Design and Chemical Synthesis of a Sequence-Specific DNA-Cleaving Metalloprotein . Ni (II) GGH-γδ(141-183)

Thesis by

Kenneth Scott Graham

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1993

(Submitted October 29,1992)

To Mike, Molly, Mitzi

and

Mom and Dad

Acknowledgements

I would like to thank the support staff at Caltech, particularly members of the mechanics and instrument shops without whose expertise and assistance many experiments would have gone uncompleted. I would also like to thank the members of the Dervan group and other graduate students in the Caltech community for providing a friendly and stimulating environment in which to do research. Thanks go to Dale Johnson for his aid in editing Chapter 1. I would particularly like to thank George Best for his hatred of rhetorical questions and his aid in editing Chapter 2. Elizabeth Burns, Susan Johnson, and Annie Simakauskas are to be commended for their friendship and aid in preventing my sanity from falling into question. I need to thank the members of the Caltech faculty for many useful discussions and their guidance throughout my pursuit of my doctorate. Special thanks go to Harry Gray for many useful discussions. Special thanks go to Peter Dervan for providing support and a stimulating, if not interesting, environment in which to do research. Finally, thanks go to Mo for holding the ladder. Thanks go to Martin Looser for the synthesis of the C3'-deuterated analog of thymidine and to Mark Greenberg for the synthesis of the 1, 4, and 5'-deuterated analogs of thymidine.

Abstracts

Chapter I: Affinity-Cleaving Studies Examining the Orientation and Specificity of the DNA-Binding Domain from γδ Resolvase

The DNA binding domain of gd resolvase, residues 141-183, is thought to bind DNA by the helix-turn-helix motif based on sequence similarities with other known DNA binding proteins. Incorporation of the DNA cleaving moiety, EDTA•Fe at the amino and carboxy termini of γδ (141-183), allows the positions of these residue relative to the DNA bases at three resolvase binding sites, each consisting of inverted copies of an imperfectly conserved nine base pair consensus sequence, to be mapped by high-resolution gel electrophoresis. Cleavage data for EDTA-γδ(141-183) reveal that the NH^2 terminus of the DNA binding domain of γδ resolvase is bound proximal to the minor groove of DNA near the center of the resolvase binding sites. Cleavage by EDTA•Fe attached to a lysine side chain at (Asn^{183}-Lys^{183}) at the COOH terminus of γδ (141-183) reveals that the putative recognition helix is in the adjacent major groove on the same face of the helix, oriented towards the center of the inverted repeats. Subsequent studies utilized affinity cleavage to analyze the effects of changes in amino acid composition of the recognition helix on the binding characteristics of γδ (141-183). In a systematic helix-switch experiment, the DNA contact residues of the putative recognition helix were exchanged for the DNA contact residues from another putative helix-turn-helix protein, Hin recombinase. Substitution of the amino acid in the putative recognition helix of γδ resolvase resulted in five proteins that showed no DNA binding specificity, six proteins that showed no significant alterations in DNA

binding specificity, and four proteins that showed altered DNA binding specificity.

Chapter II: Design of a Sequence-Specific DNA-Cleavage Metalloprotein Ni(II) GGH-γδ (141-183)

A forty-six residue chymeric protein that combines the cuperic binding domain of serum albumin, GGH, with the DNA binding domain of γδ resolvase (residues 141-183), was synthesized. This protein, in the presence of Ni(II) and monoperoxyphthalic acid, cleaves DNA sequence specifically at one nucleotide adjacent to each of the six γδ resolvase binding sites contained within *res*. Cleavage occurs at the center of each of the dimeric binding sites, indicating that the position of the Ni(II) GGH is proximal to the minor groove. The characteristics of the DNA products created by Ni(GGH-γδ 141-183) cleavage of DNA indicates that cleavage is produced by some time of non-diffusible oxidizing species. ESR studies indicate that cleavage is not mediated by Ni(III)•GGH. Mechanistic studies indicate that DNA cleavage likely results from abstraction of the hydrogen atom at the C4' position of the deoxyribose backbone by some high valent, nickel-oxo species. Protocols for the synthesis of deuterated thymidines are included.

Table of Contents

Acknowledgements iv

Abstracts v

Chapter I: Affinity-Cleaving Studies Examining the Orientation and Specificity of the DNA-Binding Domain from γδ Resolvase 1

Introduction 1

Principles of Sequence-Specific Recognition of Double Helical DNA 2

Role of the Major Groove of B-Form DNA in Sequence-Specific Protein Recognition 3

Role of the Minor Groove of B-Form DNA in Sequence-Specific Protein Recognition 7

The Role of B-Form DNA Conformation and Protein Binding 8

Protein Structural Motifs Used in DNA Binding 10

EcoRI-DNA Co-crystal 10

Zinc Finger DNA Binding Domains 13

The Leucine Zipper Motif 14

The β-Ribbon Motif 15

The Helix-Turn-Helix DNA Binding Motif 18

High-Resolution X-Ray Crystal Structures of Selected Helix-Turn-Helix Protein-DNA Co-crystals 22

Analysis of Protein and DNA Binding in Solution 32

Previous Studies of the Interactions Between γδ Resolvase and DNA 35

The Orientation of γδ (14-183) Determined by Affinity Cleaving 40

Synthesis 40

Footprinting 44

Position of the Amino Terminus 49

Position of the Carboxy Terminus 52
Conclusions 58
The Effects of Amino Acid Substitution on DNA-Binding Specificity 58
"Helix-Switch" Studies Using γδ (141-183) and Hin (139-190) 60
Results 68
Binding of Helix-Switch Hybrid Peptides to the Hind III/Sal I Fragment from PRW 80 Containing *Res* 70
Binding of Helix-Switch Hybrid Peptides to the Xba I/Eco RI Fragment from pMFB 36 Which Contains the Hin Recombinase Sites Hix L and Hix R 76
Conclusions 81
Experimental 84
General 84
Synthesis of tBoc-γ-aminobuteric acid: t-Boc-GABA 84
Synthesis of Tricyclohexal EDTA (TCE) 85
Peptide Synthesis 86
BEG and TCE Coupling to the Amino Terminus of Resin Bound Peptides 91
Coupling of TCE to the e-NH_2 Side Chain of a Lysine Residue at the Carboxy Terminus of a Resin-Bound Peptide 92
Peptide Deprotection and Purification 93
Labeling of the Hind III/Sal I Restriction Fragment from Plasmid PRW 80 94
Labeling of the Xba I/Eco RI Restriction Fragment from pMFB36 95
MPE Footprinting Reactions 95
Affinity-Cleavage Reactions Using EDTA Tagged γδ (141-183) Peptides and Related Mutants 95

Sequence-Specific Marker Lanes....96

References....96

Chapter II: Design of a Sequence-Specific DNA Cleaving Metalloprotein. Ni(II) GGH γδ (141-183)....103

Introduction....103

Gly-Gly-His....104

Ni(II) GGH γδ (141-183) Cleavage of DNA....106

Synthesis....106

Footprinting....108

Nickel-Mediated Cleavage Reactions....111

End Product Analysis....117

5′-End Products....119

3′-End Products....119

Summary of End Product Analysis....119

Cleavage Conditions....124

Concentration Studies....124

Time Courses....126

pH Profiles....128

ESR Studies of Ni(II)-GGH and Ni(III)-GGH....130

Kinetic Isotope Studies of the Mechanism of DNA Cleavage by Ni(II) GGH γδ (141-183)....134

Experimental Design....137

Availability of Hydrogens for Abstraction....140

Results of Kinetic Isotope Studies....142

Synthesis....142

Measurement of Kinetic Isotope Effects....151

Quantitation and Statistical Analysis of Kinetic Isotope Effects....156

Interpretation of Kinetic Isotope Effects....158

Conclusions....167

Experimental....168

General....168

Peptide Synthesis and Purification....168

Labeling of the Hind III/Sal I Restriction Fragment From Plasmid PRW 80....169

Ni(II) GGH γδ (141-183) Cleavage Reactions....169

MPE Footprinting Reactions....170

Oligonucleotide Synthesis and Purification....170

5'-End Labeling of Oligonucleotides....171

3'-End Labeling of Oligonucleotides....171

Analysis of DNA Termini by Gel Electrophoresis....172

Electron Paramagnetic-Resonance Spectroscopy....172

Sequence-Specific Marker Lanes....173

Protocols Modified for Kinetic Isotope Studies....173

Oligonucleotide Synthesis and Purification....173

5'-End Labeling of Oligonucleotides....175

T-Specific Cleavage Reactions....176

Conditions for Ni(II) GGH γδ (141-183) Cleavage....176

Gel Analysis....177

Data Collection and Work-Up....177

Statistical Treatment and Analysis of Data....178

Chemical Synthesis of Deuterated Thymidines....181

Phosphitylation of Deuterated Thymidines....202

References....202

Appendix A: Phosphorimager Data for the Ni(II) GGH γδ (141-183) Cleavage-Condition Studies....207

Appendix B: Normalized Cleavage Values Used to Calculate KIE.....214

Appendix C: Mass Spectrometry of Peptides.....214

List of Figures

Chapter I

Figure 1 The hydrogen bond donors and acceptors presented by Watson Crick base pairs to the major and minor groove.....4

Figure 2 A schematic representation of one subunit from the dimeric EcoRI DNA co-crystal.....11

Figure 3 A schematic representation of the Zif-268-DNA complex solved by x-ray crystallography.....14

Figure 4 A schematic representation of the Y-shaped model for the dimer of the DNA-binding domain of GCN4 bound to its recognition sequence of DNA.....16

Figure 5 A schematic representation of the DNA operator O_{R3} in the right-handed B-form together with representations of the dimeric crystal structures of Cro, the carboxy terminal domain of *cap*, and the amino terminal domain of λ-repressor.....17

Figure 6 A sequence alignment of selected helix-turn-helix proteins.....19

Figure 7 A schematic representation of the helix-turn-helix regions of proteins bound to their operator sequences taken from protein DNA co-crystals.....23

Figure 8 A schematic representation of the high-resolution assay used to interpret affinity-cleavage patterns.....33

Figure 9 Cleavage patterns produced by a diffusible oxidant generated by Fe(II)•EDTA located in the major and minor grooves of right-handed DNA.....34

Figure 10 The DNA sequences contained within *res* that are recognized and bound by γδ resolvase and γδ (141-183).....36

Figure 11 A consensus of inhibitory modifications displayed relative to the 9 base pair consensus sequence.....37

Figure 12 A schematic representation of the 140 residue amino terminal domain resulting from chymotrypsin cleavage of γδ resolvase.....38

Figure 13 The optimum DNA-binding sequence for γδ (141-183).....39

Figure 14 The 43 amino acid DNA-binding domain of γδ resolvase.....42
Figure 15 The synthetic scheme for the attachment of the tricyclohexal ester of EDTA to the ε-amino group of Lys^{83}.....43
Figure 16 Autoradiogram of a high-resolution denaturing polyacrylamide gel containing ^{32}P-end labeled fragments from PRW80.....45
Figure 17 Histograms of footprinting and affinity-cleavage data from the gel contained in Figure 16.....47
Figure 18 A model for γδ (141-183) binding to site I of *res*.....50
Figure 19 Autoradiogram of a high-resolution denaturing polyacrylamide gel containing ^{32}P-end labeled fragments from PRW80.....53
Figure 20 Histograms of the affinity-cleavage data from the gel in Figure 19.....55
Figure 21 A schematic representation of the two models of γδ (141-183)-EDTA•Fe bound to the right half of site I.....57
Figure 22 The recognition sequences of the half sites bound by Hin(139-190) (a) and γδ (141-183) (b).....62
Figure 23 An autoradiogram of a high-resolution gel containing ^{32}P-end labeled fragments of DNA from the Sal I/Hind III fragment from PRW80 containing *res*.....63
Figure 24 An autoradiogram of a high-resolution gel containing ^{32}P-end labeled fragments of DNA from the Xba I/EcoRi fragment of PMFb36 containing the Hin recombinase binding sites, Hix R and Hix L.....65
Figure 25 An alignment of the amino acid sequences in the helix-turn-helix domain of λ cro, Hin recombinase and γδ resolvase.....67
Figure 26 The amino acid sequences of the recognition helices from Hin, γδ, and the hybrid peptides.....69
Figure 27 An autoradiogram of a high-resolution denaturing polyacrylamide gel containing 5'-^{32}P-end labeled fragments from PRW80.....73
Figure 28 An autoradiogram of a high-resolution denaturing polyacrylamide gel containing 5'-^{32}P-end labeled fragments from PMFb36.....78

Chapter II

Figure 1 Model for the GGH ligand bound to Ni(II) 105
Figure 2 The amino acid sequence of the protein GGH γδ (141-183).....107
Figure 3 Autoradiogram of a high-resolution denaturing gel containing ^{32}P-end labeled DNA fragments from the Hind III/Sal I restriction fragment of PRW80 that contains *res*.....109
Figure 4 Autoradiogram of a high-resolution denaturing gel containing ^{32}P-end labeled fragments from the Hind III/Sal I restriction fragment of PRW80 that contains *res*.....112
Figure 5 Histograms of footprinting and cleavage data.....114
Figure 6 Schematic representation of the model for the designed metalloprotein Ni•GGH γδ (141-183) bound to *res* site I.....115
Figure 7 The sequence of the synthetic oligonucleotide duplex containing site I from *res* 118
Figure 8 Autoradiogram of high-resolution denaturing gel used for analysis of the 5'-end products produced by Ni(II) GGH γδ (141-183).....120
Figure 9 Autoradiogram of high-resolution denaturing gel used for analysis of the 3'-end products produced by Ni(II) GGH γδ (141-183).....122
Figure 10 DNA cleavage produced by Ni((I)•GGH gd (141-183) as a function of monoperoxyphthalic acid or hydrogen peroxide concentration.....126
Figure 11 Time course of DNA-cleavage by 2.0 μM Ni(II)•GGH γδ (141-183).....127
Figure 12 pH profiles of DNA-cleavage reaction mediated by Ni(II)•GGH γδ (141-183).....128
Figure 13 The ESR spectrum obtained at -150 °C from the oxidation of 5 mM Ni(II)•GGH γδ (141-183) with one equivalent of $IrCl_6^{2-}$ (IV).....132
Figure 14 The ESR spectrum obtained at -150 °C from the oxidation of 5 mM Ni(II)•GGH γδ (141-183) in the presence of 2.5 M hydrogen peroxide.....133
Figure 15 A schematic representation of a kinetic isotope experiment using Ni(II)•GGH γδ (141-183).....138
Figure 16 A stereo view of the sugar phosphate backbone of DNA from the minor (a) and major (b) groove.....141

Figure 17 The sequence of a synthetic oligonucleotide duplex used for kinetic isotope experiments.....143
Figure 18 An autoradiogram of a high-resolution denaturing polyacrylamide gel containing ^{32}P-end labeled fragments.....152
Figure 19 A graphic representation of the cleavage intensities at the reference and experimental sites obtained from phosphorimager analysis.....155

List of Tables

Chapter I

Table 1 Overall Yields for the Synthetic Peptides.....44
Table 2 Synthetic Yields for the Proteins used in theHelix-Switch Experiments.....71
Table 3 Retention Times for the Purified Helix-Switch Hybrid Proteins.....71
Table 4 Protein binding to the Hind III/ Sal I fragment from PRW80.....75
Table 5 Protein binding to the Xba I/ EcoR I fragment from pMFB36.....80

Chapter II

Table 1 Kinetic Isotope Effects (k_H/k_D) for Deuterated Thymidine.....158

Appendix A

Table I Phosphorimager Data from the Concentration Studies of DNA Cleavage in the Presence of 2.0 mM Ni(II) GGH γ δ (141-183) 208
Table II. Phosphorimager Data from the Concentration Studies of DNA Cleavage in the Presence of 1.0 mM Ni(II) GGH γ δ (141-183) 209
Table III. Phosphorimager Data from the Concentration Studies of DNA Cleavage in the Presence of 0.5 mM Ni(II) GGH γ δ (141-183).....210
Table IV Densitometer Data from the Concentration Studies of DNA Cleavage in the Presence of 2.0 mM Ni(II) GGH γ δ (141-183) 211
Table V Phosphorimager Data from the Time Course of DNA Cleavage by Ni(II) GGH γ δ (141-183) 212
Table VI Phosphorimager Data from the pH Profiles of DNA Cleavage by Ni(II) GGH γ δ (141-183) 213

Appendix B

Table I. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) 215
Table II. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) 216
Table III. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) 217
Table IV. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) 218
Table V. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) 219
Table VI. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) 220

Appendix C

Table I. Mass Spectroscopy of γδ resolvase Affinity Cleaving Peptides. Values are in atomic mass units and expected mass is calculated based on normal isotopic distributions.....**223**

List of Schemes

Chapter I

Scheme I. Synthesis of C1'Deuterated Thymidine.....144
Scheme II. Synthesis of C2'Dideuterated Thymidine.....146
Scheme III. Synthesis of C3'Deuterated Thymidine.....147
Scheme IV. Synthesis of C4'Deuterated Thymidine.....149
Scheme V. Synthesis of C5'Dideuterated Thymidine.....150

Chapter I:

Affinity Cleaving Studies Examining the Orientation and Specificity of the DNA Binding Domain from γδ Resolvase

Introduction

DNA-binding proteins interact with DNA to perform a wide range of functions necessary to living organisms. These interactions could be divided into three categories: structural, synthetic, and regulatory. Proteins and DNA interact structurally on two levels, locally and globally. Examples where DNA structure is affected on a local level, through protein interactions, include modification at specific sites, as in the case of site-specific DNA methylases, or cleavage at specific DNA sequences, as in the case of endonucleases. Examples where DNA structure is affected on a global level by protein interactions include packing by nucleosome proteins to form chromatin, or changes in the superhelical characteristics of DNA mediated by topoisomerases. Cases where proteins interact with DNA to serve synthetic functions include the actions of DNA polymerases that mediate replication, or RNA polymerases that mediate transcription. The interactions of enzymes responsible for the repair of damaged DNA, either by the excision of lesions like thymine dimers or correction of base pair mismatches, would also fall into this category. The category of regulatory interactions between proteins and DNA involves a series of interactions that can affect gene expression. In some cases, these interactions can be as simple as the binding of a protein, for example, in the case where the protein may bind adjacent to a promoter and block transcription, or a different

protein could bind adjacent to the same promoter and enhance or potentiate transcription. Other cases are more complicated in which a gene can be removed from the physical proximity of its promoter as in the case of site-specific recombinases. The one thing all these different types of interactions have in common is a requirement for an association, or binding, to take place between the protein and DNA. Within a cell, DNA contains all the information needed for the cell to grow, divide, respond to its environment, and to differentiate. The management of this huge data base, as well as the conversion of its information into useful cellular functions, is mediated by the interaction of proteins with DNA. Transcription, replication, restriction, and genomic regulation are but a few of the cellular activities that are under the control of DNA binding proteins. The studies described herein will focus on the sub problem of regulation and investigate the nature of the DNA-protein interactions responsible for the sequence-specific binding of a regulatory protein. For regulation to occur, the regulatory sequences associated with a specific gene must be identifiable in the presence of thousands or hundreds of thousands of other genes (Schleif, 1988). The question remains: How do proteins recognize the specific sequences of DNA?

Principles of Sequence-Specific Recognition of Double Helical DNA. B-DNA, which exists under physicological conditions, is probably the dominant form of DNA *in vivo* (Ollis and White, 1987). The important helical parameters that define B-DNA are the presence of approximately 10.3 to 10.6 base pairs per turn, with a rise per base pair of approximately 3.4Å and an overall diameter of 22Å for the duplex (Saenger, 1984). This form of DNA contains two distinct grooves positioned on opposite sides of the helix, a large major groove with a 11.7Å

width and 8.5Å depth and a smaller minor groove with a 5.7Å width and a 7.5Å depth.

Structural, biochemical, and molecular genetic studies have identified two sources of sequence specificity in the interactions between proteins and B-DNA (Steitz, 1990). The first of these is direct hydrogen bonding and van der Waals interaction between the protein side chains and the exposed edges of base pairs, in the major, and to a lesser extent, the minor groove of B-DNA. The second form of interaction is based on the sequence- dependent bendability, or deformability, of the DNA duplex, which provides selectivity through the ability of certain nucleic acid sequences to adopt a particular structure that is required for protein binding at a lower free energy cost relative to other DNA sequences.

Role of the Major Groove of B-Form DNA in Sequence-Specific Protein Recognition. The floors of both the major and minor grooves of B-form DNA have specific patterns of hydrogen bond donors and acceptors, in addition to sites for potential van der Waals interactions. Seeman, et al (1976) have examined these patterns and have observed that the pattern of hydrogen-bond donors and acceptors, in combination with the location of potential van der Waals contacts is unique for all four possible base pair combinations in the major groove (Figure 1). Therefore, in the major groove it is possible to distinguish between TA, AT, GC, and CG base pairs. Additionally, Seeman, et al (1976) observed that there were fewer features presented by the base pairs in the minor groove that allow for discrimination between the different base pairs (Figure 1). In the minor groove the hydrogen-bond acceptors (N3 on guanine and adenine and O2 on cytoseine and thymine) occur at almost identical positions *regardless of*

the base pairing. Only the exocyclic amino group of guanine, N2, exists to differentiate AT from GC . The end result is that direct minor groove recognition can distinguish only a binary code of GC or CG versus AT or TA.

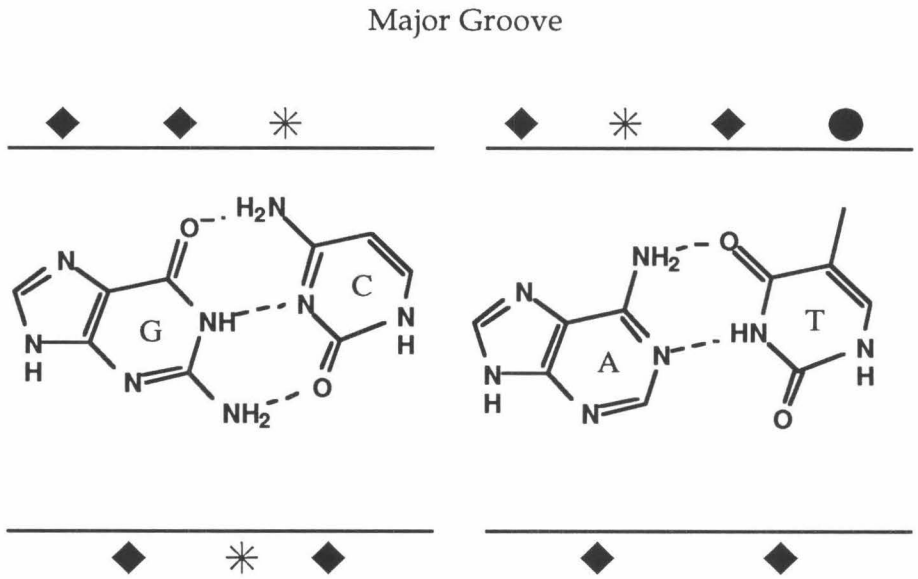


Figure 1 The hydrogen bond donors and acceptors presented by Watson Crick base pairs to the major and minor groove (adapted from Lewis, et al., 1985). ✳ represents hydrogen bond donors and ◆ represents hydrogen bond acceptors. ● represents a methyl group capable of providing van der Waals interactions. The varied pattern presented by the base pairs to the major groove demonstrates that it is possible to distinguish among AT, TA, CG, and GC base pairs. The pattern of hydrogen bond donors and acceptors in the minor groove demonstrate that it is possible to distinguish only between AT and GC containing base pairs.

The ability of the two different grooves of B-DNA to provide sites for hydrogen bonding and van der Waals interactions is not the only consideration in examining potential determinations for DNA-binding specificity. The minor groove is not only limited by potential hydrogen bonding patterns, but it is also limited sterically. The minor groove of B-form DNA is too deep and too narrow to allow direct contact by amino acid side chains alone (Steitz, 1990). The situation for the major groove is quite different, however. As pointed out by Zubay and Doty (1959), the major groove of B-form DNA can readily accommodate a protein α-helix. In their model, hydrogen bonding contacts could occur between the amino acid side chains on one face of the α-helix and the DNA base pairs in the floor of the major groove.

The predictions and observations of Seeman, et al (1976) and Zubay and Doty (1959) have been borne out by the high-resolution crystal structures of several DNA-protein complexes (Aggarwal, et al., 1988; Jordan and Pabo, 1988; Wolberger, et al., 1988; McClarin, et al., 1986; Kissinger, et al., 1990). These high resolution crystal structures have shown extensive hydrogen bonding and shape complementarity between the major groove of DNA and the surfaces of the sequence-specific DNA-binding proteins. Generally, the structural complementarity between a protein and a specific DNA sequence occurs in an idiosyncratic manner, and there does not appear to be a defined code for nucleic acid sequence recognition by a given type of protein (Pabo and Sauer, 1984; Matthews, 1988). In the initial model proposed by Seeman, et al (1976) there was a series of preferential interactions that were thought to occur between specific amino acid side chains and specific base pair combinations. It does not appear that this model is entirely correct, although there does appear to be some degree of preference for the types of interactions produced by

certain amino acids. One of the interactions proposed by Seeman, et al (1976), a bidentate interaction between the N7 and O6 of guanine and the guanidium group of arginine, has been observed in the DNA cocrystals of EcoRI (McClarin, et al., 1986) and Trp repressor (Otwinowski, et al., 1988). However, other interactions with the side chain of arginine and DNA have also been observed. In the cocrystal structure of EcoRI, the guanidium group of arginine was found to interact with the N7's of two adjacent adenines (McClarin, et al., 1986). Additionally, the cocrystal structure of the homeodomain protein, engrailed, has shown that arginines can interact with the phosphodiester backbone (Kissinger, et al., 1990). Examination of the interactions that the side chain of glutamine makes with DNA shows that it also possesses the ability to form several different types of hydrogen bonding patterns. This amino acid side chain was originally proposed to interact with the hydrogen bond donors and acceptors of a single adenine (Seeman, et al., 1976). In some cases, the glutamine side chain does interact with the hydrogen bond donors and acceptors of a single adenine (Aggarwal, et al., 1988; Jordan and Pabo, 1988). In other cases the glutamine side chain interacts with the O6 and N7 of guanine (Aggarwal, et al., 1988). This side has even been observed to form a van der Waals interaction with a thymine methyl group (Kissinger, et al., 1990). An interesting and unexpected bidentate interaction was observed for another amino acid side in the cocrystal of EcoRI. In this structure a glutamic acid side chain simultaneously interacted with the N6's of two adjacent adenines (Frederick, et al., 1985). Seeman, et al (1976) originally pointed out that the ability of amino acid side chains to form bidentate interactions with DNA greatly enhanced their ability to provide sequence-specific recognition.

The number and type of interactions that could be used to provide sequence-specific recognition of DNA were tremendously expanded when it became apparent that the amino acid side chains could mediate their interactions with the DNA through solvent molecules, and could modify their interaction by simultaneously interacting with other amino acid side chains or solvent molecules and the DNA. In l repressor, glutamine 44 interacts simultaneously with an adenosine and with glutamine 33, which in turn is also interacting with the backbone of the phosphate (Jordan and Pabo, 1988). The crystal structure of Trp repressor shows an interesting variation in which hydrogen-bond contacts between the protein and the major groove are mediated through water molecules (Otwinowski, et al., 1988). In the Trp repressor DNA complex, two water molecules appear to make hydrogen bonds between the protein and specific bases in the DNA operator. The crystal structure of 434 repressor shows that a water molecule is used to orient the carboxamide side chain of glutamine 33 so that it is directed towards the O4 of a thymine (Aggarwal, et al., 1988). In addition to the hydrogen-bonding contacts involving more than a single amino acid side chain, van der Waals contacts of this type have been observed. The cocrystal of the homeodomain protein, engrailed, contains a van der Waals interaction that involves the simultaneous interaction between the side chain of isolusine 47, a portion of the α-carbon backbone, and the methyl group of a thymine in the operator sequence.

