Developing technologies for real-time whole-organism imaging via FTIR spectromicroscopy

Thesis by

Elizabeth Anne Holman

In Partial Fulfillment of the Requirements for

the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2022

(Defended July 6, 2021)

© 2022

Elizabeth Anne Holman
ORCID: 0000-0002-5158-4689

ACKNOWLEDGEMENTS

Thank you for the positive and constructive support from the Caltech, Advanced Light Source, BSISB imaging program, CAMERA, Kavli Nanosciences Institute, and Molecular Foundry communities, all of which were wonderful environments for learning and troubleshooting project challenges on the fly.

Specifically, thank you to my friends, family, collaborators, and labmates who were very understanding about the travel schedule involved with these projects. Thank you to Bert and Nathan from KNI who were willing to go above and beyond for exploring aspects of my research prior to the need to focus on peptoid technologies. Thank you to Barbara who was willing to watch over my plants at Caltech during experiments away from Los Angeles. Thank you to Ron for his mentorship while at the Molecular Foundry, and to Isaac, who willingly taught me the nuances of Langmuir trough experiments. Thank you for the support from the graduate office and registration (Tess, Angelica, Felicia, Natalie, Doug), which minimized the complications that may have arisen. Thank you to Corie and other LBNL staff who facilitated my access to LBNL for ongoing peptoid-related experiments at the Molecular Foundry during COVID-19. Thank you to my thesis committee members for their patience, wisdom, and constructive and honest advice.

Without funding support from the NSF GRFP and the DOE SCGSR, these projects would not have been possible. Without the mentorship of and support from Doug (co-advisor), who urged that I approach Paul (advisor) for a project discussion, and Paul, who told me that my job as a graduate student was to become the best scientist that I could possibly be, I would never have been able to freely pursue my research interests in an unfettered manner.

I am thankful and grateful for the opportunities that I have had over the course of my time as a graduate student. Notably, although I am not certain about the identity of reviewer number two, thank you for your candid commentary during COVID-19 quarantine which made me remember (and subsequently follow through with) the importance of divulging the complete truth behind my initial enthusiasm of investigating scientific phenomena.

Lastly, I am thankful for my life-changing yet brief interaction with the late Professor John H. Richards, who heavily influenced my approach to graduate research. Accessible, personable, and scientifically versatile with perpetual excitement for intellectual discussions even during our last meeting as his TA, he taught me that if I strongly believe in the importance of a specific scientific pursuit, then that “is the science worth doing” even if it requires learning aspects of scientific fields outside of my expertise. I still feel extremely guilty about the extensive wisdom and insights that were either inaccurately documented or undocumented due to his sudden and unexpected passing. I sincerely wish that I had more time to learn from him and still miss his guidance even now.

ABSTRACT

FTIR spectromicroscopy heavily resides in the domain of cell-based and tissue-based studies when focusing on its direct application to biological systems. The goal of the reported graduate research is to extend FTIR spectromicroscopy to multicellular whole-organism imaging, ideally for non-invasive, non-destructive, and label-free spatiochemical imaging of biological model *Caenorhabditis elegans* (*C. elegans*). With modern optics, detector, and light source technologies implemented at synchrotron facilities, this thesis focuses on exploring the feasibility of multicellular whole-organism imaging while identifying challenges and presenting working solutions for them. A brief introduction and explanation of infrared spectroscopy and the instrumentation enabling the experimental technique's development into one of spatiochemical imaging capabilities (**Chapter 1**) precedes the performed graduate research. The selection and preliminary assessment of *C. elegans* as our biological multicellular whole-organism model (**Chapter 2**), a solution for performing efficient scanning hyperspectral imaging of complex biological organisms (**Chapter 3**), and the design and synthesis of biomimetic nanomaterials for IR imaging applications (**Chapter 4**) are described in subsequent chapters.

PUBLISHED CONTENT AND CONTRIBUTIONS

Holman, E.A., Fang, YS., Chen, L. *et al*. Autonomous adaptive data acquisition for scanning hyperspectral imaging. *Commun Biol* **3**, 684 (2020). https://doi.org/10.1038/s42003-020-01385-3

E.A.H. participated in implementing the algorithm at ALS Beamline 1.4.3. and in designing and testing the IR processing module. E.A.H. designed and performed proof-of-principle experiments. L.C. and E.A.H. performed IR data processing. E.A.H. performed IR spectral analysis, gathered materials from all authors, and wrote the manuscript.

TABLE OF CONTENTS

NOMENCLATURE

IR. Infrared region of the electromagnetic spectrum.

MIR. Mid-infrared region of the electromagnetic spectrum.

Absorption band. A range of energies, often described in the form of frequency or wavelength for spectroscopy, that are characteristic of a specific transition from an initial to final state of a sample.

Thermopile. A device using the thermoelectric effect to convert thermal energy into electrical energy through the use of connected thermocouples.

Wheatstone bridge. The resistance bridge circuit that calculates an unknown resistance by balancing two arms of a bridge circuit. It consists of three known resistors and one unknown resistor and is used to precisely measure low resistances.

Wavenumber. A unit of spatial frequency that is defined as number of wavelengths per unit distance, typically centimeters for applications in spectroscopy.

FTIR. Fourier-transform infrared spectroscopy, a technique used to obtain a high-resolution infrared spectrum of absorption or emission of a sample.

NMR spectroscopy. Nuclear magnetic resonance spectroscopy.

NEP. Noise equivalent power.

BLIP. Background-limited infrared photodetector.

MCT. Mercury cadmium telluride (usually used in the case of detector identification).

SIR. Synchrotron infrared radiation.

SR-FTIR. Synchrotron radiation Fourier transform infrared (usually used with spectromicroscopy).

PCA. Principal component analysis.

MCR-ALS. Multivariate curve resolution alternating least square analysis.

VIS. Visible light region of the electromagnetic spectrum.

C h a p t e r 1

INTRODUCTION

"Who sees farther, a dwarf or a giant? Surely a giant for his eyes are situated at a higher level than those of a dwarf. But if the dwarf is placed on the shoulders of the giant, who sees further? Surely the dwarf, for now the eyes of the dwarf are situated at a higher level than those of the giant. So too, we are dwarfs, astride the shoulders of giants. We master their wisdom and move beyond it. Due to their wisdom we grow wise and are able to say all that we say, but not because we are greater than they."

Isaiah di Trani (c. 1180–1250)

Brief history of mid-infrared spectroscopy[1]

The infrared (IR) region of the electromagnetic spectrum was first discovered by Sir William Herschel in 1800 and emerged from his observation that heat along with light was generated on the lenses that he used for his solar observations. Through the early 1800s, sensitive temperature-measurement methods and alternative sources of solar heat radiation were developed and discovered respectively. This lay the foundation for experiments of the 1840s and 1850s, which supported John Tyndall's speculation that the origins of IR absorption are vibrations within molecules.

Macedonio Melloni performed the first measurements of the mid-infrared (MIR) absorption bands using a rocksalt crystal for dispersion experiments. By 1850, he completed his slit-containing spectrometer design that incorporated a lens to image the slit and a method by which the slit-shaped thermopile could sample across the spectrum created by a fixed prism. This became the instrument for MIR radiation study for several decades, and its spectral resolution was limited by the amount of radiation needed to fill the detector.

Samuel P. Langley addressed the resolution issue by implementing three concepts into the form of a smaller detector, termed a bolometer. By constructing the detector with low heat capacity materials, he created a chamber in which temperature changes would be greatly increased in comparison to previous devices; by utilizing narrow slits, he reduced the heat lost by the device prior to assessing the device's change in temperature. Incorporating A. F. Svanberg's methodology of heat detection from 1851, Langley used a blackened platinum wire as a material of unknown electrical resistance in a Wheatstone bridge—the platinum wire was blackened to better absorb the infrared radiation, and its resistance would change as a function of temperature. The bolometer would be used to lay the foundation for further developments in the IR detector field until the application of electronic amplification methods and semiconductor technology during World War II.

Although Langley also was responsible from the first calibration of IR spectra in terms of wavelength, Willem Henri Julius recorded the MIR spectra of 20 organic liquids in 1892 and found that certain compounds with the same functional group always had a conserved absorption band at the same wavenumber[2]. His work highlighted multiple instances of spectral regularity among the 20 compounds and drew attention to the strengths of the fundamental region of infrared, namely the amount of detail present in the spectra when systematically measuring across the MIR region. Years later and inspired by W. H. Julius' results, William Coblentz obtained spectra of 112 organic compounds up to the transmission limit of his rocksalt prism of 667 cm^{-1} (15 μm) in comparison to the transmission limit of Julius' at 1000 cm^{-1} (10 μm). In the 1920s, Jean Lecomte introduced the concept of viewing an IR spectrum as an overall pattern of absorption bands that could serve as a molecular fingerprint when studied with the compound's distinguishable

isomers[3]. By the 1930s, Reinhard Mecke developed a notation for approximately describing vibrational motions of group characteristic vibrational modes, namely the bond-stretching (ν), angle-bending (δ), and out-of-plane (γ) modes. The cumulative results of the scientists around the world during this time and were applied during World War II, showing the usefulness of IR spectroscopy for molecular structure analysis via group frequencies by 1947[1].

With advancements in electronic amplification methods and double beam spectrometers, the level of automation and ability to correct the data for light-source behavior respectively greatly increased the accessibility and ease of IR spectroscopy during and after World War II. The introduction of computers to spectrometers in the 1970s enabled exploitation of interferometry and subsequent development of Fourier transform infrared (FTIR) spectrometers[1]. The advantages obtained in higher sensitivity coupled with digital processing of IR data drastically increased the opportunities of IR spectroscopic applications, setting the stage for modern-day FTIR imaging and its use in agriculture, polymer science, and biomedicine[4, 5].

Instrumentation for FTIR spectromicroscopy

New experiments that could be performed after 1945 due to technological advancements in instrumentation broadened our understanding of spectroscopic phenomena via IR spectroscopy until NMR spectroscopy of the 1950s and mass spectrometry of the 1960s were able to serve as physical methods for determining structural information of compounds. However, this only led IR spectroscopy to be explored for further applications towards areas such as microscopy. IR spectromicroscopy combines infrared spectral analysis with the spatial resolution of a microscope. Three major technologies emerged that supported

the potential of FTIR to be used as a spatiochemical imaging technique: the sensitive infrared detectors, brighter light sources and optics, and increased computational power of computers used with the spectral microscopes.

Well-designed infrared detectors try to achieve generally reasonable responsivity and noise equivalent power (NEP) while minimizing all noise sources to the point where thermal background shot noise is the dominant noise component for the detector. IR detectors that operate in this regime successfully are called background limited infrared photodetectors (BLIPs). Where an ideal IR detector generates no noise, and hence, the only other noise component detectable by the detector should emerge from the shot noise due to photons reaching the detector from objects at ambient temperatures that are within the detector's field to view, a detector operating in the background limited regime or BLIP is the best practical form of a detector in the real-world experimental space. In the case of infrared detectors, responsivity is determined by the ratio of electrical output divided by the incident infrared power in order to quantitatively determine the detector's ability to convert infrared radiation into electrical signal. Importantly, this parameter neither determines the repeatability of the detector's measurement nor the minimum magnitude of signal that the detector can resolve. To address this, NEP is used to quantitatively assess the detector's ability to resolve small signals and is the ratio of noise power density at the detector's output divided by the its responsivity[6]. NEP is incorporated into specific detectivity (D*), which corrects for NEP's dependency on the detector's active area as

$$D^* = \frac{(A_{det})^{0.5}}{NEP},$$

where A_{det} is the area of the detector, and D^* also represents the signal-to-noise ratio of a detector of unit area in a unit bandwidth when a radiant power of 1 W is hitting the detector.

With the emergence of the photon detectors and cryogenic cooling options, detection capabilities increased for IR spectroscopy in the MIR region for two reasons: (1) photon detectors respond over a narrow section of the spectrum in comparison to thermal detectors to reduce thermal background noise and reach operational BLIP conditions with appropriate cooling, and (2) photon detectors operate on a much faster time scale than thermal detectors, the response time being on the order of microseconds instead of milliseconds respectively[6]. Specifically, the introduction single-element photoconductive mercury cadmium telluride (MCT) detectors greatly enabled MIR spectroscopy to enter its era of time-resolved spectroscopy when paired with an interferometer.

As a direct result of computational abilities and the MCT detector's response time, co-addition of frames or mirror scans could easily be performed per spectrum as well as background subtraction per spectrum in order to increase the signal-to-noise ratio (S/N); the higher sensitivity of the photoconductive single-element MCT detectors also contributed to the overall increase of signal with minimal noise[7]. To further sample visualization with FTIR spectroscopy, the commercial development of multichannel IR detectors became possible through a hybrid detector approach, in which one layer performs the photon detection and the second layer executes signal collection, multiplexing, and amplification. This permitted the development of single-element detector arrays to increase the spatial coverage per sampling event—64 x 64 pixel MCT detectors (3500 cm^{-1} to 900 cm^{-1}) and 256 x 256 MCT arrays (4000 cm^{-1} to 900 cm^{-1}) became commercially available in FTIR instruments[4]. Today, IR detectors can effectively cover the MIR region of 4000 cm^{-1} to 650

cm^{-1} while spatially resolving scientific phenomena. The main detection limitation occurs in the temporal domain, where the speed of photon-to-signal conversion in MCT detectors occurs on the millisecond timescale. It should be noted that in 2021, single-photon detection was successful achieved in the MIR region for wavelengths between 3 to 10 microns using tungsten silicide superconducting nanowire detectors, suggesting that developments in nanotechnology may enable future detector developments for FTIR spectromicroscopy[8].

In a similar manner to detector technology limiting the broad application of FTIR spectroscopy, light sources and their associated brightness became increasingly important when aiming for high spectral as well as spatial resolution for FTIR spectromicroscopy. Light of a specific wavelength is often characterized by its brightness, bandwidth, and power. Brightness is defined as the following

$$B(\overline{v}) = \frac{power}{area \times solid\ angle},$$

and its relationship to the S/N in an IR region where an FTIR instrument is limited by detector noise is

$$\frac{S}{N} = \frac{B(\overline{v}) \times (\delta\overline{v}) \times D^* \times C}{(\Delta f)^{0.5} \times A^{0.5}},$$

where $B(\overline{v})$ is source brightness, $\delta\overline{v}$ is resolution, D^* is detectivity, C is optical efficiency multiplied by the limiting étendue of the experiment, Δf is the electronic bandwidth, and A is detector area. Since brightness cannot be well-controlled for IR thermal sources outside of either altering emissivity or changing temperature, brightness maximization was typically ignored until photodetectors and alternative light sources became available and accessible.

Government-funded electron particle accelerators enabled the generation and use of synchrotron radiation as a high-brightness and broadband IR light source for diffraction-

limited FTIR spectromicroscopy. The Advanced Light Source at Lawrence Berkeley National Laboratory is the nearest government synchrotron IR (SIR) radiation source to Pasadena, CA. At this synchrotron, electrons are accelerated close to the speed of light and are circulating in a storage ring. Since the electron is traveling relativistically in a magnetic field with a motion that describes a circle, the light energy emitted from the electron generates a sharp pulse (in time), its resulting Fourier transform yielding a broadband spectrum[9]. Specifically, SIR can be up to 1000 times brighter than conventional thermal IR sources in the molecular fingerprint region of the MIR region[10]. When paired with custom beamline optics, this brightness can approach a diffraction-limited spot size of 2 to 10 μm, which directly contrasts with the globar IR source spot size of greater than 100 μm aperture in order to obtain similar S/N values[11].

Although more developments in materials science and engineering allowed scientists to harness SIR as an alternative light source in FTIR spectromicroscopy, advancements in computer hardware and software catalyzed the technique's application to complex biological systems. Automation of the Michelson interferometer in FTIR spectrometers permitted rapid acquisition of IR spectra, and utilizing broadband light sources increased the detector's S/N by obtaining multiplexed measurements—as described by Fellgett's advantage[1]. By increasing the speed with which IR spectra could be gathered from a hardware standpoint, it became possible to not only observe biological systems in a non-invasive and label-free manner but also to observe their phenomena and how these organisms may change under perturbative effects, such as introducing environmental stimuli. An initial primary concern emerged from the intensity of the MIR photons generated from the synchrotron when applying SR-FTIR spectromicroscopy to living biological systems, though the energy of the

MIR photons theoretically was too low in energy to cause bond breakage. This question was addressed by assessing temperature heating of a fully hydrated lipid layer. The mechanism of heating emerges from infrared heating of water, which then causes the hydrated lipid layer to undergo a phase change at 315 K during heating. This experiment showed that negligible sample heating at 0.5 ± 0.2 K occurred over the course of a 30-minute acquisition timeframe at a single point[12]. With these findings along with experiments monitoring *E. coli* biofilm response to antibiotics, SR-FTIR spectromicroscopy became widely accepted as a non-invasive, label-free spatiochemical imaging technique that could potentially track *in situ* biochemical phenomena in real-time[13].

Toward applications for whole-organism spatiochemical imaging

SR-FTIR spectromicroscopy is a non-destructive, label-free, highly sensitive spatiochemical technique that can perform real-time monitoring of biological systems ranging from single cells to biofilms and tissues. Over the last decade, MIR-tunable quantum cascade lasers were developed further to serve as a more accessible broadband light source alternative to a synchrotron[14, 15]. With broadband MIR light sources to be implemented in the near future for FTIR spectromicroscopy, interest in actively applying this technique in the biomedical field for precision theranostics and disease research has renewed and grown.

Currently, FTIR spectromicroscopy heavily resides in the domain of cell-based and tissue-based studies when focusing on its direct application to biological systems. The goal of this reported graduate research is to extend FTIR spectromicroscopy to multicellular whole-organism imaging via the biological model *Caenorhabditis elegans* (*C. elegans*). With modern optics, detector, and light-source technologies implemented at synchrotron facilities, this thesis focuses on exploring the feasibility of multicellular whole-organism

imaging while identifying challenges and presenting working solutions for them. The selection and preliminary assessment of *C. elegans* as our biological multicellular whole-organism model (**Chapter 2**), a solution for performing efficient scanning hyperspectral imaging of complex biological organisms (**Chapter 3**), and the design and synthesis of biomimetic nanomaterials for IR imaging applications (**Chapter 4**) are described in subsequent chapters.

References

1. Sheppard, N., *The Historical Development of Experimental Techniques in Vibrational Spectroscopy*, in *Handbook of vibrational spectroscopy*, J.M. Chalmers and P.R. Griffiths, Editors. 2002, J. Wiley: New York.
2. Brand, J.C.D., *Lines of light : the sources of dispersive spectroscopy, 1800—1930*. 1995, Australia: Gordon and Breach. xiv, 266 p.
3. *J. Lecomte, Father of French Spectroscopy, Pioneer of Infrared, French Scientific Ambassador*. European Spectroscopy News, 1977. **10**.
4. Kidder, L.H., A.S. Haka, and E.N. Lewis, *Instrumentation for FT-IR Imaging*, in *Handbook of vibrational spectroscopy*, J.M. Chalmers and P.R. Griffiths, Editors. 2002, J. Wiley: New York.
5. Wetzel, D.L., J.A. Sweat, and D.D. Panzer, *Synchrotron powered FT-IR microspectroscopy enhances spatial resolution for probing and mapping of plant materials.* AIP Conference Proceedings, 1998. **430**(1).
6. Theocharous, E. and J.R. Birch, *Detectors for Mid- and Far-infrared Spectrometry: Selection and Use*, in *Handbook of vibrational spectroscopy*, J.M. Chalmers and P.R. Griffiths, Editors. 2002, J. Wiley: New York.
7. Smith, G.D. and R.A. Palmer, *Fast Time-resolved Mid-infrared Spectroscopy Using an Interferometer*, in *Handbook of vibrational spectroscopy*, J.M. Chalmers and P.R. Griffiths, Editors. 2002, J. Wiley: New York.
8. Verma, V.B., et al., *Single-photon detection in the mid-infrared up to 10 mu m wavelength using tungsten silicide superconducting nanowire detectors.* Apl Photonics, 2021. **6**(5).
9. WIlliams, G.P., *Synchrotron and Free Electron Laser Sources of Infrared Radiation*, in *Handbook of vibrational spectroscopy*, J.M. Chalmers and P.R. Griffiths, Editors. 2002, J. Wiley: New York.
10. Duncan, W.D. and G.P. Williams, *Infrared Synchrotron Radiation from Electron Storage-Rings.* Applied Optics, 1983. **22**(18): p. 2914-2923.
11. Carr, G.L., J.A. Reffner, and G.P. Williams, *Performance of an Infrared Microspectrometer at the Nsls.* Review of Scientific Instruments, 1995. **66**(2): p. 1490-1492.
12. Martin, M.C., et al., *Negligible sample heating from synchrotron infrared beam.* Applied Spectroscopy, 2001. **55**(2): p. 111-113.
13. Holman, H.Y., et al., *Synchrotron IR spectromicroscopy: chemistry of living cells.* Anal Chem, 2010. **82**(21): p. 8757-65.
14. Du, Z.H., et al., *Mid-Infrared Tunable Laser-Based Broadband Fingerprint Absorption Spectroscopy for Trace Gas Sensing: A Review.* Applied Sciences-Basel, 2019. **9**(2).
15. Bandyopadhyay, N., et al., *Ultra-broadband quantum cascade laser, tunable over 760 cm(-1), with balanced gain.* Opt Express, 2015. **23**(16): p. 21159-64.

C h a p t e r 2

WHOLE-ORGANISM FTIR SPECTROMICROSCOPY: A FEASIBILITY STUDY IN *CAENORHABDITIS ELEGANS*

Abstract

We assess the feasibility of mapping the biochemical energy landscape in the roundworm *C. elegans* as the first step towards developing experiments to elucidate biochemical cues driving the decision to different animal states, such as either young larval (L1) arrest or a strain-modeled disease state. Our preliminary results show that the bright SR-FTIR offers high signal-to-noise spectra with diffraction-limited spatial resolution in *C. elegans*; our fine SR-FTIR spectral features can describe spatially compartmentalized chemistries in the context of the whole organism when complemented with other orthogonal analytical techniques such as mass spectrometry. This suggests that further developments in experimental techniques to perform non-invasive, label-free spatiochemical imaging of *C. elegans* in real time can be used for future studies testing the hypothesis that global biochemical phenotypes may play a deterministic role in subsequent gene expression.

Introduction

Nutritional control of gene expression has come to the forefront of biological and health research due to increasing evidence supporting a nutritional parameter behind aging as well as prevalent disease states such as diabetes, obesity, and cancer[1]. *C. elegans* provides a superb model system to study nutritional control due to its ability to halt development very early in response to environmental cues that are independent of population density, which in this case is regulated commonly through pheromone signaling. L1 arrest is

triggered by an absence of food, is reversible with feeding, and confers a high resistance to a variety of environmental stresses without inducing morphological changes in the larvae—contrasting with other alternative developmental programs. When coupling the lack of morphological modification in the L1 arrest state with the principle of stress response hormesis[2], a preliminary biochemical investigation of L1 and L1-arrested larvae requires a technique that permits non-invasive and real-time probing of *C. elegans* to exclude technique-induced free radical production and propagation that alter observable biochemistries[3-5]. This is especially true in the case of expanding biological studies toward investigating the biochemistry behind the concept of maternal stress memory in *C. elegans* progeny after their mother undergoes stress-induced L1 arrest[6].

Synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectromicroscopy meets the non-invasive spatiotemporal and spatiochemical baseline requirements necessary to achieve our experimental goal[7-9]. This is achieved due to SIR being a broadband and very bright infrared light source, which addresses S/N considerations mentioned in **Chapter 1** with respect to multiplexed measurements improving S/N as well as the direct relationship between instrument S/N and light source brightness. Non-destructive FTIR spectromicroscopy is established as a useful tool to study nematodes[10-12]. Starting in 2004, the Doglia group demonstrated that their highly reproducible FTIR spectromicroscopy data revealed distinct principal absorption band differences among the pharynx, intestine, and tail regions of *C. elegans*. With their single-point measurements from a single-element infrared detector of photoconductive mercury cadmium telluride (MCT) using an aperture size of 60μm x 80μm, they obtained a 4 cm^{-1} spectral resolution over the course of 256 scan co-additions[10]. Their detector selection and data acquisition method

theoretically favors spectral-chemical resolution over spatial resolution with coverage of the mid-infrared region with a lowest cut-off of ~650 cm^{-1}.

The subsequent decade following the Doglia study was marked with technological advancements in not only computer hardware and software used to control data acquisition but also focal plane array (FPA) detectors based upon photovoltaic MCT designs rather than photoconductive MCT designs. The first application of a 64 x 64 pixel FPA detector to effectively cover a 170 μm x 170 μm sample area was applied to FTIR investigations *C. elegans* in 2016[11], and the FTIR-FPA imaging data revealed that rapid mapping speed came at the cost of fine spectral information encoded in the FTIR signal. Each pixel's S/N in MCT FPA detectors is determined by less sensitive photovoltaic MCT detectors when compared to the photoconductive MCT detectors,[13] which motivated the development of our approach to overcome this challenge of signal-to-noise ratio (S/N) decrease through experimental design.

Results and discussion

Our solution to maintain rapid chemical and diffraction-limited spatial mapping of *C. elegans* for potential live, multicellular organism imaging at the cellular and tissue level is to simply increase the brightness of the broadband source—in other words, replace the thermal infrared emitter IR source with a much brighter source in the form of synchrotron infrared radiation. Our preliminary results verified that although the FPA imaging system permits an increased image acquisition speed by multiplexing the data acquisition, low S/N limits its applicability to living *C. elegans* measurements when restricted to using a thermal infrared emitter IR source. Next, we tested our solution of substituting the thermal infrared emitter with synchrotron radiation as our IR source and found that our expected S/N

advantage encoding fine spectral feature resolution among different tissue regions of the *C. elegans* sample is achievable with this strategy, as depicted in **Figure 1**.

