the mol object.
  """
  mol = Chem.MolFromSmiles(smiles)

  # Embeds the molecule and runs MMFF optimization (note some molecules fail)
  try:
    conformers = rdDistGeom.EmbedMultipleConfs(
        mol, numConfs=num_confs, pruneRmsThresh=rms_thresh,
        useSmallRingTorsions=True, useMacrocycleTorsions=True, numThreads=0)

    # Embed conformers, keep track of lowest energy
    min_energy = None
    min_conf = None
    mp = AllChem.MMFFGetMoleculeProperties(mol)

    for conf in conformers:
      # Force field optimize new conformers and calculate energies
      rdForceFieldHelpers.MMFFOptimizeMolecule(mol, confId=conf)
      ff = AllChem.MMFFGetMoleculeForceField(mol, mp, confId=conf)
      conf_energy = ff.CalcEnergy()
      if min_energy is None or (conf_energy < min_energy):
          min_energy = conf_energy
          min_conf = conf

    # Prepares voxel grids and encodes shape
    # Could rework this to create grid for each atom type (C, N, O, etc.)
    grid = rdGeometry.UniformGrid3D(*box_dim, spacing=spacing)
    Chem.rdShapeHelpers.EncodeShape(
        mol, grid, confId=min_conf, vdwScale=vdw_scale,
        stepSize=shell_scale, maxLayers=n_shells)

    # Set voxel array and embed grid
    vox_dim = tuple(int(d/spacing) for d in box_dim)
```

```
    voxels = np.zeros((vox_dim[0]**3,), dtype=np.int32)
    cDataStructs.ConvertToNumpyArray(grid.GetOccupancyVect(), voxels)
    voxels = np.reshape(voxels, (*vox_dim, 1))
  except:
    if return_mol:
      return None, None
    return

  # Return if array was not filled
  if (voxels.max() == 0) or (np.isnan(voxels).any()):
    if return_mol:
      return None, None
    return

  # Adds channel dimension to be passed to tf datasets
  # Could rework this to stack conformers as channels
  if reshape:
    voxels = np.expand_dims(voxels, axis=0)

  # In some cases will want the mol object for additional calculations
  if return_mol:
    return voxels, mol

  return voxels
```

Storing large sets of voxelized compounds can be data intensive; 100k molecules at 32^3 resolution requires roughly 13 GB of RAM. Even in cases where the datasets can be stored in memory, wrapping such examples as, e.g., `tf.data.Dataset` objects can duplicate memory and exceed RAM.[26] Therefore, serving voxel data to VNN models by way of generator functions can be advisable, particularly for augmented datasets where each data point is expanded to many instances. Generator functions for passing in-memory voxel data, voxel data computed on the fly from SMILES sets, and voxel data augmented with all possible right-angle rotations (see main text) are included below:

```
# Keeping imports for smiles_to_voxels(). See above.
from typing import Tuple
import numpy as np

from rdkit import Chem
from rdkit.Chem import AllChem, rdDistGeom, rdForceFieldHelpers
from rdkit.Geometry import rdGeometry
from rdkit.DataStructs import cDataStructs

from scipy import ndimage  # For voxel rotations.
```

```
# Simple generator to load full dataset un-augmented. X is a voxel array.
def data_generator(X, y):
  for x, y_ in zip(X, y):
    yield x, y_

# Generator for on-the-fly voxelization. X is a list of SMILES.
def smiles_generator(
    X, y, batch_size=None, vox_dim=(32, 32, 32), vox_params={}):
  if batch_size:
    evox = np.zeros([batch_size, *vox_dim, 1])
    ey = np.zeros([batch_size, y.shape[-1]])
  count = 0
  for x, y_ in zip(X, y):
    vox = smiles_to_voxels(x, **vox_params)
    if vox is None:
      continue
    if batch_size:
      evox[count] = vox
      ey[count] = y_
      count += 1
      if count == batch_size:
        yield evox, ey
        count = 0
    else:
      yield vox, y_