Role of the Minor Groove of B-Form DNA in Sequence-Specific Protein Recognition. Just as x-ray crystallographic studies have demonstrated some unexpected findings in major groove recognition by proteins, they have also demonstrated some unexpected findings in terms of minor groove recognition. The initial model predicted that there would be few sequence-specific interactions

between the amino acid side chains of proteins and the base pairs in the floor of the minor groove (Seeman, et al., 1976). It appears, however, that sequence-specific interactions can also occur in the minor groove. The crystal structure of the homeodomain protein engrailed identified two arginines that contact DNA bases in the floor minor groove (Kissinger, et al., 1990). The side chain of arginine 5 in this protein reaches directly into the minor groove and hydrogen bonds with the O2 of thymine. The side chain of arginine 3 is also positioned in the major groove; however, its exact hydrogen-bonding pattern is uncertain. It could potentially hydrogen-bond with the O2 of thymine and/or hydrogen bond with a sugar oxygen from an adjacent adenosine base pair. Water-mediated interactions between arginine guanidinium groups and AT sequences in the minor groove have been observed for the 434 repressor (Aggarwal, et al., 1988). It is interesting to note that the enzyme DNAse I, a protein that is considered to bind DNA "nonspecifically," exhibits preferences for cleaving DNA that are thought to arise from interactions in the minor groove (Suck, et al., 1988).

The Role of B-Form DNA Conformation and Protein Binding. While the role of hydrogen bonding and van der Waals interactions in sequence-specific recognition of DNA was predicted, the extent to which the sequence-dependent conformational variability of B-form DNA would be involved in nucleic acid sequence recognition was not anticipated. This conformational variablity appears to play an extremely important role in most, though perhaps not all, protein DNA complexes (Steitz, et al., 1990). Significant distortion of DNA structure has been observed in the crystal structures of DNA complexes with EcoR I (Frederick, et al., 1985; McLaren, et al 1986), 434 repressor (Aggarwal, et al., 1988), Trp repressor (Otwinowksi, et al., 1988), and DNAse I (Suck, et al.,

1988). The changes in duplex DNA structure that havebeen observed in these complexes include changes in twist (Aggarwal, et al., 1988), alterations of groove width (Suck, et al., 1988), and kinking of the DNA (McClarin, et al., 1986; Otwinowski, et al., 1988; Aggarwal, et al., 1988). These crystal structures suggest that when a protein binds DNA in solution, the DNA is not always a passive partner. The conformational flexibility that has been observed for B-form DNA is dependent on the nucleic acid sequence in the region. Generally, AT sequences have a narrower minor groove than GC sequences (Saenger, 1984), and AT sequences typically favor bending into the minor groove (Drew and Travers, 1984). GC rich sequences, however, tend to favor the formation of kinks that tend to narrow the major groove (Drew and Travers, 1984).

Examination of the types of contacts that form between proteins and DNA and the types of conformational changes that DNA can undergo has demonstrated that the process by which proteins recognize specific DNA sequences is not a simple one. It involves more than the recognition of a simple code of hydrogen-bond donors and acceptors found in the major groove of B-form DNA. The amino acid side chains of DNA binding proteins can interact with the base pairs at the floor of the minor and major grooves in a wide variety of ways. The interactions can be mediated through water molecules, or even by other amino acid side chains found on the protein. The result of this is that there is no simple code for sequence-specific recognition, in which a given amino acid will always recognize a specific DNA base pair. To complicate the situation, it has become apparent that the conformational flexibility of DNA can also play a role in the sequence-specific recognition of DNA by proteins. The result of this is that the selective binding of a protein to a particular DNA

sequence requires not only recognition by the protein of an ensemble of steric and chemical features, but also the accommodation of the DNA to the protein.

Protein Structural Motifs Used in DNA Binding. Many DNA-binding proteins contain small discrete domains, which they use to recognize DNA in a sequence specific manner. In several cases, a specific type of domain, or motif, is used by several different proteins to bind to DNA. Examples of some of the structural motifs used by proteins to bind DNA in a sequence-specific manner are the α-helical structure of EcoRI, the zinc finger, the leucine zipper, the β-ribbon, and the helix-turn-helix. Each of these motifs possesses its own unique mode of DNA recognition. The characteristics of these different motifs will be discussed below.

EcoRI-DNA Co-crystal. The DNA sequence recognized by this restriction endonuclease is centered on a diad axis of symmetry, which contains the palindromic sequence 5'-GAATTC-3'. EcoRI binds to the palindromic sequence as a dimer of two identical sub-units, and in the presence of magnesium, catalyzes the cleavage of the palindromic sequence between the GA resulting in a 5'-phosphate. In the absence of magnesium, the EcoRI endonuclease apoprotein was co-crystallized with a DNA sequence containing its palindromic recognition site (McClarin, et al., 1986; Frederick, et al., 1984). It is important to note that the DNA sequence used to form the cocrystal was identical to the deodecamer originally crystallized by Dickerson and Drew except for the addition of a thymidine overhang at the 5'-end (Dickerson and Drew, 1981; Dickerson, 1983). Figure 2 contains a schematic representation of the cocrystal structure of EcoRI bound to its palindromic binding site. Comparison of the DNA structure in the EcoRI co-crystal with that of the

Dickerson deodecamer (Dickerson and Drew, 1981) shows that there are two distinct kinks, or distortions, in the DNA (Frederick, et al., 1984; McClarin, et al 1986). The first kink occurs at the center of the dyad axis of symmetry. The

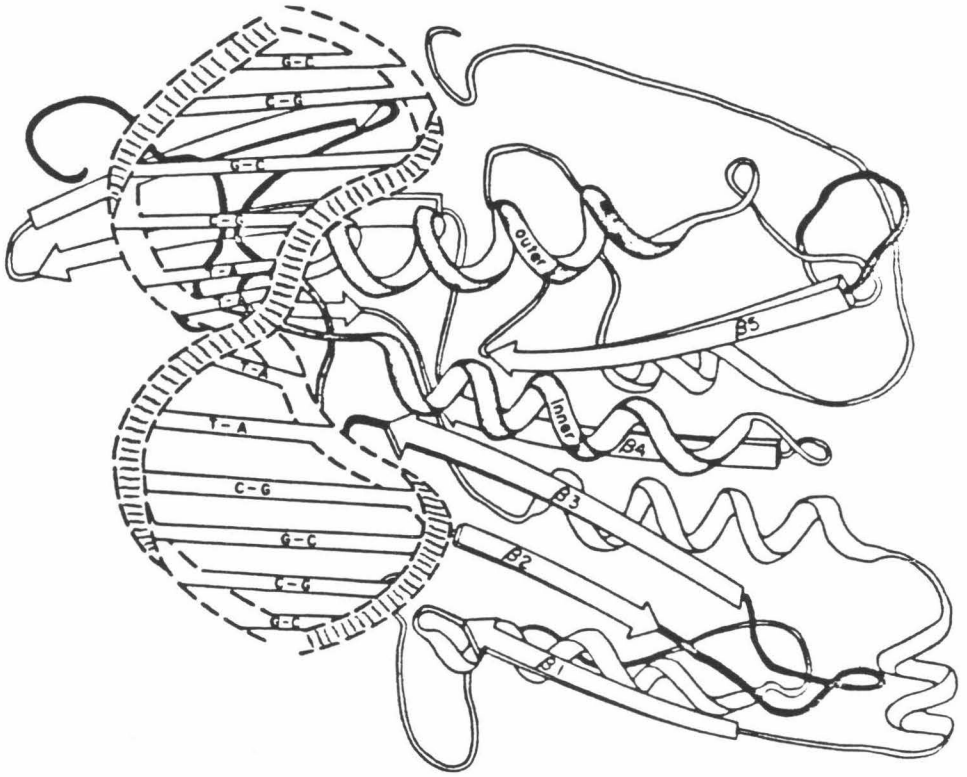


Figure 2 A schematic representation of one subunit from the dimeric EcoRI DNA co-crystal (McClarin, et al 1986). The sequence of the DNA oligomer is 5'-TCTCGAATTCGCG-3'. The other subunit of the complex would be rotated 180° about the horizontal line through the center of the first subunit in the plane of the page.

second kink is observed three base pairs from the dyad axis on both sides of the diad. When the structure of the EcoRI protein itself is examined, it is found to

contain a five-stranded β sheet surrounded on either side by α helices (McClarin, et al., 1986). In the co-crystal, the amino ends of the α helices bind to the center of the major groove of DNA. This positions the positive end of the helix dipole in contact with the DNA. Three amino acids, two arginines, and one glutamic acid emanate from these α-helices to form hydrogen bonds with DNA bases on the floor of the major groove (McClarin, et al., 1986). Arginine 200 forms a hydrogen bond with the N7 and O6 of guanine. This pattern of hydrogen bonding is consistent with the model originally suggested by Seeman, et al (1976). Arginine 145 forms hydrogen bonds with the N7's of two adjacent adenines, while glutamic acid 144 from the other subunit accepts hydrogen bonds from the N6 amino groups located on the same two adenines. It has been suggested that while this hydrogen-bonding pattern may account for much of the sequence-selectivity of the enzyme, it may not be sufficient to account for its low rate of cleavage at incorrect sequences (Steitz, 1990). It is possible that the low rate of cleavage at incorrect sequences is not a result of the binding specificity of this enzyme. This notion is supported by the binding characteristics observed by another restriction endonuclease, EcoRV (Vermote, et al., 1992; Vermote and Halford, 1992). EcoRV was found to bind many different DNA sequences with equal affinity. However the enzyme was able to cleave DNA at only one particular sequence, 5'-GATATC-3'. This indicates that it is possible that the sequence-selectivity, in DNA cleavage, exhibited by restriction endonucleases may not result soley from the protein's ability to recognize and bind one specific sequence of DNA. The sequence-selectivity in DNA cleavage may result from a combination of the protein's ability to recognize a DNA sequence and the ability of that sequence to adopt a conformation suitable for strand scission to occur.

Zinc Finger DNA Binding Domains. The zinc finger motif was originally identified as a 30-residue repeating sequence in the transcription factor IIIA (TF IIIA) (Miller, et al., 1985). The motif was christened the "zinc finger" because it was found that the coordination of a zinc ion to a set of invariant cysteines and histidines, within each repeating unit, was vital to determining the protein's tertiary structure (Diakun, et al., 1986; Evans and Hollenberg, 1988). Subsequently, a second type, or subclass, of zinc finger proteins was identified (Evans and Hollenberg, 1988; Beato, 1989). This class is exemplified by the yeast GAL 4 transcriptional activator and mammalian steroid receptors. In this class of zinc fingers, the zinc ion is coordinated with four invariant cysteine residues. Mutagenesis experiments have shown that replacement of one pair of cysteins, with a pair of histidines, completely abolishes the DNA-binding activity of the GAL 4-like proteins (Evans and Hollenberg, 1988).

Recently, the DNA co-crystal of a TF IIIA-type protein, Zif268, has been solved to 2.1 Å resolution (Pavletich and Pabo, 1991). A schematic representation of the structure is shown in Figure 3A. Crystallography confirmed the structure originally predicted by 2D NMR measurements (Parraga, et al., 1988; Lee, et al., 1989) in which each zinc finger consisted of two antiparellel β-sheets and an α helix surrounding centrally bound zinc ion. In the co-crystal, the zinc finger wraps around the DNA double helix, making contacts in the major groove through the use of the amino terminal portions of the α helices. It appears that each of the three zinc fingers in this protein act as independent units, each recognizing a three base pair subset of the consensus recognition sequence. Figure 3B shows a summary of the base pair contacts made by Zif286 with its DNA-binding sequence. The five arganines that contact DNA in the major groove form hydrogen bonds with the N7 and O6 of guanines in a manner that

was consistent with the original predictions of Seeman, et al (1976). An additional interaction occurs between histidine 49 and a guanine in the major groove. Currently, no detailed, three-dimensional structures are available for the GAL 4-type zinc fingers.

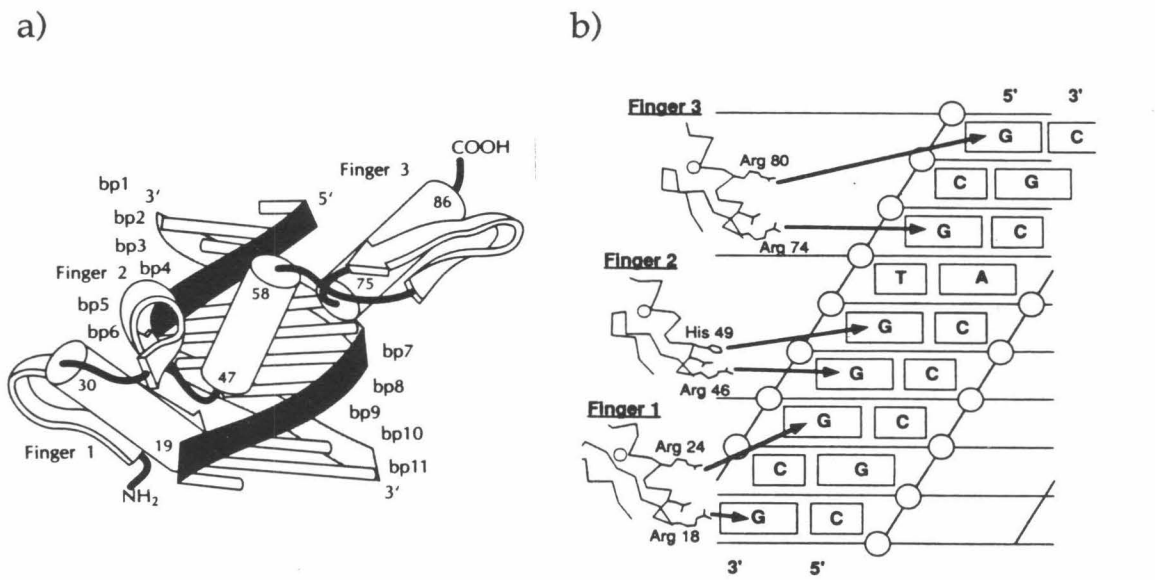


Figure 3 a) A schematic representation of the Zif 268-DNA complex solved by x-ray crystallography. The terminal residues of each α helix and β sheet and the DNA base pair numbers are indicated, adapted from Pavletich and Pabo (1991). b) A schematic summary of the base pair contact observed in the co-crystal of the Zif 268 zinc finger domain, adapted from Kaptein (1991).

The Leucine Zipper Motif. This type of DNA-binding element was originally identified by its dimerization motif, an α-helical segment referred to as the leucine zipper (Landshultz, et al., 1988). The dimerization domain of these peptides is a 30 amino acid segment comprising the carboxal terminal half of

the protein, which forms an ampiphilic α helix containing a leucine at every seventh residue. The amino terminal portion of these proteins contains a basic region that is thought to be responsible for DNA binding. Members of this class of proteins include the GCN4 transcriptional activator, and the *jun* and *fos* ocoproteins (Landshultz, et al., 1988). These proteins are proposed to bind to DNA, using a "scissor-grip" model in which the ampiphilic helices of two monomers join to form the base of a Y-shaped dimeric molecule. The basic amino terminal regions of the monomers form the bifurcating arms of the "Y," which are thought to bind to DNA by tracking in opposite directions along the major groove (Vinson, et al., 1989). Currently, this structural motif has not been characterized by crystallographic techniques. However, solution studies employing footprinting (Vinson, et al., 1989) and affinity-cleaving (Oakley and Dervan, 1990) have found the protein to bind to DNA in a manner consistent with the scissor-grip model shown schematically in Figure 4.

The β-Ribbon Motif. Two types of procaryotic proteins that contact DNA through a β-ribbon motif have been identifed (Phillips, 1991). The first type of β-ribbon proteins includes the protein HU, the *E. coli* integration host factor, IHF, and the B.*sbutilis* phage-transcription factor, TF I. HU forms the condensed nucleo-protein structures observed in many procaryots. It binds to DNA in a non-specific manner through interactions with the minor groove (White, et al., 1989; Yang and Nash, 1989). Crystallography has shown the core structure of a dimer of the protein HU to contain two α-helices, and two projecting ribbons that originate from each monomer (Tanaka, et al., 1984). The β-ribbons are formed by the carboxal terminal third of the polypeptide chain, and are thought to be responsible for forming the protein-DNA contacts. Structural information for the other two members of this group of β ribbon

proteins, IHF and TF I, is limited. However, it is known that these two proteins, unlike HU, bind to DNA in sequence-specific manners (Phillips, 1991).

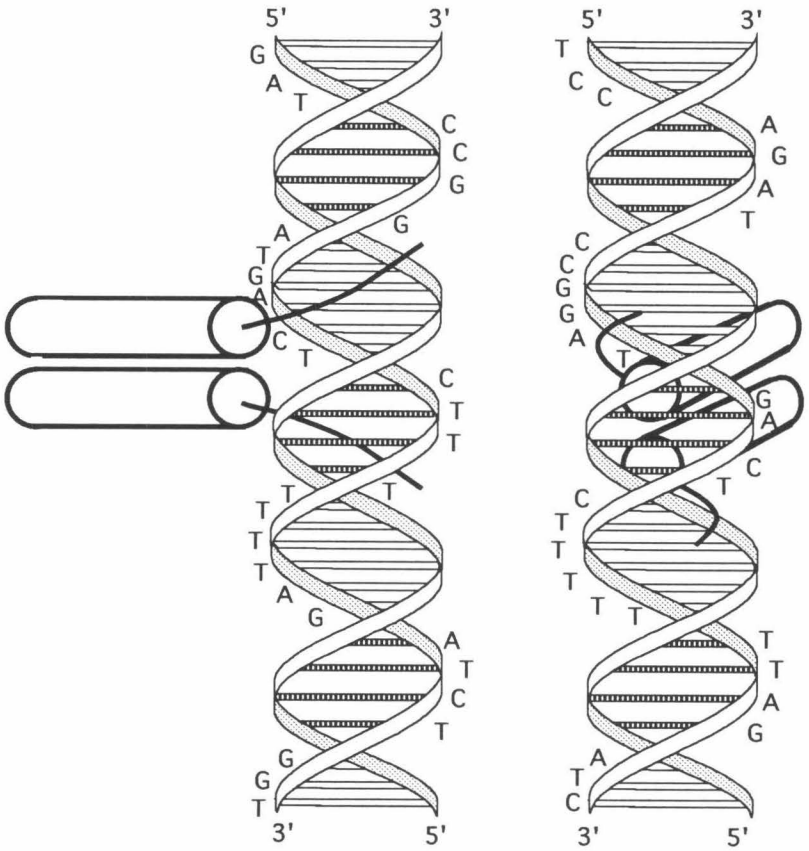


Figure 4 A schematic representation of the Y-shaped model for the dimer of the DNA-binding domain of GCN4 bound to its recognition sequence of DNA. The views represent side and front views of the complex. The leucine zipper dimerization domain is represented by a pair of cylinders projecting out of the major groove of DNA. Adapted from Oakley and Dervan (1990).

The second type of β-ribbon protein includes the met j repressor from E. coli and the arc and MMT repressors from salmonella phage P22. The structure of met j has been determined by crystallography (Rafferty, et al., 1989). The

protein is a dimer with a core comprised of four α-helices, two from each subunit. In met j an adjacent carboxy terminal helix protrudes from each monomer. An anti-parallel β-ribbon is formed from the end terminal segment of each monomer and protrudes from the core. In a DNA complex of met j that has been determined crystallographically, the β-ribbon lies in the major groove of the DNA (Phillips, 1991). The structure of the Arc repressor has been examined by

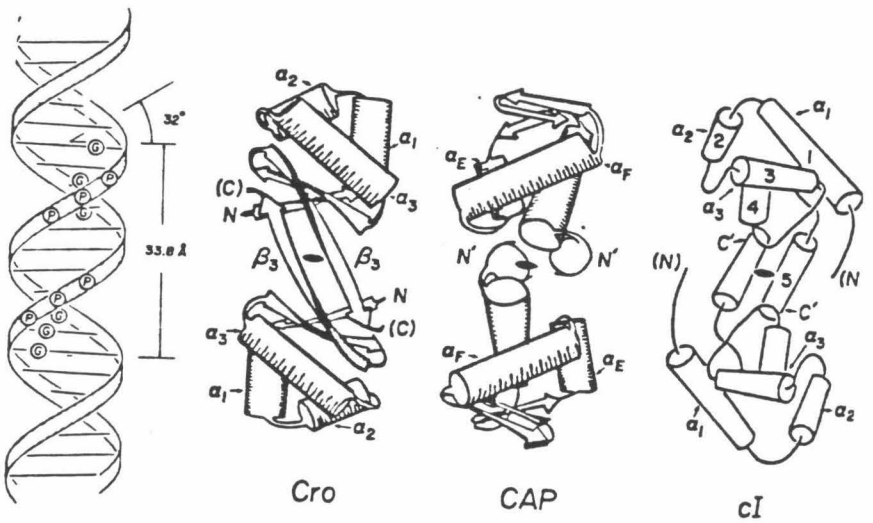


Figure 5 A schematic representation of the DNA operator O_{R3} in the right-handed B-form together with representations of the dimeric crystal structures of Cro, the carboxy terminal domain of *cap*, and the amino terminal domain of λ–repressor (cI). The location of the carboxyl terminus of the proteins is designated by the letter C; the location of the amino terminus of the proteins is designated by the letter N. Letters in parentheses indicate that the termini are disordered. The helix-turn-helix units of the proteins are α_2-α_3 or α_e-α_f. The circled phosphates and guanines on the O_{R3} operator are sites where ethylation or methylation inhibit cro binding. Adapted from Ohlendorf and Matthews (1983).

NMR (Breg, et al., 1990) and found to be consistent with the structure determined from met j, except that there is an absence of the C-terminal α-helices.

The Helix-Turn-Helix DNA-Binding Motif. The helix-turn-helix DNA-binding motif is perhaps the best characterized of all the protein motifs known to bind DNA. Originally, this structure was identified by comparisons of the crystal structures of CAP and λ-Cro (Steitz, et al., 1982). The existence of this motif was subsequently confirmed by comparison with the crystal structure of another DNA-binding protein, λ-repressor (Pabo and Sauer, 1984; Ohlendorf and Matthews, 1983). Figure 5 shows a schematic representation of these proteins based on their x-ray crystal structures. Currently, the DNA co-crystals of at least seven proteins that use this motif to bind to DNA have been solved to high resolution (Harrison, 1991). Comparison of protein-sequence homology has identified at least thirty more DNA-binding proteins that could potentially recognize DNA, using the helix-turn-helix motif (Sauer, et al., 1990). A sequence alignment of several of these proteins is shown in Figure 6. Initially, this DNA-binding motif was thought to be used only by procaryots; however, structural studies on the eukaryotic homeodomain proteins, antenapedia (Antp) and engrailed, have demonstrated the existence of this motif in eukaryots (Qian, et al., 1989; Kissinger, et al., 1990).

The helix-turn-helix motif is defined by several common structural features. The core of this domain generally consists of twenty amino acids (Harrison, 1991). This region contains two α-helices linked by a three amino acid turn. The α -helices lie across one another at approximately a 120° angle (Harrison, 1991). The relative position of these two helices is maintained by a series of

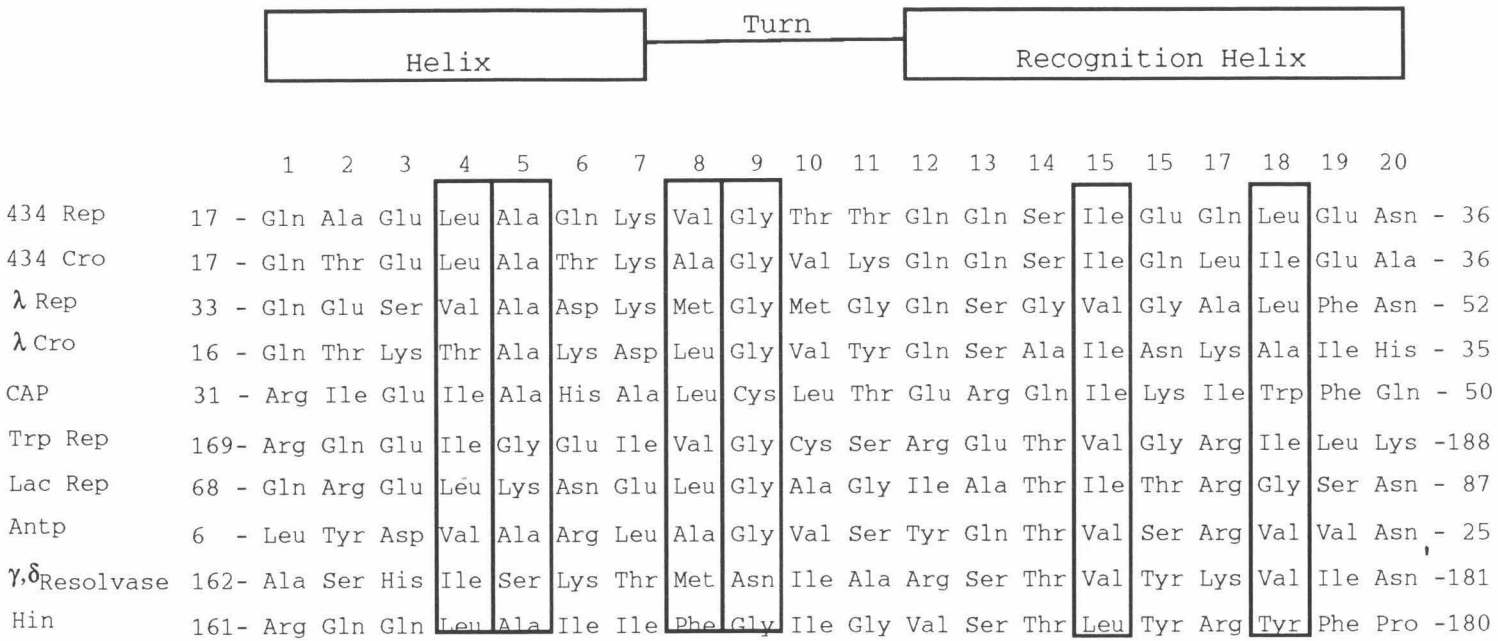


Figure 6 A sequence alignment of selected helix-turn-helix proteins. Amino acids in boxes are invariant hydrophobic residues used to determined the sequence alignment. Adapted from Sauer, et al., 1990.

hydrophobic residues that occur at invariant positions along the α-helices (Figure 6) (Sauer et al., 1990). A comparison of the α-carbon backbones for a series of these proteins shows that the backbones are superimposable on each other with a root mean-square difference in positions of the atoms ranging from 0.4 to 1.0Å (Zhang, et al., 1987; Brennan and Matthews, 1989; Qian, et al., 1989).

Proteins using the helix-turn-helix motif to bind DNA should share common features in the sequence-specific recognition of DNA. Initially, sequence-specific recognition of DNA by these proteins was thought to be achieved solely by the second helix of the helix-turn-helix domain, which was termed the recognition helix (Wharton, et al., 1984). In the initial model, this helix was postulated to lie in the major groove of DNA and to form sequence-specific hydrogen-bond contacts between the amino acid side chains, on one face of the helix, and the bases in the floor of the major groove. The high-resolution crystal structures of several protein-DNA complexes have shown that protein-DNA contacts of this type do occur (Kissinger, et al., 1990; Aggarwal, et al., 1988; Jordan and Pabo, 1988; Wolberger, et al., 1988). In the original model the first helix of the helix-turn-helix domain was to anchor the recognition helix in place by making a series of non-specific contacts with the phosphate backbone of DNA. These types of interactions have also been observed in the co-crystals of helix-turn-helix proteins.

As the structures of more co-crystals of helix-turn-helix proteins became available, it became evident that the original model was not entirely correct and that the amino acid composition of the recognition helix might not be the sole determining factor in DNA-binding specificity of the protein (Pabo, et al., 1990).