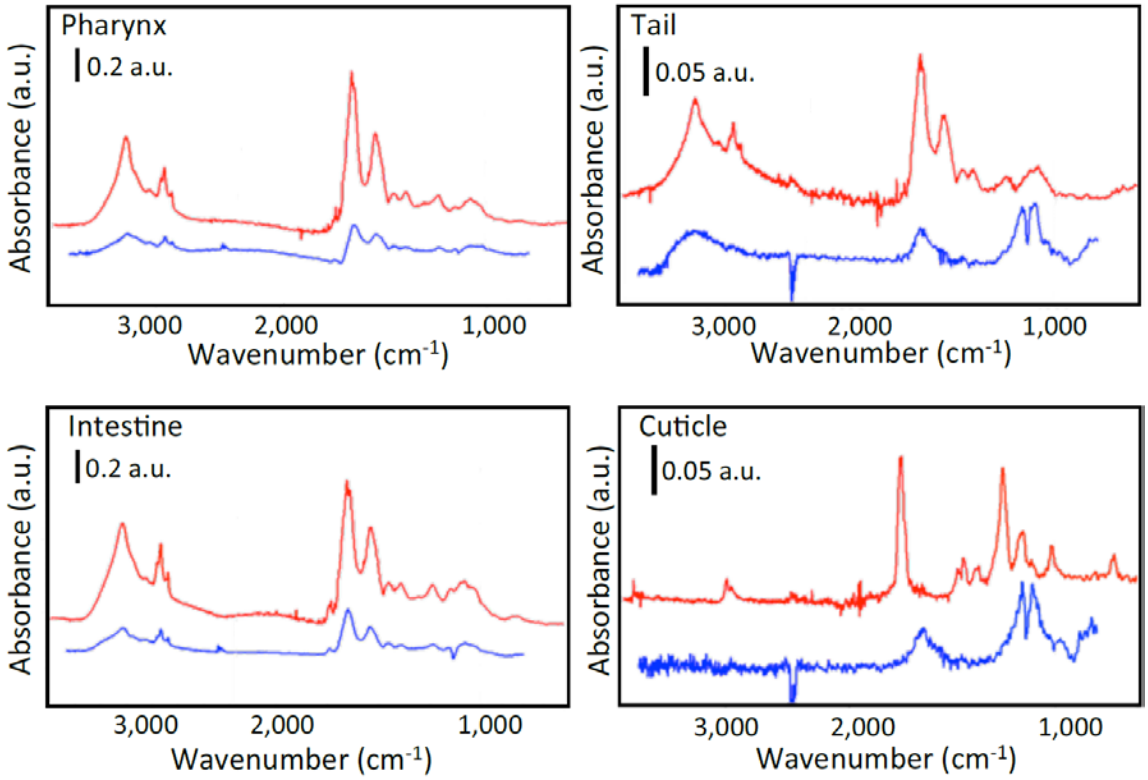


Figure 1. A comparison of FTIR spectromicroscopy results obtained from using the synchrotron versus a thermal infrared emitter as our IR source when targeting different anatomical regions of a *C. elegans* sample. The SR-FTIR spectromicroscopy data (in red) was collected over 5 seconds in 4 scans while the FTIR-FPA data (in blue) was collected over 26 minutes in 256 scans.

The superior quality of the spectral data allows us to (1) fully mine the infrared metadata, (2) identify defining spectral regions of interest, and (3) determine the key differences in molecular family composition among the assayed phenotypes modeled by *C. elegans*. The quality of SR-FTIR metadata directly guides our appropriate selection of complementary analytical techniques for obtaining corroborative and additionally informative data to characterize more comprehensively our observable, spatially compartmentalized chemistries in the context of the whole organism.

After showing that SIR brightness and broadband attributes when coupled with the sensitive photoconductive single-element MCT detector addressed concerns raised by Sheng et al. in 2016, we proceeded to investigate the best approach for laying the foundation for linking mass spectrometry results to non-invasive, label-free SR-FTIR spatiochemical and spatiotemporal imaging of the living, whole-organism *C. elegans* animal. To test the viability of creating either a tissue- or organ-specific FTIR library, we isolated a 5-micron-thick section of the cuticle from a flash-frozen gravid adult *C. elegans* hermaphrodite for subsequent SR-FTIR spatiochemical imaging. To avoid signal contamination, we achieved this sectioning in a purely mechanical manner using a two-armed micromanipulator (AxisPro SS; Micro Support Co., Ltd.; Japan) without using either chemical fixation or glue for the removal of the cuticle. As shown in **Figure 2,** the chemical signature of the partial section of the *C. elegans* cuticle FTIR was easily distinguishable from the background with only four co-added scans necessary, implying that generation of SR-FTIR *C. elegans* tissue libraries

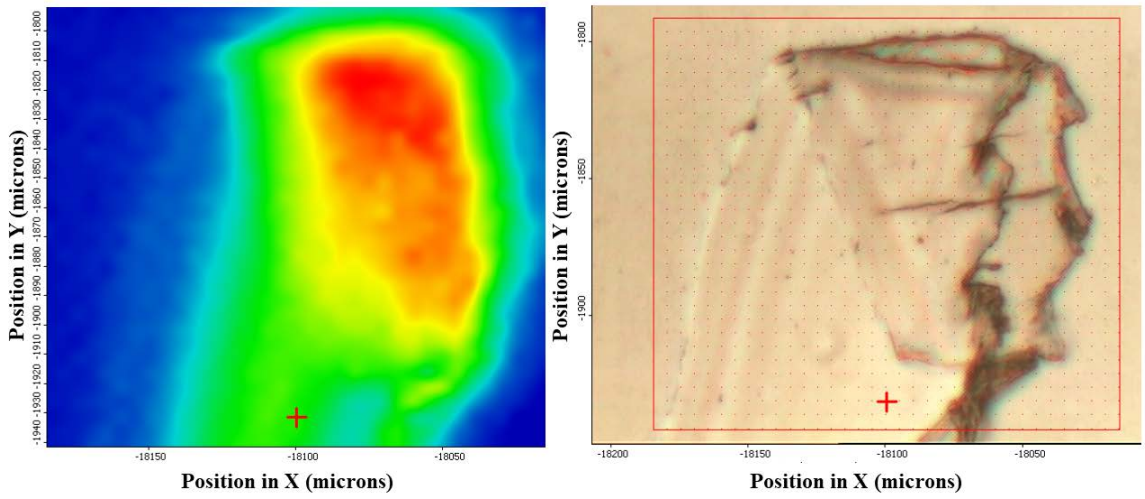


Figure 2. (Left) Heat map of sectioned cuticle at 1015 cm^{-1}, an absorption band associated with alkyl phosphate stretching modes. (Right) The bright-field image of the sample through a 32x IR objective. The red rectangle correlates to the mapped region, each dot indicates the location at which a broadband FTIR spectrum was acquired, and the red cursor marks spatial location of the epicuticle FTIR spectrum shown in *Figure 3*.

can be performed quickly. Theoretically, library generation for applications of SR-FTIR spectroscopy would involve dissecting tissues, arranging the tissues into an array, and then scanning through each sample with approximately 4 co-added scans—this translates to approximately 1.25 seconds between each tissue sample. These data would ideally be stored in a *C. elegans* database such as Wormbase for ease of reference, and the tissue samples should retain hydration in order to preserve the structural information that can be obtained from the FTIR spectra.

With the feasibility of tissue-based library generation for *C. elegans* studies affirmed, we next addressed concerns of signal deconvolution from the cuticle data as well as the necessary steps performed to remove synchrotron noise from the MIR region (**Figure 3**). Prior to March 2019, the maximum mirror velocity of the ALS Beamline 1.4.3 spectral

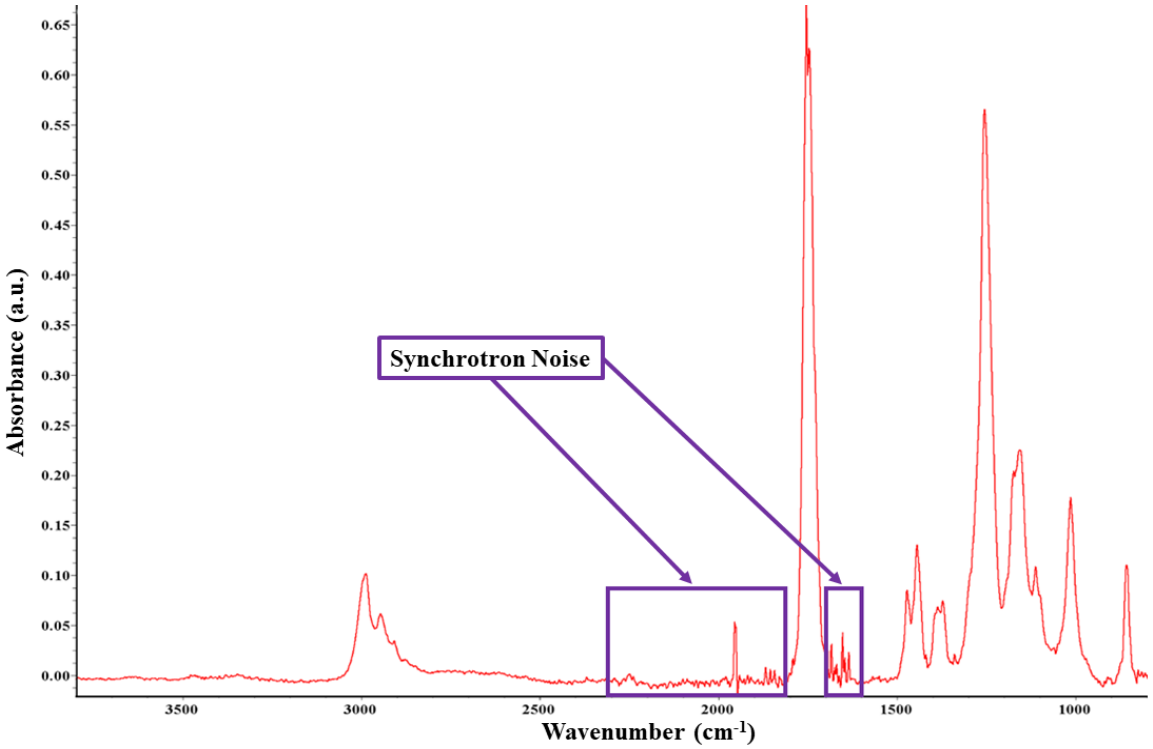


Figure 3. SR-FTIR spectrum displayed in OMNIC 9.8 from sampled point 88 of 1085, spatially identified by the red cursor in *Figure 2*. When using a mirror velocity of approximately 1.8 cm/s, synchrotron noise emerging from synchrotron-specific behavior can be detectable in the fingerprint region of the MIR spectra.

microscope was 1.8 cm/s. By upgrading the computer, the OMNIC software that drives the spectral microscope, and the Michelson interferometer, we were able to achieve mirror velocities up to 6.3 cm/s. With these velocities at a slight cost to total signal, we could shift the synchrotron noise from the MIR fingerprint region to a variable region below 900 cm^{-1}, with the majority of the noise restricted to 800 cm^{-1} and below. Due to the signal contamination specifically within the 1800 to 1700 cm^{-1} range, we remain conservative in our following proof-of-concept FTIR spectral analysis to support the potential physico-chemical information that SR-FTIR spectral information can provide to the *C. elegans* community.

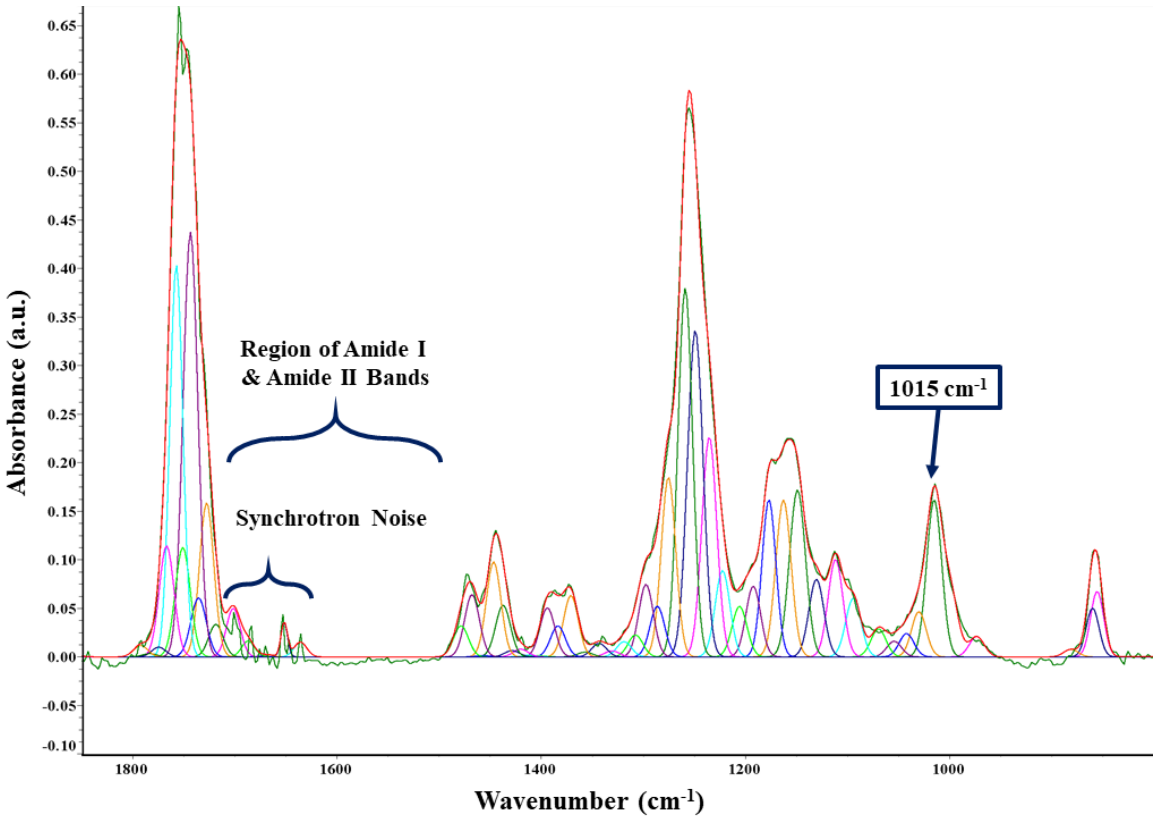


Figure 4. Gaussian approximations performed in OMNIC 9.8 for SR-FTIR spectrum of epicuticle mechanically isolated from gravid adult hermaphrodite. Although synchrotron noise is also likely present in the sharp peak approximately at 1754 cm^{-1}, the major synchrotron noise contributions in the amide regions are labeled for clarity of epicuticle identification.

To accurately interpret the SR-FTIR spectra of the partial cuticle, we perform curve fitting while assuming Gaussian distributions as shown in **Figure 4**. The most notable feature of the spectrum is the complete absence of amide I and amide II absorption bands—meaning that this sample neither contains collagen nor glycoproteins in detectable quantities relative to the nitrogen-less lipid content. This strongly suggests that we isolated solely the epicuticle of the cuticle with the micromanipulator, which is known to be lipid-rich while being an exception to the extensive collagen presence of the nematode's flexible exoskeleton[14]. The lack of C-N stretching modes also confirms the removal of the glycoprotein surface coat of the cuticle, further supporting the sample's identity as an epicuticle isolate.

Viewing the fitted Gaussian distributions in the 1800 to 1700 cm^{-1} region, we conservatively can comment on the strong presence of saturated aliphatic esters and dicarboxylic acids via the ν(C=O) modes, which include but are not limited to sn-1 ester C=O (free C=O) and sn-2 ester C=O (bound C=O) around 1741 and 1723 cm^{-1} respectively[15, 16]. Specifically, ν(C=O) component bands at 1743 and 1727 cm^{-1} that emerge strongly in our peak deconvolution are known to be commonly shifted in hydrated diacylglycerolipids that exhibit phase-state dependence[17]. When the strong absorption peak at 1015 cm^{-1} is interpreted in the context of the aforementioned ester ν(C=O) modes and the P=O associated vibrations emerging from associated and unassociated ν(P=O) modes that are embedded between approximately 1310 and 1150 cm^{-1}[15], we can conclude that this lipid-rich epicuticle sample contains alkyl phosphates, likely in the form of phospholipids. It is also important to note that the presence of ethers is supported by the Gaussian deconvolution, specifically the ν_{asym}(C-O-C) modes of vinyl ethers with a band at approximately 1205 cm^{-1} and of saturated aliphatic ethers with bands ranging from 1150 to 1070 cm^{-1}[15].

This preliminary chemical information on the epicuticle would ideally be supplemented with mass spectrometry to address detection limitation concerns, especially in the context of nitrogen-containing compounds in the epicuticle. However, several testable hypotheses emerge from this preliminary data associated with further studies to better understand the relatively uncharacterized epicuticle layer of the *C. elegans*' cuticle[18, 19]. Currently, the epicuticle is functionally associated with water loss reduction and acting as a physical barrier; yet, the data suggests that the ether and ester lipid-rich layer could also potentially serve as a tunable gel or crystalline liquid depending upon the temperature and addition of cholesterol-like secretions into the epicuticle. Ether lipids have already been established as necessary for optimal fertility, lifespan, survival at cold temperatures, and resistance to oxidative stress in *C. elegans* via loss-of-function mutation experiments[20]. The latter two concepts can potentially be relevant to the epicuticle. Survival at cold temperatures can be related to tuning the gel-like epicuticle to maintain optimal fluidity as temperature changes. Oxidative stress response can be related if the epicuticle is viewed not only for its exoskeletal role but also as a temporary storage and readily available phospholipid sink to serve potentially as free radical termination sites, molecular precursors, and lipid-based free radical scavengers such as plasmalogens[21].

After having established the usefulness of SR-FTIR spectromicroscopy as a guiding tool to generate evidence-driven hypotheses to increase our understanding of the biochemistry of *C. elegans* at a structural and functional level, we moved towards the final stage in assessing the feasibility of live whole-organism imaging. Accurate spectral interpretation of *C. elegans* in the MIR region became possible with the FTIR spectrometer upgrade in March 2019, and a more detailed proof-of-concept application of multivariate

analysis for the head region of a L2 larval stage *C. elegans* animal SR-FTIR spectromicroscopy is located in **Chapter 3** for a spectral range of 3500 to 2600 cm^{-1}. Therefore, we will apply multivariate curve resolution alternating least squares (MCR-ALS)

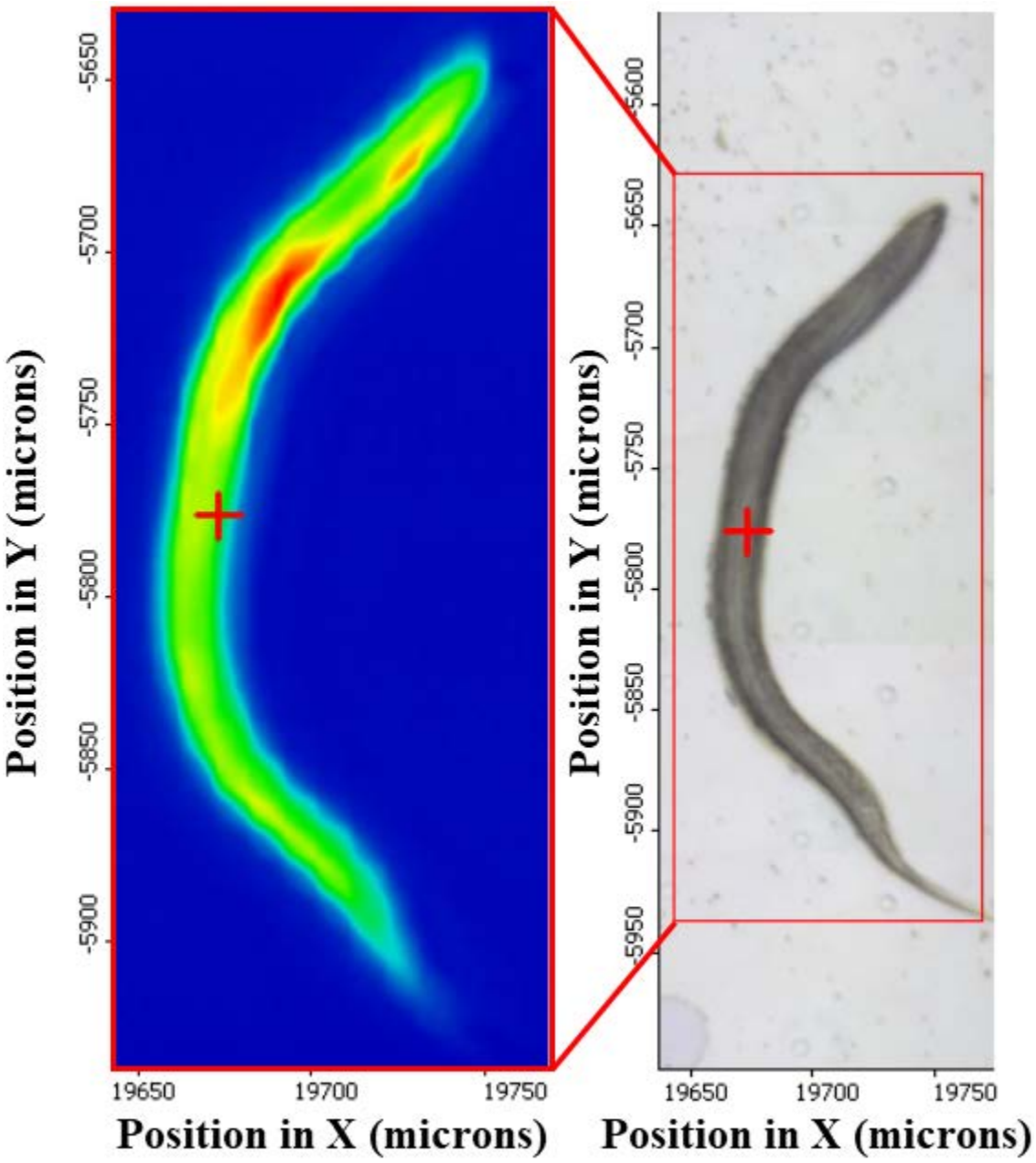


Figure 5. Plotted using OMNIC 9.8. (Left) Intensity plot of SR-FTIR map at 2959 cm^{-1} performed as a quality assessment for spatial resolution. An intensity plot at major peak 1641 cm^{-1} (not shown) confirms that absorption spectra did not exhibit saturation due to muscle-dense structures. (Right) Bright field image the sample through a 32x IR objective, and the red rectangle correlates to the mapped region shown on left.

analysis of a similar L2 larval stage *C. elegans* animal over the approximate spectral range of 1800 to 875 cm^{-1} to ensure avoidance of synchrotron noise contamination and appropriately address the multicomponent aspect of this biological system that univariate analyses have difficulty in comprehensively addressing[22].

To perform a data quality assessment of the acquired SR-FTIR spectral map detailing our L2 larval stage *C. elegans*, we examine the univariate spectra in the alkane ν(C-H) modes for alkane functional groups for spatial and spectral resolution examination[15]. As shown in **Figure 5**, we successfully achieve fine diffraction-limited spatial resolution with our 2.0 by 2.0 mapping step size while avoiding spectral signal saturation at our most intense spectral peak—the amide I band which primarily is governed in our case by H-bonded primary amide C=O stretching modes exhibited by proteins[23].

In selecting our multivariate analysis method, we select MCR-ALS over principal component analysis (PCA) due to the application of the non-negative least squares solver used to determine the eigenvector solutions in MCR-ALS[24]. This means that our loading vectors acquired through executing MCR with initial estimates in OMNIC 9.8 can undergo subsequent model-fitting approximations for potentially more accurate identification of component-associated identities when assuming non-negative components do reflect spectral ground truth[22, 25]. We set our eigenvector number to 9 parameters, basing this on the calculated χ^2-value of 0.0188 in OMNIC 9.8. Our parameter selection was further confirmed by the residual plot which revealed coefficients for pixels outside of our region of interest (the worm). MCR component 5 was identified as instrument-specific, and MCR components 6 and 9 emerged from wave-like behavior exhibited by IR light as it interacts with particles close to the size of its wavelength (**Appendix A**).

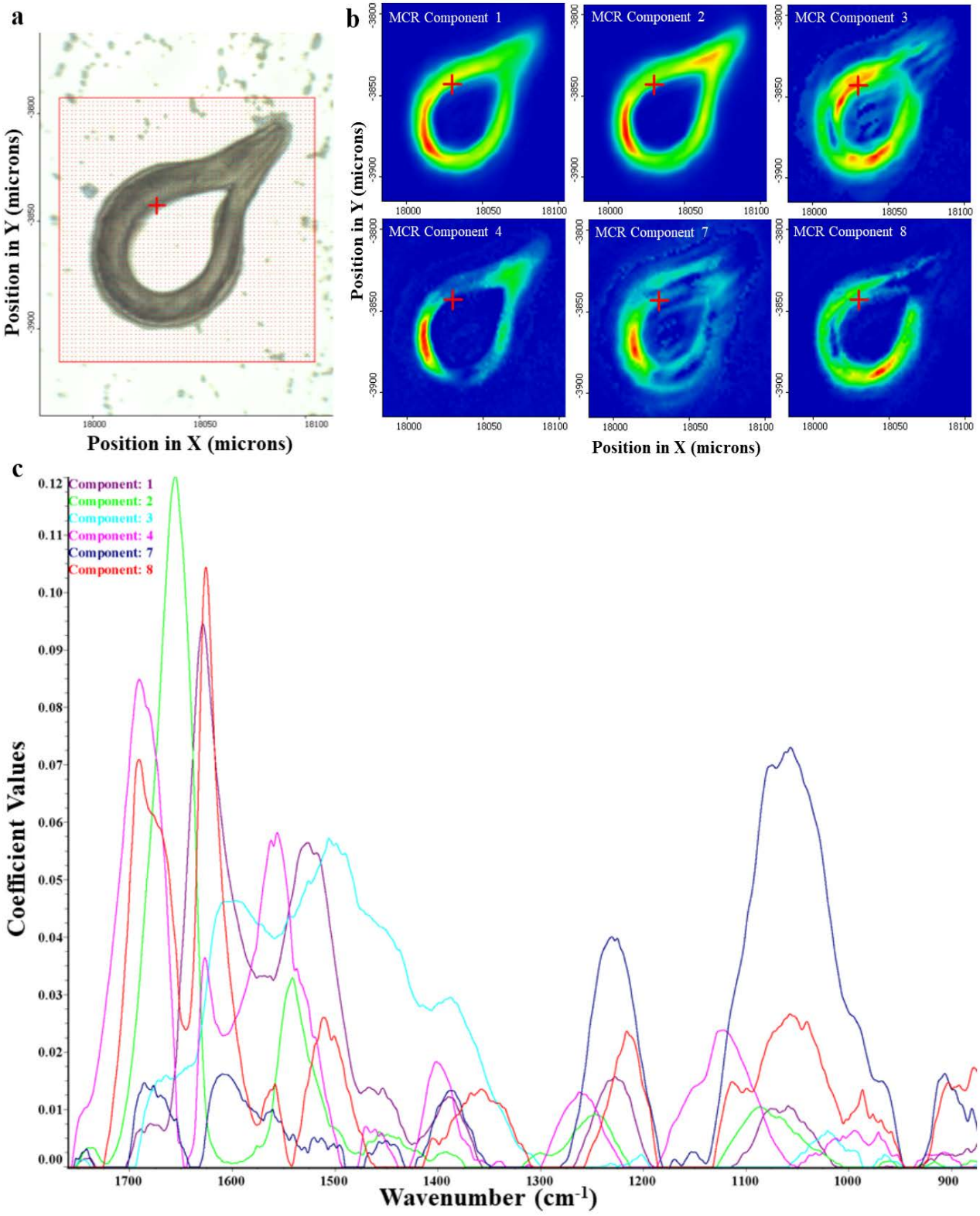


Figure 6. MCR-ALS performed on larval stage L2 *C. elegans* in Omnic 9.8. **(a)** Bright field image of SR-FTIR mapped region. **(b)** MCR components 1, 2, 3, 4, 7, and 8 coefficients displayed as individual heat maps per component. **(c)** Each individual MCR component-loading vector plotted as a function of wavenumber for ease of peak identification and spectral analysis.