# 10-fold rotational data augmentation. X is a voxel array.
def data_augmentor(X, y):
  for x, y_ in zip(X, y):
    x_aug = np.squeeze(x)
    # Capture 9 rotations of interest
    rots = [x_aug,
            ndimage.interpolation.rotate(
                x_aug, angle=90, axes=(0, 1), reshape=False),
            ndimage.interpolation.rotate(
                x_aug, angle=180, axes=(0, 1), reshape=False),
            ndimage.interpolation.rotate(
                x_aug, angle=270, axes=(0, 1), reshape=False),
            ndimage.interpolation.rotate(
                x_aug, angle=90, axes=(1, 2), reshape=False),
            ndimage.interpolation.rotate(
                x_aug, angle=180, axes=(1, 2), reshape=False),
            ndimage.interpolation.rotate(
                x_aug, angle=270, axes=(1, 2), reshape=False),
            ndimage.interpolation.rotate(
                x_aug, angle=90, axes=(0, 2), reshape=False),
            ndimage.interpolation.rotate(
                x_aug, angle=180, axes=(0, 2), reshape=False),
            ndimage.interpolation.rotate(
                x_aug, angle=270, axes=(0, 2), reshape=False)]
    rots = [np.rint(np.reshape(rot, (x.shape))) for rot in rots]
    for rot in rots:
      yield rot, y_
```

Using these generators, data can be passed to `tf.Model`'s through use of the `tf.data.Dataset.from_generator` API as follows:

```
# Keeping imports for smiles_to_voxels(). See above.
from typing import Tuple
import numpy as np

from rdkit import Chem
from rdkit.Chem import AllChem, rdDistGeom, rdForceFieldHelpers
from rdkit.Geometry import rdGeometry
from rdkit.DataStructs import cDataStructs

from scipy import ndimage  # For voxel rotations in generator. See above.

import tensorflow as tf

# Set data splitting parameters
train_frac = 0.8
valid_frac = 0.1
train_size = int(math.floor(train_frac*X.shape[0]))
valid_size = int(math.floor(valid_frac*X.shape[0]))
batch_size = 32

# Setting no augmentation as default. X is a dataset voxel array.
generator = data_generator

# Construct tf datasets
train = tf.data.Dataset.from_generator(
    generator=generator, args=[X[:train_size], y[:train_size]],
    output_signature=(tf.TensorSpec(shape=X[0].shape),
                      tf.TensorSpec(shape=y[0].shape))).batch(batch_size)
valid = tf.data.Dataset.from_generator(
    generator=generator, args=[X[train_size:train_size+valid_size],
                               y[train_size:train_size+valid_size]],
    output_signature=(tf.TensorSpec(shape=X[0].shape),
                      tf.TensorSpec(shape=y[0].shape))).batch(batch_size)
test = tf.data.Dataset.from_generator(
    generator=generator, args=[X[train_size+valid_size:],
                               y[train_size+valid_size:]],
    output_signature=(tf.TensorSpec(shape=X[0].shape),
                      tf.TensorSpec(shape=y[0].shape))).batch(batch_size)
```

These `tf.data.Dataset.from_generator` wrappers can take either pre-generated voxel arrays or simply lists of SMILES as input, depending on the generator used. An example loop to generate a voxel dataset from a pandas `pd.Dataframe` with SMILES and target labels is included below:

```
# Keeping imports for smiles_to_voxels(). See above.
from typing import Tuple
import numpy as np

from rdkit import Chem
from rdkit.Chem import AllChem, rdDistGeom, rdForceFieldHelpers
from rdkit.Geometry import rdGeometry
```

```
from rdkit.DataStructs import cDataStructs

import pandas as pd
from tqdm import tqdm  # For tracking voxelization progress

# Load the example data from csv (shuffling as well)
df = pd.read_csv(<path_to_data_file>).sample(frac=1, random_state=42)

# Set parameters for voxelization. Using defaults from main text here.
num_vox = 10000  # Number of examples to prepare
vox_params = {
    'box_dim': (16, 16, 16),
    'spacing': 0.5,
    'vdw_scale': 0.3,
    'shell_scale': 0.15,
    'n_shells': 4,
    'num_confs': 5,
    'rms_thresh': 0.1,
    'reshape': False,
    'return_mol': True,
}

# Choose modeling targets to include. Example here from PCQCPM6 dataset.
targets = [
           'B3LYPHOMO',
           'B3LYPLUMO',
           'B3LYPGAP',
           ]
smiles_column = 'SMILES'  # Adjust if needed for custom dataset.