While the positions proposed for the two helices in the model were correct, it underestimated the importance of the contacts with the DNA made by the first helix of the motif. It considered these contacts to be non-specific and placed little importance on their ability to influence sequence specificity of the protein. Subsequently, the phosphate contacts, which are formed by the first helix and turn of the helix-turn-helix domain, were found to determine the alignment of the recognition helix with respect to the DNA major groove (Pabo, et al., 1990). The exact nature of this alignment is thought to be a determining factor in the type of hydrogen bonds that can occur between the recognition helix and the floor of the major groove, and as a result of this influence, the specificity of the DNA-binding protein (Pabo, et al., 1990; Harrison and Aggarwal, 1990). This has led to the idea that a series of subfamilies of helix-turn-helix proteins may exist. Members of the same subfamily would share common orientation of the recognition helix relative to the DNA. An example of helix-turn-helix proteins that share a common recognition pattern, and as a result would belong to the same family, are 434 and λ-repressors (Pabo, et al., 1990). In these proteins, their relative positions with respect to the DNA are very similar, and so are the patterns of hydrogen bond contacts that form between the recognition helices and the DNA. This is exemplified by the DNA contacts formed by the glutamine residue that is present at the amino terminus of the recognition helix in both of these proteins. In the operator sequences for both of these proteins, there is an adenine located towards the 5'-end. A virtually identical hydrogen-bonding pattern was observed in the co-crystals of these proteins between this similarly positioned adenine and the conserved glutamine located in the recognition helix. An example of a different and unrelated family of helix-turn-helix proteins would be the homeodomains. The positioning of the homeodomain helix-turn-helix on DNA as observed in crystal structures and

by NMR in solution (Otting, et al., 1990; Kissinger, et al., 1990) is quite different from that found in the complexes of prokaryotic helix-turn-helix proteins such as 434 and λ-repressors (Brennan, 1991; Pabo, et al., 1990). In the homeodomain proteins the entire motif is shifted and reoriented with respect to the DNA.

Another factor not initially considered in the binding model for helix-turn-helix proteins, was the DNA conformation itself. In many of the protein-DNA co-crystals of helix-turn-helix proteins, the DNA is highly distorted (Harrison and Aggarwal, 1990; Brennan, 1991). This bending, or distortion, of the DNA at and adjacent to the binding site has been suggested to play a role in determining the binding specificity of these proteins (Harrison and Aggarwal, 1990; Brennan, 1991).

High-Resolution X-Ray Crystal Structures of Selected Helix-Turn-Helix Protein-DNA Co-Crystals. A schematic representation of λ-repressor bound to its DNA operator site is shown in Figure 7a (Jordan and Pabo, 1988). An examination of the general structural features of the protein-DNA complex shows that the protein binds as a dimer with one helix from each subunit positioned in the major groove. The overall complex is approximately symmetrical. The position of the twofold axis of symmetry from the protein dimer corresponds with the position of the twofold axis of symmetry from the DNA-binding site, or operator. The operator was found to be a relatively straight segment of B-form DNA with an average helical twist corresponding to 10.5 base pairs per turn. This is extremely close to the average parameters predicted for B-form DNA (Saenger, 1984). As observed in the crystal structure of the λ-repressor alone, (Pabo and Lewis, 1982) the protein in the co-crystal contained five α-helices. Helices 2 and 3 form the helix-turn-helix unit. Helix 3, the recognition

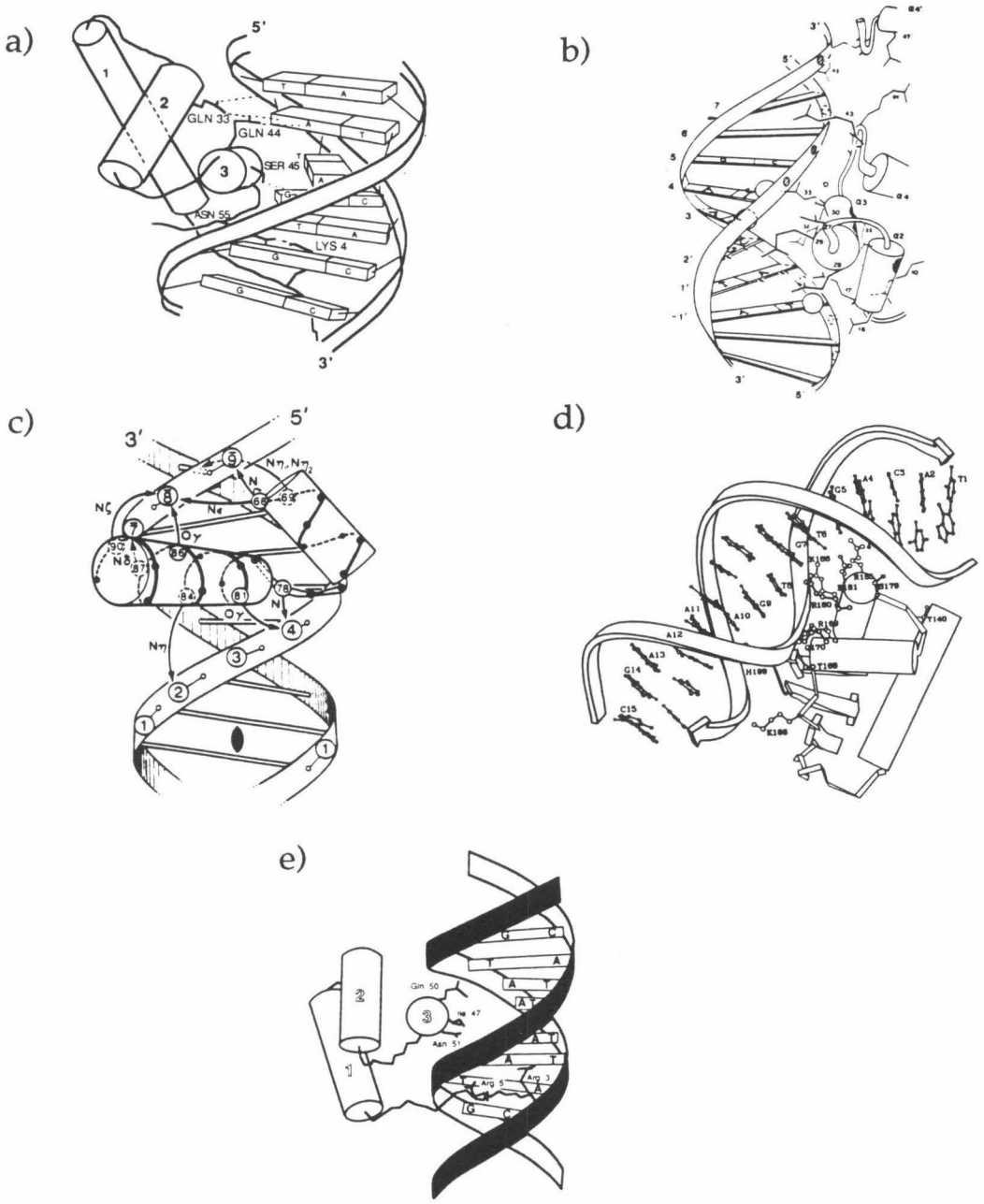


Figure 7 A schematic representation of the helix-turn-helix regions of proteins bound to their operator sequences taken from protein DNA co-crystals. The proteins represented are (a) λ-repressor (Jordan and Pabo, 1988), (b) 434-repressor (Aggarwal, et al., 1988), (c) TRP-repressor (Otwinowski, et al., 1988), (d) CAP (Schultz, et al., 1991), and (e) the homeo domain protein engrailed (Kissinger, et al., 1990). In these representations, the α-helices are represented as cylinders.

helix, was positioned in the major groove. The protein-protein contacts responsible for the dimerization of the proteins were formed between the fifth helix of each monomer. Specific hydrogen-bonding interactions were observed between the major groove of DNA and the side chains of two amino acids located near the

amino terminus of helix 3, the recogntion helix. Glutamine 44 was observed to form a bidentate hydrogen bond between the N7 and N6 of an adenine. This type of bidentate hydrogen-bonding pattern was predicted to be a determinant of protein-DNA binding specificity (Seeman, et al., 1976). The interaction between glutamine 44 and the adenine appears to be stabilized by a hydrogen bond that occurs between glutamine 33, from helix 2, and glutamine 44. Serine 45, which is also located at the amino end of the recognition helix, formed a hydrogen bond with the N7 of a guanine contained in the operator sequence. Other sequence-specific hydrogen-bonding, involved asparagine 55, which is found just past the carboxyl end of helix 3, and lysine 4, which is present in the amino terminal segment of the protein. In this interaction asparagine 55 and lysine 4 combined to recognize a guanine near the dimeric center of the operator sequence. Asparagine 55 simultaneously formed a hydrogen bond with the N7 of guanine and lysine 4, which in turn formed a hydrogen bond with the O6 of the same guanine. Sequence-specific hydrophobic interactions were observed to occur between the β-carbon of alanine 49 and the terminal carbon of isoleucine 54 and the methyl groups of two different thymines in the major groove. For each monomer a series of five hydrogen bond interactions were observed with the phosphate sugar backbone. These interactions flanked the 5'- and 3'- edges of the point in the major groove that was occupied by the recognition helix. The amino terminal end of this protein was observed to track along the major groove.

Figure 7b contains a schematic representation of 434 repressor bound to its operator sequence (Aggarwal, et al., 1988). 434 repressor, a protein containing 5 α-helices, binds as a dimer to its operator sequence with an approximate two-fold axis asymmetry. The DNA in this complex is B-form; however, it is distorted by bending, creating an arc of 65Å in radius (Aggarwal, et al., 1988). The width of the minor groove changes throughout the operator sequence. At the edges of the operator sequence, it is approximately 14Å. Towards the middle of the operator sequence, near the axis of pseudosymmetry, the width narrows to 8.4Å. These are significant alterations in groove width when compared with the normal 11.7Å width of the minor groove of B-form DNA (Saenger, 1984). Comparison of the structure of 434 repressor from the co-crystal (Aggarwal, et al., 1988) with the crystal structure of the protein alone (Mondragon, et al., 1989) shows that no large conformational changes occur in the protein upon DNA binding. However, this comparison reveals that significant adjustment in the positions of some amino acid side chains occurs coincident with DNA binding. The protein-protein contacts in the dimer are mediated by a series of hydrophobic interactions that occur between helices 4 and 5 of the monomers, in addition to a salt bridge in this region. Helices 2 and 3 form the helix-turn-helix unit, with helix 3 serving as the recognition helix. Helix 3 lies in the floor of the major groove, with the amino termini of the adjacent helices, helices 2 and 4, lying in close proximity to the sugar phosphate backbone of the DNA. In this co-crystal, as in the one of λ-repressor, the amino end of the recognition helix makes specific contacts with DNA bases in the floor of the major groove (Pabo and Lewis, 1988; Aggarwal, et al., 1988). Hydrogen-bonding contacts between the recognition helix of 434 repressor and the major groove of DNA are mediated by glutamines 28 and 29. Glutamine 28 forms a

bidentate hyrodgen bond with an adenine located at the 5'-end of the recognition sequence, and glutamine 29 hydrogen bonds to the guanine of the adjacent GC base pair. A non-polar interaction is observed between the methyl group of a thymine and a hydrophobic pocket formed by a portion of the side chain from glutamine 29 and from threonine 27. Hydrogen bond contacts between the protein and the DNA sugar phosphate backbone occur at the 5'-edge of the recognition sequence with the amino acid side chains found at the amino end of α-helix 2. Contacts between the protein and the sugar phosphate backbone of the DNA are also observed between the amino acids in the loop between α-helices 3 and 4, and the phosphate groups at the 3'-edge of the recognition sequence. It is interesting to note that some of these contacts were mediated by solvent molecules. An additional solvent-mediated protein-DNA contact is observed between arginine 43, which is located in the loop between helices 3 and 4, and the minor groove. In this contact arginine 43, through three bridging water molecules, simultaneously contacts an AT base pair and two sugar phosphates.

A schematic representation of the helix-turn-helix region of Trp repressor bound to its operator sequence is shown in Figure 7c. Initially, this structure was solved to 2.4Å resolution (Otwinowski, et al., 1988) and later it was more highly refined to 1.9Å resolution (Luisi and Sigler, 1990). The protein binds to its operator DNA as a dimer with a pseudo C2 axis of symmetry. The structure-bound Trp repressor contained 6 α-helices (A-F). Helices D and E form the helix-turn-helix unit of the protein. In the co-crystal, the amino end of the recognition helix, helix E, and the loop between helices D and E point into the major groove, with helix D positioned across the major groove. The structure of the DNA operator is approximately B-form (Otwinowski, et al.,

1988). While the DNA structure does not exhibit a large bend, as was observed in the case of 434 repressor (Anderson, et al., 1987), it does show some structural differences from normal, uncomplex B-form DNA (Otwinowski, et al., 1988; Luisi and Sigler, 1990). In the structure of the protein-DNA complex, the phosphates that are involved in protein-DNA contacts show marked deviations in position relative to unbound DNA (Otwinowski, et al., 1988). Some of these phosphates have moved by as much as 9.6Å The adjustments in the positions of these phosphates give the appearance of the DNA reaching and grabbing the protein as if to hold it in place (Luisi and Sigler, 1990). When the structure of complex Trp repressor is compared with that of uncomplex Trp repressor, some interesting differences are observed (Otwinowski, et al., 1988). The amino region of complex Trp repressor representing helices A through C shows no significant conformational differences from the structure of unbound repressor. However, the region of Trp repressor, which is involved in DNA binding, containing helices D and E, the helix-turn-helix unit shows a major reorganization upon binding to the DNA. When the protein-DNA co-crystal is examined for sequence-specific contacts, only one direct contact is observed to occur between the protein and the major groove. In this contact, arginine 69, located near the amino terminus of helix D, forms a bidentate hydrogen bond with a guanine at the 5'-end of the operator sequence. This interaction is stabilized by a water molecule that bridges from the arginine to an adjacent phosphate. While there are no other direct contacts between the protein and the major groove of the operator sequence, there are three contacts mediated by six water molecules (Luisi and Sigler, 1990). Two of the water-mediated contacts involve the protein-amide backbone and the third involves the side chain of lysine 72. Eleven hydrogen-bonding contacts were observed with the sugar phosphate backbone (Otwinowski, et al., 1988; Luisi and Sigler, 1990).

The Trp repressor operator complex provided new insights into the determinants of protein-DNA recognition. In the co-crystal structures of λ- and 434 repressor there were a significant number of sequence-specific interactions formed between the protein and the DNA bases in the floor of the major groove. Only one interaction of this type was observed in the Trp repressor DNA complex. This raises the question of what factors determine the binding specificity of Trp repressor? Luisi and Sigler, et al. (1990) have postulated that the sequence-specific recognition of the Trp repressor operator by Trp repressor is mediated by a combination of bridging water molecules, DNA conformational plasticity, and protein conformational plasticity. They postulated that the hydrogen-bonding interactions mediated by the water molecules would play the largest single role in determining protein-binding specificity, with the shape-selectivity of the protein and DNA for one another providing the balance. They noted that the Trp repressor protein would be unable to contact simultaneously the phosphates on each side of the major groove unless the DNA was able to bend and conform to the protein.

A schematic representation of CAP bound to its DNA operator sequence is shown in Figure 7d (Schultz, et al., 1991). CAP binds to its DNA operator as a dimer with a pseudo dyad axis of symmetry. There is a single molecule of cyclic AMP found between each CAP monomer and the DNA operator sequence. The overall structure of the DNA operator sequence shows a 90° bend, which results from two 40° kinks between base pairs 5 and 6 on each side of the dyad axis of symmetry. In the co-crystal, the amino terminal end of helix D, the recognition helix of the helix-turn-helix unit, penetrates the major groove with its axis parallel to the plane of the bases. In other helix-turn-helix proteins,

the recognition helix is parallel with the plane of the groove. In this complex, three amino acid side chains emanating from the recognition helix appear to hydrogen-bond directly to three base pairs in the major groove. Arginine 180 forms a bidentate hydrogen bond with the N7 and O6 of a guanine at the center of the recognition sequence. Arginine 185 hydrogen bonds to a guanine from a GC base pair at the 3'-end of the recognition sequence. Glutamic acid 181 forms a hydrogen bond with the N4 of the cysteine from the same GC base pair. As with other helix-turn-helix proteins, a significant number of hydrogen bonding interactions were observed between the phosphate backbone and the protein on either side of the portion of the major groove occupied by the recognition helix. In the case of CAP, sequence-specific DNA binding is thought to be a result of a combination of hydrogen bond interactions between the protein and the floor of the major groove and the sequence-dependent bendability or deformability of the DNA duplex (Schultz, et al., 1991).

A schematic representation of the last protein-DNA complex to be discussed, the homeodomain protein engrailed, is shown in Figure 7E (Kissinger, et al., 1990). Unlike its prokaryotic brothers, the eukaryotic protein engrailed binds to its DNA recognition sequence as a monomer. The protein contains three α-helices and an amino terminal arm. Helices 2 and 3 form the helix-turn-helix unit. Helix 3, the recognition helix, is positioned in the major groove, with helices 1 and 2 lying across the major groove perpendicular to helix 3. The DNA duplex in the co-crystal is a relatively straight segment of B-form DNA with an average helical twist of 34.2° and approximately 10.53 base pairs per turn. The only distortion appears to be a slight widening of the major groove in the region where helix 3 binds (Kissinger, et al., 1990). Extensive sequence-specific contacts occur between the floor of the major groove and the side

chains of amino acids located near the amino end of helix 3. In the crystal structure, a bidentate hydrogen bond is observed between asparagine 51 and the N7 and N6 of an adenine located at the 3'-end of the binding site. A hydrophobic interaction is observed to occur between isoleucine 47 and the methyl group of the thymine in the adjacent AT base pair. Glutamine 50 also forms a van der Waals contact with the methyl group of a thymine in the major groove. As in the prokaryotic proteins, extensive hydrogen-bonding interactions are observed between the protein and the sugar phosphate backbone on either side of the portion of the major groove that is occupied by the recognition helix. Unlike the prokaryotic proteins, the homeodomain protein engrailed makes multiple sequence-specific contacts with DNA bases in the adjacent minor groove. The amino terminal arm of the protein lies along the minor groove. The side chains of two arginines, located in the amino terminal arm, reach into the minor groove and form sequence-specific hydrogen-bonding interactions with thymines. When the structure of homeodomain is compared to the structure of the prokaryotic protein λ-repressor, the α-carbon backbones from the helix-turn-helix portions of these proteins superimpose with a root mean square difference in atomic positions of only 0.84Å (Kissinger, et al., 1990). Although the structure of the helix-turn-helix units in λ-repressor and engrailed are very similar, examination of the protein-DNA co-crystals reveals that they are used in very different ways to recognize the DNA. If the λ-repressor operator complex is used as a starting point and the helix-turn-helix units of the proteins are superimposed, it appears that the DNA in the homeodomain complex has been shifted towards the C-terminal end of the second helix in the helix-turn-helix unit. This difference in positioning shifts the register of amino acids in the recognition helix that are in position to make critical contacts with DNA base pairs. While the helix-turn-

helix motif is similar in engrailed and prokaryotic proteins, it appears to be used in a different manner to recognize DNA.

Comparison of the DNA co-crystals from several helix-turn-helix proteins reveals that there are several potential mechanisms used by this class of proteins to recognize DNA. There is direct hydrogen bonding, as in the case of λ-repressor or engrailed, direct hydrogen bonding in combination with alternations in DNA conformation, as in the case of 434 repressor and CAP, and water-mediated hydrogen-bonding in combination with changes in structural conformation of both the protein and DNA, as in the case of Trp repressor. It has become apparent that even if the same mechanism of sequence-specific recognition is used, for example, direct hydrogen-bonding in the case of λ-repressor and engrailed, variations in the relative positioning of the protein and DNA can significantly alter the amino acid side chains that are positioned to interact with the DNA. A more graphic example of differences in the relative positioning of helix-turn-helix proteins and DNA is provided by lac repressor. The orientation of the recognition helix of this protein is opposite with respect to the recognition helices of λ-repressor and the other helix-turn-helix proteins that have been discussed (Boelens, et al., 1987; Lehming, et al., 1987; Shin, et al., 1991). While a large body of structural information exists on binding of DNA by helix-turn-helix proteins, our understanding is limited because a great number of these proteins are uncharacterized. Since the determination of crystal structures can be problematic for some proteins, and because multi-dimensional NMR can require a large time investment, a need for techniques that can rapidly provide information about protein-DNA complexes is obvious. To meet this need, a series of solution-phase techniques that can rapidly

provide information on the nature of protein-DNA complexes have been developed.

Analysis of Protein and DNA Binding in Solution. Three solution techniques frequently used to examine the nature of protein-DNA interactions are chemical or enzymatic footprinting (Galas and Schmitz, 1978; Dervan, 1986; Tullius, et al., 1987), interference experiments (Siebenlist and Gilbert, 1980), and affinity cleavage (Dervan, 1991). The first two techniques are extremely useful in describing the DNA binding site for a given protein. Footprinting maps the region of DNA that is protected by a bound protein. If a chemical agent is used, it can be mapped to nucleotide resolution. Interference experiments using methylation and ethylation can identify specific bases, in the case of methylation, and specific phosphates, in the case of ethylation, which are involved in protein binding. However, neither of these techniques provides information on the orientation of the protein with respect to the DNA. In order to obtain information about the protein's orientation with respect to the DNA, a technique called affinity-cleaving has been developed (Dervan, 1991). The affinity-cleaving technique involves incorporation of the DNA-cleaving moiety, EDTA•Fe, at discrete amino acid residues within a protein (Sluka, et al., 1987). The cleavage pattern generated following chemical activation, with a reducing agent such as dithiothreitol (DTT), allows the position of the protein relative to the DNA to be mapped to nucleotide resolution (Sluka, et al., 1987). EDTA•Fe localized at a specific DNA-binding site cleaves both DNA strands, typically over a four to six base pair region, via a diffusible species (Figure 8) (Shultz, et al., 1982; Taylor, et al., 1984; Dervan, 1986). The central point of the cleavage pattern marks the portion of the DNA sequence that is proximal to the amino

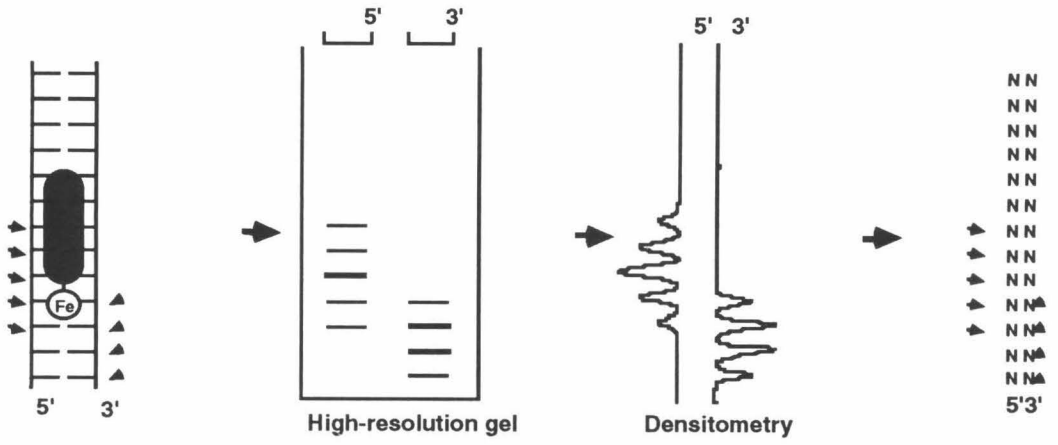


Figure 8 A schematic representation of the high-resolution assay used to interpret affinity-cleavage patterns.

acid at which the EDTA has been covalently attached (Sluka, et al., 1987). Due to the right-handed nature of double-helical DNA, the groove in which the EDTA iron is located can also be identified by analysis of the cleavage patterns (Dervan, 1991). An EDTA•Fe located in the minor groove generates an asymmetric cleavage pattern with the maximal cleavage loci shifted to the 3' side on opposite strands (Figure 9) (Taylor, et al., 1984; Dervan, 1986). When the EDTA is located in the major groove, the maximal cleavage loci are 5' shifted; in addition cleavage of lower efficiency occurs on the distal strands of the adjacent minor grooves (Moser and Dervan, 1987; Griffin and Dervan, 1989; Oakley and Dervan, 1990). This results in a pair of 3' shifted asymmetric loci of unequal intensity on opposite strands (Figure 9). This pair of patterns is consistent with a diffusible species, generated by an EDTA iron located above the major groove, that reacts in the major and minor grooves of DNA with unequal rates, preferentially (although not necessarily exclusively) in the minor groove. Tullius and Dombrowski (1985) obtained a cleavage pattern that was

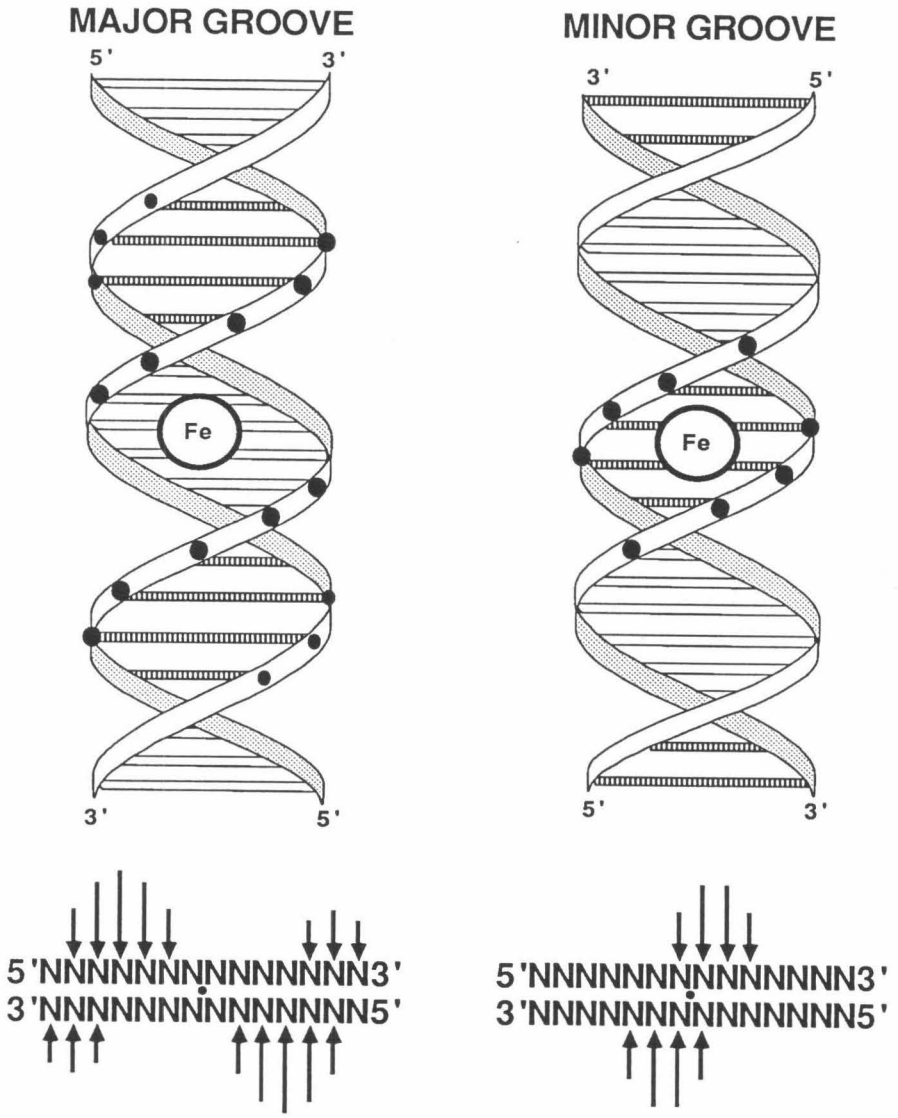


Figure 9 Cleavage patterns produced by a diffusible oxidant generated by Fe(II)•EDTA located in the major and minor grooves of right-handed DNA. Filled circles represent points of cleavage along the phosphodiester backbone. The sizes of the circles represent the extent of cleavage. Beneath each helical representation is the type of densitometric cleavage pattern observed when EDTA is positioned in the respective groove.

sinusoidal in intensity when DNA, bound to a precipitate of calcium phosphate, was allowed to react with EDTA iron. This sinusoidal cleavage pattern was a direct result of the differences in reactivity of the major and minor grooves of DNA towards the oxidative species generated by the EDTA iron. The studies described in this section detail the use of affinity cleavage to elucidate the mechanism by which γδ resolvase binds to DNA.