Our MCR-ALS spectral analysis of larval stage L2 wildtype (N2) *C. elegans* detailed in **Figure 6** reveals potential compartmentalization of chemistry even within the same organs. MCR component 1 captures the overall majority of the extended chain polypeptides (amide I band ~1630 cm^{-1} and amide II band at 1525 cm^{-1}) and carbohydrates (multiple from C-H and O-H deformation vibrations between 1460 and 1200 cm^{-1}) in the animal, which mainly distinguishes between spectral chemistry exhibited by the organism in contrast to the OP50-contaminated ZnSe crystal upon which the sample was mounted. MCR component 2 co-localizes with head, neck, and body muscle[26, 27] while exhibiting chemical spectral signatures of hydrated proteins in the form of a broadened amide I band (~1700 to 1600 cm^{-1}), broadened amide II bands (1575 to 1480 cm^{-1}), and amide III bands (1308 to 1230 cm^{-1}), of which all are peptide-characteristic peaks[28, 29]. The amide-rich, peak-broadened MCR component 3 suggests that the amide-containing compounds described by this loading vector are hydrated, and the co-localization with the nerve ring suggests that these may reflect amide-containing glycophospholipids, especially when taking into account the ν(P-O-C) mode commonly found in alkyl phosphates between 1050 and 990 cm^{-1}[15, 16]. MCR component 7 corresponds to large amounts of hydrated inorganic phosphate species and phosphorus-containing compounds[15, 30], which co-localize to a very specific sub-region within the intestine. MCR components 4 and 8 also co-localize with the intestinal region of the *C. elegans*[27], suggesting that perhaps there is specialization of biochemical processes specific to sub-regions of the intestine at the L2 larval stage of *C. elegans* development. Notably, this potential specialization of biochemical processing within an organ could also be present in the adult animal, although this hypothesis should be assessed with orthogonal analytical techniques.

Conclusion

Our results support that whole organism imaging in *Caenorhabditis elegans* is not only feasible but also an under-utilized analytical tool in guiding the formation of evidence-driven testable hypotheses in complex biological systems. Furthermore, this study shows that living multicellular whole-organism spatiochemical imaging is possible with SR-FTIR spectromicroscopy. With biological studies pushing the boundary more towards an idealized concept of real-time omics, we show that non-invasive and label-free SR-FTIR spectromicroscopy is a promising realistic alternative to "real-time" omics in generating a longitudinal dataset on the same animal if there are resources invested into its development.

Several challenges need to be addressed in order to increase the accessibility of this technique to whole organism real-time imaging applications. First, mapping regions and sampling methods were limited to a rectangular uniform grid sampling methodology due to software restrictions with the FTIR spectrometer. This led to high-resolution spatiochemical map acquisition times of 8-12 hours, depending upon the larval stage of the *C. elegans* sample. With the maximum allowable sample acquisition time being 12 hours due to the cryogen cooling needs of the detector, sampling must be performed efficiently with respect to physical space and sample chemistries—our solution to this problem can be found in **Chapter 3**. Second, monitoring living organisms responding to controlled introduction of environmental stimuli require analytical chambers that are relatively IR transparent to minimize signal attenuation through the infrared light path. We investigate alternative optically transparent windows to the current 500-micron ZnSe crystals to address this challenge in **Chapter 4**.

Methods

Sample Preparation. For the cuticle sample, adult hermaphrodites aged by 3 days were flash-frozen with liquid nitrogen. A single animal was selected and moved to the micromanipulator stage, in which 1 probe arm was exchanged with a blade. The initial cut was performed length-wise at the thickest portion of the nematode. Due to the nematode's curvature, the second cut was performed 5 microns below the initial horizontal slicing plane to generate the cuticle sample. No chemical fixatives were used in this sample preparation to reduce signal contamination.

Whole-organism samples were mounted on 0.5-mm-thick ZnSe crystals, which were cleaned with Milli-Q water, 5% acetic acid, acetone, then Milli-Q water sequentially in order to remove organics while minimizing crystal damage. Each *C. elegans* (N2; Caenorhabditis Genetics Center) animal used was selected at their specific developmental stage based on morphology. Each animal was moved from their agar growth plates to 1 μL of 0.25 mM Levamisole (CAS# 16595-80-5, Sigma-Aldrich) on the ZnSe crystal and rinsed three times with Milli-Q water before mounting the sample onto the microscope stage for imaging.

Instrumentation. Synchrotron FTIR measurements were performed on a Nicolet Nic-Plan IR microscope with a ×32, 0.65 numerical aperture objective with a Thermo Scientific Nicolet iS50 FTIR spectrometer using a KBr beamsplitter and MCT (HgCdTe) detector at Beamline 1.4.3 of the Advanced Light Source at Lawrence Berkeley National Laboratory (Berkeley, CA, USA). Scanning SR-FTIR diffraction-limited (2–10 μm) spectra were collected in transmission mode between 650 and 4000 cm^{-1} at 4 cm^{-1} spectral resolution and recorded with 8 co-added scans at interferometer mirror velocities of 1.8 cm/s (only for case in **Figure 1**) and 6.3 cm/s.

References

1. Baugh, L.R., *To grow or not to grow: nutritional control of development during Caenorhabditis elegans L1 arrest.* Genetics, 2013. **194**(3): p. 539-55.
2. Gems, D. and L. Partridge, *Stress-response hormesis and aging: "that which does not kill us makes us stronger".* Cell Metab, 2008. **7**(3): p. 200-3.
3. Holman, H.Y., et al., *Synchrotron IR spectromicroscopy: chemistry of living cells.* Anal Chem, 2010. **82**(21): p. 8757-65.
4. Holman, H.Y., et al., *Synchrotron infrared spectromicroscopy as a novel bioanalytical microprobe for individual living cells: cytotoxicity considerations.* J Biomed Opt, 2002. **7**(3): p. 417-24.
5. Holman, H.Y., M.C. Martin, and W.R. McKinney, *Synchrotron-Based FTIR Spectromicroscopy: Cytotoxicity and Heating Considerations.* J Biol Phys, 2003. **29**(2-3): p. 275-86.
6. Hibshman, J.D., A. Hung, and L.R. Baugh, *Maternal Diet and Insulin-Like Signaling Control Intergenerational Plasticity of Progeny Size and Starvation Resistance.* PLoS Genet, 2016. **12**(10): p. e1006396.
7. Birarda, G., et al., *Synchrotron infrared imaging of advanced glycation endproducts (AGEs) in cardiac tissue from mice fed high glycemic diets.* Biomed Spectrosc Imaging, 2013. **2**(4): p. 301-315.
8. Chen, L., et al., *Synchrotron infrared measurements of protein phosphorylation in living single PC12 cells during neuronal differentiation.* Anal Chem, 2012. **84**(9): p. 4118-25.
9. Holman, H.Y., et al., *Real-time molecular monitoring of chemical environment in obligate anaerobes during oxygen adaptive response.* Proc Natl Acad Sci U S A, 2009. **106**(31): p. 12599-604.
10. Ami, D., et al., *Fourier transform infrared microspectroscopy as a new tool for nematode studies.* FEBS Lett, 2004. **576**(3): p. 297-300.
11. Sheng, M., A. Gorzsas, and S. Tuck, *Fourier transform infrared microspectroscopy for the analysis of the biochemical composition of C. elegans worms.* Worm, 2016. **5**(1): p. e1132978.
12. Zhang, D., et al., *Depth-resolved mid-infrared photothermal imaging of living cells and organisms with submicrometer spatial resolution.* Sci Adv, 2016. **2**(9): p. e1600521.
13. Norton, P., *HgCdTe infrared detectors (Reprinted from Electrochem Soc. Proc., pg 48-70, 1999).* Opto-Electronics Review, 2002. **10**: p. 159-174.
14. Page, A.P. and I.L. Johnstone, *The cuticle*, in *WormBook*, J.M. Kramer and D.G. Moerman, Editors. 2007.
15. Socrates, G. and G. Socrates, *Infrared and Raman characteristic group frequencies : tables and charts.* 3rd ed. 2001, Chichester ; New York: Wiley. xv, 347 p.
16. *Infrared Spectroscopy of Biomolecules*. 1995: Wiley-Liss.
17. Lewis, R.N., et al., *Components of the carbonyl stretching band in the infrared spectra of hydrated 1,2-diacylglycerolipid bilayers: a reevaluation.* Biophys J, 1994. **67**(6): p. 2367-75.

18. Peixoto, C.A., J.M. Kramer, and W. de Souza, *Caenorhabditis elegans cuticle: a description of new elements of the fibrous layer.* J Parasitol, 1997. **83**(3): p. 368-72.
19. Peixoto, C.A. and W. De Souza, *Freeze-fracture and deep-etched view of the cuticle of Caenorhabditis elegans.* Tissue Cell, 1995. **27**(5): p. 561-8.
20. Shi, X., et al., *A Caenorhabditis elegans model for ether lipid biosynthesis and function.* J Lipid Res, 2016. **57**(2): p. 265-75.
21. Wallner, S. and G. Schmitz, *Plasmalogens the neglected regulatory and scavenging lipid species.* Chem Phys Lipids, 2011. **164**(6): p. 573-89.
22. Motegi, H., et al., *Identification of Reliable Components in Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS): a Data-Driven Approach across Metabolic Processes.* Scientific Reports, 2015. **5**(1): p. 15710.
23. Parker, F.S., *Amides and Amines*, in *Applications of Infrared Spectroscopy in Biochemistry, Biology, and Medicine*. 1971, Springer US: Boston, MA. p. 165-172.
24. Felten, J., et al., *Vibrational spectroscopic image analysis of biological material using multivariate curve resolution–alternating least squares (MCR-ALS).* Nature Protocols, 2015. **10**(2): p. 217-240.
25. Dodd, J.G. and L.K. DeNoyer, *Curve-fitting: Modeling Spectra*, in *Handbook of vibrational spectroscopy*, J.M. Chalmers and P.R. Griffiths, Editors. 2002, J. Wiley: New York.
26. Altun, Z.F. and D.H. Hall, *Muscle system introduction*, in *WormAtlas*. 2009.
27. Altun, Z.F. and D.H. Hall, *Introduction to C. elegans anatomy*, in *WormAtlas*. 2009.
28. Kong, J. and S. Yu, *Fourier transform infrared spectroscopic analysis of protein secondary structures.* Acta Biochim Biophys Sin (Shanghai), 2007. **39**(8): p. 549-59.
29. Ji, Y., et al., *DFT-Calculated IR Spectrum Amide I, II, and III Band Contributions of N-Methylacetamide Fine Components.* ACS Omega, 2020. **5**(15): p. 8572-8578.
30. Miller, F.A. and C.H. Wilkins, *Infrared Spectra and Characteristic Frequencies of Inorganic Ions - Their Use in Qualitative Analysis.* Analytical Chemistry, 1952. **24**(8): p. 1253-1294.

Chapter 3

AUTONOMOUS ADAPTIVE DATA ACQUISITION FOR SCANNING HYPERSPECTRAL IMAGING

Material from:

Holman, E.A., Fang, YS., Chen, L. et al. Autonomous adaptive data acquisition for scanning hyperspectral imaging. *Commun Biol* **3**, 684 (2020). https://doi.org/10.1038/s42003-020-01385-3 Published by Springer Nature.

Abstract

Non-invasive and label-free spectral microscopy (spectromicroscopy) techniques can provide quantitative biochemical information complementary to genomic sequencing, transcriptomic profiling, and proteomic analyses. However, spectromicroscopy techniques generate high-dimensional data; acquisition of a single spectral image can range from tens of minutes to hours depending on the desired spatial resolution and the image size. This substantially limits the timescales of observable transient biological processes. To address this challenge and move spectromicroscopy towards efficient real-time spatiochemical imaging, we developed a grid-less autonomous adaptive sampling method. Our method substantially decreases image acquisition time while increasing sampling density in regions of steeper physico-chemical gradients. When implemented with scanning Fourier transform infrared (FTIR) spectromicroscopy experiments, this grid-less adaptive sampling approach outperformed standard uniform grid sampling in a two-component chemical model system and in a complex biological sample of *Caenorhabditis elegans* (*C. elegans*). We

quantitatively and qualitatively assess the efficiency of data acquisition using performance metrics and multivariate infrared spectral analysis, respectively.

Introduction

Advancements in optical microscopy, especially fluorescence microscopy, have enabled biologists to observe multiplexed living cellular events with ever higher spatial and temporal resolutions[1]. The use of targeted fluorescent indicators provides spatial and temporal context to omics analyses[2-4], resulting in discoveries of dynamic spatial architecture in disease pathogenesis[5], organogenesis[6], and wound healing[7]. These advances inspired the drive towards high-dimensional image-based profiling[8], which requires high information-content and rapid, robust measurements of as many living cell or tissue phenotypes as possible to capture time-dependent spatial heterogeneities in structure and morphological patterning.

One solution is to introduce another complementary dimension of label-free observation one that focuses on the spatiochemical mapping of biological systems. This information can be used to guide fluorescence microscopy, its real-time imaging capabilities limited to a few features of interest identified *a priori*, and to improve the interpretation of omics data and information from advanced transmitted light microscopy images. Non-invasive and label-free multiplexed imaging techniques, such as scanning synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectromicroscopy, can identify and monitor spatial heterogeneity in chemical composition that is indistinguishable when using the visible region of the electromagnetic spectrum; however, a major challenge in using these techniques for real-time characterization of time-dependent biochemical processes is the substantial image acquisition time that ranges from minutes to hours. This complication

emerges from the high dimensionality of the generated data set, which contains not only spatial but also spectral information, and the utilization of uniform grid (UG) sampling as the current standard sampling method, which historically is objective and computationally inexpensive.

With advancements in the accessibility of computing technology, we find that grid-less autonomous adaptive data acquisition (AADA) is a viable and more efficient alternative to UG sampling. AADA maintains a systematic and reproducible approach while capturing spectral and spatial heterogeneity with fewer sampled points and shorter experimental time frames. We discuss the significance of this method for studying time-sensitive living systems and its future development toward monitoring time-dependent phenomena in biological systems prior to expanding our discussion towards AADA's applicability to other fields and workflow processes.

Results

To implement an autonomous adaptive sampling algorithm (**Figure 1**) for data acquisition, we prioritized optimization of spatiotemporal and spatiochemical sampling efficiency while operating under experimental parameters that nonetheless yield comprehensive and informative spectral map data. We assume that less predictable yet detectable phenomena emerging from spatiochemical heterogeneity are primary regions of interest, informational "hot spots" that should be spatiochemically resolved with subsequent sample points after initial detection. To achieve this, our adaptive sampling is driven by leave-one-out cross-validation (LOOCV)[9] to facilitate rapid and accurate approximations of the experimentally mapped space[10] for predictive error calculations from which the

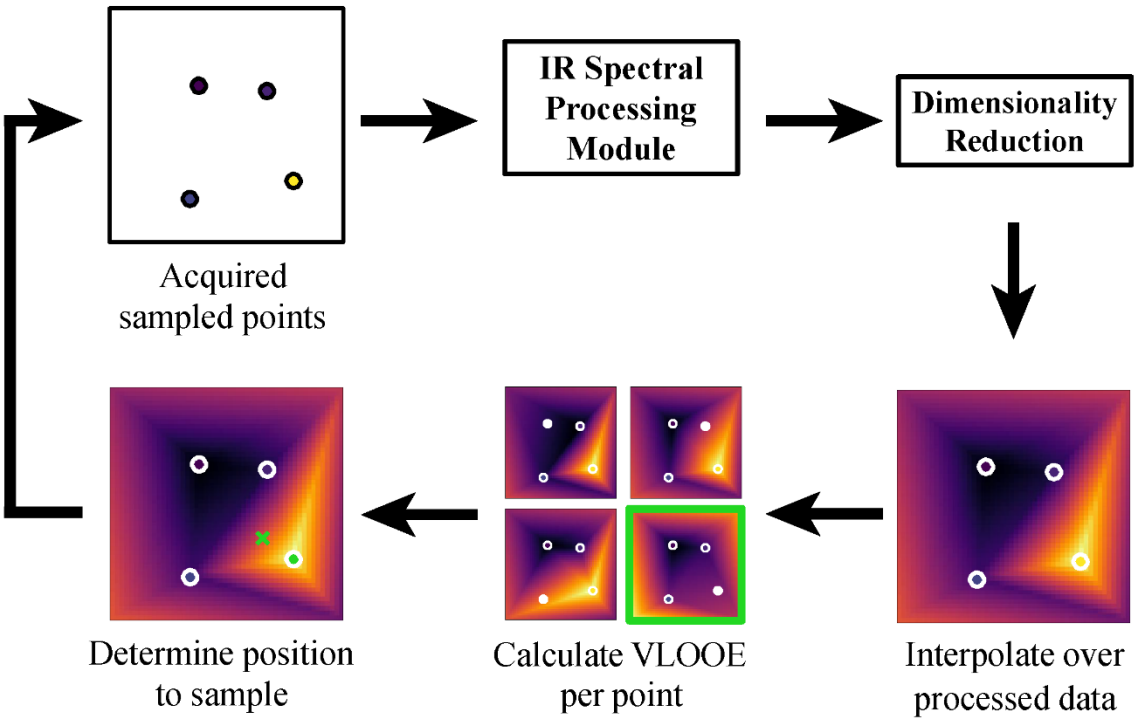


Figure 1. Autonomous adaptive sampling workflow.

algorithm can rapidly identify regions for subsequent sample exploitation[11-13]. We build our surrogate models using a hybrid sequential sampling strategy closely related to other established methods[11, 14, 15] by combining two-dimensional (2D) barycentric linear interpolation with Voronoi tessellation (LIV). With LIV, the relative importance of a sampled point is determined by its adjusted Voronoi-weighted leave-one-out error (ϵ_{LOO})[12, 16], which is calculated by normalizing and equally weighing the leave-one-out error with the Voronoi predictive error. Since collected IR spectra often form continuous and multimodal regions in our input space per sampled point[12], we introduced an IR spectral preprocessing module upstream of our surrogate model construction and LOOCV to conserve the spectral resolution while the algorithm determines where to subsequently sample.

As our first experimental demonstration, we designed a two-component chemical model system of blue permanent marker and high vacuum grease for spatiochemically

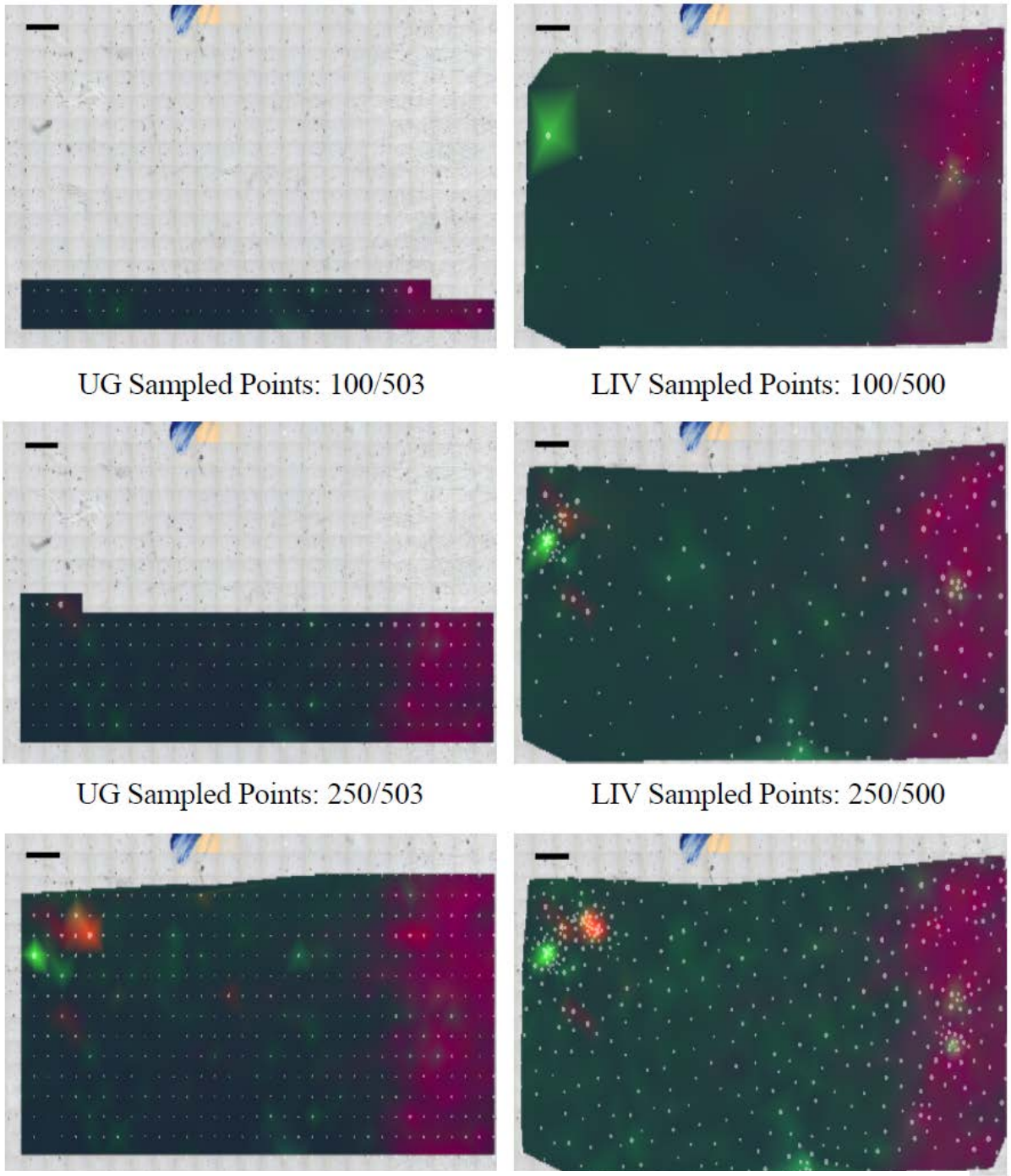


Figure 2. Autonomous sampling of abiotic two-component sample for current standard UG and adaptive LIV methods (scale bar, 200 μm). First three PCA components are shown as an RGB overlay with sampled points (white circles) and their predictive error (circle diameter).

resolved characterization using scanning FTIR spectroscopy. This complete sample characterization enabled quantitative evaluations and comparisons between adaptive LIV and widely utilized, non-adaptive UG data acquisition (**Figure 2**). In this visibly featureless case, the mapped domain was selected with minimal experimenter knowledge input to guide the autonomous adaptive data acquisition. Under these experimental conditions, we quantitatively and qualitatively determine data acquisition performance using mathematical and spectral metrics. When using mean Voronoi-weighted LOO error $\langle \epsilon_{LOO} \rangle_V$ to quantify modeling accuracy, we found that adaptive LIV data acquisition outperformed the non-adaptive data acquisition methods (**Figure 3**) in this experimental system.

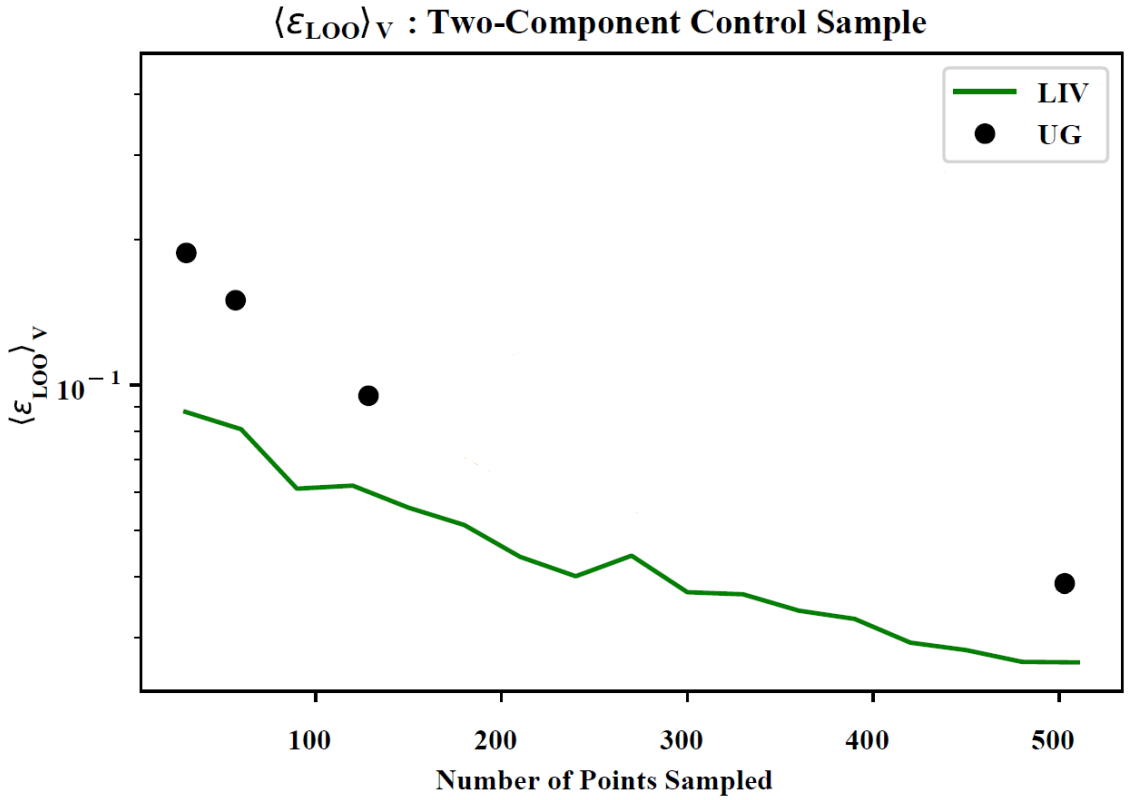


Figure 3. Mean Voronoi-weighted leave-one-out error plotted for non-adaptive and adaptive data acquisition experiment for the abiotic two-component sample. Adaptive LIV data acquisition of map regions defined without domain knowledge outperforms current gold-standard non-adaptive data acquisition.

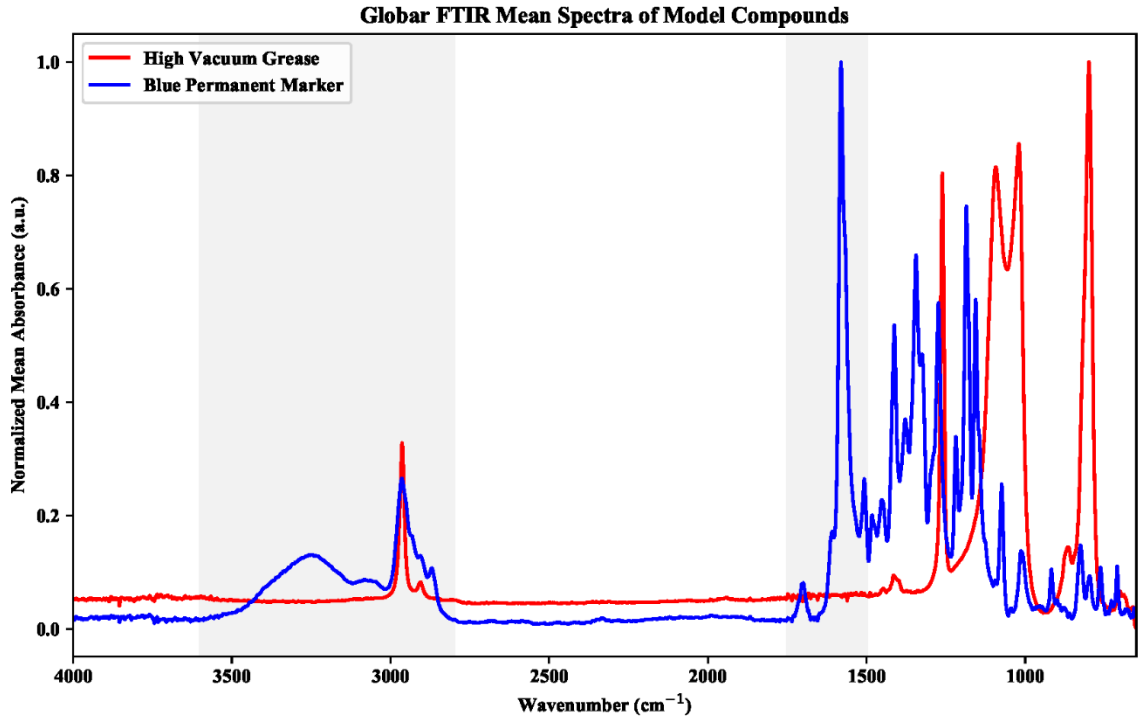


Figure 4. Globar FTIR mean absorbance identity fingerprint spectra (N = 8) of each model compound eventually used in synthetic two-component control sample. On-target ratios were calculated using spectral peaks at 1580 cm^{-1} and 798 cm^{-1}, which were the major identification peaks of permanent marker and high vacuum grease respectively. PC-LDA regions (gray) were determined from mean and variance spectra.