# Create voxel and target arrays
X = []
y = []
for i, (_, row) in tqdm(enumerate(df.iterrows()), total=num_vox):
  if i >= num_vox:
    break
  vox, _ = smiles_to_voxels(row[smiles_column], **vox_params)
  if vox is not None:
    X.append(vox)
    y.append(np.array([row[target] for target in targets]))

# Convert to single arrays
X = np.stack(X)
y = np.stack(y)
```

Alternatively, if pre-training is being conducted utilizing the 3D shape features as auxiliary targets as proposed in Section 4.5, these descriptors can be computed during voxel construction using the `get_descriptors()` function and voxelization loop below (note also that this example modifies the array construction to avoid mass appending for large scale voxelization as is needed for pre-training):

```
# Keeping some imports for smiles_to_voxels(). See above.
from typing import Any, List, Tuple
import numpy as np

from rdkit import Chem
from rdkit.Chem import AllChem, rdDistGeom, rdForceFieldHelpers
from rdkit.Chem import Descriptors, Descriptors3D
from rdkit.Geometry import rdGeometry
from rdkit.DataStructs import cDataStructs
from rdkit.ML.Descriptors import MoleculeDescriptors

from tqdm import tqdm  # For tracking voxelization progress

def get_descriptors(
    mol: Chem.Mol, descriptor_list: List[Tuple[str, Any]]) -> np.ndarray:
  """Calculates molecular/geometric descriptors for a given molecule.

  Args:
    mol: rdkit mol object for molecule of interest.
    descriptor_list: list of descriptors to compute. See example below for
format.

  Returns:
    np array containing computed descriptors

  Note are passing mol and not SMILES here. Purpose is that expect mol object
  to contain pre-embedded conformers, as required for geomtric descriptor
  calculation.

  Note some rdkit descriptors are passed as their function calls and some are
  passed as string names depending on the module they're computed with. This is
  handled within the function.
  """
  if (mol is None) or (mol.GetNumConformers() == 0):
    return {
        name: np.nan for name, _ in descriptor_list
    }

  str_descs = []
  func_descs = []
  calcs = np.zeros(len(descriptor_list), dtype=np.float32)
  for nd, (name, desc) in enumerate(descriptor_list):
    if type(desc) == str:
        str_descs.append((nd, name, desc))
    else:
        func_descs.append((nd, name, desc))

  calcr = MoleculeDescriptors.MolecularDescriptorCalculator(
      [desc for _, _, desc in str_descs])
  str_calcs = calcr.CalcDescriptors(mol)
  for (nd, _, _), calc in zip(str_descs, str_calcs):
    calcs[nd] = calc

  for nd, _, calcr in func_descs:
    calcs[nd] = calcr(mol)

  return calcs


# 3D shape descriptors from main text
desc_list = [
    ('Asphericity', Descriptors3D.Asphericity),
```

```
    ('Eccentricity', Descriptors3D.Eccentricity),
    ('InertialShapeFactor', Descriptors3D.InertialShapeFactor),
    ('NPR1', Descriptors3D.NPR1),
    ('NPR2', Descriptors3D.NPR2),
    ('PMI1', Descriptors3D.PMI1),
    ('PMI2', Descriptors3D.PMI2),
    ('PMI3', Descriptors3D.PMI3),
    ('RadiusOfGyration', Descriptors3D.RadiusOfGyration),
    ('SpherocityIndex', Descriptors3D.SpherocityIndex),
]