Previous Studies of the Interactions Between γδ Resolvase and DNA. γδ resolvase is a 183 amino acid protein encoded by the transposable element, gamma delta, a member of the TN3 family of bacterial transpons (Grindley and Reed, 1983). Resolvase is both a site-specific recombinatorial protein and a repressor. γδ resolvase interacts with 120 bp site named *res*, which lies within the region between the divergently transcribed TNP A and TNP R genes of the γδ transposon. Resolvase protects all three of the dimeric binding sites within *res* from nuclease digestion (Grindley, et al., 1982; Kitts, et al., 1983). Each binding site consists of a pair of inverted repeats of nine base pairs (Figure 10). The consensus sequence for these binding sites is TGTCYNNTA (where Y is pyrimidine and N is any base) (Grindley, et al., 1982). The binding sites are separated by a variable spacer of 7, 10 or 16 base pairs. Site I, which has a 10-bp spacer, contains the recombination, or crossover point. γδ resolvase has been shown to induce a structural change corresponding to a bend in the DNA at site I (Hatfull, et al., 1987; Salvo and Grindley, 1988).

All three dimeric binding sites contained in *res* are required for recombination (Grindley, et al., 1982; Kitts, et al., 1983; Wells and Grindley, 1984). When two identically oriented copies of the *res* site are present on a superhelical plasmid, γδ resolvase in the presence of magnesium will recombine these sites to

Site	Sequence
IL	TGTGCGATA
IR	CGTCCGAAA
IIL	TGTCTATTA
IIR	TGTCTGTTA
IIIL	TGTCCGATA
IIIR	TGTATCCTA
Consensus	TGTCYNNTA

Figure 10 The DNA sequences contained within *res* that are recognized and bound by γδ resolvase and γδ(141-183). In all cases, the DNA sequences are written in the 5' to 3' direction. The consensus sequence was derived from comparison of the sequences that are recognized by the proteins. In the consensus sequence Y = a pyrimidine and N = any base.

produce two caternated circular pieces of DNA (Reed and Grindley, 1981; Reed, 1981). Recombination proceeds via a staggered mechanism in which DNA cleavage occurs at site I (Reed and Grindley, 1981). During recombination the DNA between sites I and II is bent, or looped out, due to interactions between the different resolvase dimers within the complex (Salvo and Grindley, 1988). The unique structure obtained by this complex was termed a "resolvasome". In the absence of magnesium, the DNA is cut but not religated and an intermediate can be isolated in which the resolvase protein is covalently linked to the 5' phosphate of the DNA via serine 10 (Reed and Grindley, 1981; Reed and Moser, 1984; Hatfull and Grindley, 1986).

The interactions between the 183 amino acid protein γδ resolvase and DNA have been examined by methylation and ethylation interference studies (Figure

11) (Falvey and Grindley, 1987). Major groove methylations within the 9 bp recognition sequence, as well as ethylation of phosphates within and adjacent to this region were found to inhibit resolvase binding. Minor groove methylations of adenine at and adjacent to the 3'-end of the recognition sequence also inhibited resolvase binding. Resolution was inhibited by methylation of adenine at the center of site I, suggesting that minor groove contacts near the crossover may be required for this activity.

Figure 11 A consensus of inhibitory modifications displayed relative to the 9 base pair consensus sequence. R = purine, Y = pyrimidine, N or n = any base. (n is used for positions outside the 9 bp-conserved segment.) Arrows indicate inhibitory ethylated phosphates, asterisks indicate major groove inhibitions (i.e., methylations where a guanine occurs in this position), and open circles indicate minor groove inhibitions (i.e., inhibitory methylations where an adenine occurs at this position). Adapted from Falvey and Grindley (1987).

γδ resolvase can be cleaved by chymotrypsin into two fragments, one 140 amino acids long and one 43 amino acids long (Abdel-Meguid, et al., 1984). The 140 residue amino-terminal domain does not bind DNA, but does contain residues known to be important for catalysis (Newman and Grindley, 1984) and to be responsible for protein-protein interactions (Abdel-Meguid, et al., 1984). Recently, the crystal structure of the catalytic domain of γδ resolvase has been solved at 2.7 Å resolution (Figure 12) (Sanderson, et al., 1990). The resolvase catalytic domain is an αβ structure consisting of a central β sheet made up of one antiparallel and four parallel β strands surrounded by four α

helices. The serine at position 10, which becomes covalently linked to the DNA during the recombinational event, is located in a loop between β strand 1 and α helix a. It was surprising to find this active site serine present on the surface of the protein. Enzyme-catalytic residues are usually positioned in an enzyme-active site pocket and not on the surface of the protein. A mutational study has identified four resolvase residues, within this 140 amino acid domain, that are responsible for the protein-protein interactions required for recombination (Hughes, et al., 1990). Mutations at residues 2, 32, 54, and 56 produce a phenotype that is characterized by an inability to catalyze recombination, and that shows a loss of cooperative binding to the *res* DNA sequence. The side chains from these four residues are clustered on the surface of the protein that mediates the interactions between resolvase dimers in the crystal structure.

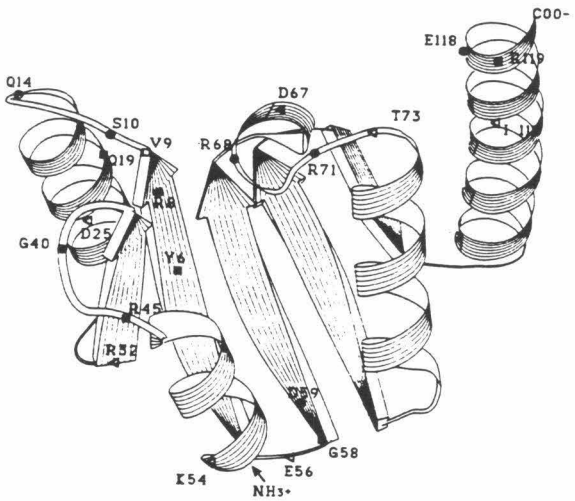


Figure 12 A schematic representation of the 140 residue amino terminal domain resulting from chymotrypsin cleavage of γδ resolvase. Adapted from Sanderson, et al. (1990).

The 43 amino acid carboxyl terminal domain, γδ (141-183), binds specifically to each half site contained within *res*. However, unlike the native resolvase, which binds to all the sites with equal affinity, γδ (141-183) binds each of the six half sites with different affinities (Abdel-Meguid, et al., 1984). Ethylation interference experiments reveal that the phosphate contacts made by the carboxyl terminal, DNA-binding domain are similar to the intact resolvase, with the exception of a single phosphate at 3' edge of each contact region (Rimphanitchayakit, et al., 1989). Phosphate contacts extend across adjacent major and minor grooves on one face of the DNA helix. The DNA is contacted over a 12 base pair sequence that includes the 9 base pair inverted repeat.

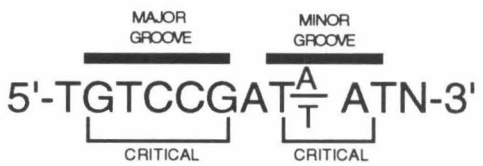


Figure 13 The optimum DNA-binding sequence for γδ (141-183) determined from mutagenesis experiments. Portions of the sequence that are contacted in the major and minor groove of the protein are indicated. Nucleotide substitution at positions that are labeled critical can result in a 100-fold increase k_D of the protein. Adapted from Rimphanitchayakit and Grindley (1990). N = any base.

Saturation mutagenesis of the DNA site bound by γδ (141-183) has defined the sequence requirements for binding of γδ resolvase to its recognition site (Rimphanitchayakit and Grindley, 1990). Saturation mutagenesis provided 31 of the possible 36 single base mutations within the twelve base pair minimum binding sequence. Binding assays *in vitro* with γδ (141-183) show that substitutions at eight of the twelve positions strongly inhibit complex formation, increasing the dissociation constant by 100 times or more. The

critical positions fall into two groups: positions 1-6 towards the 5'-end of the binding site and positions 9-10 towards the 3'-end of the binding site (Figure 13). These positions correspond to regions where the DNA domain spans the major and minor grooves of its respective binding site (Falvey and Grindley, 1987; Rimphanitchayakit, 1989). Base substitutions at the intervening positions, 7 and 8, have modest effects on binding, less than a 20-fold increase in K_D. Substitutions at positions 11 and 12 corresponding to the 3'-end of the DNA sequence have no effect on binding. The results of this study have identified an optimal sequence for γδ resolvase binding, 5'-TGTCCGAT(A/T)ATN-3' (where N equals any base) (Rimphanitchayakit, 1990).

γδ (141-183) contains a high degree of sequence similarity with the helix-turn-helix regions of several DNA binding proteins (Pabo and Sauer, 1984). Given sequence similarities with other DNA binding proteins, it has been proposed that resolvase uses the helix-turn-helix motif in the major groove for sequence-specific DNA-binding (Abdel-Meguid, et al., 1984). The proposed orientation of the DNA binding protein γδ (141-183) is one in which the carboxyl end of the protein lies in the major groove, and the amino terminus is oriented toward the center of each dimeric binding site (Abdel-Meguid, et al., 1984; Rimphanitchayakit, et al., 1989).

The Orientation of γ,δ (141-183) Determined by Affinity Cleaving.

Synthesis. Four 43 amino acid proteins based on the DNA-binding domain of γδ resolvase (residues 141-183) were synthesized by a combination of automatic or manual, solid-phase techniques using tBoc-protected amino acids (Kent, 1988; Sarin, et al., 1981; Tam, et al., 1979; Merrifield, 1969). One of the proteins was γδ

(141-183) with no modifications; it was a synthetic protein corresponding to the natural fragment produced by chymotrypsin cleavage of native γδ resolvase (Abdel-Meguid, et al., 1984). Two of the peptides contained one EDTA molecule located at a discrete position within γδ (141-183) (Figure 14A). The first of these, EDTA-γδ (141-183), had EDTA derivatized to the amino terminus of the protein (Figure 14B). The EDTA was attached to the amino terminus of the protected resin-bound peptide using tribenyzl EDTA-γ-aminobuteric acid (BEG) as described by Sluka, et al (1987; 1990). The second peptide containing one EDTA molecule, γδ (141-183)-EDTA, had the EDTA molecule derivatized near the carboxy terminus (Figure 14C). To provide a site of attachment for the EDTA near the carboxy terminus, the sequence of γδ (141-183) was altered. At position 183, the asparagine, which is found in the native molecule, was replaced by a lysine, whose ε-amino group provided the point of attachment for the EDTA molecule. Attachment of the EDTA molecule near the carboxy terminus of the protein was accomplished by an orthoganol synthetic scheme that combined the use of tBoc and Fmoc protection (Figure 15) (Mack, et al., 1990; Sluka, et al., 1990). The N-ε-Fmoc-protecting group on the lysine on position 183 was removed selectively from the side chain using piperidine and DMF. The tricyclohexal ester of EDTA was then attached to the side chain via an HOBt ester. The remaining portion of γδ (141-183)-EDTA was then synthesized using standard tBoc chemistry. The fourth γδ resolvase derivative that was synthesized, EDTA-γδ (141-183)-EDTA, contained two EDTA's attached at discrete positions within the peptide. The first EDTA was attached adjacent to the carboxy terminus on the ε-amino group of lysine 183, and the second EDTA molecule was attached to the amino terminus of the protein at position 141. The overall synthetic yield for these peptides is found in Table 1.

A)

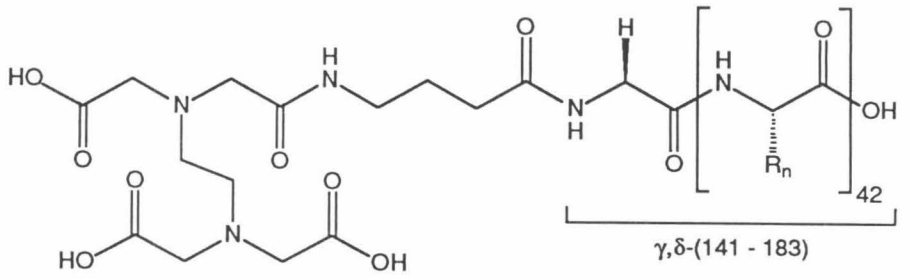
GRKRKIDRDAVLNMWQQG

α_1

LGASHISKTMNIARSTVYKVINESN

α_2 α_3

B)



C)

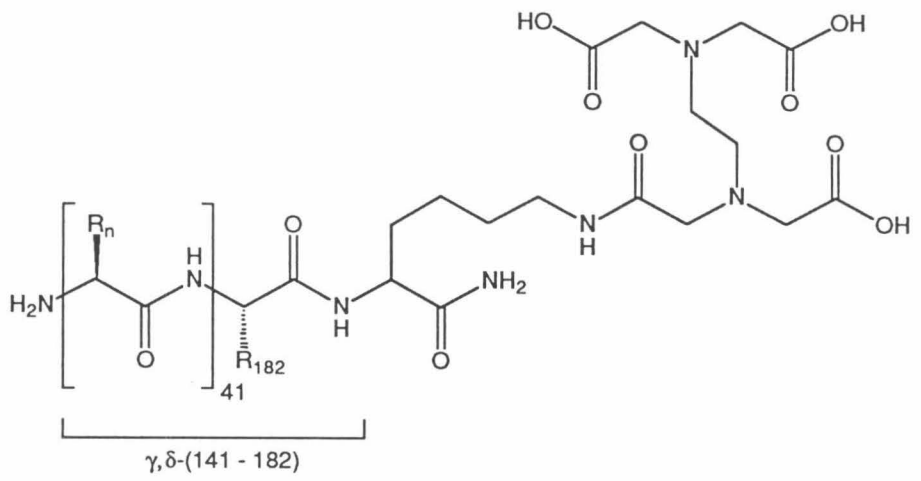


Figure 14 a) The 43 amino acid DNA-binding domain of γδ resolvase. Underlined regions are possible α-helices assigned according to a secondary structure predicting algorithm (Garnier, et al., 1978). In this structure helices 2 and 3 would form the helix-turn-helix unit of the protein, with helix 3 representing the recognition helix. b) The analog with EDTA attached to the amino terminus, EDTA-γδ(141-183). c) The analog with EDTA attached near the carboxy terminus containing the point mutation Asn^{183} to Lys^{183}, γδ(141-183)-EDTA.

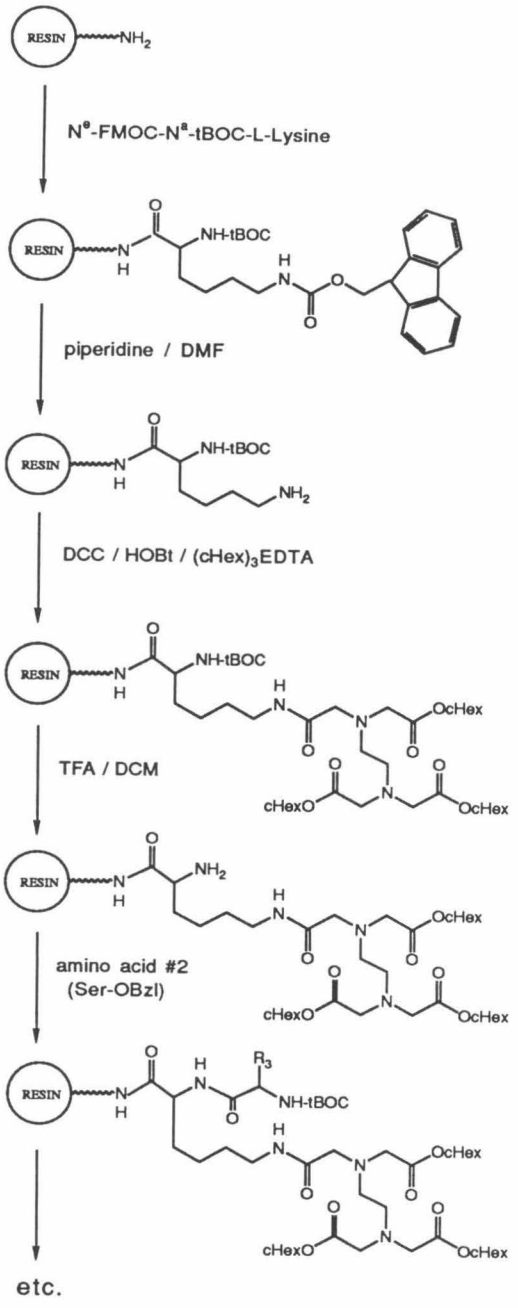


Figure 15 The synthetic scheme for the attachment of the tricyclohexal ester of EDTA to the ε-amino group of Lys^{183}, on *P-methylbenzylhydrylamine* resin. DCC = dicyclohexalcarbodimide, (TFA) = trifluoroacetic acid; Bzl represents benzoyl and *chex* = cyclohexyl.

Following synthesis the proteins were deprotected and cleaved from the resin using HF treatment (prior to HF treatment the histidine-protecting, dinitrophenyl, was removed by thiolysis, and terminal t-Boc groups were removed with trifluoroacetic acid). The proteins were purified on a C8 reverse phase HPLC column using a 0-60% acetonitrile gradient. Sequences of the proteins were confirmed, using amino acid analysis and Edman degradation. Footprinting and affinity-cleaving patterns obtained with these four peptides were used to evaluate the orientation and binding characteristics of γδ resolvase.

Table 1. Overall Yields for the Synthetic Peptides Based on the DNA-Binding Domain Obtained by Chymotrypsin Digestion of γδ Resolvase

Peptide	Position of EDTA Attachment	Number of Residues Coupled	Overall Yield
γδ(141-183)	none	42	74.5%
EDTA-γδ(141-183)	NH2-terminus	43	73.3%
γδ(141-183)-EDTA	COOH-terminus	43	80.2%
EDTA-γδ(141-183)-EDTA	NH2,COOH-termini	45	78.2%

Footprinting. Footprinting studies of the synthetic protein γδ (141-183) demonstrated that at 2.0 μM concentrations the 43-mer binds to all six half sites contained within the *res* binding sites (Figures 16 and 17). At each half site the γδ (141-183) protects a twelve base pair region of DNA centered on the 9 base pair consensus sequence (Figures 16 and 17). These observations are consistent with DNAse footprinting studies (Abdel-Meguid, et al., 1984) and ethylation

Figure 16 An autoradiogram of a high-resolution denaturing polyacrylamide gel containing ^{32}P-end labeled fragments from PRW80. Bars on the left indicate the positions of the three binding sites, I, II, III (each consisting of a pair of imperfectly conserved, inverted repeats or half-sites) for gd resolvase with *res*. Odd and even numbered lanes 1-12 contain 5'- and 3'-labeled DNA, respectively. Lanes 1 and 2 contain intact DNA, lanes 3 and 4, EDTA-γδ (141-183) at 0.5 μM concentration; lanes 5 and 6, EDTA-γδ (141-183) at 2.0 μM concentration; and lanes 7 and 8, EDTA-γδ (141-183) at 10.0 μM concentration. Lanes 9 and 10 are MPE footprinting lanes containing γδ (141-183) at 2.0 μM concentration. Lanes 11 and 12 are MPE•Fe control lanes. Lanes 13 and 16 are 5' and 3' chemical sequencing G reactions, respectively (Maxam and Gilbert, 1980); lanes 14 and 15 contain 5' and 3' chemical sequencing A reactions, respectively (Iverson and Dervan, 1987).

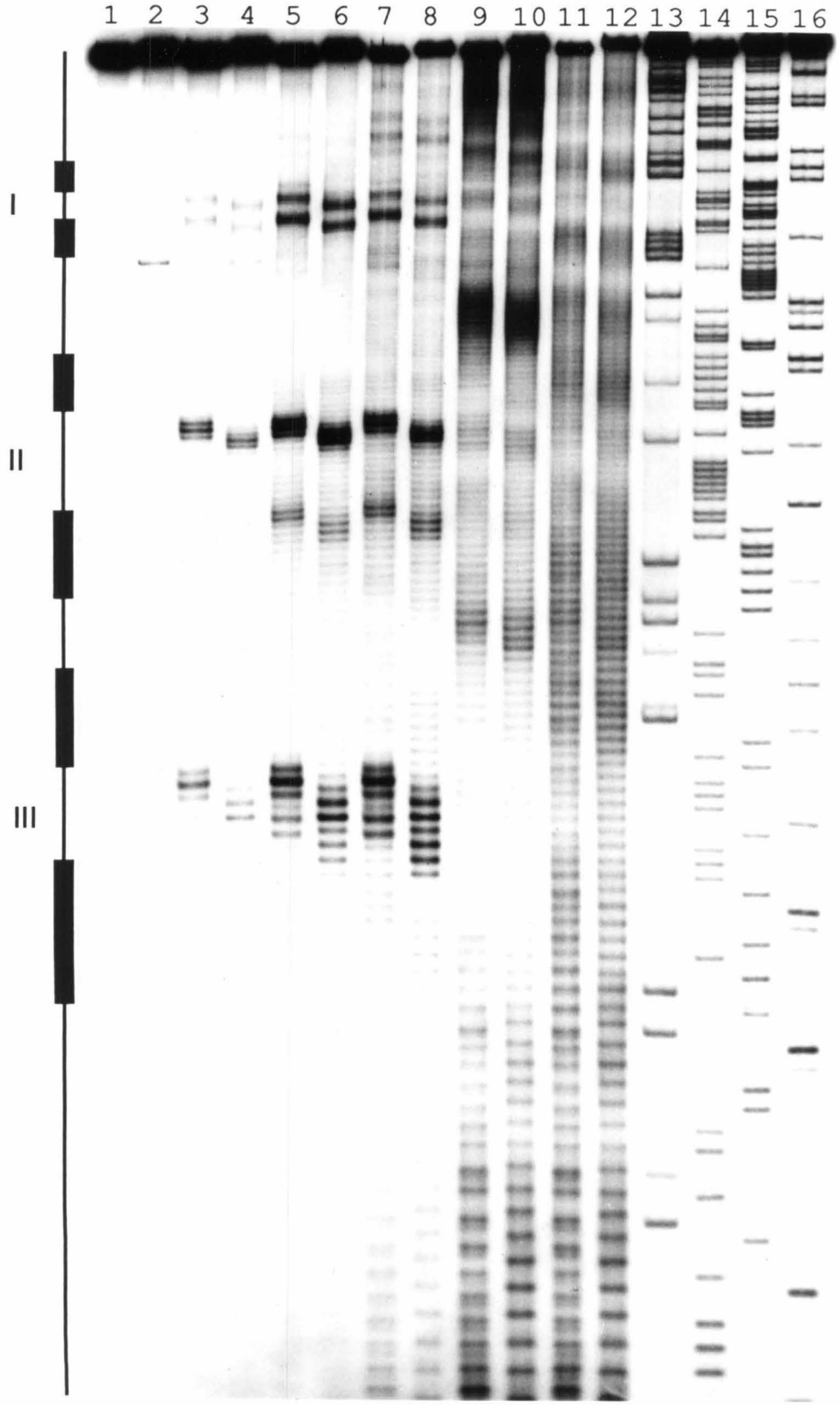
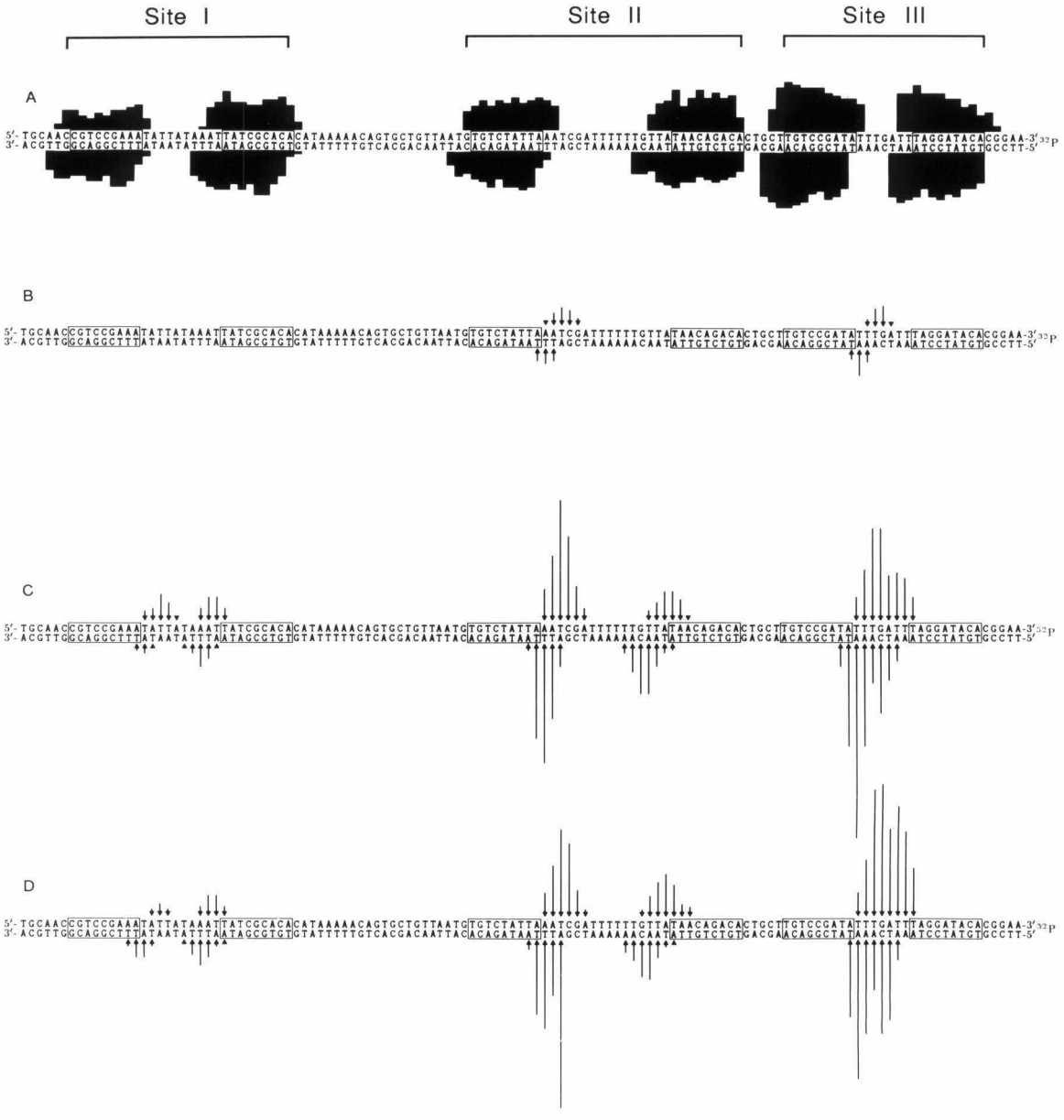

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
I
II
III

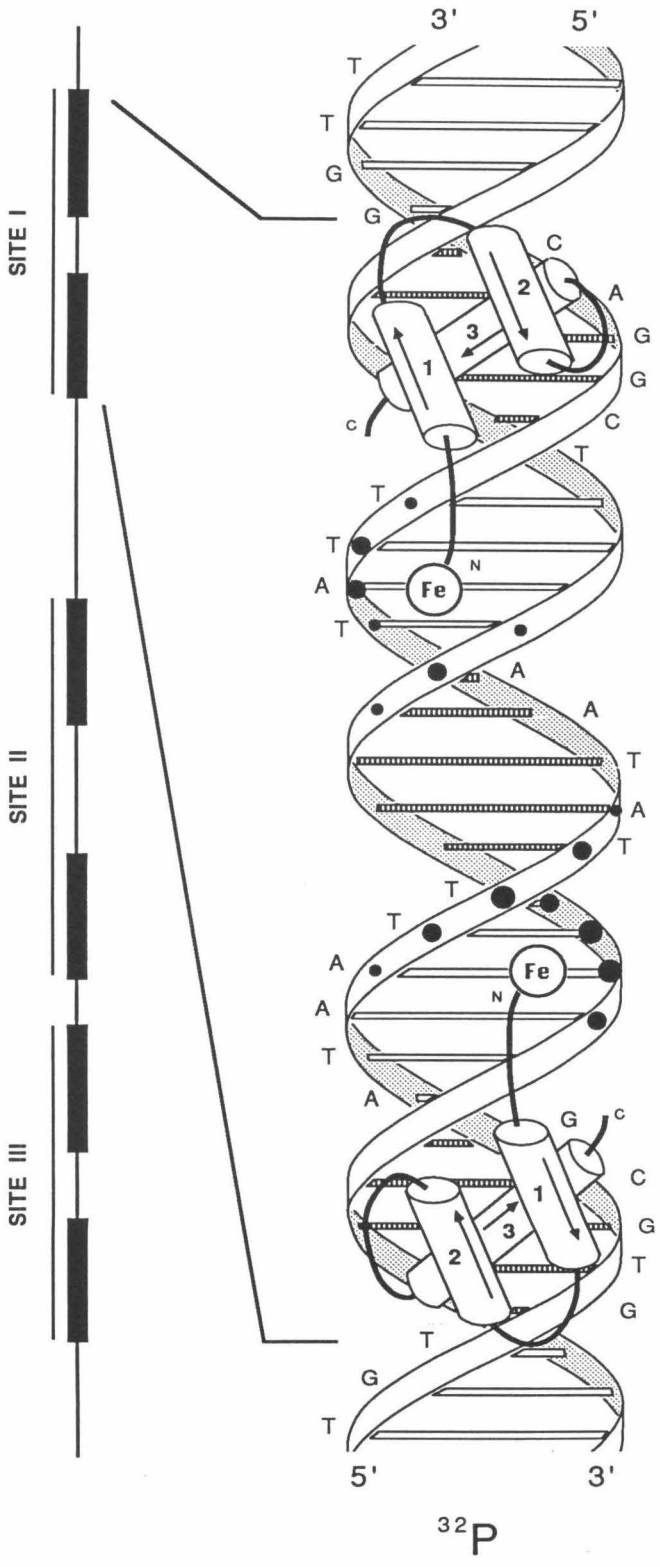
Figure 17 Histograms of footprinting and affinity-cleavage data from the gel contained in Figure 16. The sequence, left to right, corresponds to the DNA sequence for site I to site III (top to bottom of the gel). I, II, and III are indicated by brackets. Boxes represent the 9 base pair binding sites assigned for γδ resolvase (Grindley, et al., 1982). a) bars represent the extent of protection from MPE cleavage in the presence of γδ (141-183) (2.0 μM). b) Arrows represent the extent of cleavage by Fe•EDTA-γδ (141-183) at 0.5 μM. C) 2.0 μM and D) 10.0 μM.