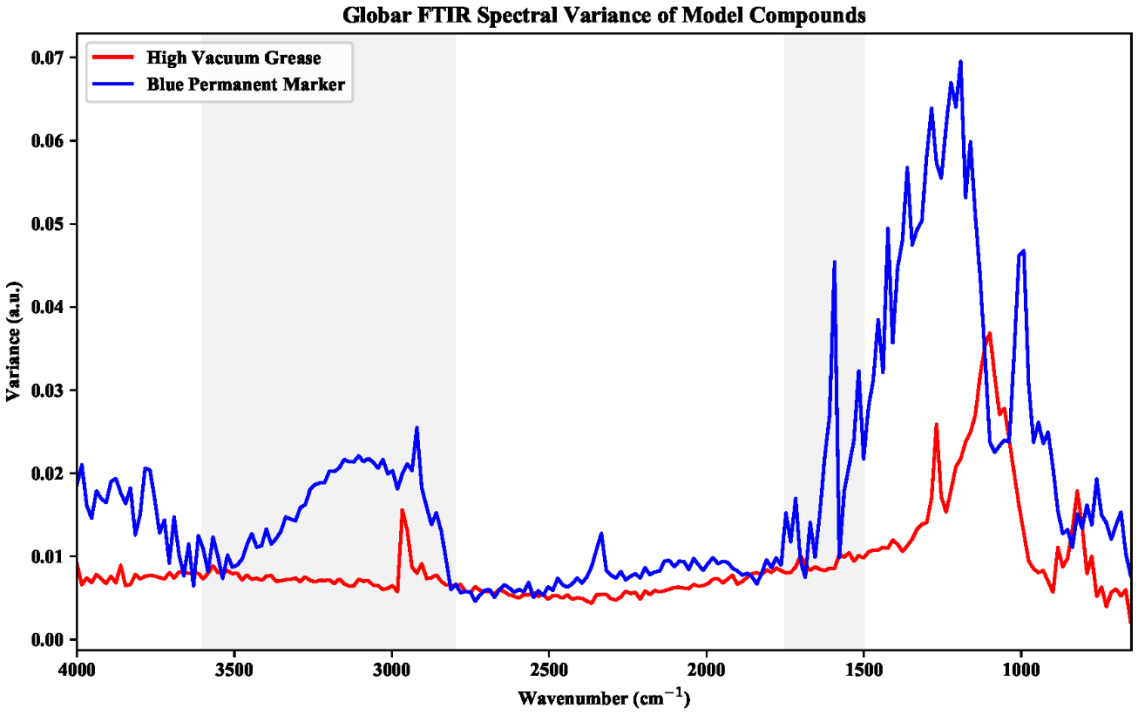


Figure 5. Globar FTIR variance fingerprint spectra of each model compound eventually used in synthetic two-component control sample. The variance was used to determine the domains selected for exploratory PC-LDA (gray).

To verify this conclusion, we tuned the spectral, on-target ratio (OTR) assessment by selecting the major contributing peak per component using our normalized mean standard spectra (**Figure 4**) and variance spectra (**Figure 5**); peak selection guided by normalized spectra emphasize chemical identification[17] over concentration in spectral interpretation. For high vacuum grease, we referenced the symmetric stretching mode of ν(Si-O-Si) at

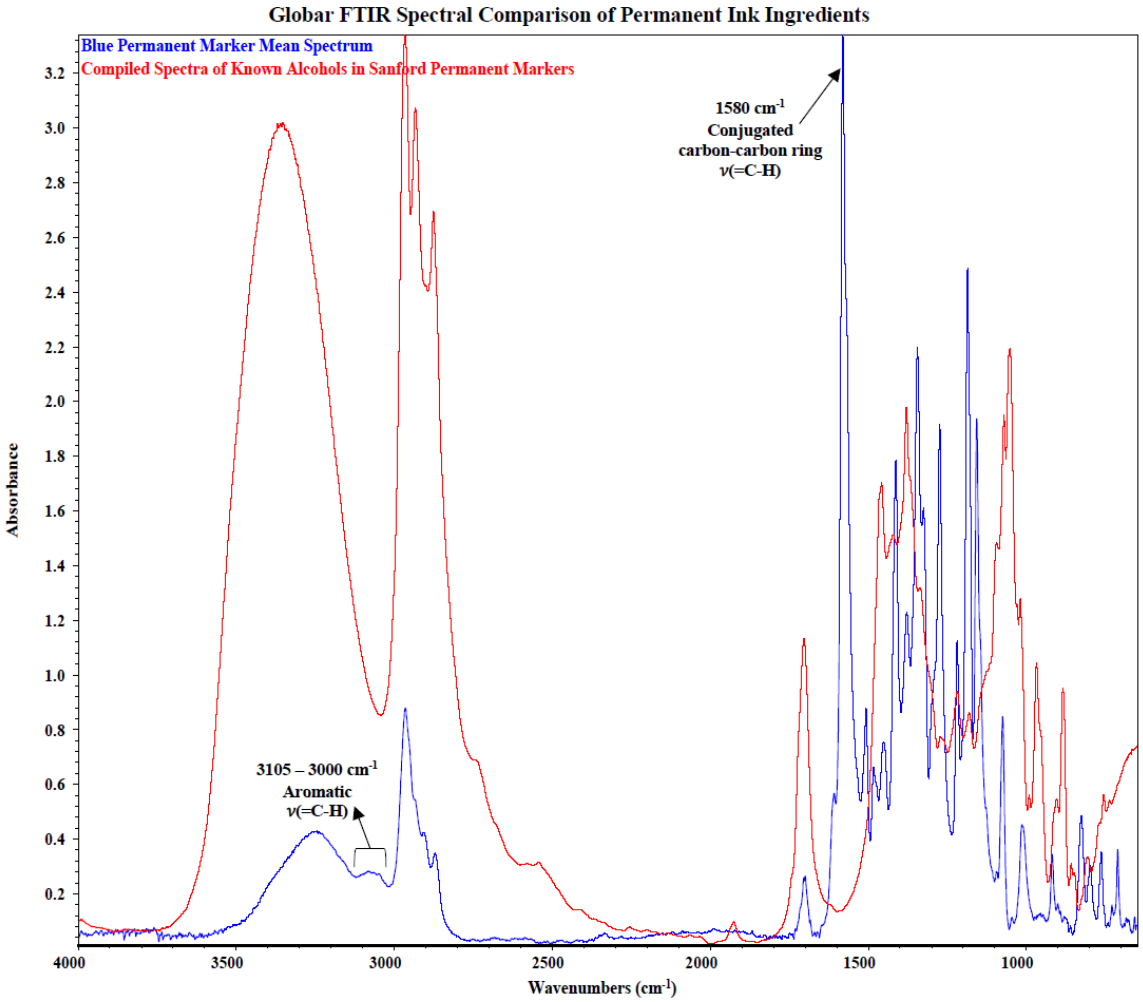


Figure 6. Known alcohol ingredients of Sanford Permanent Marker ink shown as cumulative IR spectrum generated using OMNIC 9.8 alcohol libraries: ethyl alcohol, 1-propanol, 1-butanol, and diacetone alcohol in red to contrast against our normalized permanent marker ink mean spectrum obtained using globar FTIR spectromicroscopy. Identified peaks support the discussion and analysis of the main text presenting 1580 cm^{-1} and region 3105 cm^{-1} to 3000 cm^{-1} vibrational modes as unique to the pigments or dyes present in the deposited ink. Plotted in OMNIC 9.8.

798 cm^{-1} emerging from its fumed amorphous silica[18] composition. For permanent ink presence, we used the major peak at 1580 cm^{-1} stemming from conjugated carbon–carbon ring ν(-C=C-) stretching modes[19] in pigment compounds[20], which was further substantiated by the presence of aromatic ν(=C-H) vibrations between 3105 cm^{-1} and 3000 cm^{-1}[19] (**Figure 6**). All spectra were evaluated for non-adaptive UG and adaptive LIV experiments prior to processing the OTR as the proportion of on-target sampled points to total sampled points. Using this spectrally-based metric for enhanced real-world fidelity[21], we confirmed that adaptive LIV data acquisition (OTR = 0.95) outperforms non-adaptive UG

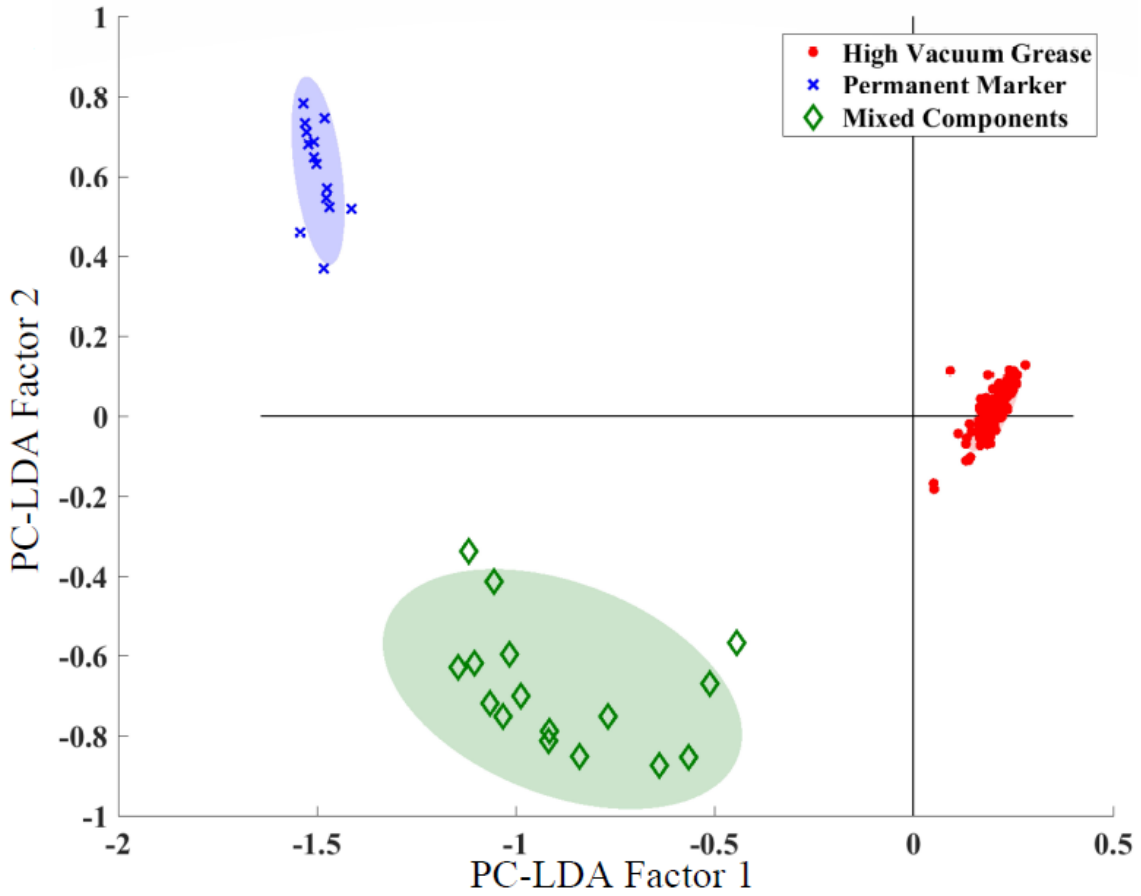


Figure 7. Exploratory 2D PC-LDA score plot of noise-filtered and baseline-corrected acquired data 95% confidence intervals as shaded areas.

(OTR = 0.19) data acquisition in experimental cases where domain knowledge is either limited or unavailable.

To verify that our acquired adaptive LIV data is interpretable through multivariate analysis from an experimenter's standpoint, we performed principal component analysis (PCA) followed by linear discriminant analysis (LDA)[22] on the noise-removed IR data to discriminate between the permanent marker and high vacuum grease present in the spatiochemical map (**Figure 7**). We see that the first PC-LDA factor distinguishes between permanent ink-containing spectral regions and pure high vacuum grease, while the second PC-LDA factor separates between pure permanent ink and regions containing both permanent ink and high vacuum grease. This conclusion is further supported by the mean spectra plotted per cluster (**Figure 8**); we see the imine ν(C=N-H) from 3400 cm^{-1} to 3300 cm^{-1} and intermolecular hydrogen-bonded ν(O-H) at 3550 cm^{-1} and 3230 cm^{-1}

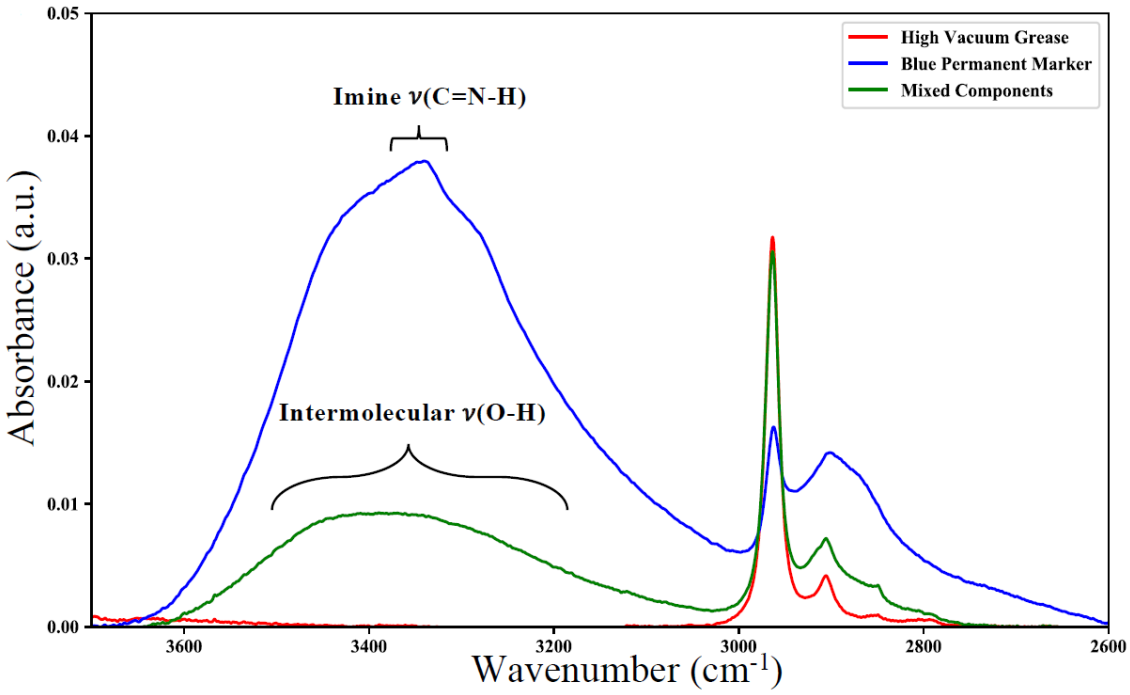


Figure 8. Globar FTIR spectral mean of each identified PC-LDA cluster showing the peak identifications used to corroborate chemical identity of high vacuum grease (red), permanent marker (blue), and mixed components (green).

contributions[19] from permanent ink's pigment compounds and alcohols, respectively. The identified high vacuum grease cluster matches the standard mean spectra expectations with vibrational silence in frequencies >3000 cm^{-1}, while peak broadening and the change in peak ratio between the imine and intermolecular hydrogen bonding regions of the permanent ink and mixed component clusters suggest that permanent ink alcohols experienced inhibited evaporation in the mixed component regions due to the ink's deposition under the high vacuum grease during sample preparation (**Figure 9**).

For our proof-of-principle bioimaging case, we applied scanning broadband SR-FTIR spectromicroscopy to overcome signal-to-noise limitations when characterizing a young L2 *C. elegans* animal. *C. elegans* are well characterized in genetics, microscopy, and omics-related fields while also representing a large, whole-organism experimental model

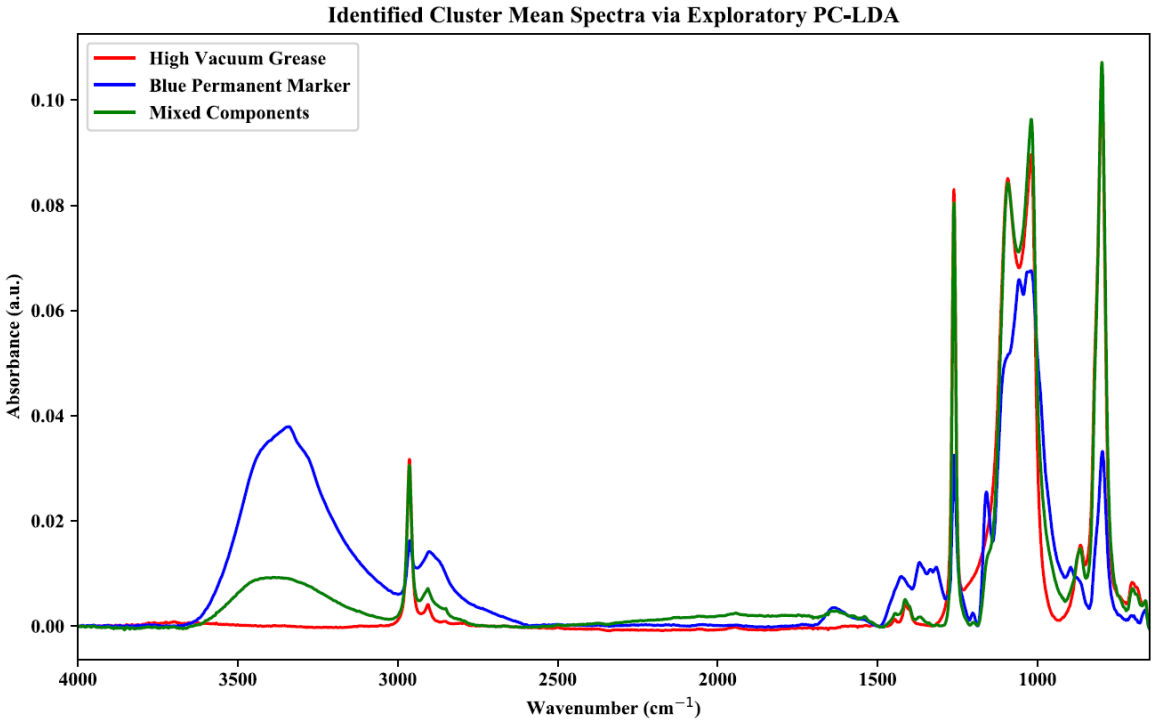


Figure 9. Full spectral region for identified cluster mean spectra of baseline-corrected cluster spectra. Spectra were not normalized in order to conserve concentration information.

containing known compartmentalized chemistry. Relative to the diffraction-limited spatial resolution (2–10 μm) of scanning SR-FTIR spectromicroscopy, their large size of 100 μm to 1 mm in length when coupled with current mapping region software restrictions often lead to temporally inefficient spatiochemical mapping of unfixed samples. With our implemented graphical user interface at ALS Beamline 1.4.3 (Lawrence Berkeley National Laboratory, Berkeley, CA), we were able to apply domain knowledge in spatial and spectral restrictions to better optimize our adaptive data acquisition of *C. elegans* (**Figure 10**) for comparison

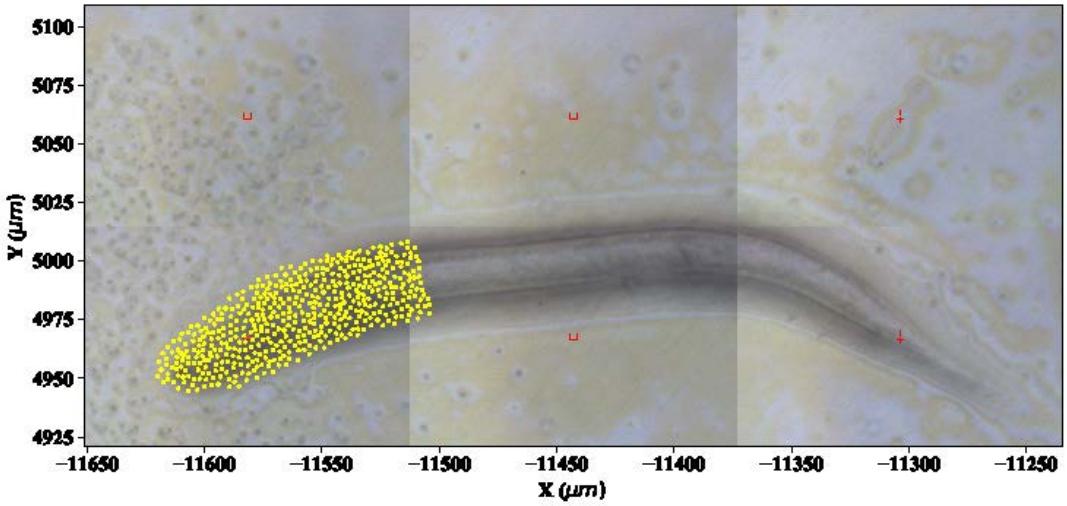


Figure 10. Adaptive LIV data acquisition of an early-stage L2 *C. elegans* over a spatial map defined with domain knowledge to include the pharyngeal, nerve ring, and intestinal regions of the animal.

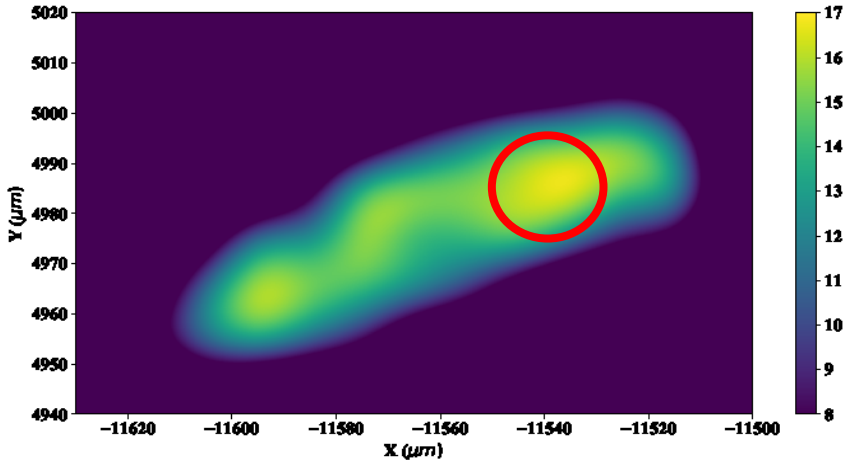


Figure 11. Density plot of sampled areas from adaptive LIV sampling for ease of more frequently sampled regions (red circle).

against the high-spatial (step-size 1.5 μm) resolution map of the same sample. We found that increased adaptive LIV sampling in the spatial domain (**Figure 11**) identified regions characterized by chemistries consistent with those of known anatomical features. Sampling increased in either transitional or overlapping anatomical regions between pharyngeal, head,

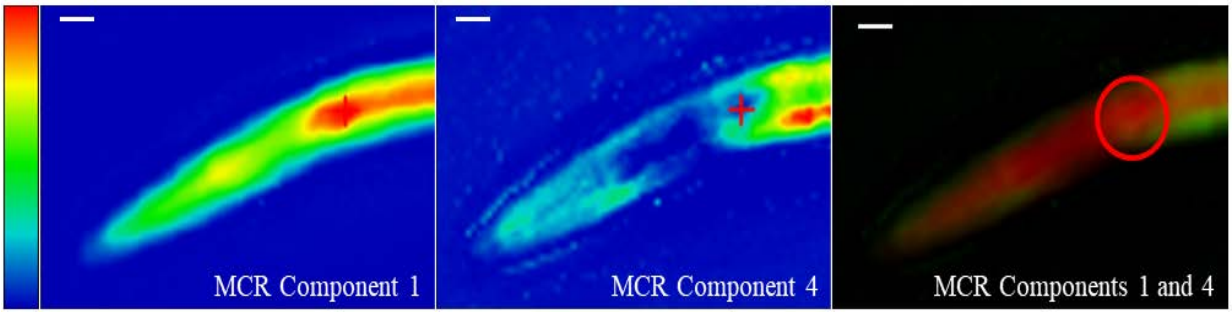


Figure 12. Coefficient heat maps of MCR component 1 (intensity range: 0.00–5.08), MCR component 4 (intensity range: 0.00–1.13), and the overlaid RG coefficient maps of both MCR components 1 and 4. Red cursor indicates same pixel; red circle indicates same region of dense sampling shown in *Figure 11* (scale bar, 10 μm).