# Set parameters for voxelization
num_vox = 10000  # Number of examples to prepare
vox_params = {
    'box_dim': (16, 16, 16),
    'spacing': 0.5,
    'vdw_scale': 0.3,
    'shell_scale': 0.15,
    'n_shells': 4,
    'num_confs': 5,
    'rms_thresh': 0.1,
    'reshape': False,
    'return_mol': True,
}
targets = desc_list

# Create voxel and target arrays to assign into (instead of appending)
x_arr_shape = (num_vox, *tuple(int(d/vox_params['spacing']) for
                      d in vox_params['box_dim']), 1)
y_arr_shape = (num_vox, len(targets))

X = np.zeros(x_arr_shape, dtype=np.int32)
y = np.zeros(y_arr_shape)

# Set counter to slice final array for failed examples
successful_vox = num_vox
for i, (_, row) in tqdm(enumerate(df.iterrows()), total=num_vox):
  if i >= num_vox:
    break
  vox, mol = smiles_to_voxels(row['SMILES'], **vox_params)
  if vox is not None:
    X[i] = vox
    y[i] = get_descriptors(mol, desc_list)
  else:
    successful_vox -= 1

# Slice out failed examples
X, y = X[:successful_vox], y[:successful_vox]
```

When using these particular descriptors, it can be useful to normalize each target in the range [0, 1] since they are computed on drastically different scales and can be challenging for multi-task models to predict simultaneously. We utilize scikit-learn's `preprocessing.MinMaxScaler` for this purpose as follows:

```
from sklearn import preprocessing

scaler = preprocessing.MinMaxScaler()
scaler.fit(y)
y = scaler.transform(y)
```

Using the data generation procedures and model implementations above, training and testing can be accomplished by the following example protocol:

```
import numpy as np

from sklearn import metrics

import tensorflow as tf

# Set training specifications
epochs = 100

# Learning rate function
lri = 0.001
lrf = 0.0001
decay_steps = int(epochs*(train_size/batch_size))
learning_rate = tf.keras.optimizers.schedules.PolynomialDecay(
    lri, decay_steps, lrf)

# Set model parameters
model_params = {
    'channel': 8,
    'n_targets': y.shape[-1],  # See data preparation above
    'embed_dim': 256,  # Chosen to roughly match classic descriptor encodings
}

# Initialize and compile model
model = QMCVNet(**model_params)
model.compile(
    loss=tf.keras.losses.MeanAbsoluteError(),
    optimizer=tf.keras.optimizers.Adam(learning_rate=learning_rate),
    metrics=[tf.keras.metrics.MeanAbsolutePercentageError()],
    )

# Set up checkpoint to save best model
experiment_name = 'vox_to_gap'
checkpoint_filepath = os.path.join(
    os.getcwd(), 'experiments', experiment_name, 'model')
model_checkpoint_callback = tf.keras.callbacks.ModelCheckpoint(
    filepath=checkpoint_filepath, monitor='val_loss', mode='min',
    save_best_only=True)

# Train model (see above for preparation of train, valid, and test)
model.fit(
    train, epochs=epochs, callbacks=[model_checkpoint_callback],
    validation_data=valid)

# Get training results
```

```
history = model.history.history

# Load model weights from best epoch
model.load_weights(checkpoint_filepath)

# Make test set predictions
preds = model.predict(test)

# Pool augmented predictions if necessary
if generator == data_augmentor:
  preds = np.reshape(
      preds, (1, len(preds), len(targets))).astype(np.float32)
  preds = tf.keras.layers.AveragePooling1D(
      pool_size=10, strides=10)(preds)
  preds = np.reshape(preds.numpy(), (preds.shape[1]))
preds = np.squeeze(preds)