Site I
Site II
Site III
A
B
C
D
5′- TGCAACGCGTCCGAAATATTATAAATTATCGCACACATAAAAACAGTGCTGTTAATGTGTCTATTAAATCGATTTTTTGTTATAACAGACACTGCTTGTCCGATATTTGATTTAGGATACACGGAA-3′ 32P
3′- ACGTTGGCAGGCTTTATAATATTTAATAGCGTGTGTATTTTTGTCACGACAATTACACAGATAATTTAGCTAAAAAACAATATTGTCTGTGACGAACAGGCTATAAACTAATCCTATGTGCCTT-5′

interference studies (Rimphanitchayakit, et al., 1989) using γδ (141-183) derived from a chymotrypsin digest of native resolvase. This indicates that the synthetically derived γδ (141-183), and its related EDTA containing analogs, should provide useful information about the orientation and binding characteristics of the DNA-binding domain of γδ resolvase.

Position of the Amino Terminus. Affinity-cleaving studies with Fe•EDTA-γδ (141-183) yield a 3' shifted cleavage pattern adjacent to the 3'-end of the each half site contained within the *res* DNA sequence (Figures 16 and 17). These cleavage patterns are located at the center of each dimeric binding site for γδ resolvase. The 3' shift of the cleavage pattern indicates that the Fe•EDTA group attached to the amino terminus of γδ (141-183) is located proximal to the minor groove of DNA near the center of the dimeric binding site (Figure 18). This is in agreement with the ethylation interference studies conducted by Rimphanitchayakit, et al (1989). Ethylation of the 3' most phosphates contained within the 9 base pair consensus sequence, as well as two additional phosphates that are positioned across the minor groove, inhibited binding of the 43 amino acid carboxy terminal domain of γδ resolvase. Given this information and comparing it with other helix-turn-helix proteins, Rimphanitchayakit, et al., (1989) proposed a model that positioned the amino terminus of γδ (141-183) in the minor groove adjacent to the 3'-end of the 9 base pair consensus sequence for binding. The affinity-cleaving studies with Fe•EDTA-γδ (141-183) support this model by establishing the position of the amino terminus of the protein as proximal to the minor groove at the 3'-end of the 9 base pair consensus sequence.

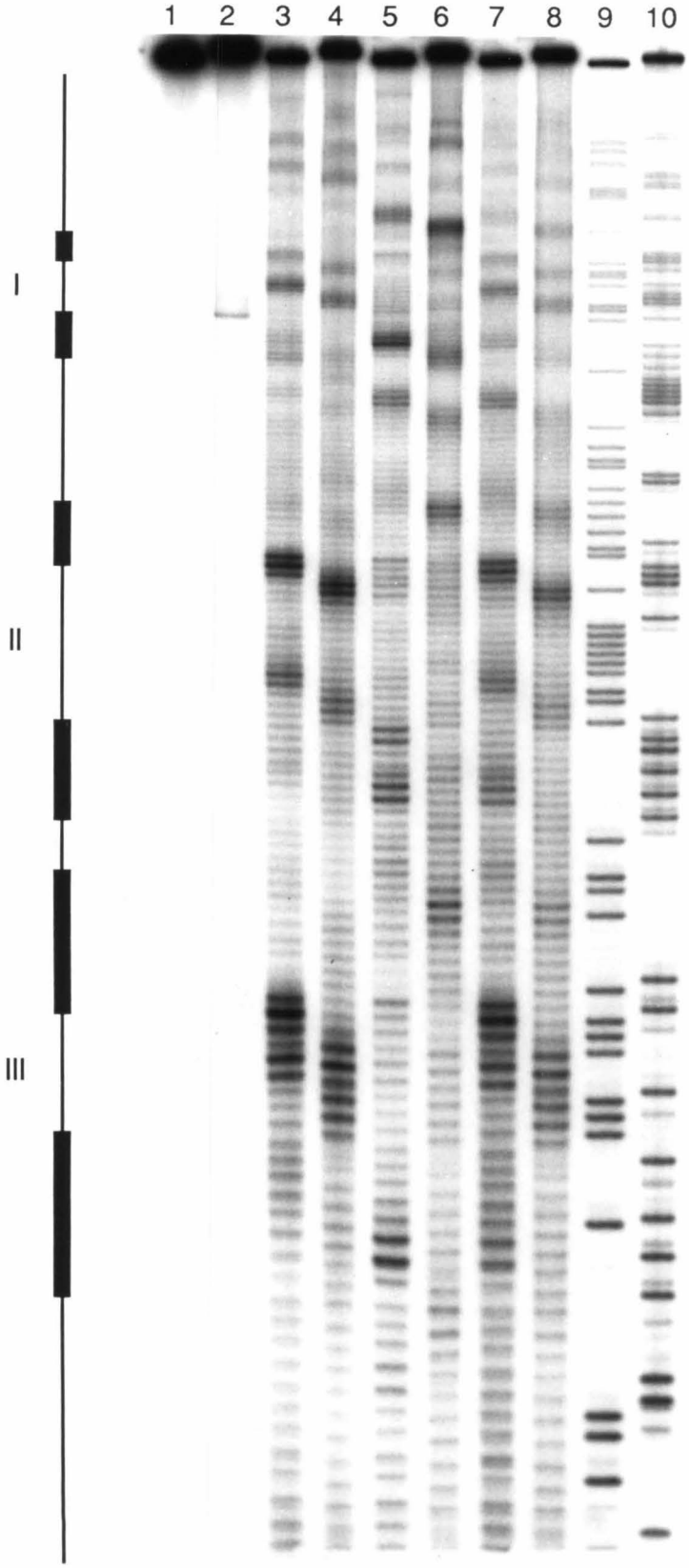
Figure 18 A model for γδ (141-183) binding to site I of *res*. The location of the EDTA•Fe moiety attached to the amino terminus of γδ (141-183) was assigned from the affinity-cleavage patterns contained in Figure 17. Filled circles represent the positions of cleavage along the phosphodiester backbone. The sizes of the circles represent the extent of cleavage at the indicated base position.


3'
5'
SITE I
SITE II
SITE III
Fe
N
1
2
3
5'
3'
^{32}P

The cleavage of *res* by Fe•EDTA-γδ (141-183) at different concentrations (in the range of 0.5 to 10 μM) shows that the six half sites have different affinities for the DNA-binding domain (Figures 16 and 17). At 0.5 μM protein concentration, sites II-L and III-L are cleaved. At a fourfold higher concentration (2.0 μM), cleavage at sites I-L, I-R, II-R, and III-R appear with unequal intensity (Figures 16 and 17). If the amount of cleavage is proportional to the extent of occupancy, the data suggest that the relative affinities of the DNA-binding domain of γδ is II-L, III-L > III-R, II-R > I-L, I-R. This is somewhat different with the relative binding affinities assigned from footprinting studies of the DNA-binding domain obtained by the chymotrypsin digest. Abdel-Meguid, et al (1984) assigned the relative orders of affinity as II-L, III-L, > I-R, > III-R, I-L > II-R. The difference in relative affinities is probably a result of the method by which the binding affinities were obtained.

Position of the Carboxy Terminus. The specific cleavage patterns produced by γδ (141-183)-EDTA•Fe are shifted to the 3' side and indicate that the EDTA•Fe, attached near the carboxy terminus of the putative recognition helix, is positioned within the γδ-binding site above the major groove adjacent to the 5' end of the 9 base pair recognition sequence (Figures 19 and 20). In the case of the right half of site I, the cleavage pattern is centrally located about the sequence 5'-TGTGC-3' (Figure 21). An examination of the structural data available for prokaryotic regulatory proteins that bind DNA, using the helix-turn-helix motif, identifies two possible orientations for the putative recognition helix of γδ (141-183) (Figure 21). High-resolution, x-ray crystal structures of the λ repressor and 434 repressor protein-DNA complexes show that the carboxy termini of the recognition helices in these proteins are pointing toward the center of the dimeric binding sites (Jordan and Pabo, 1988;

Figure 19 An autoradiogram of a high-resolution denaturing polyacrylamide gel containing ^{32}P-end labeled fragments from PRW80. Bars on the left-hand side of the gel mark the location of the three dimeric-binding sites contained within *res*. Odd numbered lanes contain 5'-end labeled DNA and even numbered lanes contain 3'-end labeled DNA. Lanes 1 and 2 contain intact DNA; lanes 3 and 4 contain 10 μM Fe•EDTA-γδ(141-183); lanes 5 and 6 contain 10 μM γδ(141-183)-EDTA•Fe; lanes 7 and 8 contain 10 μM Fe•EDTA-γδ (141-183)-EDTA•Fe; lanes 9 and 10 are A-specific cleavage lanes.


1
2
3
4
5
6
7
8
9
10
I
II
III

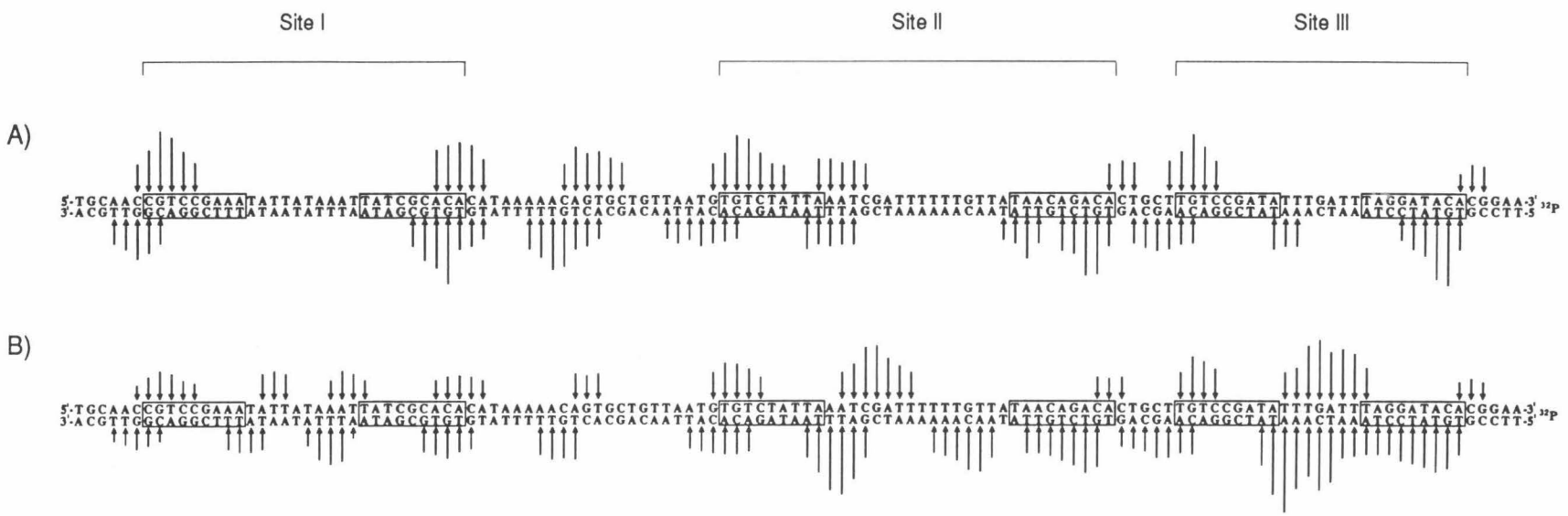


Figure 20 Histograms of the affinity-cleavage data from the gel in Figure 19. The sequence left to right represents sites I, II, and III (top to bottom of the gel). Arrow heights indicate the extent of cleavage at the indicated bases. A) Cleavage by γδ (141-183)-EDTA•Fe; B) Cleavage by Fe•EDTA-γδ (141-183)-EDTA•Fe.

Aggarwal, et al., 1988). The second possible orientation is obtained from NMR studies of the *lac* repressor headpiece bound to its operator DNA (Boelenes, et al., 1987). The studies show that the putative recognition helix from *lac* repressor is oriented with the carboxy terminus in the opposite direction, pointing away from the center of the dimeric-binding site. When the position of the EDTA•Fe attached to the ε amino group of lysine 183 is considered for each possible orientation, relative to the cleavage data, a better fit is obtained with the orientation shown in Figure 21A. This orientation is similar to the orientation of the recognition helix of λ and 434 repressors obtained from crystal structures (Aggarwal, et al., 1988; Jordan and Pabo, 1988). This orientation is also similar to that assigned for the 52 residue DNA-binding domain of Hin recombinase (Mack, et al., 1990). It is opposite to the orientation assigned to the *lac* repressor headpiece recognition helix by affinity-cleaving techniques (Shin, et al., 1991). The fact that the cleavage pattern is seen predominantly on one, but not on both, minor grooves adjacent to the major groove location of the helix-turn-helix motif suggests that the EDTA•Fe moiety (and hence the putative reocgnition helix of γδ (141-183)) is not positioned symmetrically in the major groove. These data would be consistent with the carboxyl end of the recognition helix projecting outward from the floor of the major groove and tilting away from the center of the inverted repeat-binding site. Clearly, refinement of these models must await more definitive x-ray crystallographic and NMR analysis of the protein-DNA complex.In controls, γδ (141-183) equipped with EDTA at both the amino and carboxyl termini, Fe•EDTA-γδ (141-183)-EDTA•Fe produces a cleavage pattern consistent with the combination of patterns from Fe•EDTA-γδ (141-183) and γδ (141-183)-EDTA•Fe (Figures 19 and 20). The fact that the cleavage patterns obtained for

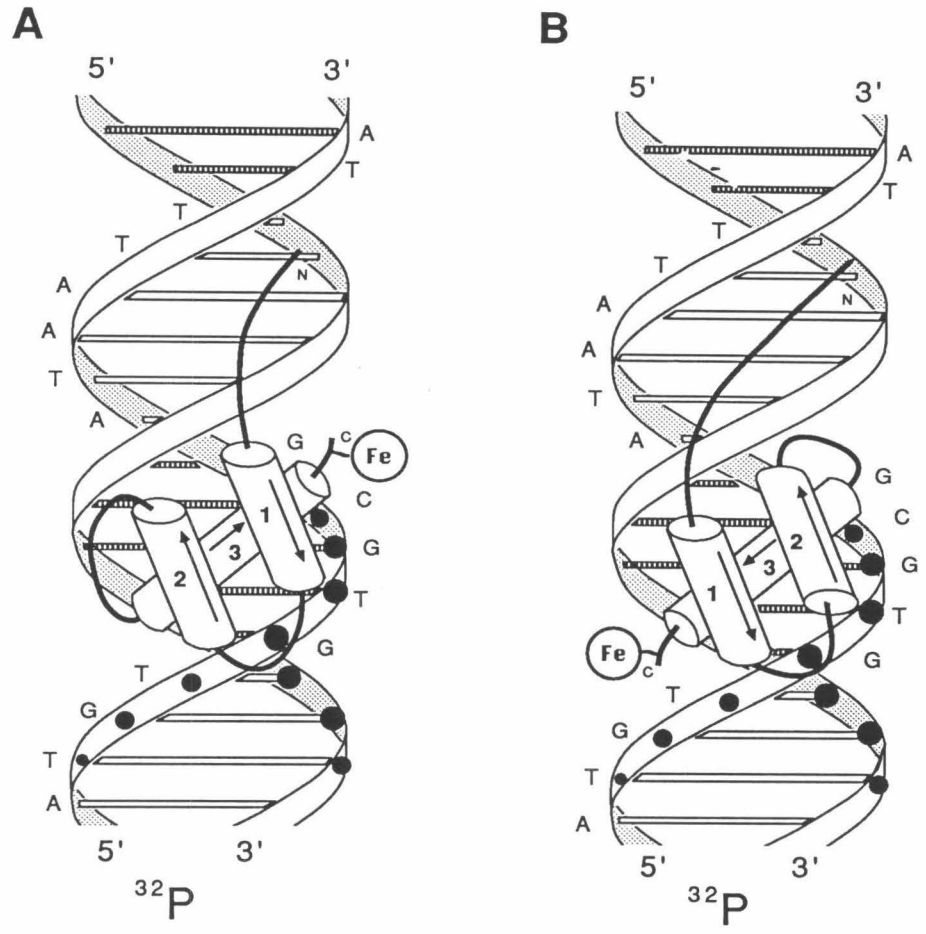


Figure 21 A schematic representation of the two models of γδ (141-183)-EDTA•Fe bound to the right half of site I. Sites of DNA cleavage along the helix are marked by dots, the sizes of the dots correspond to the magnitude of cleavage at that position. a) The orientation of the putative recognition helix based on 434 and λ-repressor DNA co-crystals. b) The orientation of the putative recognition helix based on *lac* repressor NMR studies. Model A seems to provide the best fit with the experimental cleavage data.

the EDTA•Fe derivatized analogs of γδ (141-183) remain unchanged whether or not EDTA•Fe is present or absent at other positions in the molecule suggests that the structure of γδ (141-183) is not significantly perturbed by the attachment of EDTA•Fe at position 141 or position 183.

Conclusions. MPE•Fe(II) footprinting studies have shown that the synthetic protein, γδ (141-183), protects a twelve base pair region of DNA centered on the γδ recognition sequence. Affinity-cleavage studies with EDTA on the amino terminus have located the amino terminus of γδ (141-183) in the minor groove at the center of each dimeric binding site. Cleavage studies with EDTA attached at the caboxy terminus of γδ (141-183) reveal that the putative recognition helix is in the adjacent major groove, oriented (N to C) toward the center of each binding site. The binding model for γδ (141-183), which is similar to that proposed for Hin (139-190) (Sluka, et al., 1987, 1990; Mack, et al., 1990), places the helix-turn-helix motif of the protein in the major groove with residues at the amino terminus extending across the DNA phosphodiester backbone and making specific contacts on the same face of the helix to the adjacent minor groove. This model for γδ (141-183) binding is consistent with methylation and ethylation interference studies, using the native resolvase (Falvey and Grindley, 1987) and ethylation interference studies using γδ (141-183) obtained from a chymotrypsin digest (Rimphanitchayakit, et al., 1989). γδ and Hin DNA-binding domains may be examples of modular DNA and protein interactions with two adjacent DNA sites, major and minor grooves, bound on the same face of the helix by two separate parts of the protein.

The Effects of Amino Acid Substitution on DNA-Binding Specificity.

Experiments performed by Wharton, et al (1984) and Wharton and Ptashne (1985) have shown that by changing the amino acid composition of the recognition helix from a protein that binds DNA using the helix-turn-helix motif can alter its DNA binding specificity. The study performed by Wharton, et al (1984) involved the proteins 434 repressor and 434 cro. Through the use of

sequence homologies, the putative recognition helices in each of these proteins were identified and aligned. Comparison of the putative recognition helices from 434 cro and 434 repressor showed that they differed in only 5 amino acids. A "helix switch" was performed by substituting the recognition helix of 434 repressor with the recognition helix of 434 cro. The resulting hybrid protein was named repressor*. Repressor* possesed the same sequence-specific DNA contacts and binding characteristics as the 434 cro protein (Wharton, et al., 1984). Exchanging the recognition helices had succeeded in changing the DNA-binding characteristics of the protein.

In a second experiment, Wharton and Ptashne (1985) were able to alter the binding specificity of 434 repressor by replacing only the amino acids on the solvent-exposed surface of the recognition helix. By aligning and superimposing the sequences of 434 and P22 repressors with the helix-turn-helix unit observed in the x-ray crystal structure of λ-repressor, the residues on the solvent-exposed, or DNA contact, surface of the recognition helices were identified. The sequence of 434 repressor was then mutated to incorporate these residues from P22 repressor. The resulting hybrid molecule possessed the DNA-binding specificity of P22 repressor. In order to determine which portion of the recognition helix was responsible for determining DNA-binding specificity, the recognition helix was divided into three segments. The first segment, the amino terminal end of the recognition helix contained three "contact residues," the second segment, the center of the recognition helix contained one "contact residue," and the third segment, the carboxyl end of the recognition helix contained one "contact residue." Amino acid substitution within the carboxy segment of the recognition helix was found to have no effect on the DNA-binding specificity of the hybrid protein. Simultaneous

substitution of the three "contact residues" in the amino segment of the recognition helix produced a protein that showed no binding specificity towards either 434 or P22 binding sites. Substitution of only the contact residue in the central segment of the recognition helix also produced a protein that showed no binding specificity towards either 434 repressor or P22 repressor binding sites. Only simultaneous substitution of all the "contact residues" located in the amino and center segments created a protein that exhibited sequence-specific DNA binding. Since the amino segment of the recognition helix contained three "contact residues" that were always substituted simultaneously, it was impossible to determine if all three of these residues are involved in DNA binding. A systematic study that would examine all possible single and multiple amino acid exchanges between the contact surfaces of two helix-turn-helix proteins might provide further insight into the molecular basis of protein-DNA recognition by helix-turn-helix proteins.

"Helix Switch" studies using γδ (141-183) and Hin(139-190). γδ resolvase and Hin recombinase are both thought to bind to DNA, using the helix-turn-helix motif (Abdel-Meguid, et al., 1984; Sluka, et al., 1987). Hin recombinase and γδ resolvase show a 35% homology of amino acid composition (Simon, et al., 1980). Both proteins contain two functional domains, an amino-terminal, catalytic domain and a carboxy terminal DNA-binding domain (Abdel Meguid, et al., 1984; Sluka, et al., 1987). The carboxy terminal DNA-binding domains, γδ (141-183) and Hin (139-190), are synthetically accessible, using solid-phase tBoc peptide synthesis (Sluka, et al., 1987; Graham and Dervan, 1990). Affinity-cleaving studies have demonstrated that the amino and carboxy termini of these proteins share similar orientations with respect to their DNA-binding sites (Sluka, et al., 1987, 1990; Mack, et al., 1990; Graham and Dervan, 1990).

The model for DNA-binding by both of these includes a helix-turn-helix motif in the major groove with residues at the amino terminus extending across the DNA phosphodiester backbone and making specific contacts on the same face of the helix to the adjacent minor groove (Sluka, et al., 1987, 1990; Mack, et al., 1990; Graham and Dervan, 1990; Glasgow, et al., 1989; Bruist, et al., 1987; Rimphanitchayakit, et al., 1989; Falvey and Grindley, 1987).While the orientation of the proteins relative to their DNA-binding sites is similar and while they possess a high degree of amino acid sequence homology, their DNA-binding sites are significantly different in the 5' region (Figure 22). Methylation interference studies have demonstrated that the major groove in this region is involved in protein binding (Falvey and Grindley, 1987; Rimphanitchayakit, et al., 1989; Glasgow, et al., 1989; Bruist, et al., 1987). The models for DNA binding by γδ resolvase and Hin recombinase position the recognition helices of the proteins in the major groove in this region of the recognition sequence. This indicates that the recognition helix from γδ resolvase is designed to bind to the sequence 5'-TGTCC-3', while the recognition helix from Hin recombinase is designed to bind 5'-TTCTY-3'. Consequently, it is not surprising that these two peptides show little cross reactivity towards each other's binding sites (Figure 23 and 24). The combination of synthetic accessibility, similar DNA binding orientations, and low cross reactivity for DNA binding sites make γδ (141-183) and Hin(139-190) excellent candidates for a systematic helix-switch experiment. *Experimental Design.* A sequence alignment of the helix-turn-helix region from λ-cro, Hin(139-190), and γδ (141-183) is shown in Figure 25. Sequence alignment was made based on the residues that are thought to control the packing (and therefore the three-dimensional structure) of the helix-turn-helix unit, and not on the residues that contact the DNA. This type of sequence alignment makes

a) DNA sequences recognized by Hin recombinase.

Site	Sequence
HixLR	TTATCAAAAA
HixLL	TTCTTGAAAA
HixLR	TTCTTCCTTA
HixRL	TTCTCCTTTA
Consensus	TTCTYNNAAA

b) DNA Sequences Recognized by γδ resolvase

Site	Sequence
IR	TGTGCGATA
IL	CGTCCGAAA
IIR	TGTCTATTA
IIL	TGTCTGTTA
IIIR	TGTCCGATA
IIIL	TGTATCCTA
Consensus	TGTCYNNTA

Figure 22 The recogniton sequences of the half sites bound by Hin(139-190) a) and γδ (141-183) b). In all cases the DNA sequences are written in the 5' to 3' direction. The consensus sequences are derived from comparison of the sequences from the individual half sites. N = any base. Y = pyrimidine.

it possible to indentify differences in amino acid composition at positions in the recognition helix that are thought to be in contact with the DNA. These differences in amino acid composition may be responsible for producing the differences in DNA-binding specificity observed between the proteins. Based on this alignment the sequences of the helix-turn-helix regions of γδ (141-183) and Hin (139-190) were superimposed upon the α carbon backbone of λ-cro,

Figure 23 An autoradiogram of a high-resolution gel containing ^{32}P-end labeled fragments of DNA from the Sal I/Hind III fragment from PRW80 containing *res*. Lanes 1 through 5 contain 5'-end labeled DNA, lanes 6 through 10 contain 3'-end labeled DNA. Lanes 2 through 9 are affinity-cleavage lanes. Lanes 2 and 9 contain EDTA•Fe(II)-γδ (141-183) at 2.0 μM concentration. Lanes 3 and 8 contain EDTA•Fe(II)-Hin(139-190) at 40.0 μM concentration, lanes 4 and 7 contain EDTA•Fe(II)-Hin(139-190) at 20.0 μM concentration, lanes 5 and 6 contain EDTA•Fe(II)-Hin(139-190) at 10.0 μM concentration. Lanes 1 and 10 contain A-specific cleavage lanes (Iverson and Dervan, 1987).