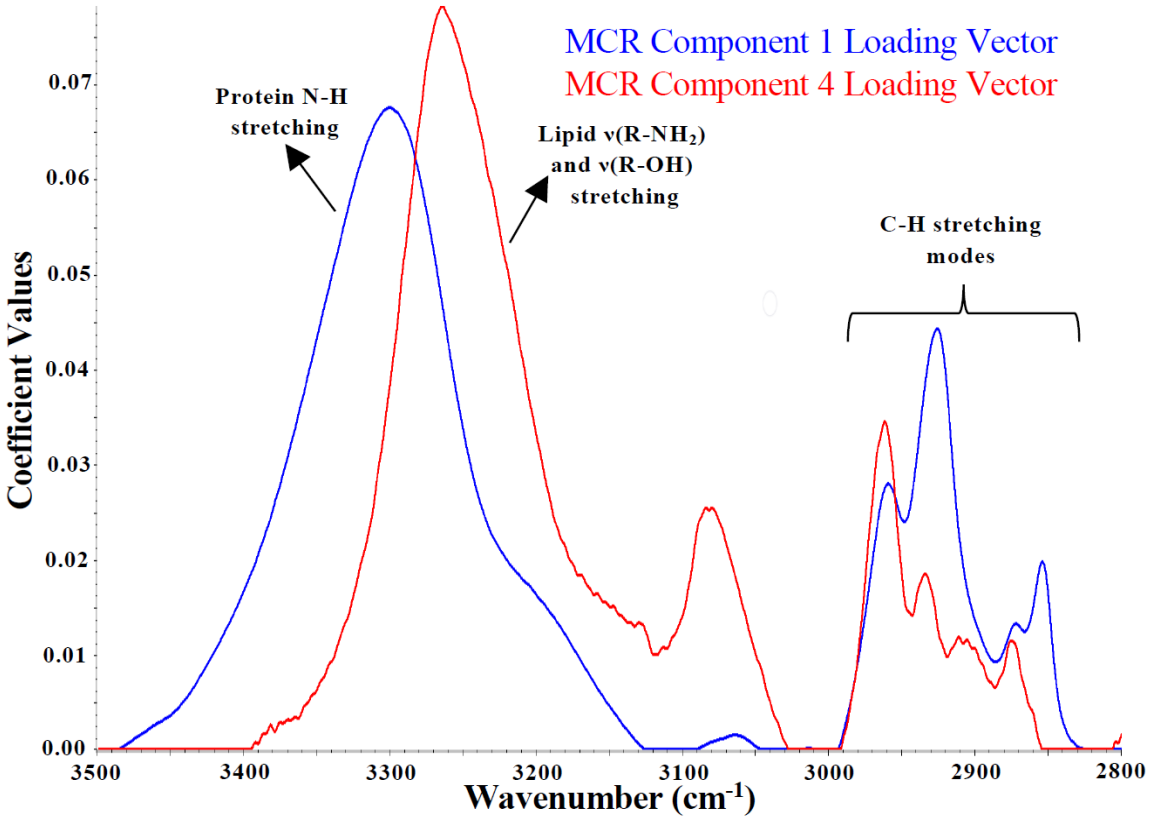


Figure 13. Loading vectors for MCR components 1 (blue) and 4 (red) over evaluated frequency domain of 3500 cm^{-1} to 2800 cm^{-1}. Vibrational stretching assignments are labeled as discussed in the main text and *Figure 14*.

neck, and body wall muscle[23], regions of the nerve ring[24], and the lipid-rich intestine[25]. Our qualitative post validation of adaptive data acquisition using multivariate curve resolution[26] (MCR) and Fourier self-deconvolution[27] (FSD) SR-FTIR analyses further confirmed these anatomical co-localization results with reliable MCR components[28] 1 and 4 (**Figure 12**) corresponding to hydrated proteins (amino acid ν(N-H) stretching modes)[19] and hydrated lipid assemblies (N-H, O-H, methyl, and ν(-$(CH_2)_n$-)

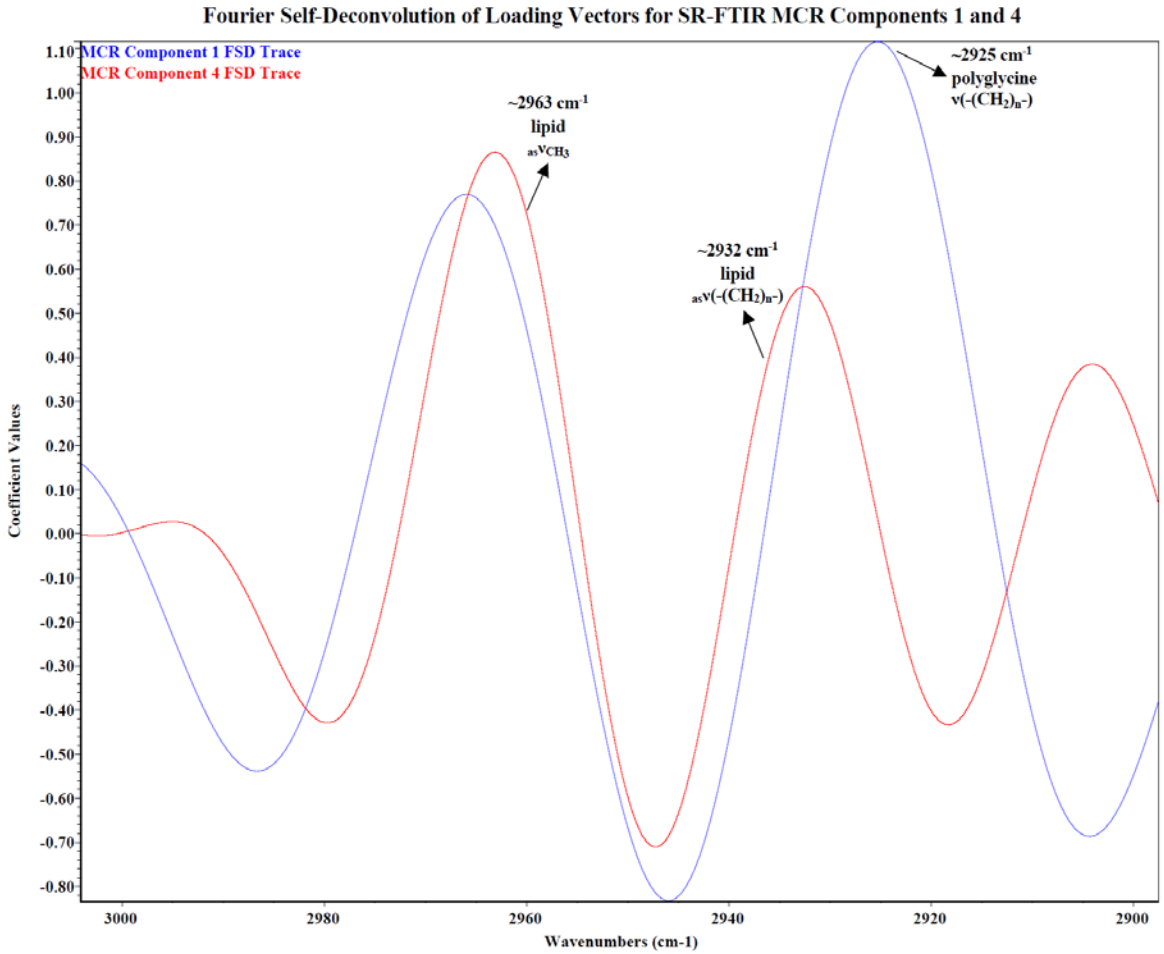


Figure 14. FSD plots of the 3000 cm^{-1} to 2900 cm^{-1} spectral region for MCR components 1 and 4 for accurate peak identification in a high signal-to-noise region of C-H stretching vibrations. To be conservative, we only used major peaks to support assignments and co-localization with chemistry of known anatomical structures. For component 4, we referenced the asymmetric stretching mode of characterized lipid methyl groups (~2963 cm^{-1}) and lipid antisymmetric stretching -$(CH_2)_n$- modes (2916 cm^{-1} – 2936 cm^{-1}). For component 1, we referenced the asymmetric stretching -$(CH_2)_n$- mode of characterized biological polyglycines (~2925 cm^{-1}). Plotted in OMNIC 9.8.

stretching modes)[19, 29], respectively (**Figure 13** and **Figure 14**). With these two components overlapping in the more frequently sampled region, we verify that adaptive LIV data acquisition helps resolve spatiochemical gradients in a complex whole-organism model.

Discussion

We constructed and implemented grid-less adaptive LIV data acquisition to address a key challenge in the hyperspectral imaging of time-sensitive systems. Specifically, we decrease image acquisition time while improving sampling density in regions of increased spatiochemical complexity. Using this sampling strategy, we non-destructively explore the chemistry of anatomical features in living *C. elegans*. We observe that increased sampling density corresponded with known anatomical features, and these results serve as a proof-of-principle for the use of AADA on a complex, biological specimen.

In this study, all experimental LIV-based AADA cases were performed on standard hardware found with commercial high-dimensional imaging microscope designs, revealing the accessibility and computational efficiency of the algorithm for a broadened use in imaging techniques that require a sequential exploration of space, such as scanning probe techniques. We show that LIV-based AADA can operate efficiently and effectively under conditions where the map area is unconstrained, and therefore, when the main goal of a study is to characterize a system through a discovery approach. This performance implies that LIV-based AADA will still benefit an experimentalist who has a detectable, discovery aspect of their research in otherwise well-characterized biological systems that can range from single living cells to, in the case of smaller animal models like *C. elegans*, whole organisms. We also report that LIV-based adaptive sampling outperforms standard sampling methodologies in complex biological systems in which we apply domain knowledge to restrict mapping

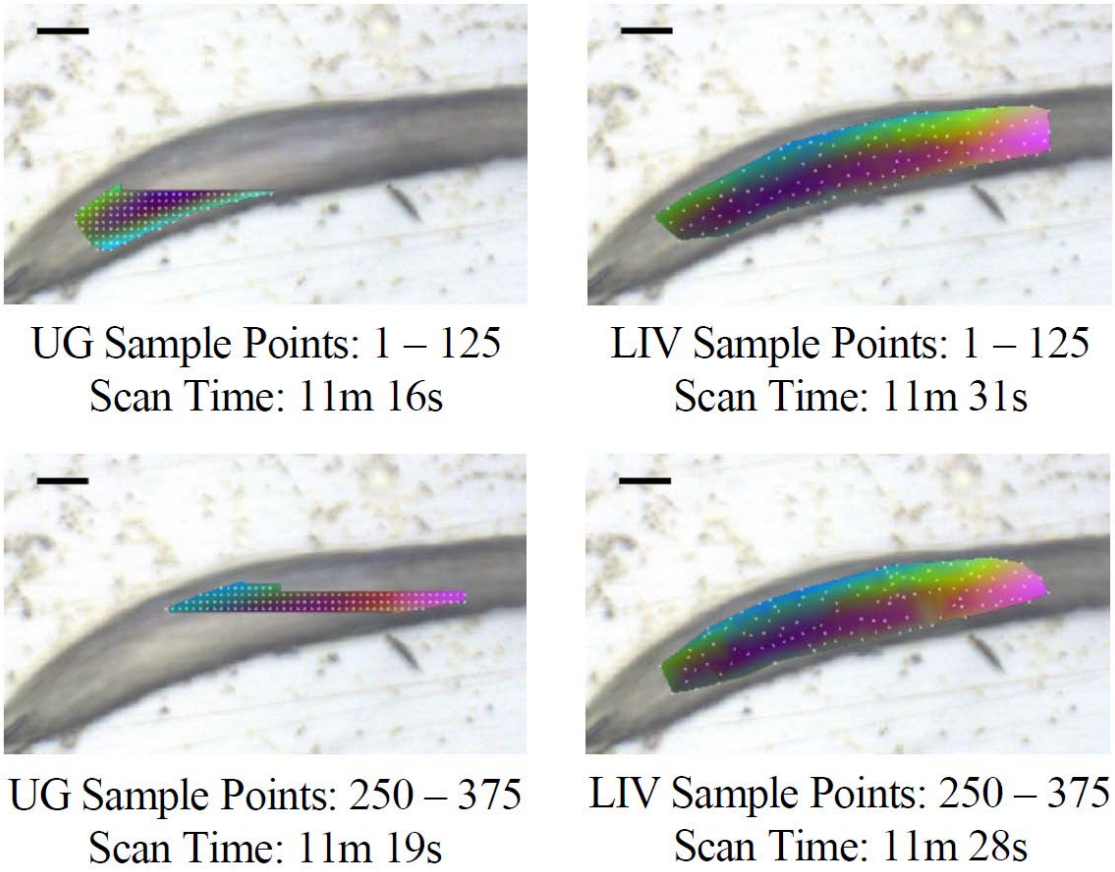


Figure 15. Standard UG sampling compared with LIV sampling over a tightly bound, pharyngeal mapping region of late-stage L1 *C. elegans* (scale bar, 10 μm) per given time interval as defined by ordered, sample point domain. The first three principal components are displayed as RGB false color composites.

regions. When specifically referencing instrument time usage to spatiochemically image the young L2 *C. elegans* experimental case, we were able to map the head region in 45 min with the LIV-based AADA software in comparison to ~4.9 h with the commercially available software. Lastly, we find that LIV-based AADA provides more comprehensive spatiochemical understanding of the total map domain at any given time interval in comparison to the established and standard UG sampling (**Figure 15**), suggesting that this aspect can be harnessed for further development of AADA to achieve adaptive high-dimensional real-time, non-invasive, label-free imaging through modular additions to the sampling algorithm.

This advance in hyperspectral imaging offers the biological community an orthogonal perspective into the dynamic physico-chemical architectures of studied tissues and model organisms. Critically, this information can potentially guide an investigator towards time-points and regions of interest for follow-up omics characterization, which is important in but not limited to the areas of carcinogenesis and developmental biology. Particularly in cases of discovery-based experimental design, AADA enables unbiased assessment of spatially resolved chemical changes between biological samples that differ by genotype, drug treatment or substance exposure, and physiological state such as age. More broadly, LIV-based AADA can be applied to fields outside of biology, such as hyperspectral remote sensing and space exploration. In these cases, future development towards real-time AADA will enable rapid identification, characterization, and even surveillance of chemical spills, toxic algal bloom formation, and spontaneous solar events.

Methods

Autonomous adaptive sampling. Our adaptive sampling workflow is based upon LOOCV and begins with an initial scan of randomly distributed points. Using PCA for dimensionality reduction and our IR preprocessing module for frequency domain restriction and rubber-band baseline correction, we increased computational and temporal efficiency by calculating and operating over the first 5 principal components during our proof-of-principle, temporally intensive, high-dimensional data acquisition. A model U_0 based upon 2D barycentric linear interpolation (LIV) is constructed from this processed data set. We quantify the sensitivity of the surrogate model to the removal of an individual data point through the ϵ_{LOO}. By removing a single point X_i, model U_{-i} is rebuilt using the incomplete data set. The ϵ_{LOO} associated with the sample point is the difference between the two models evaluated at the

removed point $\delta(\mathrm{U}_0(\mathrm{X}_i), \mathrm{U}_{-i}(\mathrm{X}_i))$ with respect to the L_2 norm[30]. After this is iterated for every sampled point in the acquired data set, the region defined by the sampled point with highest ϵ_{LOO} is sampled next by picking a random point within that region. This procedure is repeated until a set criterion is reached, which in our case was 500 total sampled points.

To assess algorithm sampling performance, we aggregate the ϵ_{LOO} of all sample points in the acquired data set and quantify the self-consistency using established LOOCV[31]. We take the mean ϵ_{LOO}, $\langle\epsilon_{\mathrm{LOO}}\rangle_{\mathrm{V}}$, of all sample points in the data set and use it as an unbiased, quantitative measure of the model accuracy due to theoretical guarantees of $\langle\epsilon_{\mathrm{LOO}}\rangle_{\mathrm{V}}$ convergence to a model's generalization error[32]. Since acquired points are often neither regularly nor uniformly distributed in the case of adaptive sampling, we partition the region into a collection of cells $\{V_i\}$ containing positions closest to each point $\{X_i\}$. The mean is then weighted by the associated Voronoi area of sample point $\{X_i\}$. Explicitly, we define

$$\langle\epsilon_{LOO}\rangle_V = \frac{\sum_i |V_i| \cdot \epsilon_{LOO,i}}{\sum_i |V_i|}.$$

With the LOOCV adaptive sampling procedure, we follow the heuristic for $\langle\epsilon_{\mathrm{LOO}}\rangle_{\mathrm{V}}$ minimization, and thus, effectively achieve minimization of model generalization error by sampling near the point with the largest ϵ_{LOO}. This section includes materials provided by co-author Yuan-Sheng Fang (Department of Physics, UC Berkeley) that were modified for the related publication.

Surrogate modeling. We use the *scipy.interpolate.griddata* method from the Python Scipy package to implement 2D barycentric linear interpolation and treat each PCA component independently. Although it is computationally efficient, it does not quantify uncertainty in error estimate. To address this, we include the Voronoi area associated with each point into our calculated ϵ_{LOO} by treating it as an ad hoc regularizer. For a collection of points $\mathrm{X} =$

$\{x_i \in \mathbb{R}^d\}$, the Voronoi cell that we associate with point x_i is the region of space containing positions in Euclidean distance closest to x_i:

$$V_i = \{u \in \mathbb{R}^d : \left|\left|x_i - \boldsymbol{u}\right|\right|_2 < \left|\left|x_j - \boldsymbol{u}\right|\right|_2 \; \forall j \neq i\} \subset \mathbb{R}^d.$$

The Voronoi area is the area of the set, $V_i = ||V_i||$. This implies that if point x_i is spatially isolated from the rest of the data set, then point x_i will be associated with a greater Voronoi area. By approximating the error uncertainty using the Voronoi area, we utilize the fact that linear interpolation error tends to increase with larger distances from points used in the interpolation. To achieve this, we first normalize both ϵ_{LOO} and V_i in order to compare both quantities using a linear scaling from [0, 1]:

$$\sigma(X_i) = \frac{X_i - \min(X)}{\max(X) - \min(X)}.$$

Next, we take the regularized LIV ϵ_{LOO} to be

$$\epsilon_{LOO,i}^{(LIV)} = \sigma\left(\epsilon_{LOO,i}\right) + \sigma(V_i),$$

which is used to calculate our adjusted ϵ_{LOO} for our adaptive data acquisition in simulations and experiments[33]. Our approach is inspired by and related to the LOLA-Voronoi and CV-Voronoi surrogate modeling techniques[11, 14]. This section includes materials provided by co-author Yuan-Sheng Fang (Department of Physics, UC Berkeley) that were modified for the related publication.

Sample preparation. All samples were mounted on 0.5-mm-thick ZnSe crystals, which were cleaned with Milli-Q water, 5% acetic acid, acetone, then Milli-Q water sequentially in order to remove organics while minimizing crystal damage. The two-component control sample was prepared with high vacuum grease (2021854-1993, Dow Corning) that was lightly applied to a 0.5-mm-thick ZnSe crystal (CAS# 1315-09-9, International Crystal

Laboratories) in an area identifiable by fiducial markings drawn with a permanent marker (Item #37003, Sanford Ultra-Fine Blue Sharpie Permanent Marker). Spectral standards were acquired of both components independently prior to autonomous adaptive sample acquisition of abiotic two-component system. Spectral regions of component mixing could be identified by alcohol presence in the mixed spectra.

The first *C. elegans* (N2; Caenorhabditis Genetics Center) animal used for temporal exploratory LIV experiments was selected at the late L1 stage (based on morphology). The second animal for qualitative LIV assessment via FTIR spectral analysis was selected at the young L2 stage (based on morphology). Each animal was moved from their agar growth plates to 1 μL of 0.25 mM Levamisole (CAS# 16595-80-5, Sigma-Aldrich) on the ZnSe crystal and rinsed three times with Milli-Q water before mounting the sample onto the microscope stage for imaging.

Instrumentation. Scanning benchtop and synchrotron FTIR measurements were performed on a Nicolet Nic-Plan IR microscope with a ×32, 0.65 numerical aperture objective with a Thermo Scientific Nicolet iS50 FTIR spectrometer using a KBr beamsplitter and MCT (HgCdTe) detector at Beamline 1.4.3 of the Advanced Light Source at Lawrence Berkeley National Laboratory. Adaptive sampling was implemented using a GUI (**Supplementary Figure 6**) developed in PyQt and installed on the Beamline 1.4.3 computer (Dell Optiplex 7050: 8 GB RAM, Intel Core i5-7500 CPU @ 3.41 GHz, Windows 10 64 bit). OMNIC 9.8 software by Thermo Fisher Scientific controlled the microscope and FTIR bench, and our software communicated with OMNIC through Dynamic Data Exchange to store the OMNIC background-subtracted spectral output into our software's dataframe format.

In this study, we used two different infrared sources: an internal globar source and a synchrotron source. Although an internal globar source is readily available in commercial FTIR microscopes, an accelerator-based synchrotron source offers at least 1000 times improvement in brightness (photon counts per unit time per unit area) over the globar source[34] at the same spatial resolution. As a result, we used different total co-added scan and spatial resolutions for measurements performed on each instrument, which are detailed in the following sections.

Globar FTIR spectromicroscopy and multivariate analysis. Benchtop scanning FTIR measurements using internal globar source were performed in transmission mode with an aperture-limited spatial resolution of $75\,\mu m \times 75\,\mu m$. IR spectra between $650\,cm^{-1}$ and $4000\,cm^{-1}$ at $4\,cm^{-1}$ spectral resolution were collected with 16 co-added scans at a interferometer mirror velocity of $1.83\,cm/s$. Rubber-band baseline correction and dimensionality reduction via PCA to 5 components were performed over the entire collected spectral region during adaptive LIV data acquisition. For each experimental assessment of sampling method, the total sampled points were limited to 500 to remain below the full-resolution map of 840.

On-target ratio. We define the on-target ratio (OTR) to be the number of samples that meet the on-target criteria over the total number of sampled points. To determine the criteria by which a spectrum is considered on-target, we use our full-resolution data set and remove spectra close to the detection limitations of the instrument that violate the signal-to-noise filter criteria. Using this noise-removed subset of data, we calculate the mean spectra of the noise-removed subset. After identifying one major peak component per known component standard, we evaluated all acquired spectra per method for the presence of either

aforementioned peak above the threshold that we determined as the noise-removed mean intensity at defined frequencies to define OTR as

$$OTR_{method} = \frac{N_{a\vee b}}{N_{total}},$$

where $N_{a\vee b}$ is the number of spectra that met either the first mean peak criterion, second mean peak criterion, or both mean peak criteria, while N_{total} is the total number of spectra acquired using the referenced data acquisition method. Using this definition of spectral metric, we calculated $OTR_{LIV} = \frac{474}{500}$ (0.95) and $OTR_{UG} = \frac{95}{500}$ (0.19).

FTIR multivariate analysis. The control sample components (high vacuum grease and permanent marker) were evaluated individually as spectral standards. The data were baseline corrected and vector normalized using OMNIC 9.8, and the spectral mean was calculated over 8 spectra per standard. Referencing the normalized mean and variance spectra, we use domain knowledge to perform PCA over the frequency domains of 3600 cm^{-1} to 2800 cm^{-1} and 1750 cm^{-1} to 1450 cm^{-1} simultaneously before applying LDA to maximize interclass variance over intraclass variance of our factors[22] of our baseline-corrected and vector-normalized data in MATLAB R2017a. 2D score plots were generated in which the nearness between classes indicates similarity, whereas distance implies dissimilarity. The final mean spectrum of each cluster is shown for spectral validation of vibrational modes, resulting in segregation of classes.

Synchrotron FTIR spectromicroscopy and multivariate analysis. Scanning SR-FTIR diffraction-limited (2–10 μm) spectra were collected in transmission mode between 650 and 4000 cm^{-1} at 4 cm^{-1} spectral resolution and recorded with 8 co-added scans at an interferometer mirror velocity of 6.3 cm/s. We restricted the spectral domain adaptive LIV

sampling workflow from 900 to 3700 cm^{-1} to avoid signal contamination from detectable synchrotron noise and to decrease sample morphology[17] baseline effects, respectively, on subsequent dimensionality reduction and error calculation steps. Rubber-band baseline correction and dimensionality reduction to 5 components was performed over the restricted spectral region between 900 cm^{-1} and 4000 cm^{-1} during adaptive LIV sampling. Using domain knowledge, we restricted our mapping region to the pharynx, nerve ring, and intestine[24] of our young L2 *C. elegans* to reduce off-target sampling with respect to *C. elegans* for increased temporal efficiency in spatiochemical mapping.

SR-FTIR multivariate analysis. We restricted our analyzed MCR domain from 3500 cm^{-1} to 2800 cm^{-1} for reduction of morphological effects on the spectral baseline and for higher diffraction-limited spatial resolution, since the goal of MCR analysis was to qualitatively assess adaptive LIV data acquisition performance. Based upon the cumulative explained variance calculated by OMNIC 9.8 on our experimental data, we performed MCR analysis in OMNIC using 5 components in which 99.82% of data variance is explained. Guided by well-characterized *C. elegans* anatomy and chemistry, we identified reliable MCR components[28] that would strongly correlate with muscle and lipid assembly structures—components 1 and 4. For better accuracy in peak identification on our MCR components, we applied FSD to the C-H vibrational region. Since our analysis region was restricted, we could only broadly state the presence of protein-related stretching vibrations of ν(N-H) from amino acids between 3390 cm^{-1} and 3260 cm^{-1}[19] and polyglycine asymmetric CH_2 stretching modes at ~2925 cm^{-1}[19] (**Figure 14**) in MCR component 1. Similarly for MCR component 4 and in referencing characterized hydrated lipid assemblies, we found broad peak contributions from N-H and O-H stretching modes between 3400 cm^{-1} to 3100 cm^{-1}[29],

lipid-relevant antisymmetric $\nu(\text{-}(CH_2)_n\text{-})$ modes at ~2932 cm^{-1}[29], and lipid-related methyl antisymmetric and symmetric stretching at 2963 cm^{-1} and 2873 cm^{-1}[29], respectively.

Statistics and reproducibility. Each sample size, type, and statistical method applied is described in the relevant **Method** section. For the two-component model system, spectral standards for permanent ink and high vacuum grease were performed with sample replicates ($n = 8$). Infrared spectral data are available through the CaltechDATA repository (https://doi.org/10.22002/D1.1609)[35].

References

1. Liu, T.L., et al., *Observing the cell in its native state: Imaging subcellular dynamics in multicellular organisms.* Science, 2018. **360**(6386).
2. Chalfie, M., et al., *Green fluorescent protein as a marker for gene expression.* Science, 1994. **263**: p. 802-805.
3. Giepmans, B.N., et al., *The fluorescent toolbox for assessing protein location and function.* Science, 2006. **312**(5771): p. 217-24.
4. Bodenmiller, B., *Multiplexed Epitope-Based Tissue Imaging for Discovery and Healthcare Applications.* Cell Syst, 2016. **2**(4): p. 225-38.
5. Choi, M., S.J. Kwok, and S.H. Yun, *In vivo fluorescence microscopy: lessons from observing cell behavior in their native environment.* Physiology (Bethesda), 2015. **30**(1): p. 40-9.
6. Prahst, C., et al., *Mouse retinal cell behaviour in space and time using light sheet fluorescence microscopy.* Elife, 2020. **9**.
7. Zhao, M., et al., *Electrical signals control wound healing through phosphatidylinositol-3-OH kinase-gamma and PTEN.* Nature, 2006. **442**(7101): p. 457-60.
8. Rohban, M.H., et al., *Capturing single-cell heterogeneity via data fusion improves image-based profiling.* Nat Commun, 2019. **10**(1): p. 2082.
9. Asprey, S.P. and S. Macchietto, *Designing robust optimal dynamic experiments.* Journal of Process Control, 2002. **12**: p. 545-556.
10. Queipo, N.V., et al., *Surrogate-based analysis and optimization.* Progress in Aerospace Sciences, 2005. **41**: p. 1-28.
11. Crombecq, K., et al. *A novel sequential design strategy for global surrogate modeling*. in *2009 Winter Simulation Conference (WSC)*. 2009.
12. Li, G., V. Aute, and S. Azarm, *An accumulative error based adaptive design of experiments for offline metamodeling.* Structural and Multidisciplinary Optimization, 2010. **40**.
13. Wang, C., et al., *An evaluation of adaptive surrogate modeling based optimization with two benchmark problems.* Environmental Modelling & Software, 2014. **60**: p. 167-179.
14. Xu, S., et al., *A Robust Error-Pursuing Sequential Sampling Approach for Global Metamodeling Based on Voronoi Diagram and Cross Validation.* Journal of Mechanical Design, 2014. **136**(7).
15. Singh, P., D. Deschrijver, and T. Dhaene, *A balanced sequential design strategy for global surrogate modeling*, in *Simulation Conference (WSC), 2013 Winter*. 2013, IEEE. p. 2172–2179.
16. Elisseeff, A., T. Evgeniou, and M. Pontil, *Stability of randomized learning algorithms.* Journal of Machine Learning Research, 2006. **6**: p. 55-79.
17. Baker, M.J., et al., *Using Fourier transform IR spectroscopy to analyze biological materials.* Nature Protocols, 2014. **9**(8): p. 1771-1791.
18. Lippincott, E.R., et al., *Infrared studies on polymorphs of silicon dioxide and germanium dioxide.* Journal of Research of the National Bureau of Standards, 1958. **61**: p. 61-90.