# Calculate performance metrics (averaged over targets)
ground_truth = y[train_size+valid_size:]
test_mae = tf.keras.metrics.MeanAbsoluteError()
test_mape = tf.keras.metrics.MeanAbsolutePercentageError()
test_mae.update_state(ground_truth, preds)
test_mape.update_state(ground_truth, preds)
test_r2 = metrics.r2_score(ground_truth, preds)
```

4.8 REFERENCES

(1) Nakata, M.; Shimazaki, T.; Hashimoto, M.; Maeda, T. PubChemQC PM6: Data Sets of 221 Million Molecules with Optimized Molecular Geometries and Electronic Properties. *J. Chem. Inf. Model.* **2020**, *60*, 5891–5899.

(2) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.

(3) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.

(4) Bogojeski, M.; Vogt-Maranto, L.; Tuckerman, M. E.; Müller, K.-R.; Burke, K. Quantum Chemical Accuracy from Density Functional Approximations via Machine Learning. *Nat. Commun.* **2020**, *11*, 5223.

(5) Lu, J.; Xia, S.; Lu, J.; Zhang, Y. Dataset Construction to Explore Chemical Space with 3D Geometry and Deep Learning. *J. Chem. Inf. Model.* **2021**, *61*, 1095–1104.

(6) Keith, J. A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.-R.; Tkatchenko, A. Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chem. Rev.* **2021**, *121*, 9816–9872.

(7) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comp. Chem.* **1996**, *17*, 490–519.

(8) Maturana, D.; Scherer, S. VoxNet: A 3D Convolutional Neural Network for Real-time Object Recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 922–928, Hamburg, Germany, Sept. **2015**. IEEE.

(9) Qi, C. R.; Su, H.; Niessner, M.; Dai, A.; Yan, M.; Guibas, L. J. Volumetric and Multi-View CNNs for Object Classification on 3D Data. *arXiv:1604.03265* **2016**, *Preprint*.

(10) Brock, A.; Lim, T.; Ritchie, J. M.; Weston, N. Generative and Discriminative Voxel Modeling with Convolutional Neural Networks. *arXiv:1608.04236* **2016**, *Preprint*.

(11) Caglayan, A.; Can, A. B. Volumetric Object Recognition Using 3-D CNNs on Depth Data. *IEEE Access* **2018**, *6*, 20058–20066.

(12) Zhi, S.; Liu, Y.; Li, X.; Guo, Y. Toward Real-time 3D Object Recognition: A Lightweight Volumetric CNN Framework Using Multitask Learning. *Computers & Graphics* **2018**, *71*, 199–207.

(13) Muzahid, A. A. M.; Wan, W.; Sohel, F.; Khan, N. U.; Cervantes Villagomez, O. D.; Ullah, H. 3D Object Classification Using a Volumetric Deep Neural Network: An Efficient Octree Guided Auxiliary Learning Approach. *IEEE Access* **2020**, *8*, 23802–23816.

(14) Singh, S. P.; Wang, L.; Gupta, S.; Goli, H.; Padmanabhan, P.; Gulyás, B. 3D Deep Learning on Medical Images: A Review. *Sensors* **2020**, *20*, 5097.

(15) Kuzminykh, D.; Polykovskiy, D.; Kadurin, A.; Zhebrak, A.; Baskov, I.; Nikolenko, S.; Shayakhmetov, R.; Zhavoronkov, A. 3D Molecular Representations Based on the Wave Transform for Convolutional Neural Networks. *Mol. Pharmaceutics* **2018**, *15*, 4378–4385.

(16) Skalic, M.; Varela-Rial, A.; Jiménez, J.; Martínez-Rosell, G.; De Fabritiis, G. LigVoxel: Inpainting Binding Pockets Using 3D-Convolutional Neural Networks. *Bioinformatics* **2019**, *35*, 243–250.

(17) Skalic, M.; Jiménez, J.; Sabbadin, D.; De Fabritiis, G. Shape-Based Generative Modeling for de Novo Drug Design. *J. Chem. Inf. Model.* **2019**, *59*, 1205–1214.

(18) Sunseri, J.; Koes, D. R. libmolgrid: Graphics Processing Unit Accelerated Molecular Gridding for Deep Learning Applications. *J. Chem. Inf. Model.* **2020**, *60*, 1079–1084.

(19) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363*, eaau5631.