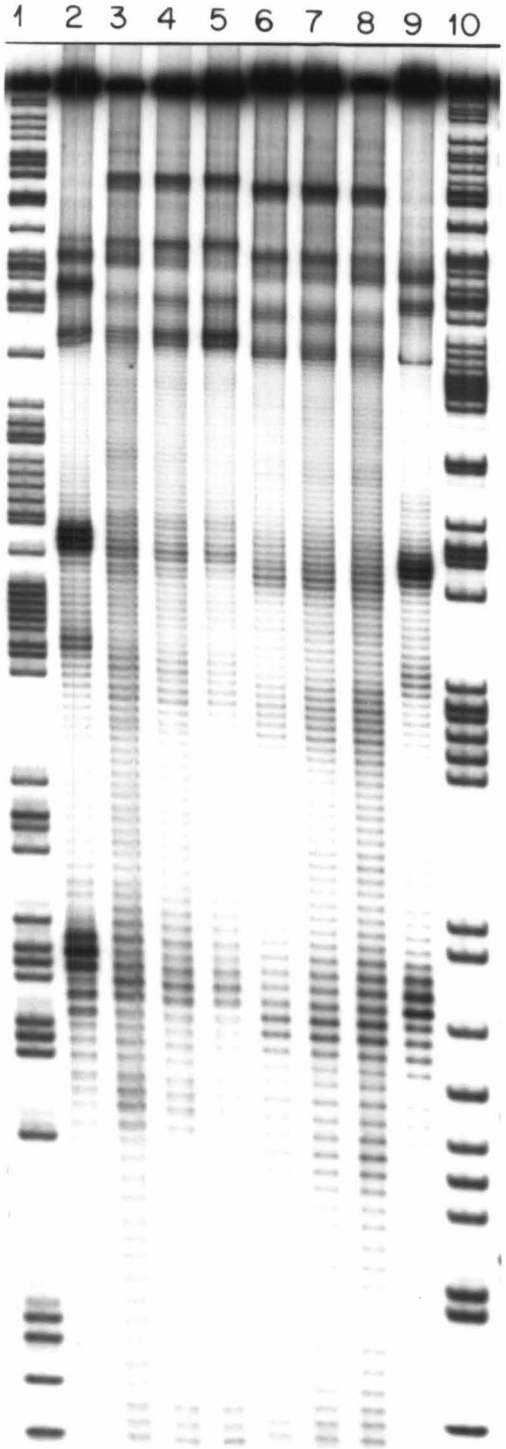

1
2
3
4
5
6
7
8
9
10

Figure 24 An autoradiogram of a high-resolution gel containing ^{32}P-end labeled fragments of DNA from XbaI/EcoRI fragment of PMFb36 containing the Hin recombinase binding sites, Hix R and Hix L. The DNA is 3'-end labeled. Lanes 1 through 4 are affinity-cleavage lanes. Lanes 1 through 3 contain EDTA•Fe(II)-γδ (141-183) at 10, 20, 40 μM concentrations, respectively. Lane 4 contains EDTA•Fe(II)-Hin(139-190) at 2.0 μM. Lane 5 is a G-specific chemical-sequencing lane (Maxam and Gilbert, 1980).

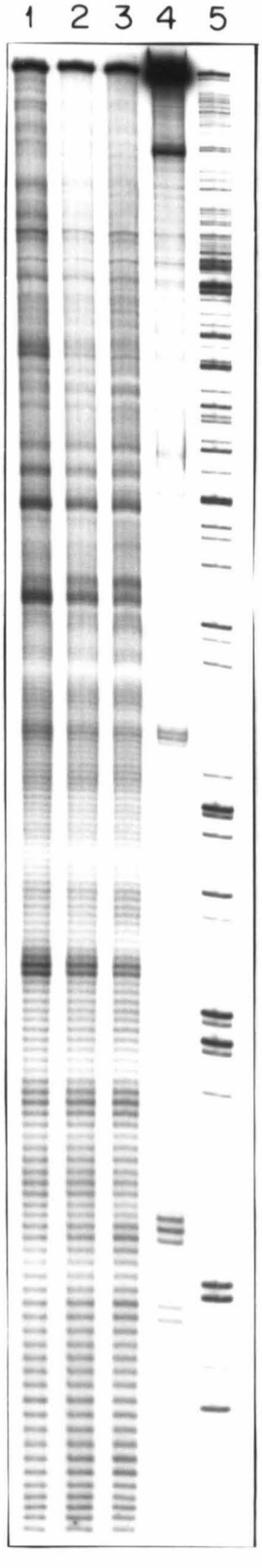

1
2
3
4
5

Figure 25 An alignment of the amino acid sequences in the helix-turn-helix domain of λ cro hidden recombinase and γδ resolvase. Residues contained within boxes are thought to provide the protein-protein contacts that define the helix-turn-helix motif. Amino acids shown in bold in the recognition helix are thought to contact the DNA based on the crystal structure of λ cro. The circled amino acids are contact residues that are variant between γδ resolvase and Hin recombinase.

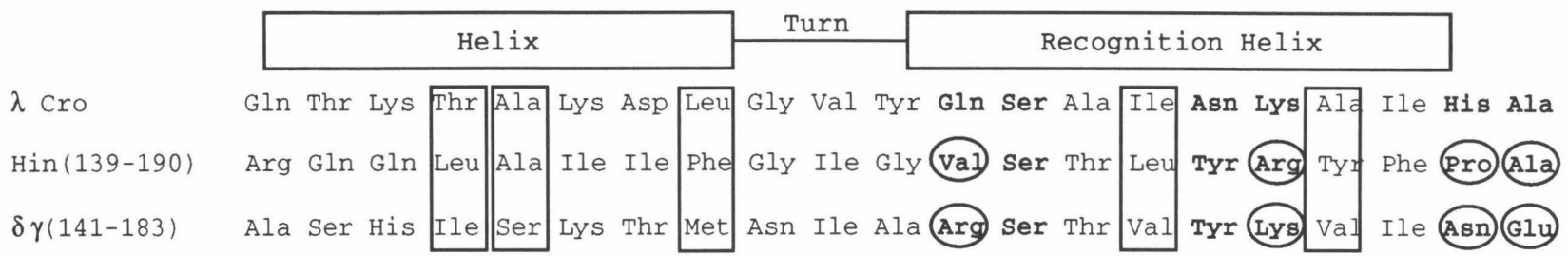


Figure 25 An alignment of the amino acid sequences in the helix-turn-helix domain of λ cro, Hin recombinase and γδ resolvase. Residues contained within boxes are thought to provide the protein-protein contacts that define the helix-turn-helix motif. Amino acid shown in bold in the recognition helix are thought to contact the DNA based on the crystal structure of λ cro. The circled amino acids are contact residues that are variant between γδ resolvase and Hin recombinase.

and the amino acid side chains that were located on the DNA contact face of the putative recognition helices were identified. Residues at six points along the recognition helix were found to have side chains that were positioned in a way that they could make contact with DNA. The amino acids at four of these positions were different in the sequences of γδ resolvase and Hin recombinase (Figure 25).

The fifteen γδ (141-183) analogs, containing single- or multiple-point mutations from the exchange of DNA "contact residues" with those found in the putative recognition helix from Hin (139-190), are shown in Figure 26. It should be noted that the proteins are designated by which amino acids are present at the four DNA contact positions that are variant between Hin and γδ. For example, natural γδ (141-183) contains an arginine, lysine, asparagine, and glutamic acid at these positions. Under this nomenclature the designation for this protein would be RKNE. If the first position is substituted with valine, which is found in Hin(139-190), the designation would become VKNE. If the first two variant-contact residues are substituted for valine and arginine, respectively, which are present in Hin(139-190), the name of the protein becomes VRNE. This nomenclature continues for substitution at any one of the four variant positions. The fifteen hybrid proteins in γδ resolvase represent all possible permutations of amino acid composition at these four variant positions. By examining all possible combinations, it is hoped that we can identify specific amino acid and DNA interactions that determine specificity.

Results. A systematic helix-switch experiment between Hin(139-190) and γδ (141-183) was conducted to determine the requirements to convert γδ (141-183) into a protein that was capable of recognizing DNA sequences bound by

	172				177			180	181
Hin	**V**	**S**	T	L	**Y**	**R**	Y	F	**P**	**A**
γδ	**R**	**S**	T	V	**Y**	**K**	V	I	**N**	**E**
Hybrids										
		S	T	V	**Y**		V	I		
	R					**R**			**P**	**A**
	V					**R**			**P**	**A***
	R					**K**			**P**	**A**
	V					**K**			**P**	**A**
	R					**R**			**N**	**A**
	V					**R**			**N**	**A**
	R					**K**			**N**	**A**
	V					**K**			**N**	**A**
	R					**R**			**P**	**E**
	V					**R**			**P**	**E**
	R					**K**			**P**	**E**
	V					**K**			**P**	**E**
	R					**R**			**N**	**E**
	V					**R**			**N**	**E**
	V					**K**			**N**	**E**
Residue Number	**172**				**177**			**180**	**181**

Figure 26 The amino acid sequences of the recognition helices from Hin, γδ, and the hybrid peptides resulting from single or multiple substitutions of amino acids within the recognition helix of γδ (141-183). In the sequences, letters shown in bold represent putative contact residues. The sequence of the hybrid proteins at the variant residues is used to name, or identify, the protein. The sequence listed with an asterisk corresponds to a complete exchange of amino acid residues between Hin and γδ resolvase. The designation for γδ resolvase under this nomenclature would be RKNE.

Hin(139-190). Since the amino acid composition of the DNA contact face of the putative recognition helices for γδ (141-183) and Hin(139-190) differed at four positions, 15 hybrid proteins of γδ resolvase were synthesized that contained one or more of the amino acids found in Hin recombinase. All 15 of these proteins were synthesized, using solid-phase tBoc methodology. EDTA was attached to the amino terminus of each of these peptides. Synthetic yields are reported in Table 2. Derivatization of the proteins with EDTA allowed the use of affinity cleaving to report both the location and the strength of protein binding. Following synthesis, the proteins were deprotected and cleaved from the resin using HF treatment (prior to HF treatment, the histidine-protecting, dinitrophenyl was removed by thiolysis, and terminal t-Boc groups were removed with trifluoroacetic acid). The proteins were purified on a C_8 reverse phase HPLC column using a 0-60% acetonitrile gradient. Sequences of the proteins were confirmed, using amino acid analysis and Edman degradation. The DNA-binding properties of the 15 hybrid peptides were examined and compared to those of γδ (141-183) and Hin(139-190) on both the Hind III/Sal I fragment of PRW 80 containing *res*, and the Xba I/Eco RI fragment of pMF B36, containing the Hin recombinase-binding sites Hix L and Hix R.

Binding of Helix-Switch Hybrid Peptides to the Hind III/Sal I Fragment from PRW 80 containing Res. The affinity-cleavage patterns produced by the EDTA-labeled, helix-switch hybrid proteins, EDTA-γδ (141-183) EDTA-Hin(139-190) when bound to the Hind III/Sal I fragment from PRW 80, which contains *res* , are shown in the autoradiogram contained in Figure 27. A summary of the strength and location of protein binding determined by densitometric analysis of the affinity-cleavage patterns from the autoradiogram in Figure 27 is found in Table 4. Examination of the data in Table 4 shows that the hybrid

Table 2. Synthetic Yields for the Proteins Used in Helix-Switch Experiments (all proteins were derivatized with EDTA at the amino terminus)

Protein	Number of Couplings	% Yield
EDTA-Hin(139-190)	53	69.8%
VRNE	44	81.5%
RRNE	44	83.1%
VKPA	44	82.9%
VKPE	44	82.2%
VRPE	44	85.9%
VRPA	44	83.0%
RKPE	44	83.6%
VKNA	44	84.7%
RRNA	44	83.7%
RRPE	44	83.1%
RKNA	44	81.5%
VRNA	44	84.3%
VKNE	44	83.5%
RKPA	44	82.6%
RRPAa	44	82.7%
RRPAb	44	82.7%
EDTA-γδ(141-183)	44	80.2%

Table 3. Retention Times for the Purified Helix-Switch Hybrid Proteins. A Brownlee C_8 Reverse Phase Column was used for the chromatography. It was eluted with a 0-60% acetonitrile gradient containing 0.1% trifuloroacetic acid. The gradient was run over a period of 60 min.

Protein	Retention Time (min)
RRPA	33.21
VRPA	35.65
RKPA	33.11
VKPA	35.53
RRNA	35.18
VRNA	38.91
RKNA	35.44
VKNA	38.85
RRPE	33.58
VRPE	35.86
RKPE	33.18
VKPE	35.49
RRNE	35.01
VRNE	38.66
VKNE	38.65

proteins could be divided into four classes.

The first class of proteins shows a strong affinity of γδ resolvase-binding sites, but no affinity for additional sites. These proteins are RKPE, RKNA, and RKPA. The only other protein that exhibits identical binding characteristics is the native peptide EDTA-γδ (141-183), or RKNE, using the same nomenclature. Comparison between these three hybrid proteins and γδ (141-183) shows that only Arg 172 and Lys 177 are invariant. The other two positions, 180 and 181, are totally variable, with all possible permutations being expressed. This seems to indicate that Arg 172 and Lys 177 are required for the recognition of γδ resolvase-binding sites.

The second class of proteins shows strong binding to the γδ resolvase sites, plus affinity for additional new sites. These proteins are RRNE, RRPE, $RRPA_{a,b}$ and RRNA. They bind all six half sites contained within *res* , plus a new site designated NI. The sequence of NI is 5'-TTATTCATTA-3'. The 5' regions of NI and Hix L (R), which is the site of the major groove contacts between the protein and the putative recognition helix, are highly homologous. The binding sequence for Hix L (R) is 5'-TTATCAAAAA-3'. When we examine the proteins that bind this series of sites, we see once again conservation of amino acids at positions 172 and 177, with variability at positions 180 and 181. In this case, positions 172 and 177 are both arginine. Apparently, Arg 177 does not affect the protein's ability to bind γδ resolvase-binding sites, but imparts an ability to bind additional sites that are more "Hin"-like such as site NI. It is interesting to note that EDTA-Hin(139-190) is also capable of binding the site NI.

Figure 27 An autoradiogram of a high-resolution denaturing polyacrylamide gel containing 5'-^{32}P-end labeled fragments from PRW80. Bars on the left indicate the position of the three binding sites I, II, and III (each consisting of imperfectly conserved inverted repeats) for γδ resolvase within *res*, Lane 1 contains control DNA. Lanes 2 through 18 are affinity-cleavage lanes containing 2.0 μM EDTA•Fe(II) amino-labeled proteins. The sequences of the proteins are written above each lane. Lane 19 contains a G-specific cleavage lane (Maxam and Gilbert, 1980) and lane 20 contains an A-specific lane (Iverson and Dervan, 1987).

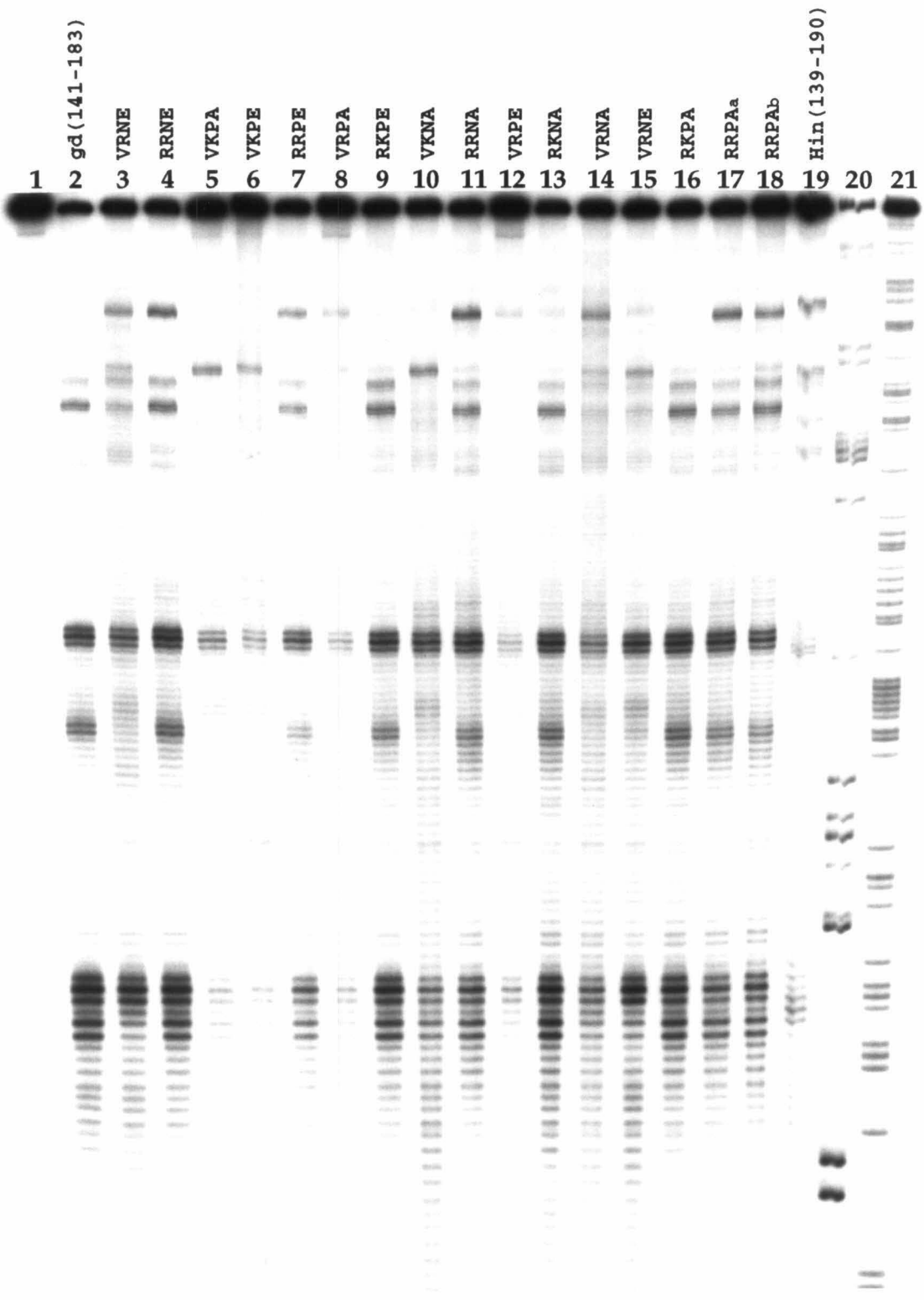

1
2 gd(141-183)
3 VRNE
4 RRNE
5 VKPA
6 VKPE
7 RRPE
8 VRPA
9 RKPE
10 VKNA
11 RRNA
12 VRPE
13 RKNA
14 VRNA
15 VRNE
16 RKPA
17 RRPAa
18 RRPAb
19 Hin(139-190)
20
21

Table 4. Protein Binding to the HindIII/Sal I Fragment from PRW80, which contains *Res* Based on Affinity Cleavage. The protein concentration is 2.0 μM. (- = no binding; + = weak binding; and ++ = strong binding)

	DNA Binding Sites						
Protein	Site IL CGTCCGAAA	Site IR TGTGCGATA	Site II L TGTCTATTA	Site II R TGTCTGTTA	Site III L TGTCCATA	Site III R TGTATCCTA	New[1] TTATTCATTA1
EDTA-γδ(141-183)	++	++	++	++	++	++	-
RKPE	++	++	++	+	++	-	-
RKNA	++	++	++	+	+	-	-
RKPA	++	++	++	++	++	-	-
RRNE	++	++	++	++	++	++	++
RRPE	++	++	++	+	++	+	+
$RRPA_a$	++	++	++	++	++	++	++
$RRPA_b$	++	++	++	+	++	++	++
RRNA	++	++	++	++	++	++	++
VKNE	-	-	+	-	+	-	-
VKPA	++	-	-	-	-	-	-
VKPE	+	-	-	-	-	-	-
VRPE	-	-	-	-	-	-	-
VKNA	+	-	+	-	+	-	-
VRNA	-	-	+	-	+	+	+
VRNE	-	+	+	-	+	++	++
VRPA	-	-	+	-	-	++	++
EDTA-Hin(139-190)	-	-	-	-	-	++	++

[1] Sequence identified by the location of affinity cleavage in combination with sequence homology with HixL(R).

The next class of proteins consists of those that possess a weak binding affinity for γδ resolvase sites, but a strong affinity for additional sites, specifically a strong affinity for site NI. These proteins are VRNA, VRNE, VRPA, and EDTA-Hin(139-190). It should be pointed out that EDTA-Hin(139-190) has the same DNA contact pattern of amino acids as the hybrid protein VRPA. However, these proteins differ in that one is a Hin recombinase, and the other is a γδ resolvase hybrid. Examination of these proteins shows once again that positions 172 and 177 are invariant within the class. In this case, position 172 is always valine and position 177 is always arginine. The other two positions, 180 and 181, are more variable.

The fourth class of proteins consists of those that retain little or no binding affinity for γδ resolvase-binding sites and express no affinity towards any additional sites, including site NI. These proteins are VKNE, VKPA, VKPE, VKNA, and VRPE. For the most part, all of these proteins possess a valine at position 172 and a lysine at position 177. The other two amino acids found at positions 180 and 181 are completely variable. The only exception to this is the protein VRPE. By analogy with the other hybrid proteins, you would expect this protein to show a binding affinity for site NI but weak or no affinity for any of the γδ resolvase-binding sites, because of the presence of a valine at position 172 and an arginine at position 177. However, the protein VRPE, for some as yet unknown reason, possesses no ability to bind DNA specifically.

Binding of Helix-Switch Hybrid Peptides to the Xba I/EcoRI Fragment from pMFB 36, which contains the Hin Recombinase Sites Hix L and Hix R.. Affinity-cleavage patterns produced by the EDTA-labeled, helix-switch hybrid proteins, EDTA-γδ (141-183) and EDTA-Hin(139-190), when bound to the XBA I/EcoRI fragment

of pMFB 36 containing Hix L and Hix R, are shown in the autoradiogram contained in Figure 28. A summary of the strength and location of protein binding determined by densitometric analysis of the affinity-cleavage patterns for the autoradiogram in Figure 28 is found in Table 5. An examination of the data contained in this table reveals that the binding characteristics of the proteins fall into three major categories.

The first category is proteins, which possess no binding affinity for Hin recombinase-binding sites, but show affinity for a novel site recognized by EDTA-γδ (141-183). These proteins are RKPE, RKNA, RKPA, RRNE, RRPE, $RRPA_{a,b}$, and RRNA. While these proteins were unable to recognize any of the Hin recombinase sites found in this restriction fragment, they were able to recognize a sequence that was also bound by EDTA-γδ (141-183). Analysis of the affinity-cleavage patterns has identified the sequence as 5'-CGCTTTTCT-3'. Within the 5'-region of this sequence, there is reasonable homology to the 5'-region of site IIL from *res*. It is not surprising that these proteins would bind a novel site recognized by EDTA-γδ (141-183) because all of them have retained strong binding affinities for the γδ resolvase-binding sites contained within *res*.

The second category of proteins includes those that contain no binding affinity for Hin recombinase-binding sites or novel sites within the restriction fragment. These proteins are VKNE, VKPA, VKPE, VKNA, and VRPE. These are the same five proteins that showed no binding specificity or affinity for sites contained within *res*. It is possible that they may not be competent to bind DNA specifically. Once again, with the exception of VRPE, they possess a valine at position 172 and a lysine at position 177.

Figure 28 An autoradiogram of a high-resolution denaturing polyacrylamide gel containing 5'-^{32}P-end labeled fragments from PMFb36. Bars on the left indicate the position of the Hin recombinase binding sites. Lane 1 contains control DNA. Lanes 2 through 18 are affinity cleavage lanes containing 2.0 μM protein with EDTA•Fe(II) attached to the amino terminus. The sequences of the proteins are written above their corresponding lanes. Lane 20 is an A-specific cleavage lane (Iverson and Dervan, 1987).

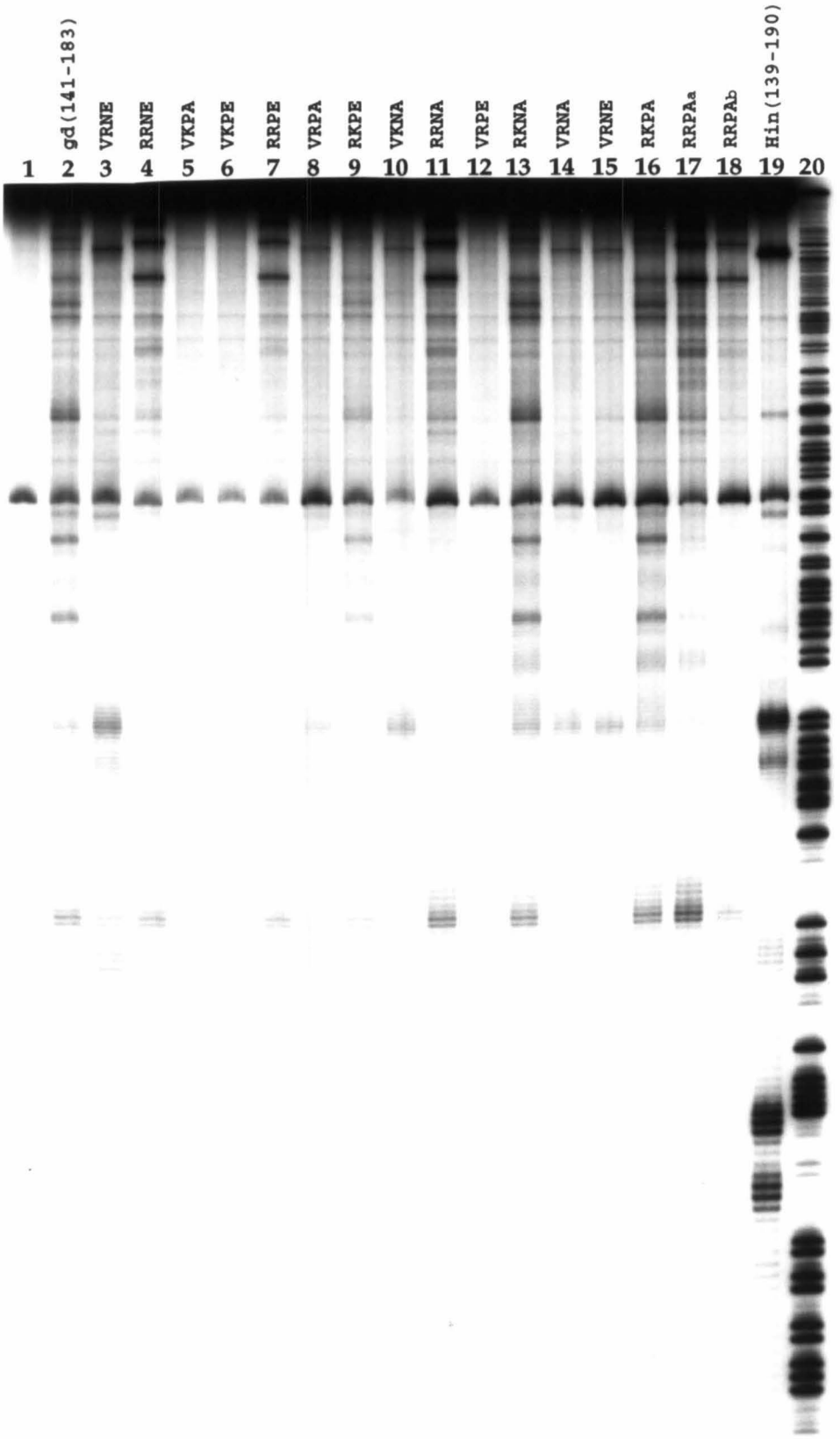

1
gd(141-183)
VRNE
RRNE
VKPA
VKPE
RRPE
VRPA
RKPE
VKNA
RRNA
VRPE
RKNA
VRNA
VRNE
RKPA
RRPAa
RRPAb
Hin(139-190)
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Table 5. Protein Binding to the XBA I/EcoRI fragment from pMFB36, which contains the Hin Recombinase-binding sites Hix L and Hix R, based on affinity cleavage. The protein concentration is 2.0 μM. (- = no binding; + = weak binding; and ++ = strong binding)

	DNA Binding Sites					
Protein	HixL(L) TTCTTGAAAA	HixL(R) TTATCAAAAA	Tertiary[1] TTCTCATGGA	Secondary L[2] TTCTTCCTTA	Secondary R1/ HixR(L)[3] TTCTCCTTTA	HixR[4]
EDTA γδ(141-183)	-	-	+[3]	-	-	-
RKPE	-	-	+[3]	-	-	-
RKNA	-	-	+[3]	-	-	-
RKPA	-	-	+[3]	-	-	-
RRNE	-	-	+[3]	-	-	-
RRPE	-	-	+[3]	-	-	-
RRPAa	-	-	+[3]	-	-	-
RRPAb	-	-	+[3]	-	-	-
RRNA	-	-	+[3]	-	-	-
VKNE	-	-	-	-	-	-
VKPA	-	-	-	-	-	-
VKPE	-	-	-	-	-	-
VRPE	-	-	-	-	-	-
VKNA	-	-	-	-	-	-
VRNA	-	-	-	+	-	+
VRNE	+	+	-	+	++	++
VRPA	-	-	-	-	+	+
EDTA-Hin(139-140)	++	++	+	++	++	++

[1] Binding sequences for the sites Tertiary, Secondary L and Secondary R are based on affinity cleavage and MPE footprinting data contained in a thesis by James P. Sluka (Caltech, 1988). [2] The binding sequences determined for Secondary R are identical to HixR(L). [3] Affinity cleavage scored at this site occurred within 3 bases of cleavage scored for Hin(139-190) and may be a result of recognition of a 5-CGCTTTTCT-3 sequence that has monology to the 5' end of site IL from *Res*. [4] Determination of exact binding site was not possible because of limits of gel resolution.