19. Socrates, G. and G. Socrates, *Infrared and Raman characteristic group frequencies : tables and charts*. 3rd ed. 2001, Chichester ; New York: Wiley. xv, 347 p.
20. Awab, H., et al., *Infrared spectroscopic technique for the forensic discrimination of marker pen inks.* Malaysian Journal of Forensic Sciences, 2011. **2**: p. 1-7.
21. Razavi, S., B.A. Tolson, and D.H. Burn, *Review of surrogate modeling in water resources.* Water Resources Research, 2012. **28**.
22. Hu, P., et al., *Metabolic phenotyping of the cyanobacterium Synechocystis 6803 engineered for production of alkanes and free fatty acids.* Applied Energy, 2013. **102**: p. 850-859.
23. Altun, Z.F. and D.H. Hall, *Muscle system introduction*, in *WormAtlas*. 2009.
24. Altun, Z.F. and D.H. Hall, *Introduction to C. elegans anatomy*, in *WormAtlas*. 2009.
25. Mak, H.Y., *Lipid droplets as fat storage organelles in Caenorhabditis elegans.* Journal of Lipid Research, 2012. **53**(1): p. 28-33.
26. Felten, J., et al., *Vibrational spectroscopic image analysis of biological material using multivariate curve resolution–alternating least squares (MCR-ALS).* Nature Protocols, 2015. **10**(2): p. 217-240.
27. Tooke, P.B., *Fourier self-deconvolution in IR spectroscopy.* Trends in Analytical Chemistry, 1988. **7**(4): p. 130-136.
28. Motegi, H., et al., *Identification of Reliable Components in Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS): a Data-Driven Approach across Metabolic Processes.* Scientific Reports, 2015. **5**(1): p. 15710.
29. *Infrared Spectroscopy of Biomolecules*. 1995: Wiley-Liss.
30. Ramirez-Lopez, L., et al., *Distance and similarity-search metrics for use with soil vis–NIR spectra.* Geoderma, 2013. **199**: p. 43-53.
31. Cawley, G.C. and N.L.C. Talbot, *Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers.* Pattern Recognition, 2003. **36**(11): p. 2585-2592.
32. Kearns, M. and D. Ron, *Algorithmic Stability and Sanity-Check Bounds for Leave-One-Out Cross-Validation.* Neural Computation, 1999. **11**(6): p. 1427-1453.
33. Bandler, J.W., et al., *Space mapping: the state of the art.* IEEE Transactions on Microwave Theory and Techniques, 2004. **52**(1): p. 337-361.
34. Holman, H.-Y.N., M.C. Martin, and W.R. McKinney, *Tracking Chemical Changes in a Live Cell: Biomedical Applications of SR-FTIR Spectromicroscopy.* Spectroscopy, 2003. **17**: p. 486940.
35. Holman, E.A., *Dataset for Autonomous Adaptive Data Acquisition (AADA)*. 2020: CaltechDATA.

Chapter 4

DESIGNING OPTICALLY TRANSPARENT PHOTOCROSSLINKED PEPTOID NANOSHEETS FOR IMAGING APPLICATIONS

Material from the in-preparation manuscript written by Elizabeth A. Holman containing information from co-authors Dong Li and Ronald N. Zuckermann.

Abstract

Peptoid bilayer nanosheets are a new class of biocompatible 2D nanomaterial with broad applications, including sensing, membrane mimicry, and pathogen binding. They are formed from the self-assembly of sequence-defined peptoid chains and can be readily functionalized to tune their properties. In order to broaden their utility, there is a need to improve their mechanical stability. Here, we create nanosheets that can potentially function as a mechanically robust, optically transparent interfacial barrier for covering microfluidic channels, microscopy grids, and other sample chambers. This barrier would increase hydration control in otherwise open-channel environments, permitting water-sensitive, non-destructive infrared spectroscopy to be incorporated into multimodal imaging platforms. Therefore, we explored new peptoid analogs containing photo-crosslinkable side chains that are capable of forming intra-monolayer as well as inter-layer covalent bonds. These constructs can be co-assembled in various ratios to achieve a range of desirable mechanical properties to provide a new route to ultrathin, mechanically robust, 2D nanomaterials that serve as optically transparent interfacial barriers for a variety of imaging techniques.

Introduction

With the advancement of high-throughput technologies in genomics and transcriptomics, a non-invasive, non-destructive, and label-free imaging technique for investigating cell-to-organism biochemical landscape changes is needed for a more complete characterization of biological systems. As early as the 1950's, modern infrared microscopy was identified as a promising technique that meets the three aforementioned requirements for biological studies interrogating cellular dynamic behavior; unfortunately, its realizable potential has always been limited by the software and hardware technology of the time. This limitation has been addressed with the introduction of mercury cadmium telluride detectors and inexpensive fast digital computers, enabling reasonable FTIR microscope design and manufacturing to be possible[1].

Fourier transform infrared (FTIR) spectromicroscopy is a powerful tool for label-free and non-destructive chemical analysis that merges visible light microscopy to match physical morphology with chemical and molecular information obtained from infrared absorption spectral data. In this manner, FTIR provides molecular insight to unlabeled chemical targets, enabling complementation with gene- and molecule-specific labeling methods in fluorescence microscopy for a more complete understanding of the organism's global biochemical environment under which concentrations of the target molecules change. However, a key challenge in live organism FTIR studies is the development of an imaging platform that both (1) ensures the organism's hydration and survival during real-time and long-term FTIR data acquisition and (2) permits the diverse set fluorescent probe emissivity detection for multimodal imaging concepts, such as complementary single-molecule-specific spatial imaging.

Maintaining hydration of the organism during measurements is difficult when attempting to decrease the IR absorbance of the device namely through minimization of the total thickness of the device and utilization of materials with maximal IR transparency. As shown in **Figure 1**, most current IR/VIS-compatible bioimaging microfluidics platforms use ZnSe windows, which do not transmit light with wavelengths smaller than 600nm (green region) and also causes MIR signal loss due to light attenuation (red region) in the typically 500-micron-thick crystal[2]. Additionally, IR-active vacuum grease is necessary when clamping the ZnSe crystals for closed-channel microfluidic imaging platforms—this introduces a potential source of strong signal contamination in the MIR fingerprint region. The emergence of these spurious bands coupled with light attenuation directly limits the

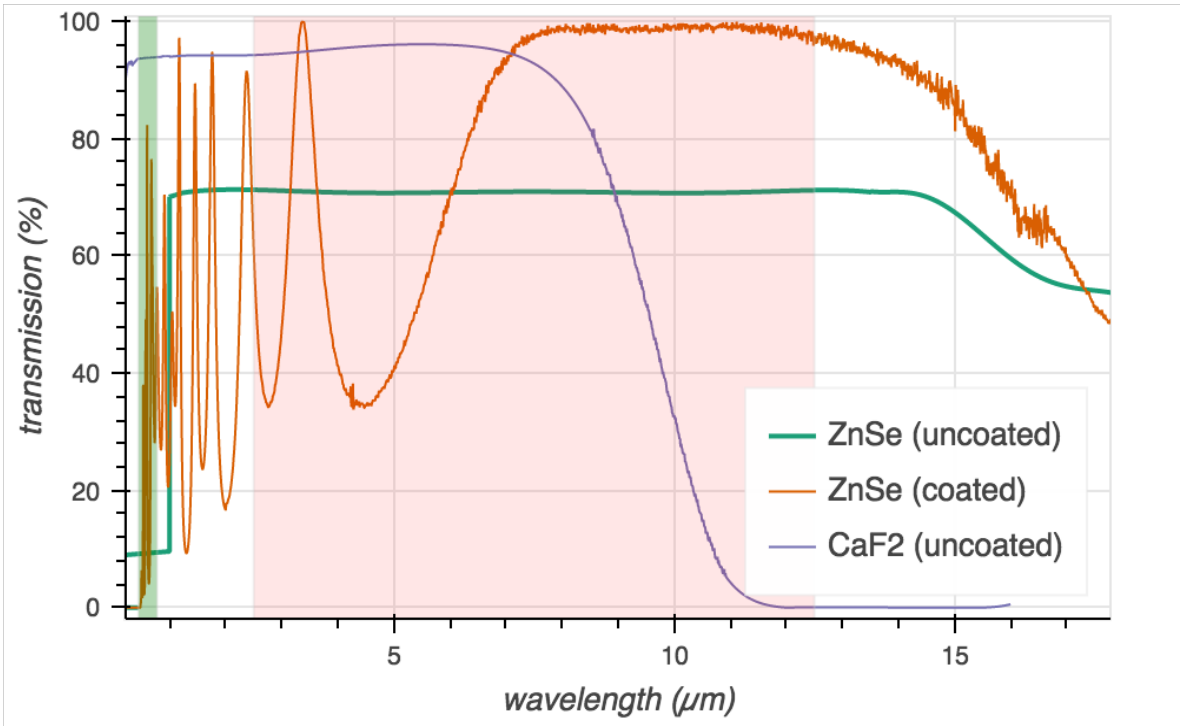


Figure 1. Light transmission through established IR windows. Fluorescence excitation/emission wavelength region displayed as green. MIR broadband spectral region indicated by red. Raw transmission data obtained from publicly disclosed ThorLabs dataset.

biological systems of interest that can be imaged to samples of specific thicknesses and IR transmissivities.

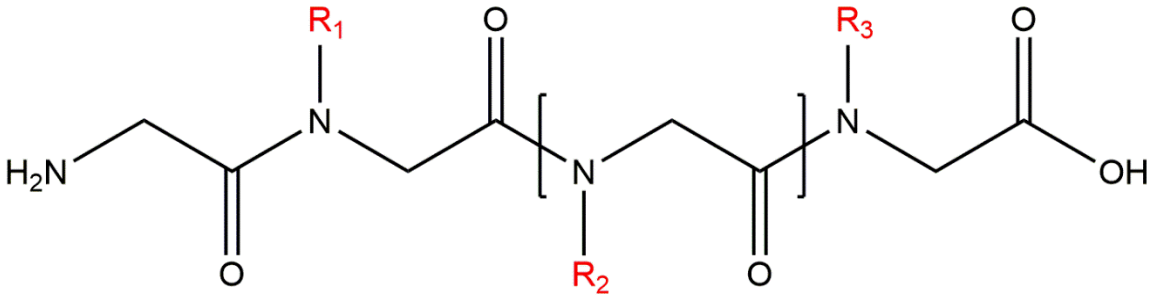


Figure 2. Generic peptoid oligomer structure. The repeating N-substituted glycine monomer is in brackets. Side-chain substituents in red can be diverse in speciation to alter peptoid oligomer structure and function.

Peptoid nanosheet technology presents potential solutions to previous optical, mechanical, and chemical challenges that have hampered multimodal bioimaging platform development for studies of living, whole organisms. Peptoids were developed in the late 1980s and are effectively regioisomers of peptides that retain the amide-based backbone for ease of synthesis[3], as displayed in **Figure 2**. Additionally, the incorporation of the side chain directly to the amide nitrogen can block protease activity, this being ideal when designing biocompatible barriers for imaging platforms. When placed at an air-water interface, peptoid oligomers can self-assemble into an ordered monolayer of peptoids. This monolayer can be either Langmuir-Blodgett or Langmuir-Schaefer transferred onto substrates for a variety of applications such as sensing[4], or it can be laterally compressed to the point of collapse, the monolayer serving as an intermediate for the eventual formation of a peptoid bilayer nanosheet[5].

In order to broaden the application of peptoid nanosheet technology to potentially address IR/VIS bioimaging platform challenges, we determined that synthesizing a new

chloro-containing analog of peptoid Block-28 (B28) was necessary. In 2016, Flood et al. showed that para-chloro analogs of B28 (B28pCl$_7$; **Figure 3a**) were able to photocrosslink through the hydrophobic core of the peptoid bilayer nanosheets for improved chemical and mechanical stability[6], and Robertson et al. showed that functionalizing the phenyl ring at the meta position with a methyl group did not disrupt of B28 nanosheet formation[7]. Combining these two concepts, we aimed to synthesize the meta-chloro analog of B28 (B28mCl$_7$; **Figure 3b**) with the goal of increasing the mechanical stability of a peptoid monolayer deposition onto a substrate by either Langmuir-Blodgett or Langmuir-Schaefer

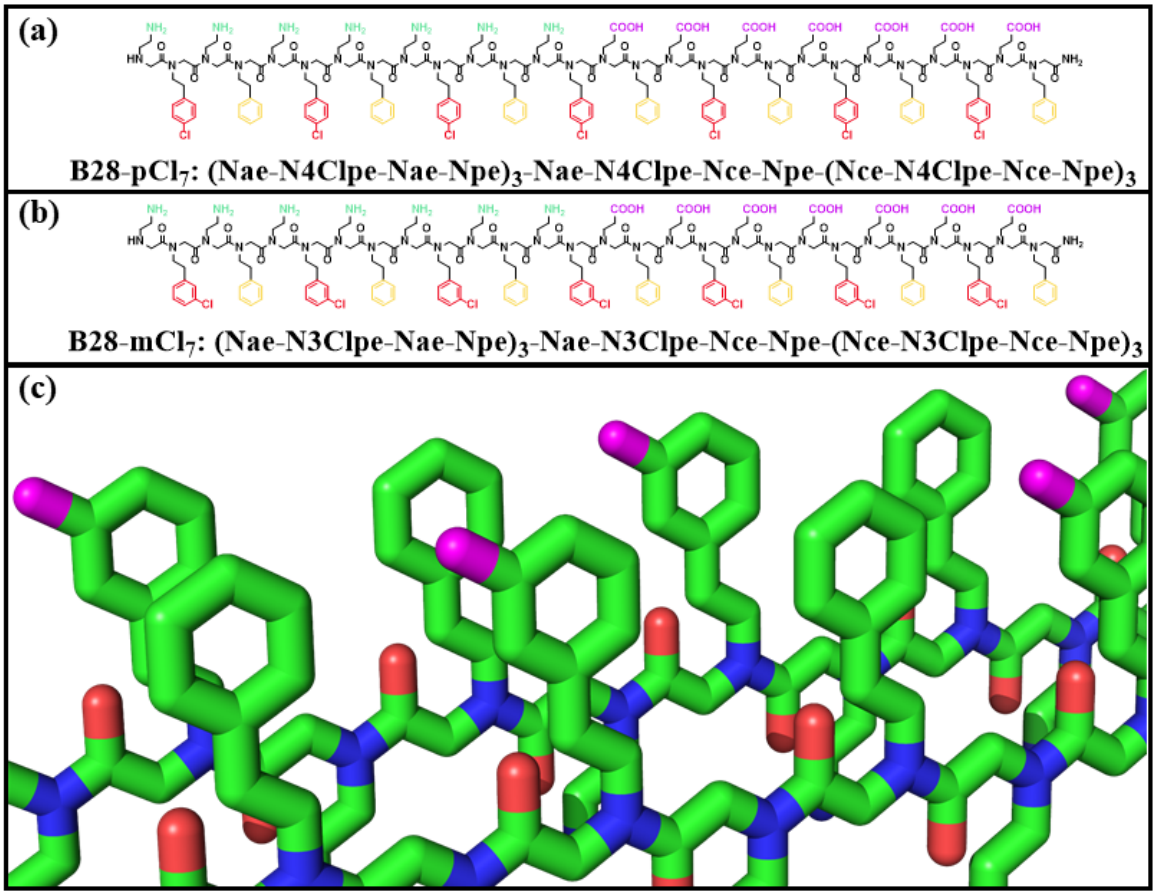


Figure 3. (a) Synthesized and published chloro-containing peptoid B28 analog has shown photo-crosslinking behavior in peptoid nanosheets. (b) Our proposed chloro-containing peptoid B28 analog. (c) Representative cartoon showing feasibility of intralayer and intermolecular covalent bond formation due to the ideal positioning of the chloride atom (pink) in peptoid B28-mCl$_7$.

methods. By functionalizing the phenyl ring at the meta position, we avoid disruption of peptoid monolayer during self-assembly. Notably, by substituting a photolabile chloride onto the phenyl ring at the meta position, we place the chlorine atom in close proximity to the neighboring polypeptoid strand for potential intermolecular crosslinking within the hydrophobic region of the peptoid monolayer (**Figure 3c**). Ideally, the incorporation of peptoid B28mCl$_7$ into the peptoid mixture that forms the peptoid monolayer on a Langmuir trough will increase the mechanical stability of the initial peptoid monolayer after UV exposure and prior to its deposition onto a silicon substrate to potentially cover etched microchannels and wells of a IR/VIS bioimaging platform.

Results and discussion

We successfully synthesized and characterized the identity (mass spectra) and purity (HPLC traces) of our peptoid stock solutions by UPLC-MS (see **Methods** and **Appendix B**).

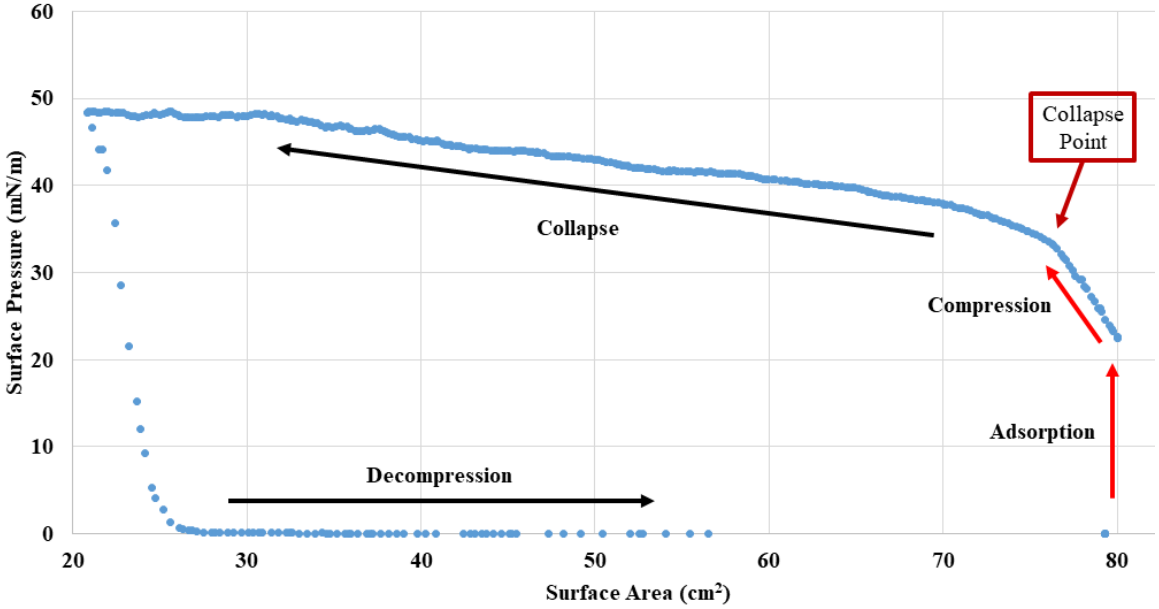


Figure 4. Peptoid block B28mCl7compression-expansion cycle shown as an isotherm. Red arrows are used to indicate at which stage the peptoid monolayer still exists prior to the collapse point between the compression and collapse phase. As the monolayer collapses, a peptoid bilayer is formed in the subphase.

To show that peptoid B28mCl$_7$ is able to form an ordered peptoid monolayer, we directly confirmed that functionalizing the phenyl ring with a chlorine did not disrupt B28mCl$_7$ monolayer formation via Langmuir trough experiments that generated a surface pressure versus area isotherm. As detailed in **Figure 4**, the presence of a collapse point in the compression-expansion cycles implied that peptoid B28mCl$_7$ successfully self-assembled into an ordered solid phase[8].

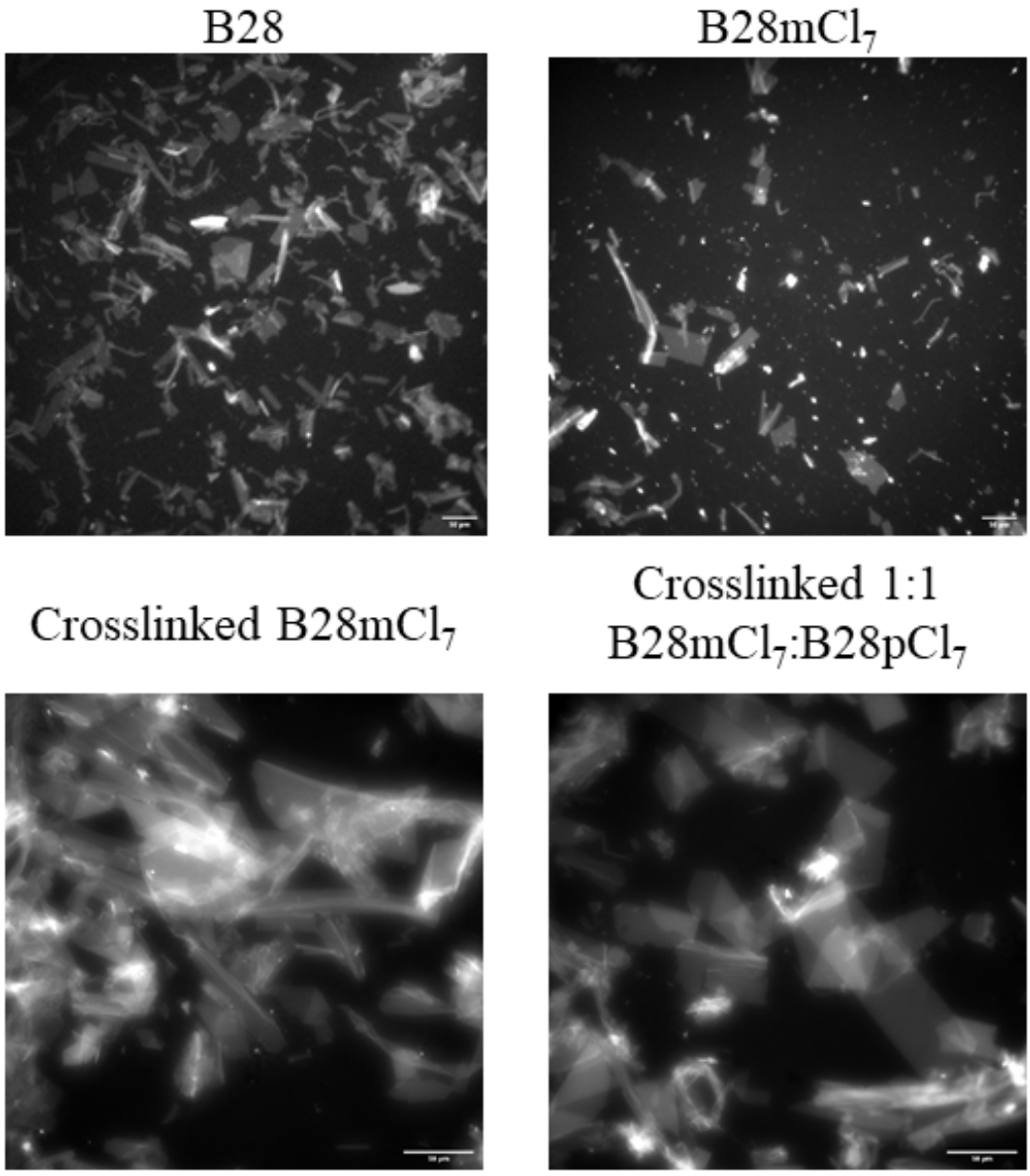


Figure 5. Epifluorescence microscopy of peptoid bilayer nanosheets formed by the vial-rocking method (scale bar = 50 µm). Crosslinked peptoid nanosheets were exposed to 254 nm light in solution prior to imaging on an agarose pad.

Depicted in **Figure 5**, epifluorescence imaging of peptoids B28 and B28mCl$_7$ further supported peptoid bilayer nanosheet formation, since the environmentally-sensitive Nile Red dye fluoresces in hydrophobic environments[5]—meaning that the dye would stably fluoresce when residing within the hydrophobic core. Crosslinked B28mCl$_7$ as well as crosslinked B28mCl$_7$ mixtures provided showed that the new peptoid B28mCl$_7$ could be used in tandem with other peptoid B28 analogs without disrupting nanosheet formation. To verify this fluorescence microscopy result, we further characterized the crosslinked peptoid B28mCl$_7$ nanosheets and compared them to control peptoid B28 and established peptoid B28pCl$_7$. As shown in **Figure 6** and **Appendix C**, the height profiles of each peptoid bilayer nanosheet were between 3 nm and 5 nm. This confirmed that incorporation of peptoid B28mCl$_7$ into peptoid B28 analog mixtures did not disrupt nanosheet formation, even with two different mechanisms of photoreactions present[6].

Having established that peptoid B28mCl$_7$ exhibits a collapsing point on its compression-expansion isotherm and successfully forms peptoid bilayer nanosheets of thicknesses that are similar to control peptoids B28 and B28pCl$_7$ after UV exposure, we tested our hypothesis that suggested B28mCl$_7$ would provide increased mechanical stability to the bilayer nanosheets into which they were incorporated. B28mCl$_7$ stereoisomer B28pCl$_7$ was already established as a peptoid block that forms nanosheets with increased mechanical stability relative to peptoid B28 bilayer nanosheets[6]. This additional stability is attributed to the photolability of the chloride that was functionalized in the *para* position on the phenyl ring and the subsequent covalent bond formation, since peptoid B28 was structurally unable to easily form new covalent bonds under the same nanosheet crosslinking conditions.