(20) Lin, M.; Chen, Q.; Yan, S. Network In Network. *arXiv:1312.4400* **2014**, *Preprint*.

(21) Cao, W.; Xue, J. Recent Progress in Organic Photovoltaics: Device Architecture and Optical Design. *Energy Environ. Sci.* **2014**, *7*, 2123.

(22) Landrum, G. A. RDKit: Open-Source Cheminformatics Software. (accessed Nov 20, 2018).

(23) Su, H.; Maji, S.; Kalogerakis, E.; Learned-Miller, E. Multi-view Convolutional Neural Networks for 3D Shape Recognition. *arXiv:1505.00880* **2015**, *Preprint*.

(24) Andrearczyk, V.; Fageot, J.; Oreiller, V.; Montet, X.; Depeursinge, A. Local Rotation Invariance in 3D CNNs. *arXiv:2003.08890* **2020**, *Preprint*.

(25) Hegde, V.; Zadeh, R. FusionNet: 3D Object Classification Using Multiple Data Representations. *arXiv:1607.05695* **2016**, *Preprint*.

(26) Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; Kudlur, M.; Levenberg, J.; Monga, R.; Moore, S.; Murray, D. G.; Steiner, B.; Tucker, P.; Vasudevan, V.; Warden, P.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: A System for Large-scale Machine Learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*, pages 265–283, Savannah, GA, USA, Nov. **2016**. ACSA.

(27) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980* **2017**, *Preprint*.

(28) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28*, 31–36.

(29) Nielsen, M. K.; Ahneman, D. T.; Riera, O.; Doyle. A. G. Deoxyfluorination with Sulfonyl Fluorides: Navigating Reaction Space with Machine Learning. *J. Am. Chem. Soc.* **2018**, *140*, 5004–5008.

(30) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186–190.

(31) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *arXiv:1704.01212* **2017**, *Preprint*.

(32) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular Graph Convolutions: Moving Beyond Fingerprints. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 595–608.

(33) Kipf, T. N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv:1609.02907* **2017**, *Preprint*.

(34) Todeschini, R.; Consonni, V. Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing / Volume II: Appendices, References, 1st ed.; Methods and Principles in Medicinal Chemistry; Wiley, **2009**; *Vol. 41*.

(35) Vogiatzis, K. D.; Polynski, M. V.; Kirkland, J. K.; Townsend, J.; Hashemi, A.; Liu, C.; Pidko, E. A. Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities. *Chem. Rev.* **2019**, *119*, 2453–2523.

(36) Barnum, G.; Talukder, S.; Yue, Y. On the Benefits of Early Fusion in Multimodal Representation Learning. *arXiv:2011.07191* **2020**, *Preprint*.

(37) Axelrod, S.; Gomez-Bombarelli, R. Molecular Machine Learning with Conformer Ensembles. *arXiv:2012.08452* **2021**, *Preprint*.

(38) Andrearczyk, V.; Fageot, J.; Oreiller, V.; Montet, X.; Depeursinge, A. Exploring Local Rotation Invariance in 3D CNNs with Steerable Filters. In *Proceedings of Machine Learning Research*, *volume 102*, pages 15–26, **2019**.

(39) Andrearczyk, V.; Oreiller, V.; Fageot, J.; Montet, X.; Depeursinge, A. Solid Spherical Energy (SSE) CNNs for Efficient 3D Medical Image Analysis. *IMVIP 2019: Irish Machine Vision & Image Processing*, Technological University Dublin, Dublin, Ireland, August 2019.

(40) Winkels, M.; Cohen, T. S. Pulmonary Nodule Detection in CT Scans with Equivariant CNNs. *Medical Image Analysis* **2019**, *55*, 15–26.

(41) Liu, L.; Chen, J.; Fieguth, P.; Zhao, G.; Chellappa, R.; Pietikäinen, M. From BoW to CNN: Two Decades of Texture Representation for Texture Classification. *Int. J. Comput. Vis.* **2019**, *127*, 74–109.