The final category of proteins is proteins possessing some binding specificity for Hin recombinase-binding sites. Three proteins fall into this category, VRNA, VRNE, and VRPA. It is important to note that these are the same three proteins that possessed Hin-like binding characteristics when the binding of the hybrid proteins to *res* was examined. Generally, while these proteins are capable of recognizing Hin recombinase-binding sites, they do it at a lower affinity than the native Hin recombinase. Additionally, they fail to recognize all the sites that are recognized by Hin recombinase. The protein VRNA shows weak affinity for the Hin recombinase secondary L and some binding in the region of Hix R. It is important to note that because of limitations in gel resolution, binding in the region of Hix R was not fully characterized, so an assignment of a binding site is not possible; therefore, these data should not be weighed as heavily as other binding sites. The protein VRPA, in which the amino acid composition of the contact face of the putative-recognition helix is identical to that of Hin recombinase. It is capable of binding weakly the Hin site secondary, R, which is equivalent to Hix R (L). This protein also demonstrates a weak binding in the region of Hix R. The remaining protein, VRNE, is more Hin-like in its binding characteristics. It possesses a weak affinity for the sites Hix L (L), Hix L (R), and secondary L. It shows a strong affinity for the site secondary R, which is equivalent to Hix R (L). In light of this, it is not surprising that the protein also shows a strong affinity in the region of Hix R.

Conclusions. While the helix-switch experiment between Hin recombinase and γδ resolvase failed to produce a hybrid protein that converted γδ resolvase into a protein with solely Hin recombinase DNA binding specificity, it did provide

useful insights into the DNA recognition of both of these proteins. Generally, it was found that DNA-binding specificity is insensitive to the amino acids found at positions 180 and 181 in the recognition helix; this is proximal to the carboxy end of the recognition helix. Positions 172 and 177, which are proximal to the amino end of the recognition helix, however, had a significant impact on the DNA-binding specificity of the proteins. This is consistent with the results obtained by Wharton and Ptashne, 1985, in their studies examining a helix-switch experiment between 434 repressor and P22 repressor. Generally speaking, arginine at position 172 is required to impart γδ resolvase-binding specificity on the hybrid proteins. The substitution of the lysine at position 177 with arginine will create a class of proteins that are capable of recognizing sequences in addition to those bound by native γδ resolvase. Placing a valine at position 172, followed by a lysine at position 177, tends to create proteins that show weak binding specificity for any given site. Placing a valine at position 172, followed by an arginine at position 177, tends to create proteins with more "Hin-like" DNA-binding specificity. The only exception to this is the protein VRPE, which expresses no binding specificity at all. It is possible that this combination of amino acids disrupts the secondary structure of the helix-turn-helix motif in such a way that it is no longer competent to bind DNA.

The failure of this helix switch to produce a protein with completely altered DNA binding specificity is not totally unexpected. The experiments conducted by Wharton and Ptashne, 1985, and Wharton, et al., 1984, which successfully produced hybrid proteins with fully altered DNA-binding specificity, were performed using proteins that showed some cross-reactivity. Our experiments used Hin recombinase and γδ resolvase, which were demonstrated to possess no cross-reactivity.

The availability of additional information has refined our view of helix-turn-helix proteins from the time when these experiments were initially conceived. Examination of the high-resolution, crystallographic data available for helix-turn-helix protein-DNA complexes (Anderson, et al., 1987; Aggarwal, et al., 1988; Otwinoski, et al., 1988; Jordan and Pabo, 1989) in addition to low-resolution, crystallographic structures or NMR studies (Wolberger, et al., 1988; Anderson, et al., 1981; McKay and Steitz, 1981; Kaptain, et al., 1985) have demonstrated that the helix-turn-helix units of the different DNA-binding proteins have extremely similar structures. Typically, the α-carbon backbones from the helix-turn-helix units can be superimposed with an average deviation of less than 1.1 Å per α-carbon (Steitz, et al., 1982; Ohlendorf, et al., 1983). The general structure of the protein-DNA co-crystals also possess some degree of similarity. Generally, the recognition helix of the helix-turn-helix unit is positioned in the major groove with the first helix of the unit lying across the major groove and the recognition helix.

Closer examination of the structures, however, discloses interesting differences in the way the proteins utilize the helix-turn-helix unit to recognize DNA. Several different mechanisms have been identified by which helix-turn-helix proteins recognize DNA. There is direct hydrogen bonding, as in the case of λ-repressor or engrailed, direct hydrogen bonding in combination with alterations in DNA conformation, as in the case of 434 repressor and CAP, and water-mediated hydrogen bonding in combination with changes in structural conformation of both the protein and DNA, as in the case of Trp repressor. It has become apparent that even if the same mechanism of sequence-specific recognition is used, for example, direct hydrogen bonding in the case of λ-

repressor and engrailed, variations in the relative positioning of the protein and DNA can significantly alter the amino acid side chains that are positioned to interact with the DNA.

The result of these "subtle differences" is that it is difficult to use the contacts for one helix-turn-helix protein to predict the pattern of side chain base interactions for another helix-turn-helix protein (Sauer, et al., 1990). There is, in effect, no simple recognition code that determines which side chains of a helix-turn-helix will interact with a particular base pair (Sauer, et al., 1990). In light of this it is easy to understand how the amino acid substitutions between Hin and γδ resolvase might fail to produce a protein of completely altered DNA-binding specificity.

Experimental

General. Nuclear Magnetic Resonance spectra were recorded on a JEOL GX 400 MHz NMR and are reported as parts per million downfield from tetramethylsilane. Ultraviolet-visible spectra were recorded on a Beckman model 25 spectrophotometer, a Perkin Elmer λ 4C spectrophotometer, or a Hewlett Packard model 8451A spectrophotometer. Mass spectrometry was performed by the Regional Mass Spec Facility at the University of California at Riverside. Laser densitometry of gel autoradiograms was performed on an LKB Ultrascan XL densitometer. Tribenzyl EDTA GABA (BEG) was synthesized using published protocols (Sluka, 1988; Sluka, et al., 1990).

Synthesis of t-Boc-γ–aminobuteric acid: t-Boc-GABA. 20.6 g (200 mmoles) of γ-aminobuteric acid was dissolved in 800 ml of a 2:1:1 mixture of

dioxane:water:1.0 N NaOH. The mixture was cooled to 0 °C in an ice bath. 48.0 g (250 mmoles) of di-t-butyldicarbonate were dissolved in 20 ml of dioxane, and this solution was added to the reaction mixture over a period of 30 min. Following the addition, the reaction was stirred for 60 min at 0 °C. Then it was allowed to warm to room temperature over a period of 90 min. The dioxane was then removed from the reaction under reduced pressure, using a rotary evaporator. The remaining aqueous solution was placed in an ice bath and cooled to 0 °C. 400 ml of ethylacetate were added and the water:ethylacetate mixture was mixed vigorously, using a vortex mixer. To this vigorously stirring solution, enough 2 M $KHSO_4$ was added to adjust the pH of the aqueous phase to 1.6. The ethylacetate was then removed, and the aqueous layer was extracted two additional times with ethylacetate (300 ml). The organic layers were pooled, extracted once with water (400 ml) and once with a saturated NaCl solution (400 ml). The ethylacetate solution was concentrated to a volume of approximately 50 ml, using a rotary evaporator. The resulting clear liquid was kept at 4 °C overnight. The following morning, the flask was full of white crystals and a small portion of liquid. The solution and crystals were taken up in approximately 200 ml of cold hexane and filtered through a glass frit. The resulting crystals were washed two times with hexane (50 ml). After drying under high vacuum, 33.1 g of t-Boc-GABA in an 82% yield were obtained. ^1H NMR (400 MHz CD CL_3) δ 5.61 (bd, s, 1H), 4.69 (bd, s, 1H), 3.20 (m, 2H), 2.40 (t, 2H, J = 6 Hz), 1.85 (m, 2H), 1.40 (s, 9H).

Synthesis of TricyclohexalEDTA(TCE). 20.56 g (68 mmoles) of EDTA-tetra acid and 4 ml of H_2SO_4 were added to a flask containing 145ml of cyclohexanol. The resulting mixture was stirred at 105 °C for 12h, at which time most of the EDTA had dissolved. The mixture was cooled to room temperature, dumped

into 300 ml of water and adjusted to pH 4 with the addition of saturated $NaHCl_3$. The resulting aqueous solution was extracted three times with diethyl ether (500 ml). The ether extracts were pooled and the diethyl ether was removed under vacuum with the use of a rotary evaporator. This afforded us with approximately 200 ml of a solution that contained traces of ether and residual cyclohexanol. The remaining cyclohexanol and any traces of ether were removed under vacuum distillation at 80 °C. The resulting yellow gum was placed on the vac line for 3 days to remove any additional cyclohexanol. This material was then purified via silica gel chromatography (methanol:CH_2Cl_2 3.1:100). 8.65 g of TCE were obtained as a yellow oil in a 23.6% yield. 1H NMR (400 MHz DMSO-d_6) δ 4.65 (m, 3H), 3.50 (s, 4H), 3.49 (s, 2H), 3.36 (s, 2H), 2.71 (s, 4H), 1.21-1.78 (m, 30H). Mass Spectroscopy +ve FAB CH_2Cl_2/NBA Observed MH^+ at M/Z 539.33 (539.33 calculated for C_{28} H_{47} O_8 N_2).

Peptide Synthesis. The synthesis of $\gamma\delta$ resolvase and the related mutant peptides was carried out, using a combination of manual peptide synthesis procedures and automated synthesis performed on an ABI model 430A Peptide Synthesizer. Manual peptide synthesis was carried out in a 20 ml vessel fitted with a coarse glass frit, as described by Kent (1988). Automated syntheses were performed on an ABI 430A Synthesizer, modified by the removal of in-line filters to the top and bottom of the reaction vessel, using a 20-ml teflon/KLF reaction vessel (Kent, et al., 1984; Kent and Clark-Lewis, 1985). The automated synthetic protocols used were developed at the California Institute of Technology (Kent and Clark Lewis, 1985), and the manual protocols used were a modification of those described by Kent (Kent, 1988). All protected amino acid derivatives were purchased from Peninsula Laboratories, with the

exception of Boc-L-his (DNP), which was obtained from Fluka and N-α-t-Boc-N-e-Fnoc-L-lysine, which was obtained from Chemical Dynamics Corporation (S. Plainfield, NJ). PAN resin with N-α-t-Boc-L-asparagine was purchased from Applied Biosystems and p-BHA resin was purchased from United States Biochemical Corporation. Dichloromethane (DCM) methanol (MeOH) and diethyl ether were purchased from Mallincroft. Trifluoroacetic acid (TFA) was purchased from Halocarbons. 1,4-butandithiol and anisol were purchased from Aldrich. Dimethylformamide (DMF), diisopropylethylamine (DIEA), dicyclohexalcarbodiimide in dichloromethane (DCC), and N-hydroxybenzotriazol in DMF (HOBt) were pruchased from Applied Biosystems. All other chemicals were purchased from Aldrich. N-α-Boc-L-amino acids were used with the following side chain protecting groups, Arg (Tos), Asp (OBzl), Glu (OBzl), His (DNP), Lys (Cl-Z), Ser (Bzl), Trp (formyl), Thr (Bzl), and Tyr (Br-Z).

The automated synthesis of proteins was conducted on the model 430A as follows. Double couplings were performed for every amino acid. The tBoc groups were removed from the amino or the resin-bound amino acid with 100% TFA. Following tBoc removal, the deprotected peptide resin was neutralized by 10% diisopropyl ethyl amine in DMF. Amino acids, with the exception of Asn, Gln, and Arg, were coupled to the free amino group as symmetric anhydride. Asn, Gln, and Arg were coupled as HOBt esters. In the first coupling, the symmetric anhydride was formed in an activating vessel with DCC and DCF. The dicyclohexalurea was removed by filtering the solution into the concentrating vessel; the volume of dichloromethane was decreased by approximately 50% and replaced with DMF. This solution was then transferred to the reaction vessel, which contained resin that had been

previously deprotected and neutralized. Prior to the second coupling, the resin was rinsed with dichloromethane and DMF and neutralized a second time with 10% diisopropyl ethyl amine in DMF. The symmetric anhydride was formed as described before, filtered into the concentrator and then added directly to the reaction vessel. DMF was added at a midpoint through the coupling cycle. In the case of amino acids that were coupled as HOBt esters, initially DCC and HOBt were added to the activator; they were allowed to mix for a period of approximately 15 min, at which time the amino acid, which had been dissolved in DMF, was added to the solution. Following activation, the solution was transferred into the concentrator, and the dichloromethane was removed and replaced with DMF. This solution was transferred to the reaction vessel, which contained previously deprotected and neutralized peptide resin. Prior to the second HOBt ester-mediated coupling, the peptide resin was rinsed with DMF and neutralized with DIEA in DMF. The activation of the HOBt ester was the same for the first coupling. At the end of the second coupling cycle, a sample resin was taken from the reaction vessel, and the coupling yield for the double coupling was determined by quantiative ninhydrin monitoring (Sarin, et al., 1981).

The coupling cycles used for manual synthesis of peptide are described below. Initially, the resin is washed with approximately 10 vessel volumes of dichloromethane. The next step is the removal of the tBoc group. This is accomplished by mixing the resin with 65% TFA/DCM. Initially, the reaction vessel is filled with 65% TFA/DCM and mixed for 1 min. The vessel is then drained and refilled with 65% TFA/DCM and mixed for 15 min. At this time, the TFA/DCM is drained from the reaction vessel, and the vessel is rinsed with 10 volumes of dichloromethane. 10% DIEA/DCM is then added to the reaction

vessel, and the vessel is mixed for 1 min. The DIEA/DCM is drained from the vessel; the resin is washed with 4 vessel volumes of dichloromethane, and refilled with 10% DIEA/DCM and mixed for another minute. The DIEA/DCM solution is then removed, and the vessel and resin are rinsed with 10 volumes of dichloromethane. At this point the resin is ready for the first coupling cycle.

Two coupling cycles are used, a symmetric anhydride cycle for all residues except Asn, Gln, and Arg, and a HOBt ester-coupling cycle, which is used for Asn, Gln, and Arg. For amino acids that are coupled with the symmetric anhydride procedure, the following is done. Prior to rinsing the resin with dichloromethane, the desired tBoc amino acid is dissolved in a separate vessel in a minimum of dichloromethane. To this solution half an equivalent of DCC in dichloromethane is added. This solution is allowed to set at room temperature for approximately 15-20 min, the typical time for deprotection and neutralization of the resin. Typically, a precipitate of dicyclohexalurea forms in the vessel containing the amino acid and DCC after about 15 min. After the resin is neutralized, the dicyclohexalurea is removed by filtration, and the symmetric anhydride is added to the peptide resin complex. After 30 min a sample of resin is taken and a yield for the coupling is determined by quantitative ninhydrin monitoring (Sarin, et al., 1981). If the coupling yield is satisfactory, > 99.7% at the beginning of the synthesis falling off to approximately 99% near the end of a 43 amino acid peptide, the resin and reaction vessel will be rinsed with dichloromethane, and the cycle will be restarted with tBoc removal and activation of the next amino acid. If the coupling yield is not satisfactory, the following steps are taken. The same amino acid is dissolved in a minimum of dichloromethane and half an equivalent of DCC in dichloromethane is added. This mixture is allowed to sit

at room temperature for approximately 15 min, or until dicyclohexalurea precipitates. The reaction vessel is then drained and the resin and vessel are rinsed with 5 volumes of DMF. A solution of 10% DIEA in DMF is then added to the reaction vessel, and it is mixed for a period of 1 min. The reaction vessel is then drained, and the resin and vessel are rinsed with 5 volumes of DMF. The dicyclohexalurea is then removed from the symmetric anhydride solution by filtration, and the symmetric anhydride is added to the reaction vessel, and the vessel is topped with DMF. The vessel is mixed for 45 min, at which time a sample of resin is taken for quantitative anhydrin monitoring. If the yield is satisfactory, the cycle is restarted with tBoc removal; if the yield is not satisfactory, a third coupling may be performed by repeating the steps used for the double coupling.

In the case of amino acids that are coupled by HOBt esters, the following procedure is used. 30 min prior to the start of deprotection of the peptide resin complex, one equivalent of DCC in dichloromethane and one equivalent of HOBt in DMF are mixed together. At this time one equivalent of the desired amino acid is dissolved in a minimum volume of DMF. After 10 min, the DCC/HOBt mixture is added to the dissolved amino acid, and this solution is allowed to sit at room temperature for approximately 40 min. At this time the peptide resin complex will have completed deprotection and neutralization. If a precipitate has formed in the HOBt/DCC/amino acid mixture, it is removed by filtration prior to addition of the mixture to the vessel containing deprotected, neutralized peptide resin. Once the HOBt amino acid complex has been added to the resin, the vessel is topped off with DMF, and the contents are mixed for 60 min. At this time a sample of resin is taken for quantitative ninhydrin monitoring. If the coupling yield is acceptable, a new

cycle is started with deprotection of the peptide resin complex; if the yield is unacceptable, a second HOBt ester coupling is performed. The DCC and HOBt solutions are mixed together and allowed to react for 10 min, at which time the amino acid dissolved in DMF is added. This mixture is allowed to sit at room temperature for 40 min. When the end of the 40 min time period is approaching, the reaction vessel is drained, and the peptide resin and reaction vessel are rinsed with 5 volumes of DMF. The vessel is then filled with 10% DIEA in DMF and mixed for 1 min. The vessel is drained and the peptide resin and vessel are rinsed with 5 volumes of DMF. The HOBt/DCC/amino acid solution is filtered if needed and added to the reaction vessel. After 2h a sample of resin is taken for quantitative ninhydrin monitoring. If the coupling yield is satisfactory, a new cycle will be started by tBoc removal and resin neutralization, followed by coupling of the next amino acid; if the yield is unsatisfactory, a third coupling can be performed using an HOBt ester (in the case of Arg, additional couplings can be performed using the symmetric anhydride method).

BEG and TCE Coupling to the Amino Terminus of Resin Bound Peptides. Approximately 200 mg of resin bound peptide (this equals approximately 20 μmoles of peptide) were placed in a 12X80 mm reactor. The resin and vessel were washed with 5 volumes of dichloromethane, the terminal tBoc protecting group was removed, and the resin was neutralized using the previously described protocols. 30 min prior to the deprotection and neutralization of the peptide, 250 μmoles of HOBt in DMF and 250 μmoles of DCC in DCM were mixed together. This mixture was allowed to sit at room temperature for 10 min, at which time 250 μmoles of either TCE or BEG dissolved in DMF were added (total volume for the solution is approximately 1.2 ml). After 20 min, the

deprotection and neutralization of the resin-bound peptide was initiated. After the peptide was neutralized, the solution containing the activated TCE or BEG was added to the reactor. The reactor was topped off with DMF and mixed for 2h. At this time a resin sample was taken for quantitative ninhydrin monitoring. If the coupling yield was < 99%, an additional 250 μmoles of TCE or BEG were activated as before, and the resin-bound peptide was reneutralized using standard procedures. The activated BEG or TCE was added to the reactor; the reactor was topped with DMF, and mixed for 2h. A sample of resin was taken for quantative ninhydrin analysis, and no additional couplings were performed.

Coupling of TCE to the ε–NH_2 Side Chain of a Lysine Residue at the Carboxy Terminus of a Resin-Bound Peptide. N-α-tBoc-Ne-FMOC-L-Lysine was coupled to a growing peptide, using the symmetric anhydride protocol with the substitution of DMF for DCM. Following this coupling, the resin-bound peptide and vessel were washed with 10 volumes of DMF. The Ne-FMOC group was removed from the lysine side chain by treating the resin-bound peptide with 20% piperidine in DMF (a 1 min treatment was followed by a 15 min treatment). Following the removal of the Ne-FMOC group, the resin-bound peptide and vessel were rinsed with 5 volumes of DMF. A solution of DMF and DCM containing 4 equivalents of activated TCE-HOBt ester was added to the reactor, and the reactor was topped with DMF. The reactor was mixed for 2h at room temperature, at which time a sample of resin was taken for quantitative ninhydrin analysis. In all cases the coupling yield was >99.8%, so no additional coupling was performed. The activated TCE-HOBt ester was formed by first mixing equal molar amounts of DCC dissolved in DCM and HOBt dissolved in DMF. The DCC/HOBt mixture was allowed to sit for 10

min at room temperature, at which time an equal molar amount of TCE dissolved in DMF was added, and the resulting mixture was allowed to sit for 40 min at room temperature.

Peptide Deprotection and Purification. 200 mg of resin bound peptide (approximately 20 μmoles) were placed in a 20x80 mm reaction vessel. The resin-bound peptide was rinsed with 5 volumes of DCM. The histidine-protecting group, dinitrilephenyl, was removed at 25 °C, using a 20% β-mercaptaethanol and 10% DIEA solution in DMF. The reactor was filled with this solution and mixed for 30 min, at which time it was drained and refilled and mixed for another 30 min. Following this treatment, the resin was rinsed with 5 volumes of DMF and 5 volumes of DCM. Next, the N-α-tBoc group was removed with a 1 min treatment with 65% TFA/DCM solution followed by a 15 min treatment with the 65% TFA/DCM solution. The resin was then rinsed with 10 vessel volumes of DCM and neutralized by two 1 min treatments with 10% DIEA/DCM. Following the neutralization with the DIEA solution, the resin-bound peptide and reaction vessel were rinsed with 10 volumes of dichloromethane, 5 volumes of 1:1 DCM:MeOH, 10 volumes of methanol and dried for 15-30 min with an aspirator. All other side chain protecting groups were removed, and the peptide-resin bond was cleaved using anhydrous HF (10 ml) in the presence of anasil (1.0 ml) and 1,4-butanedithiol (0.5 ml) as scavengers. The HF, and resin-bound peptide were mixed together vigorously for 60 min at 0 °C. The HF was removed under vacuum. The crude protein was precipitated with diethyl ether, collected on a fritted funnel, dissolved with 5% acetic acid, and washed through, leaving only the resin on the frit.

A small sample of the 5% acetic acid solution containing the peptide was removed, filtered, and subjected to analytical HPLC (Brownlee, 25 cm X 4.6 mm C_8 column, 0-60% acetonitrile, 0.1% trifluoroacetic acid over 60 min). The remaining solution was frozen and lyophilized. The residual DMP groups were removed from the resulting crude peptide by treatment in 4 M guanadidium HCL, 40 mM tris, pH 8.5, and 20% β-mercaptaethanol for 1 h at 50 °C (Kent, 1988). This solution was filtered and injected directly onto a semi-preparative C_8 HPLC column (25X1 cm). The column was flushed with water containing 0.1% trifluoroacetic acid until the guanadidium and β-mercaptaethanol had eluted. A gradient from 0-60% acetonitryl, 0.1% trifluoroacetic acid was run over 240 min and fractions were collected. Pure peptide fractions were identified by analytical HPLC. Sequencing by Edman degradation and amino acid analysis of the purified peptide were used to confirm the sequence composition. Protein concentrations were assayed based on a calculated OD_{275} (e = 6,950 based on 1 tyrosine and 1 tryptophan). Purified proteins were lyophilized for storage.

Labeling of the Hind III/Sal I Restriction Fragment from Plasmid PRW 80. Plasmid PRW 80 (Abdel-Meguid, et al., 1984) containing two copies of *Res* was digested with the restriction endonuclease Hind III to produce two fragments, 30 and 5 kilobase pairs in size. Labeling at the 3' end was accomplished using α-^{32}P-dATP with the Klinow fragment of DNA polymerase I. 5'-end labeling was accomplished by dephosphorylation of the 5'-DNA termini with calf alkaline phosphotase, followed by rephosphylation of the hydroxyl ends using γ-^{32}P-ATP and T4 polynucleotide kinase. Following end labeling, the fragments were cleaved with the restriction endonuclease Sal I. This yielded three different sized end-labeled fragments. The smallest fragment, 240 bp long, contained the

intact *Res* site. This fragment was isolated by non-denaturing polyacrylamide gel electrophoresis, using a 15 cm X 2 mm 5% polyacrylamide gel.

Labeling of the XbaI/EcoRI Restriction Fragment from pMFB36. The plasmid pMFB36 (Bruist, et al., 1987) was restricted with the endonuclease XbaI. Labeling at the 3'-end was accomplished with α-^{32}P-dATP, using the Klinal fragment of DNA polymerase I. The 5'-end was labeled with ^{32}P by dephosphorylation with calf alkaline phosphotase, followed by treatment with γ-^{32}P-ATP and P4 polynucleotide kinase. Cleavage with the restriction endonuclease EcoRI yielded two fragments. The smaller 557 bp fragment was isolated by polyacrylamide gel electrophoresis, using a 10 cm x 2 mm 5% non-denaturing gel.

MPE Footprinting Reactions. MPE footprinting reactions were performed in a buffer containing 20 mM NaCl, 100 μM in base pair calf thymus DNA, 20 mM tris, pH 8.0, and ^{32}P-end-labeled DNA. The reactions contained MPE•Fe(II) (10μM), DTT (5 mM), and γδ (144-183) (2.0 μM). For footprinting reactions, γδ (141-183) was added to the buffer and allowed to equilibrate with the DNA for 15 min. MPE•Fe(II) was subsequently added and allowed to equilibrate for 5 additional minutes. The reaction was then initiated by the addtion of DTT. After 10 min at 25 °C, the reaction was stopped by ethanol precipitation. Comparisons were run between MPE cleavage reactions in the presence of absence of γδ (141-183).

Affinity-Cleavage Reactions Using EDTA Tagged γδ (141-183) Peptides and Related Mutants. Affinity-cleavage reactions were performed in a buffer containing 20 mM tris, pH 8.0, 20 mM NaCl, 100 μM in base pair calf thymus DNA, ^{32}P-end-

labeled DNA, 1 mM dithiothreatol (DTT) and 0.5, 2.0, or 10.0 μM Fe(II)•EDTA protein. The proteins were allowed to equilibrate with the DNA for 15 min at 25 °C.; cleavage was then initiated with the addition of DTT and allowed to proceed for 45 min at 25 °C. The reaction was stopped by ethanol precipitation. The ^{32}P-labeled DNA products were analyzed by 8% denaturing polyacrylamide gel electrophoresis followed by autoradiography. Densitometric analysis of the gel autoradiogram and comparison of the individual lanes with sequenced marker lanes allowed assignment of DNA cleavage to nucleotide resolution.

Sequence-Specific Marker Lanes. Sequence-specific marker lanes were used to identify the location of affinity-cleavage sites on the labeled restriction fragments. Two types of marker lanes were used. A-specific cleavage lanes were performed using the protocol of Iverson and Dervan (Iverson and Dervan, 1987). G-specific cleavage lanes were produced using the original protocol of Maxam and Gilbert (Maxam and Gilbert, 1980).