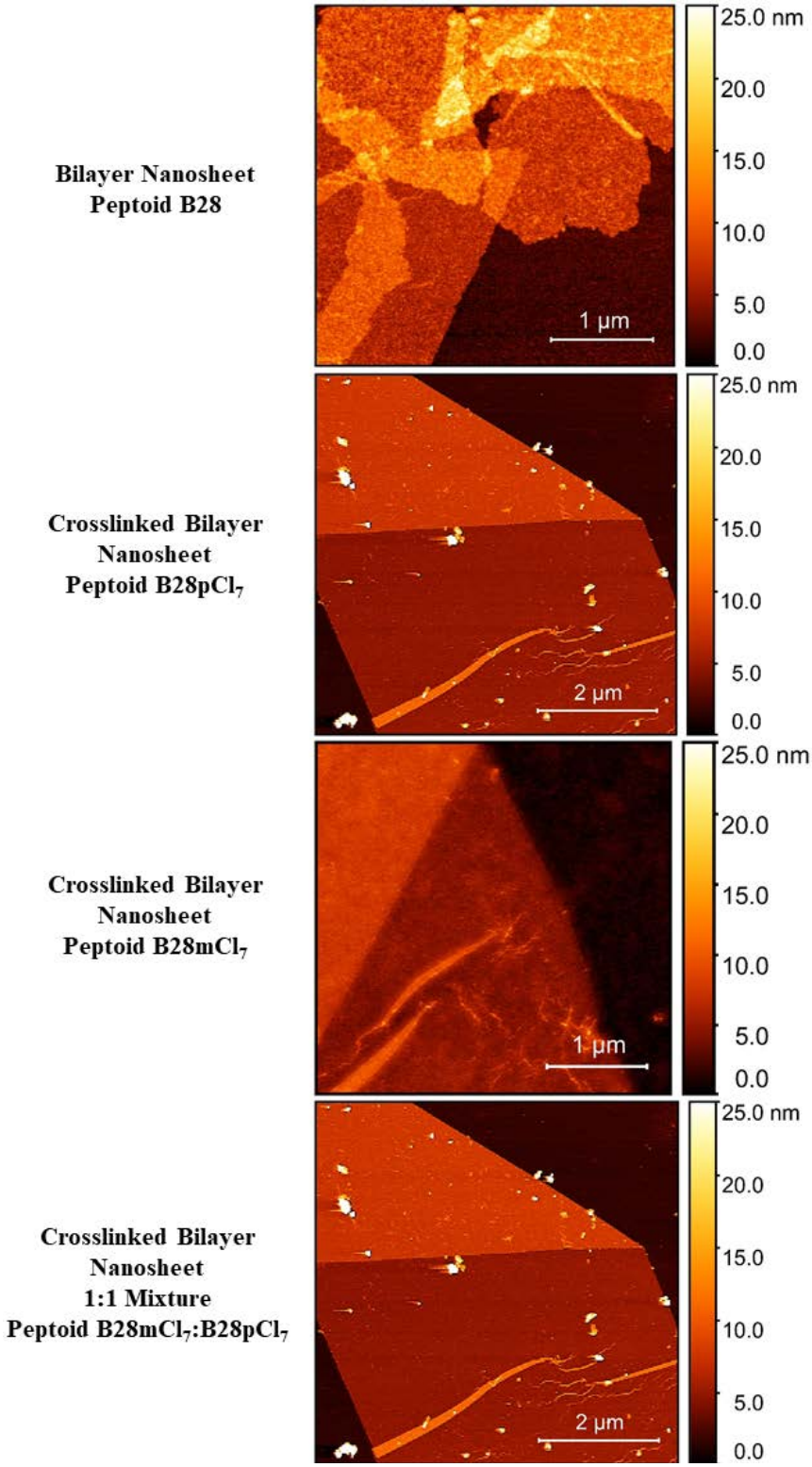


Figure 6. Atomic force microscopy of control and crosslinked peptoid bilayer nanosheets formed by the vial-rocking method. Peptoid bilayer nanosheets are all approximately 3-5 nm thick, although the topography of the nanosheets appear to vary with this sample preparation method.

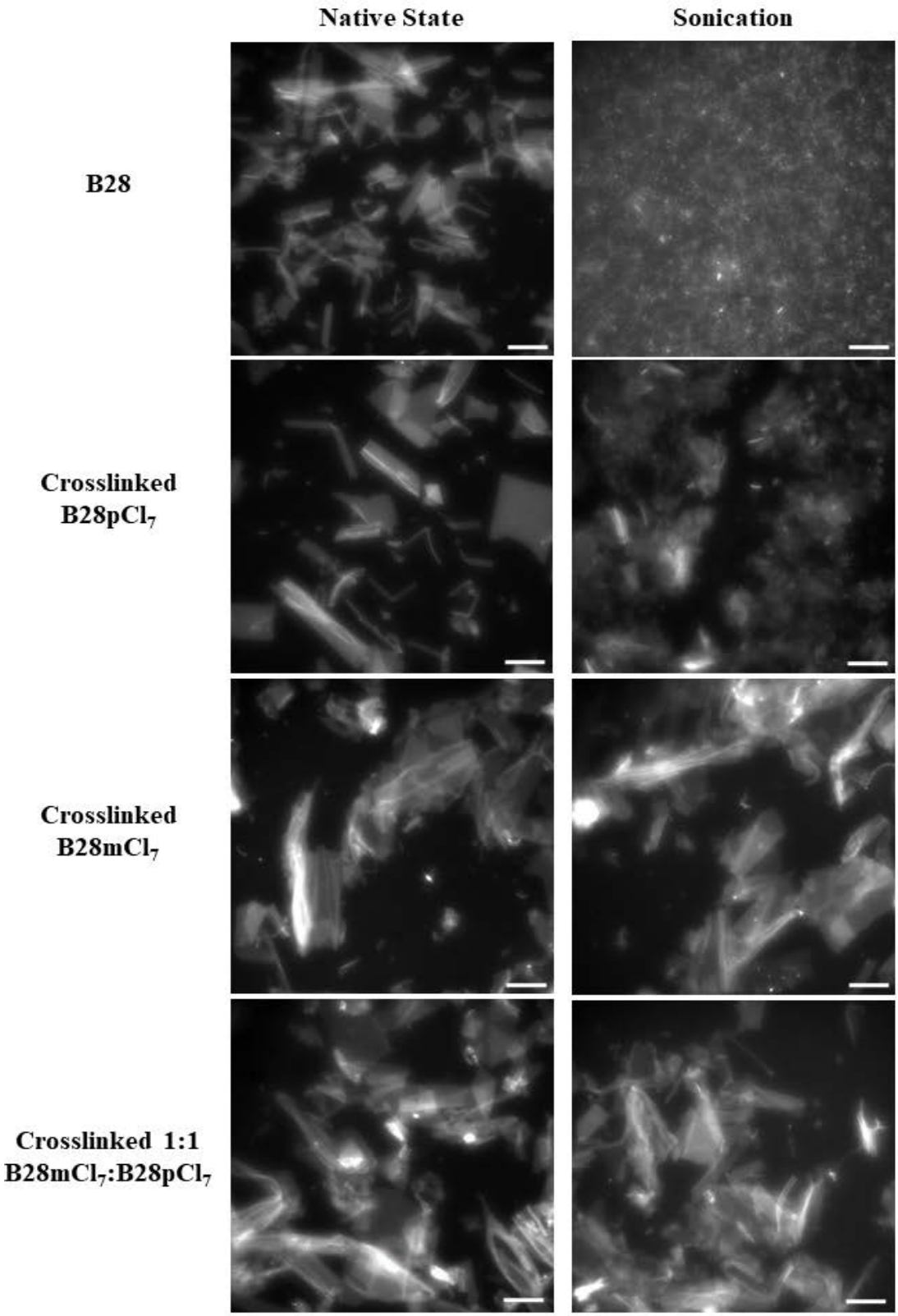


Figure 7. Epifluorescence imaging of established peptoid bilayer nanosheets and crosslinked peptoid bilayer nanosheets incorporating $B28mCl_7$ before (left) and after (right) 60 minutes of sonication (scale bar = 50 µm).

Modifying the previously established mechanical assessment of peptoid bilayer nanosheets by sonication[6], we increased the sonication time from 30 minutes to 60 minutes. As shown in **Figure 7**, we found that peptoid bilayer nanosheets incorporating peptoid B28mCl$_7$ maintained structural fidelity throughout the sonication, contrasting with peptoid bilayer nanosheets formed by peptoid B28 and even crosslinked nanosheets of peptoid B28pCl$_7$. These results suggest that future incorporation of peptoid B28mCl$_7$ into peptoid B28 analog mixtures at different ratios could be used to tune the mechanical stability of the formed peptoid nanosheet via introduction of intralayer crosslinking at the monolayer level.

To chemically assess the origin of mechanical stability in nanosheets containing peptoid B28mCl$_7$, we used FTIR spectroscopy to determine the extent of the nanosheet crosslinking reaction when the *meta* instead of the *para* position of the phenyl ring is functionalized with chloride. Previously, the extent of the peptoid B28pCl$_7$ nanosheet crosslinking reaction was experimentally determined to be approximately 40% when using Raman spectroscopy[6]. Here, we assume Gaussian approximations for peak resolving subsequently and quantify the extent of the peptoid B28mCl$_7$ nanosheet reaction using the aromatic out-of-plane C-H vibrations at 830 cm^{-1}[9], which emerge from the three adjacent hydrogen atoms to the meta-substituted aromatic. This is further supported from the experimentally acquired spectrum for peptoid standard B28 in **Figure 8**. To increase accuracy of assessing total peptoid present, we determine total amount of peptoid present using the peak area of the Gaussian approximation describing our assigned aromatic ν(C=C) mode at 1480 cm^{-1}[9], which is conserved in the peptoid standard as well as the peptoid B28mCl$_7$ nanosheets. We avoid using the spectral region between 1700 cm^{-1} and 1500 cm^{-1} to reduce error emerging from different levels of sample hydration, which is also why

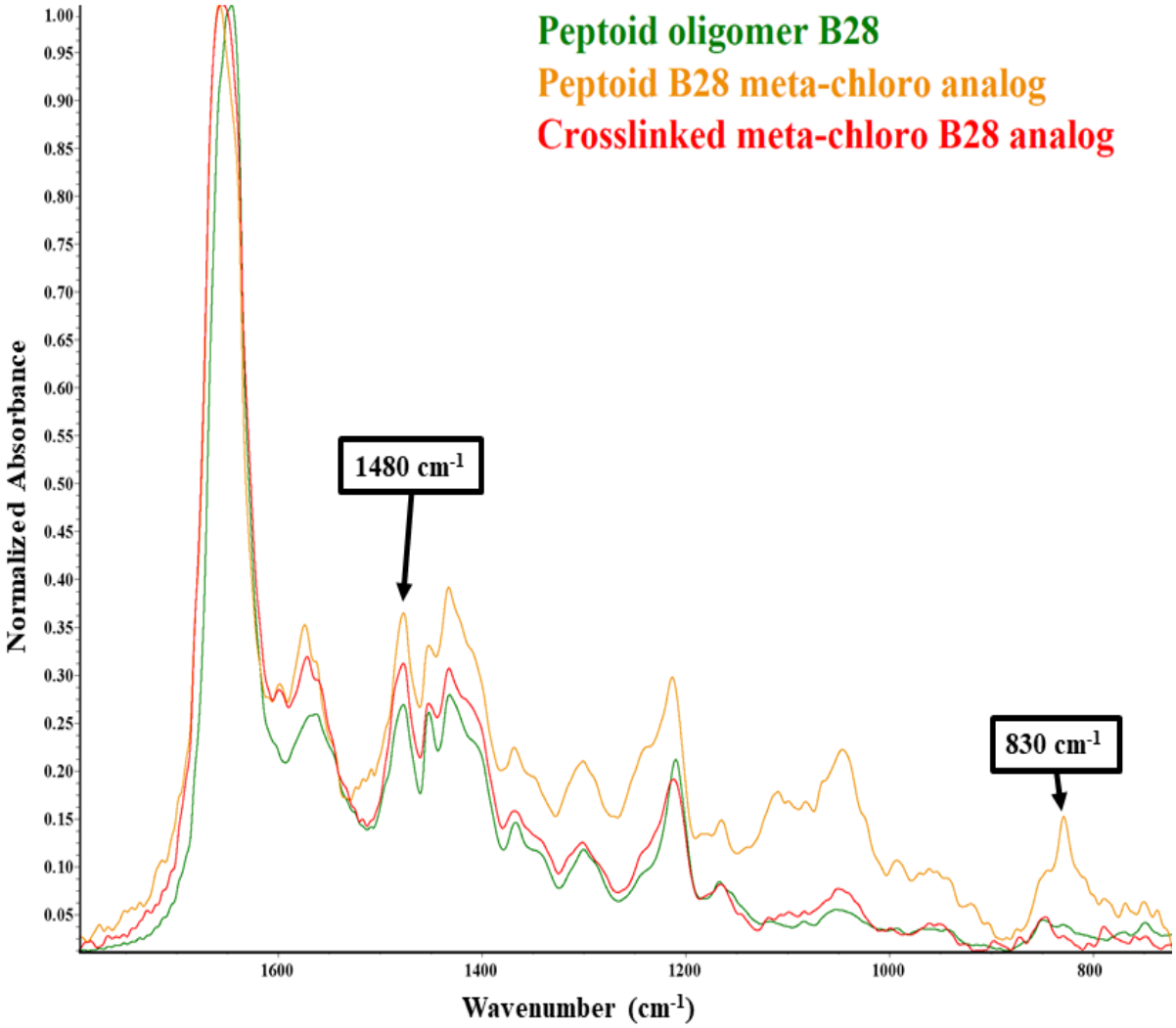


Figure 8. Normalized FTIR spectra of peptoid B28 and B28mCl$_7$ bilayer nanosheets. Spectra of crosslinked nanosheets were also obtained for extent of photolabile reaction reactivity within the peptoid B28mCl$_7$ nanosheets. Indicated peaks were used for normalization (1480 cm^{-1}) and quantification of the ν(C-Cl) disappearance (830 cm^{-1}).

normalization by the N-H and O-H stretching modes were not performed. The confounding error would take the form as broadened peaks as well as possible peak shape differences, which could drastically affect the quality of the Gaussian approximation. In this manner, we calculated the extent of the photocrosslinking reaction in the peptoid B28mCl$_7$ nanosheets to be approximately 90% (see **Methods** section).

Conclusion

We designed, synthesized, and characterized the *meta* chloro-analog of established peptoid B28. Crosslinked peptoid B28mCl$_7$ bilayer nanosheets as well as crosslinked peptoid bilayer nanosheets that incorporated B28mCl$_7$ into a peptoid block mixture showed enhanced mechanical stability. FTIR spectra support that the functionalization at the *meta* position drastically increases the extent of the nanosheet crosslinking reaction, which we suspect occurs intermolecularly between peptoid strands at the intralayer level of the peptoid bilayer nanosheets. Subsequent studies to determine the selectivity of peptoid B28mCl$_7$ covalent bond formation requires further experimentation, although the lack of maintenance of peptoid bilayer nanosheet structure when exposed to 50% acetonitrile:H_2O in the pure B28mCl$_7$ peptoid nanosheet cases (not presented due to negative correlation methodology in analysis with limited techniques for nanosheet characterization) suggest that the majority of the nanosheet crosslinking with B28mCl$_7$ may occur in an intralayer manner.

Methods

Peptoid oligomer synthesis and purification. All peptoid oligomers (B28, B28pCl$_7$, and B28mCl$_7$) were synthesized by a sub-phase monomer method previously described[6, 10], using polystyrene Rink Amide resin (100 mg per vessel) on an AAPPTec Apex 396 peptide synthesizer. Peptoids B28 and B28pCl$_7$ were purified by reverse-phase HPLC following published protocols[6, 10].

For new peptoid oligomer candidate B28mCl$_7$, the peptoid-resin was washed three times with DCM and cleaved with a 95:2.5:2.5 TFA/H_2O/TIPS (v/v/v) cocktail for 2 hours. The TFA cleavage solution was then filtered to separate the polystyrene beads from the solution prior to evaporating the filtrate to dryness on a Biotage V10 evaporator. The crude

peptoid was re-dissolved in 37% MeCN in H_2O prior to purification by reverse-phase HPLC on a semi-preparative column (C18 Vydac column). A 30-75% gradient of acetonitrile:H_2O with 0.1% TFA was used over 30 minutes with a flow rate of 17 mL/min. All purified peptoid oligomers were subsequently characterized by ultra-performance liquid chromatography mass spectrometry (UPLC-MS) via a Waters Acquity H-class Ultra High Pressure Liquid Chromatography (UPLC) system coupled to a Waters Xevo G2-XS Time-of-Flight Mass Spectrometer with an electrospray ionization (ESI) source. Mass spectra and HPLC traces of peptoid B28, B28mCl_7, and B28pCl_7 are found in **Appendix B**, where mass spectra determine the presence of our peptoid oligomers of interest while the HPLC traces are used to accurately assess purity of the stock solutions. The lyophilized peptoids were then re-dissolved in 2:1 (v:v) DMSO:H2O to create a 2mM stock solution (100X) for B28 and B28pCl_7 and a 1.33 mM stock solution for B28mCl_7.

Preparation of peptoid monolayer for compression-expansion isotherm. 15 µL of peptoid B28mCl_7 stock solution (1.33 mM in 2:1 DMSO/H_2O) was applied to the surface of a 100 mL subphase (2mM TRIS buffer, pH 8) in a Langmuir trough (KSV Nima mini trough) with surface area of 80 cm^2 in order to provide a slight excess to the amount of peptoid needed to generate the monolayer. Peptoid adsorption at the air-water interface occurred for 1800 seconds before surface area compression-expansion cycles began. The surface area was reduced at a rate of 50 cm^2/min from 80 cm (open barriers) to approximately 20 cm (closed barriers). Immediately after the barriers were closed, they were opened at a rate of 100 cm^2/min as previously described[8]. The isotherm measurement was performed in this manner to reduce the amount of utilized peptoid stock solution.

Bulk peptoid nanosheet assembly. Peptoid nanosheets were formed using the vial-rocking method as previously reported[11]. For each sample preparation, 500 µL of 20 µM peptoid solution in 2 mM TRIS buffer (pH 8) was transferred into a 4 mL dark vial and rocked with a 900 second waiting time per cycle for 350 cycles to produce nanosheets on a thermally-controlled modified rocking device at 30ºC.

Nanosheet crosslinking. 500 µL of peptoid nanosheet solution was irradiated for 9999 seconds at 254 nm in 4 mL open glass vials via a Spectrolinker XL-1500 UV Crosslinker. Samples were placed 7.5 cm below the light source.

FTIR spectroscopy instrumentation and univariate analysis. Scanning benchtop FTIR measurements were performed on a Nicolet Nic-Plan IR microscope with a ×32, 0.65 numerical aperture objective with a Thermo Scientific Nicolet iS50 FTIR spectrometer using a KBr beamsplitter and MCT (HgCdTe) detector at Beamline 1.4.3 of the Advanced Light Source at Lawrence Berkeley National Laboratory. Benchtop scanning FTIR measurements used an internal globar source and were performed in reflectance mode. IR spectra between 650 cm^{-1} and 4000 cm^{-1} at 8 cm^{-1} spectral resolution were collected with 512 co-added scans at an interferometer mirror velocity of 6.3 cm/s. Spectral acquisitions were performed by personnel of the Berkeley Synchrotron Infrared Structural Biology Imaging Program (LBNL, Berkeley, CA) due to COVID-19 restrictions governing instrument access. Gaussian approximations in OMNIC 9.8 were used for resolving peak areas for quantitative analysis. We used the following relationship to calculate extent of nanosheet photocrosslinking reaction:

$$\%\ of\ reacted\ chloride = \left(\frac{PA_{uncrosslinked} - PA_{crosslinked}}{PA_{uncrosslinked}}\right) \times 100,$$

where PA are normalized peak areas of different samples, as denoted by the subscript.

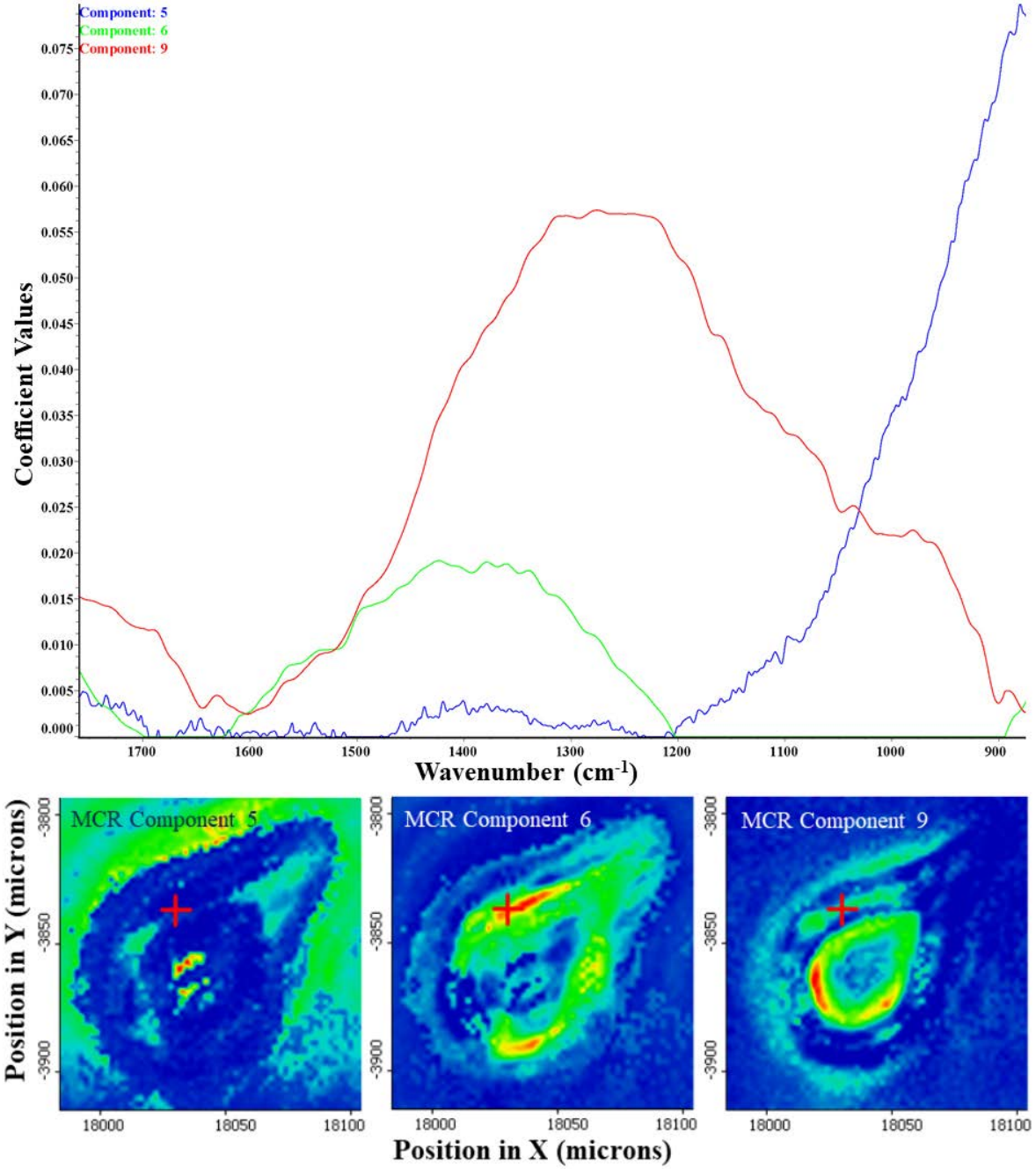
Mechanical stability analysis. Peptoid nanosheet solutions (500 µL) were sonicated for up to 60 minutes. Every 15 minutes, aliquots of each solution were removed from the sonicator and incubated with 1 µM Nile red for 15 minutes to stain the hydrophobic core of the nanosheets. From each aliquot, 1.3 µL of dye-containing nanosheets were transferred onto a 1% porous agarose pad for imaging via fluorescence microscopy. The nanosheets were imaged with epifluorescence illumination using an Olympus IX81 inverted microscope with a Hamamatsu Orca CCD camera.

Atomic Force Microscopy (AFM) analysis. All sample peptoid nanosheet solutions (500 µL) underwent buffer exchange using 100 kDa Amicon centrifugal filters to replace the 2 mM TRIS buffer (pH 8) with Milli-Q grade water. From this solution, 2 µL of buffer-exchanged nanosheet solution was transferred onto a clean mica surface and dried under N_2 flow overnight. AFM micrographs were acquired on an Asylum Cypher atomic force microscope, from which heights of individual and folded sheets were obtained. Co-author Dong Li from the Molecular Foundry (LBNL, Berkeley, CA) performed these measurements due to COVID-19 site regulations.

References

1. Howard, G.C., W.E. Brown, and M. Auer, eds. *Imaging Life: Biological Systems from Atoms to Tissues*. Illustrated edition ed. 2014, Oxford University Press.
2. Loutherback, K., et al., *Microfluidic approaches to synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectral microscopy of living biosystems.* Protein Pept Lett, 2016. **23**(3): p. 273-82.
3. Zuckermann, R.N., *Peptoid origins.* Biopolymers, 2011. **96**(5): p. 545-55.
4. Murray, D.J., et al., *Uniform, Large-Area, Highly Ordered Peptoid Monolayer and Bilayer Films for Sensing Applications.* Langmuir, 2019. **35**(42): p. 13671-13680.
5. Robertson, E.J., et al., *Design, Synthesis, Assembly, and Engineering of Peptoid Nanosheets.* Acc Chem Res, 2016. **49**(3): p. 379-89.
6. Flood, D., et al., *Improved chemical and mechanical stability of peptoid nanosheets by photo-crosslinking the hydrophobic core.* Chem Commun (Camb), 2016. **52**(26): p. 4753-6.
7. Robertson, E.J., et al., *Molecular Engineering of the Peptoid Nanosheet Hydrophobic Core.* Langmuir, 2016. **32**(45): p. 11946-11957.
8. Robertson, E.J., E.M. Nehls, and R.N. Zuckermann, *Structure-Rheology Relationship in Nanosheet-Forming Peptoid Monolayers.* Langmuir, 2016. **32**(46): p. 12146-12158.
9. Socrates, G. and G. Socrates, *Infrared and Raman characteristic group frequencies : tables and charts*. 3rd ed. 2001, Chichester ; New York: Wiley. xv, 347 p.
10. Olivier, G.K., et al., *Antibody-mimetic peptoid nanosheets for molecular recognition.* ACS Nano, 2013. **7**(10): p. 9276-86.
11. Sanii, B., et al., *Shaken, not stirred: collapsing a peptoid monolayer to produce free-floating, stable nanosheets.* J Am Chem Soc, 2011. **133**(51): p. 20808-15.