(42) Mao, H. H. A Survey on Self-supervised Pre-training for Sequential Transfer Learning in Neural Networks. *arXiv:2007.00800* **2020**, *Preprint*.

(43) Fabian, B.; Edlich, T.; Gaspar, H.; Segler, M.; Meyers, J.; Fiscato, M.; Ahmed, M. Molecular Representation Learning with Language Models and Domain-relevant Auxiliary Tasks. *arXiv:2011.13230* **2020**, *Preprint*.

(44) Elsken, T.; Metzen, J. H.; Hutter, F. Neural Architecture Search: A Survey. *arXiv:1808.05377* **2019**, *Preprint*.

(45) Schwaller, P.; Vaucher, A. C.; Laino, T.; Reymond, J.-L. Prediction of Chemical Reaction Yields Using Deep Learning. *Mach. Learn.: Sci. Technol.* **2021**, *2*, 015016.

(46) Żurański, A. M.; Martinez Alvarado, J. I.; Shields, B. J.; Doyle, A. G. Predicting Reaction Yields via Supervised Learning. *Acc. Chem. Res.* **2021**, *54*, 1856–1865.

(47) Zhao, S.; Gensch, T.; Murray, B.; Niemeyer, Z. L.; Sigman, M. S.; Biscoe, M. R. Enantiodivergent Pd-catalyzed C–C Bond Formation Enabled through Ligand Parameterization. *Science* **2018**, *362*, 670–674.

(48) Guan, Y.; Coley, C. W.; Wu, H.; Ranasinghe, D.; Heid, E.; Struble, T. J.; Pattanaik, L.; Green, W. H.; Jensen, K. F. Regio-selectivity Prediction with a Machine-learned Reaction Representation and On-the-fly Quantum Mechanical Descriptors. *Chem. Sci.* **2021**, *12*, 2198–2208.

(49) Maser, M. R.; Cui, A. Y.; Ryou, S.; DeLano, T. J.; Yue, Y.; Reisman, S. E. Multilabel Classification Models for the Prediction of Cross-Coupling Reaction Conditions. *J. Chem. Inf. Model.* **2021**, *61*, 156–166.

(50) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature* **2018**, *555*, 604–610.

(51) Buitrago Santanilla, A.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.-C.; Schneeweis, J.; Berritt, S.; Shi, Z.-C.; Nantermet, P.; Liu, Y.; Helmy, R.; Welch, C. J.; Vachal, P.; Davies, I. W.; Cernak, T.; Dreher, S. D. Nanomole-scale High-throughput Chemistry for the Synthesis of Complex Molecules. *Science* **2015**, *347*, 49–53.

(52) Perera, D.; Tucker, J. W.; Brahmbhatt, S.; Helal, C. J.; Chong, A.; Farrell, W.; Richardson, P.; Sach, N. W. A Platform for Automated Nanomole-scale Reaction Screening and Micromole-scale Synthesis in Flow. *Science* **2018**, *359*, 429–434.

(53) Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; de Diego, J. E.; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post, R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambeanu, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **2019**, *23*, 1213–1242.

ABOUT THE AUTHOR

Michael Robert Maser was born in April 1995 in the San Francisco Bay Area, where he grew up.

In 2013, Michael began his undergraduate studies at the University of California, Santa Barbara, where he developed a love for scientific research. For his undergraduate work, he was mentored by Prof. Bruce Lipshutz in synthetic methods development. After years in the Lipshutz Lab, Michael decided to continue in academic research and pursue his doctorate.

In 2017, Michael began his graduate studies under the direction of Prof. Sarah Reisman at the California Institute of Technology. There, he developed a fascination for machine learning and predictive chemistry, which ultimately became the focus of his graduate work. Following the completion of his PhD, Michael will begin as a Machine Learning Scientist at Genentech in South San Francisco.

Outside of work, Michael is happiest doing anything with his wife, Victoria.