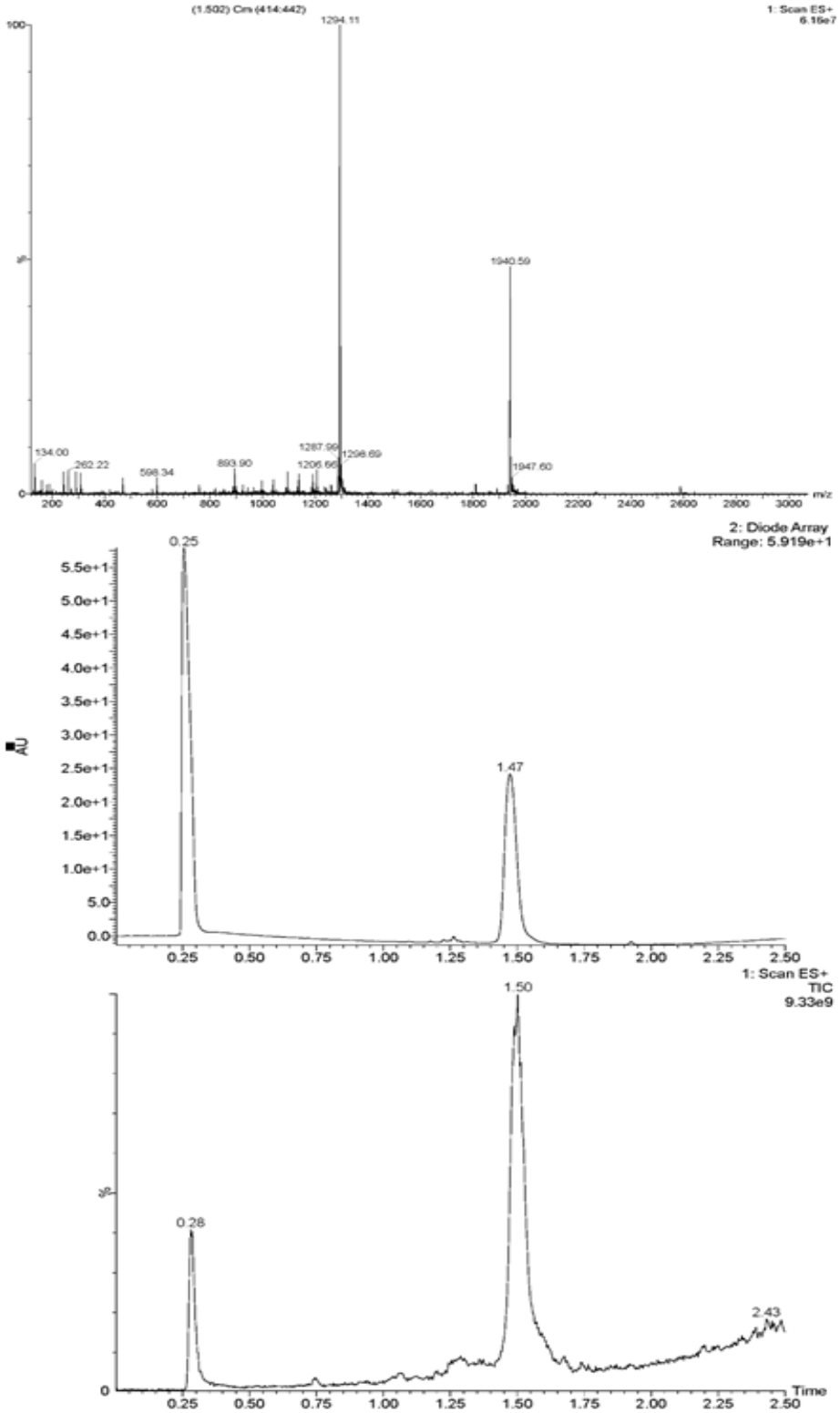
Appendix A



MCR-ALS performed on larval stage L2 *C. elegans* in Omnic 9.8. **(**Top**)** Each individual MCR component loading vector is plotted as a function of wavenumber to better exhibit the physical phenomena in spectral form incorporated into the loading vector emerging from light-matter interactions. As a result, MCR components 6 and 9 were not further mined for chemical information. **(**Bottom**)** MCR components 5, 6, and 9's coefficients displayed as individual heat maps per component, which were the components removed from spectral interpretation in **Chapter 2**.

Appendix B

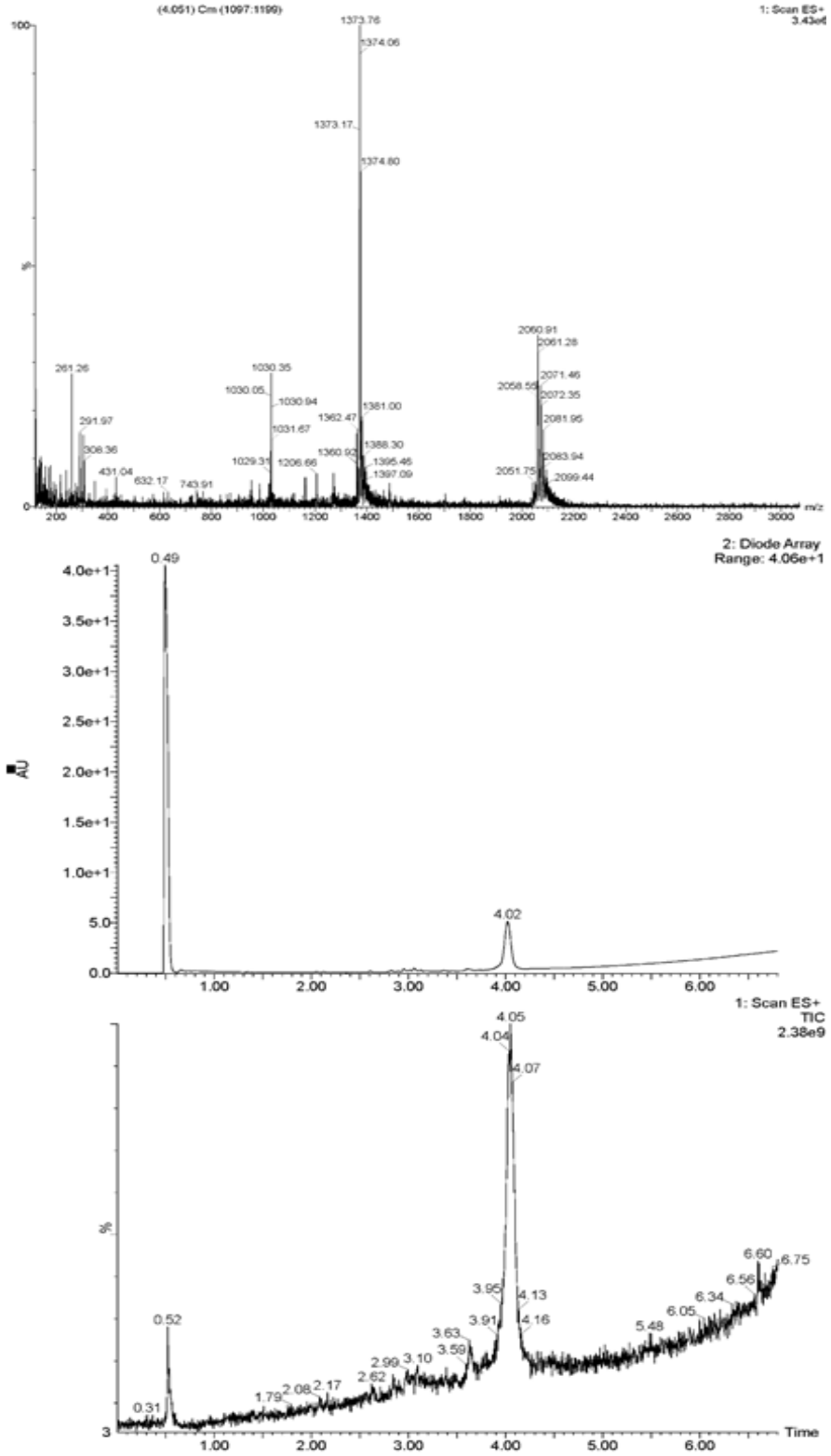
B28



UPLC-MS spectra for peptoid B28. **(**Top**)** Mass spectrum of $[M]^{2+}$ and $[M]^{3+}$ species. **(**Middle**)** HPLC trace taken at 210 nm. **(**Bottom**)** Total ion count per unit time.

Appendix B

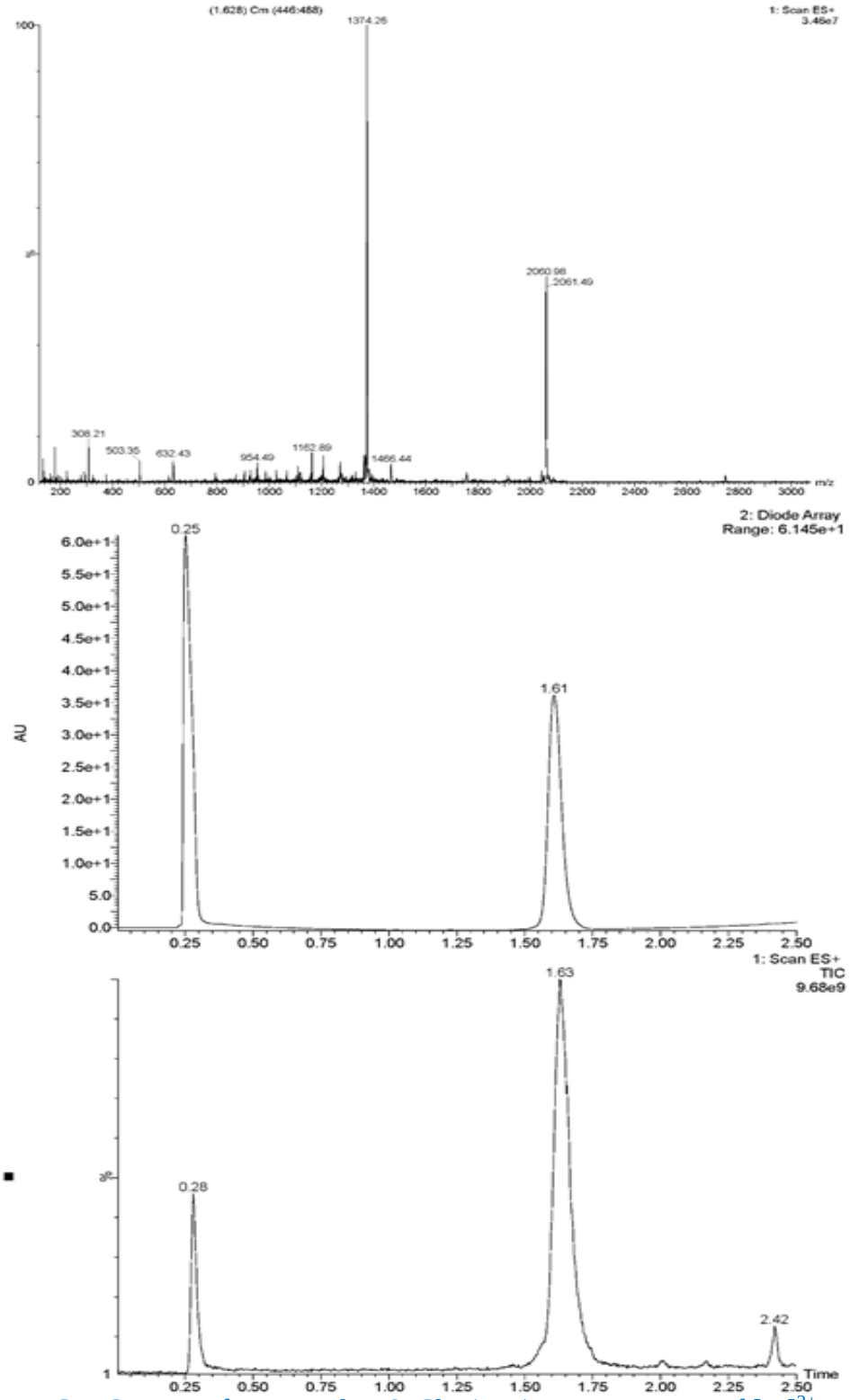
$B28mCl_7$



UPLC-MS spectra for peptoid $B28mCl_7$. **(**Top**)** Mass spectrum of $[M]^{2+}$, $[M]^{3+}$ and $[M]^{4+}$species. **(**Middle**)** HPLC trace taken at 210 nm. **(**Bottom**)** Total ion count per unit time.

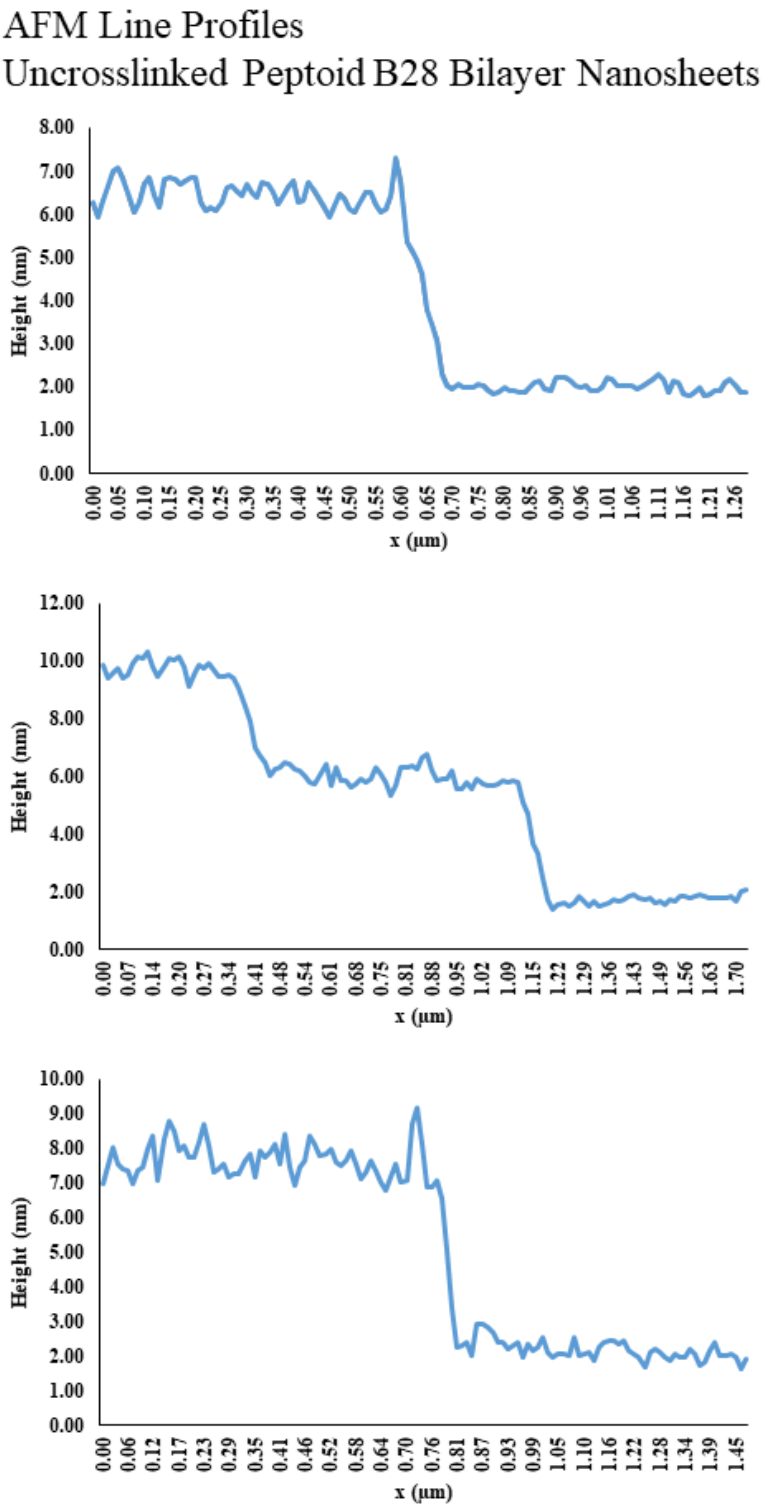
Appendix B

$B28pCl_7$



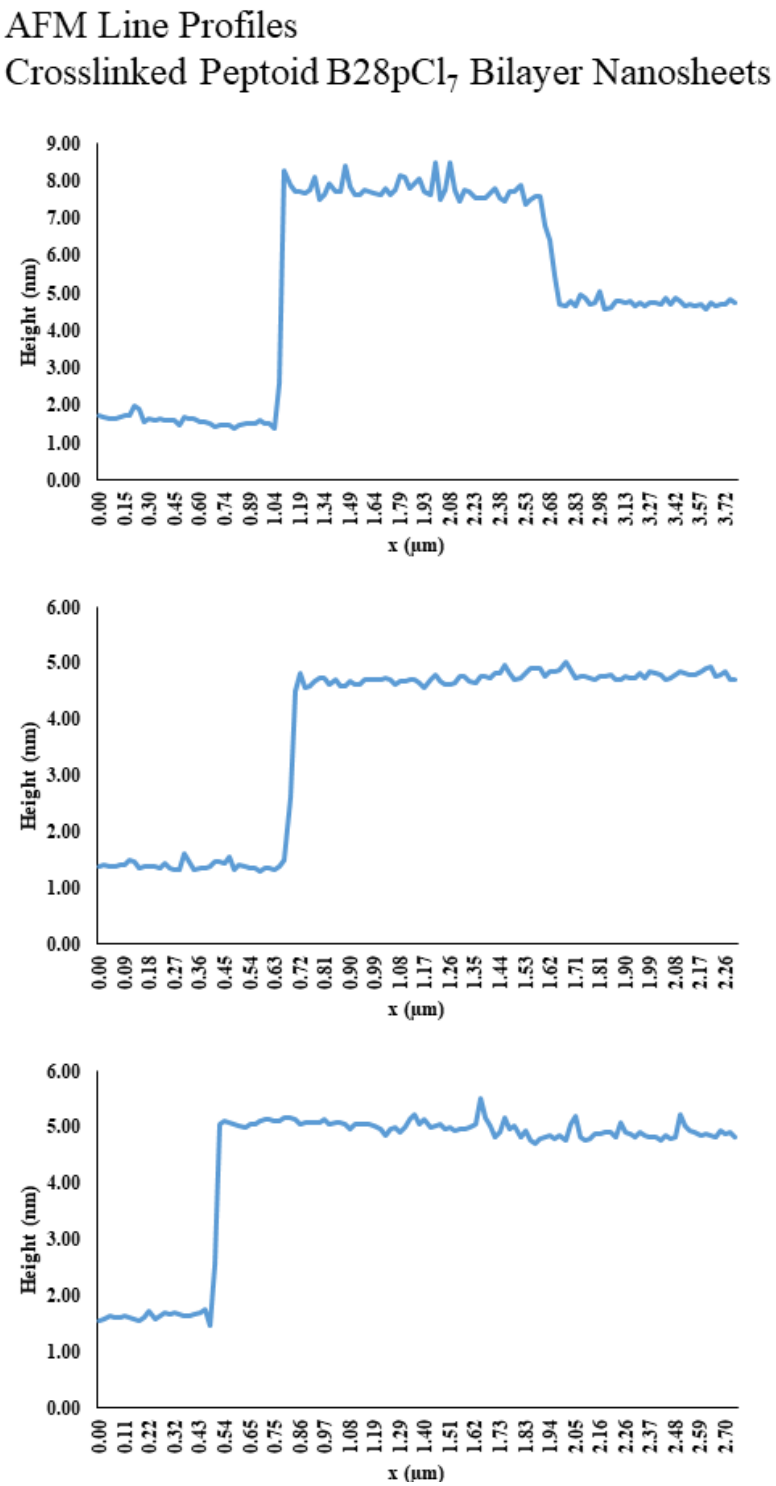
UPLC-MS spectra for peptoid $B28pCl_7$. **(**Top**)** Mass spectrum of $[M]^{2+}$ and $[M]^{3+}$ species. **(**Middle**)** HPLC trace taken at 210 nm. **(**Bottom**)** Total ion count per unit time.

Appendix C



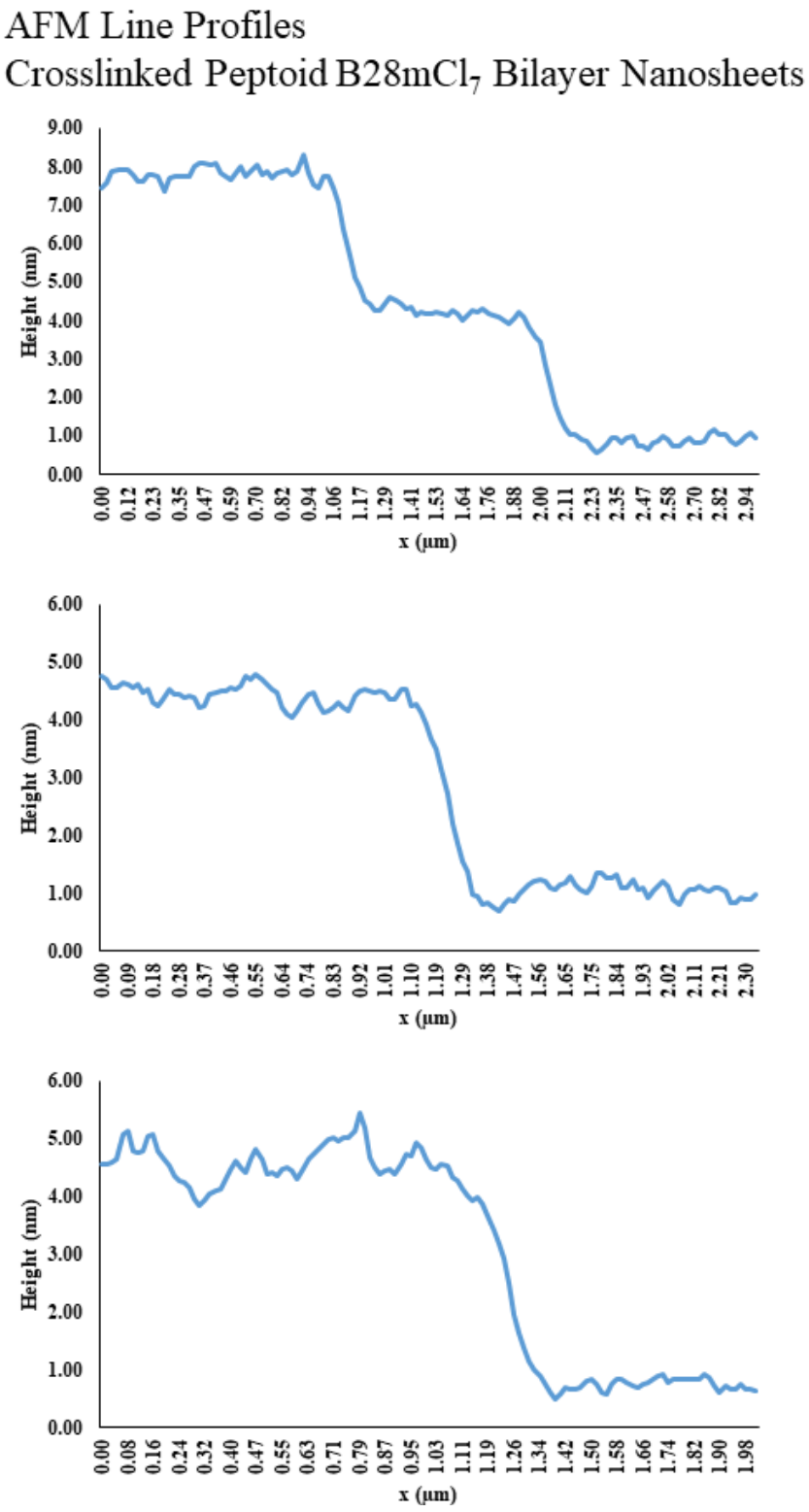
Three AFM line profiles from topographical AFM maps of uncrosslinked peptoid B28 bilayer nanosheets. Our determined nanosheet thickness reported as $mean \pm std$ is 4.779 ± 0.606 nm.

Appendix C



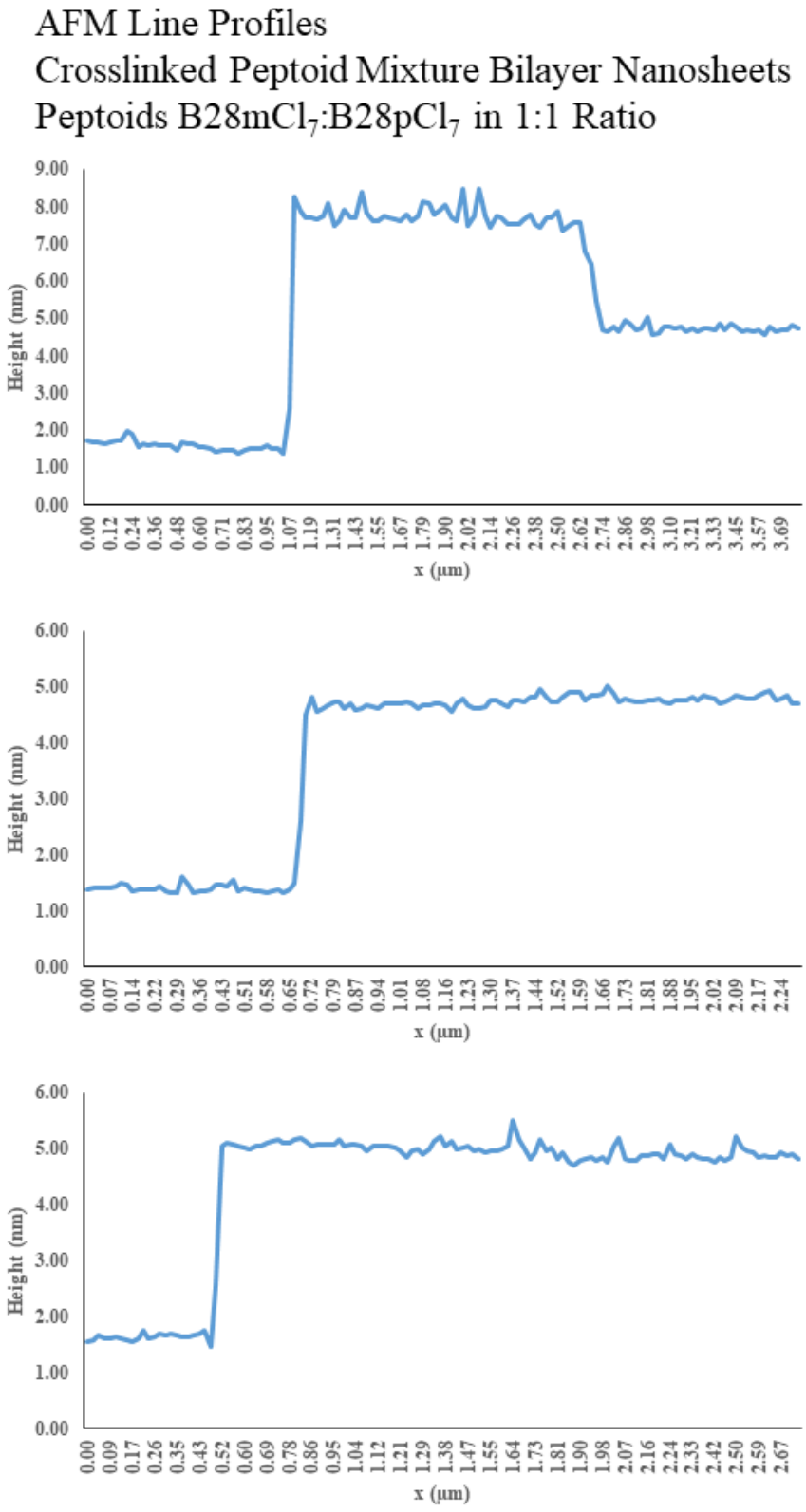
Three AFM line profiles from topographical AFM maps of crosslinked peptoid B28pCl$_7$ bilayer nanosheets. Our determined nanosheet thickness reported as $mean \pm std$ is 3.084 ± 0.082 nm.

Appendix C



Three AFM line profiles from topographical AFM maps of crosslinked peptoid B28mCl$_7$ bilayer nanosheets. Our determined nanosheet thickness reported as $mean \pm std$ is 3.763 ± 0.369 nm.

Appendix C



Three AFM line profiles from topographical AFM maps of crosslinked peptoid B28mCl$_7$:B28mCl$_7$ (1:1) mixture bilayer nanosheets. Our determined nanosheet thickness reported as $mean \pm std$ is 3.295 ± 0.075 nm.