References

Abdel-Meguid, S. S., N. D. F. Grindley, et al. (1984). Proc. Natl. Acad. Sci. U. S. A. **81**: 2001.

Abdel-Meguid, S. S., H. M. K. Murthy, et al. (1986). J. Biol. Chem. **261**: 15934.

Aggarwal, A. K., D. W. Rodgers, et al. (1988). Science **242**: 99.

Anderson, J. E., M. Ptashne, et al. (1985). Nature **316**: 596.

Anderson, J. E., M. Ptashne, et al. (1987). Nature **326**: 846.

Anderson, W. F., D. H. Ohlendorf, et al. (1981). Nature **290**: 754.

Beato, M. (1989). Cell **56**: 335.

Berg, J. N., R. Boelens, et al. (1989). Biochemistry **28**: 9826.

Boelens, R., R. M. Scheek, et al. (1987). J. Mol. Biol. **193**:213

Breg, J. N., O. J. H. J. Van, et al. (1990). Nature (London) **346**: 586-9.

Brennan, R. G. (1991). Curr. Opin. Struct. Biol. **2**: 100.

Brennan, R. G. and B. W. Matthews (1989). J. Biol. Chem. **264**: 1903.

Brennan, R. G. and B. W. Matthews (1989). Trends Biochem. Sci. **14**: 286.

Bruist, M. F., S. J. Horvath, et al. (1987). Science **235**: 777.

Buxton, R. S. and L. S. Drury (1983). Fems Microbiol. Lett. **112**: 1765.

Castell, S. E., S. L. Jordan, et al. (1986). Nucleic Acids Res **14**: 7213.

Dervan, P. B. (1986). Science, **232**: 464.

Dervan, P. B. (1991). Methods Enzymol **208**: 497.

Diakun, G. P., L. Fairall, et al. (1986). Nature **324**: 698.

Dickerson, R. E. (1983). J. Molec. Biol. **166**: 419.

Dickerson, R. E. and H. R. Drew (1981). J. Molec. Biol. **149**: 761.

Drew, H. R. and A. A. Travers (1984). Cell **37**: 491.

Droge, P., G. F. Hatfull, et al. (1990). Proc. Natl. Acad. Sci. U. S. A. **87**: 5336.

Evans, R. M. and S. M. Hollenberg (1988). Cell **52**: 1.

Falvey, E. and N. D. F. Grindley (1987). Embo J **6**: 815.

Falvey, E., G. F. Hatfull, et al. (1988). Nature **332**: 861.

Frederick, C. A., J. Grable, et al. (1984). Nature **309**: 327.

Galas, D. J. and A. Schmitz (1978). Nucleic Acids Res. **5**: 3157.

Garnier, J., J. Osguthorpe, et al. (1978). J. Mol. Biol. **120**: 97.

Glasgow, A. C., M. F. Bruist, et al. (1989). J. Biol. Chem. **264**: 10072.

Graham, K. S. and P. B. Dervan (1990). J. Biol. Chem. **265**(27): 16534.

Griffin, L. C. and P. B. Dervan (1989). Science, **245**(4921): 967

Grindley, N. D. F and R. R. Reed (1983). Cell, **32**: 3

Grindley, N. D. F., M. R. Lauth, et al. (1982). Cell, **30**: 19

Halford, S. E. (1986). Biochem. Soc. Trans. **14**: 216.

Harrison, S. C. (1991). Nature **353**: 715.

Harrison, S. C. and A. K. Aggarwal (1990). Ann. Rev. Biochem. **59**: 933.

Hatfull, G. F. and N. D. F. Grindley (1986). Proc. Natl. Acad. Sci. U. S. A. **83**: 5429.

Hatfull, G. F., S. M. Noble, et al. (1987). Cell, **49**: 103.

Hatfull, G. F., J. J. Salvo, et al. (1988). Symp. Soc. Gen. Microbiol. **43**: 149.

Hatfull, G. F., M. R. Sanderson, et al. (1989). J. Mol. Biol. **208**: 661.

Hughes, R. E., G. F. Hatfull, et al. (1990). Cell **63**: 1331.

Ishizaki, K. and E. Ohtsubo (1985). Mgg, Mol. Gen. Genet. **199**: 388.

Iverson, B. L. and P. B. Dervan (1987). Nucleic Acids Res **15**: 7823.

Iverson, B. L. and P. B. Dervan (1987). J. Am. Chem. Soc. **109**: 1241.

Jordan, R. S. and C. O. Pabo (1988). Science **242**: 893.

Kaptein, R., E. R. P. Zuiderweg, et al. (1985). J. Molec. Biol. **182**: 179.

Kaptein, R. (1991) Curr. Opin. Struct. Biol., **2**, 109.

Kent, S. B. H. (1988). Ann. Rev. Biochem. **57**: 957.

Kent, S. B. H. and I. Clark-Lewis (1985). Synthetic Peptides in Biology and Medicine. Amsterdam, Elsevier. 29.

Kent, S. B. H., L. E. Hood, et al. (1984). Peptide Chemistry 1984, Proceedings of the Japanese Peptide Symposium, Osaka, Japan, Protein Research Foundation.

Kent, S. B. H., L. E. Hood, et al. (1984). Peptides 1984; Proceedings of the 18th European Peptide Symposium, Sweden, Almqvist & Wicksell.

Kent, S. B. H., K. F. Parker, et al. (1988). Peptides: Chemistry and Biology, Proceedings of the Tenth American Peptide Symposium, Leiden, The Netherlands, ESCOM.

Kissinger, C. R., B. Liu, et al. (1990). Cell, **63**: 579.

Kitts, P. A., L. S. Symington, et al. (1983). EMBO J. **2**: 1055.

Klevit, R. E. (1988). Science **241**: 1489.

Landschultz, W. H., P. F. Johnson, et al. (1988). Science **240**: 1759.

Lee, M. S., G. P. Gippert, et al. (1989). Science **245**: 635.

Lehming, N., J. Sartorius, et al. (1987). EMBO J. **6**: 3145.

Lewis, M., J. Wang, et al. (1985). in Biological Macromolecules and Assemblies, Vol. 2, New York, John Wiley and Sons.

Luisi, B. F. and P. B. Sigler (1990). Biochmica et Biophysica Acta **1048**: 113.

Mack, D. P. and P. B. Dervan (1990). J. Am. Chem. Soc. **112**: 4604.

Mack, D. P., B. L. Iverson, et al. (1988). J. Am. Chem. Soc. **110**: 7572.

Mack, D. P., J. P. Sluka, et al. (1990). Biochemistry **29**: 6561.

Matthews, B. W. (1988). Nature **335**: 294.

Matthews, B. W., D. H. Ohlendorf, et al. (1982). Proc. Natl. Acad. Sci., USA **79**: 1428.

Maxam, A. M. and W. Gilbert (1980). Methods Enzymol. **65**: 499.

McClarin, J. A., C. A. Frederick, et al. (1986). Science **234**: 1526.

McKay, D. B. and T. A. Steitz (1981). Nature **290**: 744.

McKay, D. B., I. T. Weber, et al. (1982). J. Biol. Chem. **257**: 9518.

Merrifield, R. B. (1969). Adv. Enzymol. **32**: 221.

Michiels, T., G. Cornelis, et al. (1987). J. Bacteriol. **169**: 624.

Miller, J., A. D. McLachlan, et al. (1985). EMBO J. **4**: 1609.

Mondragon, A., S. Subbiah, et al. (1989). J. Mol. Biol. **205**: 189.

Moser, H. E. and P. B. Dervan (1987). Science **238**: 645.

Newman, B. J. and N. D. F. Grindley (1984). Cell **38**: 463.

Oakley, M. G. and P. B. Dervan (1990). Science **248**: 847.

Ohlendorf, D. H., W. F. Anderson, et al. (1983). J. Molec. Biol. **169**: 757.

Ohlendorf, D. H., B. W. Matthews (1983) Ann. Rev. Biophys. Bioeng. **12**: 259.

Ollis, D. L. and S. W. White (1987). Chem. Rev. **87**: 981.

Otting, G., Y. Q. Qian, et al. (1990). Embo J **9**: 3085.

Otwinowski, Z., R. W. Schevitz, et al. (1988). Nature **335**: 321.

Pabo, C. O., A. K. Aggarwal, et al. (1990). Science **247**: 1210.

Pabo, C. O. and M. Lewis (1982). Nature **298**: 443.

Pabo, C. O. and R. T. Sauer (1984). Ann. Rev. Biochem. **53**: 293.

Parraga, G., S. J. Horvath,, et al. (1988) Science **241:** 1489.

Pavletich, N. P. and C. O. Pabo (1991). Science **252**(5007): 809.

Phillips, S. E. V. (1991). Curr. Opin. Struct. Biol. **1**: 89.

Qian, Y. Q., M. Billeter, et al. (1989). Cell **59**: 573.

Rafferty, J. B., W. S. Somers, et al. (1989). Nature **341**: 705.

Reed, R. R. (1981). Cell **25**: 713.

Reed, R. R. (1983). Methods Enzymol : **70**: 293.

Reed, R. R. and N. D. F. Grindley (1981). Cell **25**: 721.

Reed, R. R. and C. D. Moser (1984) Cold Spring Harbor Symp. Quant. Biol. **49:** 245.

Reed, R. R., G. I. Shibuya, et al. (1982). Nature **300**: 381.

Rimphanitchayakit, V. and N. D. F. Grindley (1990). Embo J **9**: 719.

Rimphanitchayakit, V., G. F. Hatfull, et al. (1989). Nucleic Acids Res **17**: 1035.

Rogowsky, P., S. E. Halford, et al. (1985). Embo J **4**: 2135.

Saenger, W. (1984). Principles of Nucleic Acid Structure. Springer-Verlag, New York.

Salvo, J. J. and N. D. F. Grindley (1987). Nucleic Acids Res **15**: 9771.

Salvo, J. J. and N. D. F. Grindley (1988). Embo J **7**: 3609.

Sanderson, M. R., P. S. Freemont, et al. (1990). Cell **63**: 1323.

Sarin, V. K., S. B. H. Kent, et al. (1981). Anal. Biochem. **117**: 147.

Sauer, R. T., S. R. Jordan, et al. (1990). Advances in Protein Chem. **40**: 1.

Schleif, R. (1988). Science **241**: 1182.

Schultz, P. G., J. S. Taylor, et al. (1982). J. Am. Chem. Soc. **104**: 6861.

Schultz, S. C., G. C. Shields, et al. (1991). Science **253**: 1001.

Seeman, N. C., J. M. Rosenberg, et al. (1976). Proc. Nat. Acad. Sci. USA **73**: 804.

Shin, J. A., R. H. Ebright, et al. (1991). Nucleic Acids Res **19**: 5233.

Siebenlist, U. and W. Gilbert (1980). Proc. Nat. Acad. Sci. USA **77**: 122.

Simon, M. I., J. Zieg, et al. (1980). Science **209**: 1370.

Sluka, J. P. (1988). Design, Synthesis, and Characterization of Sequence-Specific DNA Cleavage Agents. California Institute of Technology.

Sluka, J. P., J. H. Griffin, et al. (1990). J. Am. Chem. Soc. **112**: 6369.

Sluka, J. P., S. J. Horvath, et al. (1987). Science **238**: 1129.

Sluka, J. P., S. J. Horvath, et al. (1990). Biochemistry **29**: 6551.

Stark, W. M., N. D. F. Grindley, et al. (1991). Embo J **10**: 3541.

Steitz, T. A. (1990) Quart. Rev. Biophys **23:** 205.

Steitz, T. A., D. H. Ohelndorf, et al. (1982). Proc. Natl. Acad. Sci, USA **79**: 3097.

Suck, D., A. Lahm, et al. (1988). Nature **332**: 465.

Tam, J. P., T. W. Wong, et al. (1979). Tetrahedron Lett. : 4033.

Tanaka, I., D. Appelt, et al. (1984). Nature **310**: 376.

Tanaka, M., T. Yamamoto, et al. (1983). Mgg, Mol. Gen. Genet. **191**: 442.

Taylor, J. S., P. G. Schultz, et al. (1984). Tetrahedron **40**: 457.

Tullius, T. D. and B. A. Dombrowski (1985). Science **230**: 679.

Tullius, T. D., B. A. Dombrowski, et al. (1987). Methods in Enz. **155**: 537.

Vermote, C. L. A. and S. E. Halford (1992). Biochem. **31**: 6082.

Vermote, C. L. M., I. B. Vipond, et al. (1992). Biochem. **31**: 6089.

Vinson, C. R., P. B. Sigler, et al. (1989). Science **246**: 911.

Wells, R. G. and N. D. F. Grindley (1984). J. Mol. Biol. **179**: 667.

Wharton, R. P., E. L. Brown, et al. (1984). Cell **38**: 361.

Wharton, R. P. and M. Ptashne (1985). Nature **316**: 601.

White, S. W., K. Appeit, et al. (1989). Proteins **5**: 281.

Wishart, W. L., C. Machida, et al. (1983). Gene **24**: 99.

Wolberger, C., Y. Dong, et al. (1988). Nature **335**: 789.

Yang, C.-C. and H. W. Nash (1989). Cell **57**: 869.

Zhang, R.-G., A. Joachimiak, et al. (1987). Nature **327**: 591.

Zubay, G. and P. J. Doty (1959). J. Molec. Biol. **7**: 1.

Chapter II:
Design of a Sequence-Specific DNA Cleaving Metalloprotein
Ni(II) GGH-γδ (141-183).

Introduction

The design of sequence-specific DNA cleavage molecules typically involves the creation of a molecule with two functional domains, one that recognizes a specific sequence of DNA and another that cleaves, or cuts, the DNA (Dervan, 1986). This typically involves the covalent attachment of "non-natural" metal chelator to a naturally occurring DNA binding molecule (Dervan, 1986; Sigman and Chen, 1990). With this type of bifunctional molecule, a specific sequence of DNA is initially recognized, bound and, subsequently, DNA cleavage occurs via oxidation of the DNA sugar phosphate backbone mediated by the action of the metal chelator. Examples of this include attachment of EDTA to drugs such as distamycin (Youngquist and Dervan, 1985) oligonucleotides (Moser and Dervan, 1987; Iverson and Dervan, 1987b) and proteins (Sluka, et al., 1987). In other examples, different metal chelators, such as copper phenanthroline, have been successfully used to create sequence-specific DNA-cleaving molecules (Sigman and Chen, 1990). Until recently, a sequence-specific cleaving molecule comprised of totally natural components was not available (Mack, et al., 1988; Mack and Dervan, 1990).

The design of a sequence-specific DNA cleaving protein consisting wholly of naturally occurring α-amino acids requires the combination of a protein-DNA

binding domain with a protein-DNA cleaving domain. The first example of this was the attachment of the tripeptide Cu-binding domain, Gly-Gly-His (GGH), to the amino terminus of the DNA-binding domain from Hin recombinase (Hin 139-190) to create a new 55 residue protein GGH (Hin 139-190) (Mack, et al., 1988). This protein is capable of binding DNA at 413 base pair sites recognized by Hin recombinase. GGH (Hin 139-190) in the presence of Cu (II), hydrogen peroxide, and sodium ascorbate cleaved DNA predominantly at one of the four Hin binding sites (Mack, et al., 1988). A marked change in the sequence specificity and efficiency of DNA cleavage by GGH (Hin 139-190) occurred when the conditions were altered (Mack and Dervan, 1990). When the protein was loaded with Ni (II) instead of Cu(II), and cleavage was initiated by monoperoxyphthalic acid in place of the hdyrogen peroxide and ascorbate, the protein, Ni•GGH (Hin 139-190), cleaved the DNA at all four of the Hin binding sites. The nickel monoperoxyphthalic acid-mediated cleavage was more efficient and rapid than the Cu hydrogen peroxide, ascorbate-mediated cleavage DNA. The Ni-mediated cleavage occurred predominantly at a single deoxyribose position adjacent to each Hin binding site (Mack and Dervan, 1990). This suggested that cleavage was mediated by some non-diffusible species (Mack and Dervan, 1990).

Gly-Gly-His. GGH was initially designed as a mimic for the Cu (II) transport site found in serum albumin (Camerman, et al., 1976; Lau, et al., 1974). The histidine containing tripeptide was found to be capable of binding both Cu(II) and Ni(II) (Camerman, et al., 1976; Lau, et al., 1974; Bryce, et al., 1966). X-ray diffraction analysis showed GGH bound to Cu(II) in a square-planar complex with coordination from an imidizole nitrogen, two deprotonated peptide amide nitrogens, and the terminal amino group (Camerman, et al., 1976). Although

the crystal structure of Ni(II) bound to GGH is not available, several lines of evidence indicate that its complex is also square-planar, as shown schematically in Figure 1. First, the crystal structures of tetraglycine with Cu(II) or Ni(II) indicate that the metal ions were bound by peptide ligands in a similar fashion; both complexes are square-planar with coordination from the terminal amino groups and three deprotonated peptide amide nitrogens (Freeman and Taylor, 1965; Freeman, et al., 1968). Further support for the square-planar nature of the Ni(II) complex with GGH is provided by an analysis of the reaction kinetics of this complex with acids and triethyl tetra amine (Bannister, et al., 1982) and analysis of the cyclic voltammetry for the Ni(III)/(II) redox couple for this complex (Bossu and Margerum, 1977).

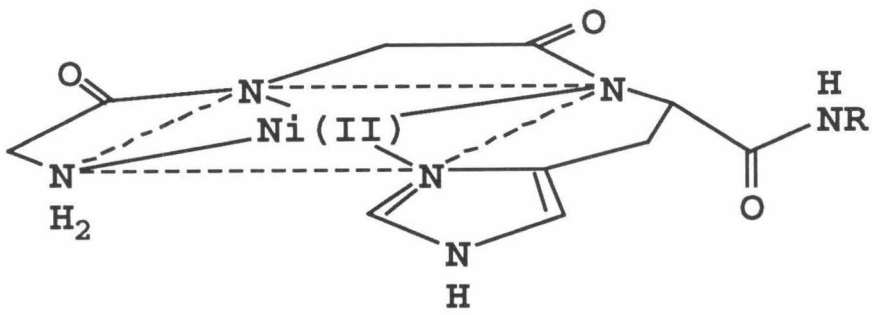


Figure 1 A model for the GGH ligand bound to Ni(II) in a square-planar complex.

The tripeptide, Gly-Gly-His, when complexed with Ni(II) was found to mediate several different types of reactions. The first type of reaction identified involved the auto oxidation of Ni (II) Gly-Gly-His with molecular oxygen (Sakurai and Nakahara, 1979; Paniago, et al., 1971). NMR studies of the products of this reaction demonstrated that the histidine in the GGH complex

had undergone an oxidative decarboxylation. The rate of this reaction was found to be greatly enhanced by the addition of oxidizing agents such as $IrCl_6^{2-}$ (IV) , H_2O_2 or $S_2O_8^{2-}$ (Sakurai and Nakahara, 1979). The Ni (II) complex of GGH was also found to mediate the epoxidation of olefins in the presence of iodosylbenzene (Mack, 1990). Reactions of this type have been observed in Ni(II) cyclam complexes (Kinneary, Albert, et al., 1988; Kinneary, Wagler, et al., 1988). Recently Ni(II) GGH and other Ni complexes were found to catalyze the oxidative cleavage of DNA (Chen, et al., 1991; Mueller, et al., 1992). These studies showed that Ni(II) GGH and other square-planar Ni complexes in the presence of oxidants such as monoperoxyphthalic acid would readily catalyze the oxidation and strand scission of DNA at looped-out, unpaired guanine residues. Interestingly, DNA bases engaged in normal Watson-Crick base pairing were found to be unreactive to this complex. Cleavage at the unpaired guanine residues was found to require a piperidine treatment and is thought to involve the formation of an 8 oxo guanine (Mueller, et al., 1992).

Studies report here the design and chemical synthesis of a new sequence-specific, DNA-cleaving 46 residue metalloprotein, GGH (γ,δ 141-183). This protein combines the DNA binding domain from γ,δ resolvase, (γ,δ 141-183), with the transition metal-binding domain, GGH (Figure 2).

Ni(II) GGH-γδ(141-183) Cleavage of DNA

Synthesis. The 46 residue protein was synthesized by stepwise, solid-phase methods (0.5 mM scale) with optimized manual protocols (Kent, 1988), on phenylacetamidomethyl(PAM) resin substituted with tBoc-asparagine. Each coupling cycle was monitored by quantitative ninhydride analysis (Sarin, et al.,

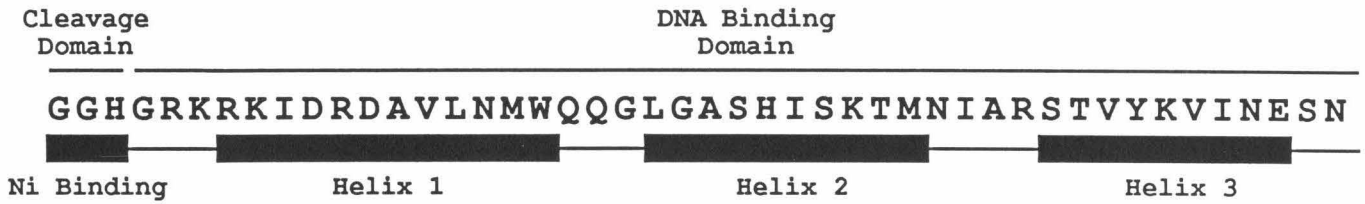
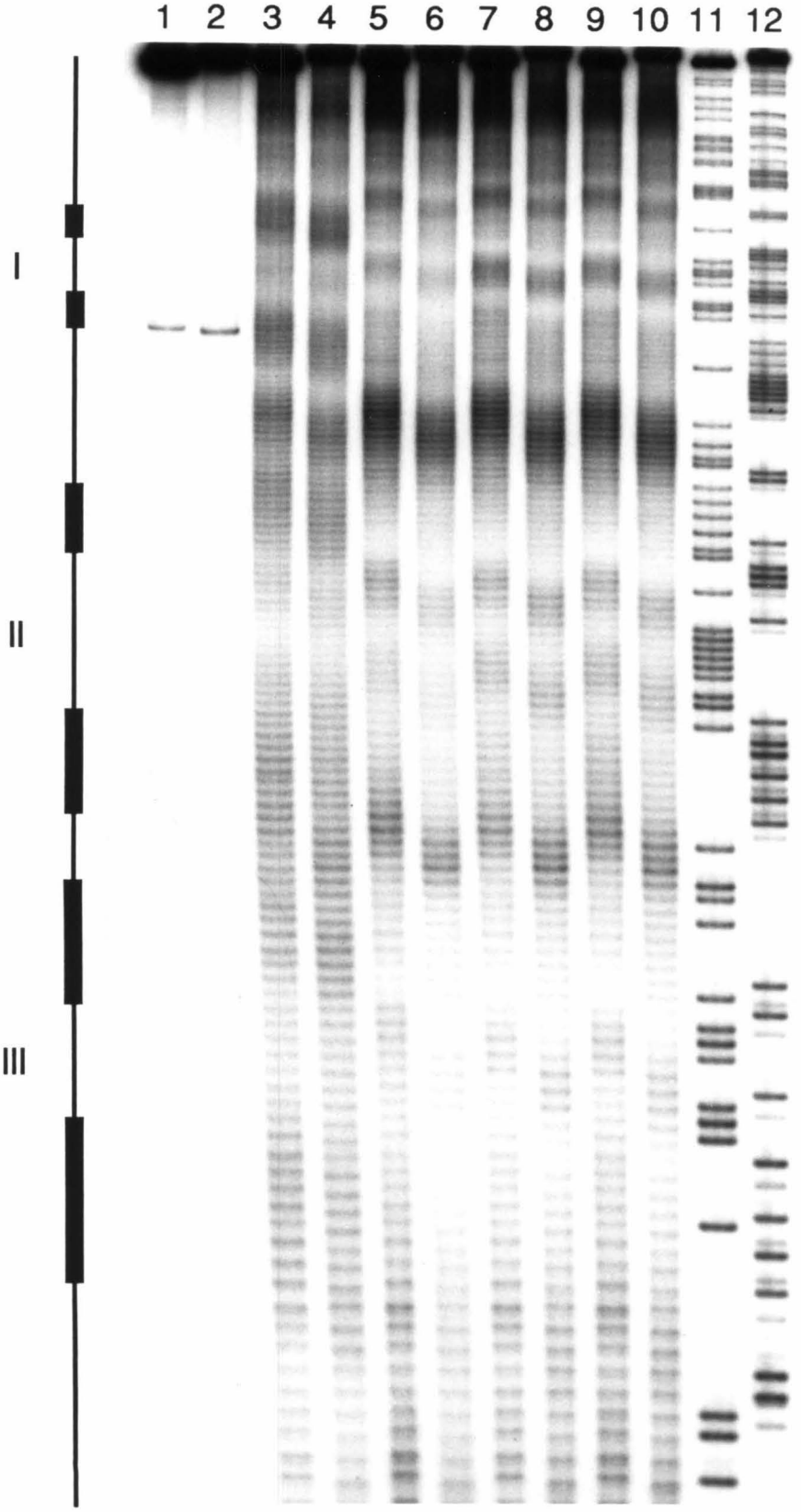


Figure 2 The amino acid sequence of the protein GGH γδ (141-183). The two functional domains, cleavage and DNA binding, are indicated. Bars indicate the site of the nickel-binding domain and the three helices that are predicted by the algorithm of Garnier et al. (1978) for γδ (141-183).

1981), and couplings were repeated until the yield was maximal for each step (99.7% early in the synthesis, decreasing to 98% near the end). The overall yield for the synthesis of GGH-γ,δ (141-183) was 73.7%. Approximately 200 mg of the resin bound peptide were deprotected and purified in the following manner. The histidine protecting group, dinitrophenyl (DNP) was removed at 25 °C using 20% 2-mercaptoethanol and 10% DIEA in DMF; this treatment was repeated twice for a period of 30 min. The final N-α-tBoc group was removed with 60% TFA in dichloromethane; the resin was neutralized with 10% DIEA in dichloromethane, rinsed, and dried. All the other side chain protecting groups were removed and the peptide-resin bond was cleaved, utilizing anhydrous HF in the presence of anisole and 1,4-butanedithiol as scavengers. The HF was removed under vacuum. The crude peptide was precipitated with diethyl ether, collected on a fritted funnel, dissolved with 5% acetic acid, and washed through the funnel, leaving the resin on the frit. This solution was frozen and lyophilized. The residual DNP groups were removed from this crude peptide by treatment in 4 molar guanadium HCL, 50 millimolar tris pH 8.5, and 20% 2-mercaptoethanol for one h at 50 °C (Kent, 1988). The crude synthetic peptide was purified by reverse-phase preparative HPLC using a Brown-Lee C8 column and an acetonitrile water gradient. This afforded approximately 8 mg of pure synthetic peptide, GGH-γ,δ (141-183) (Figure 2). The sequence of the protein was confirmed by Edman degradation, amino acid analysis and mass spectrometry (these analyses were performed by the microsequencing facility at the California Institute of Technology).

Footprinting. MPE•Fe(II) footprinting (Figure 3) reveals that the synthetic peptide GGH-γ,δ (141-183) at 2.0 μM concentrations binds all six γ,δ-resolvase half-sites contained in *Res* (lanes 7 and 8, Figure 3). Comparison of the

Figure 3 An autoradiogram of a high-resolution denaturing gel containing ^{32}P-end labeled DNA fragments from the Hind III/Sal I/restriction fragment of PRW80 that contains *res*. The bars on the left-hand side of the gel indicate the location of the γδ resolvase-binding sites. Odd number lanes contain 5'-end labeled DNA and even numbered lanes contain 3'-end labeled DNA. Lanes 1 and 2 contain intact DNA. Lanes 3 and 4 contain MPE•Fe(II) cleavage in the absence of protein. Lanes 5 and 6 are MPE•Fe(II) footprints in the presence of 2.0 μM γδ (141-183). Lanes 7 and 8 are MPE•Fe(II) footprints in the presence of 2.0 μM GGH γδ (141-183). Lanes 9 and 10 are MPE•Fe(II) footprints in the presence of 2.0 μM Ni•GGH γδ (141-183). Lanes 11 and 12 are A-specific cleavage lanes (Iverson and Dervan, 1987).

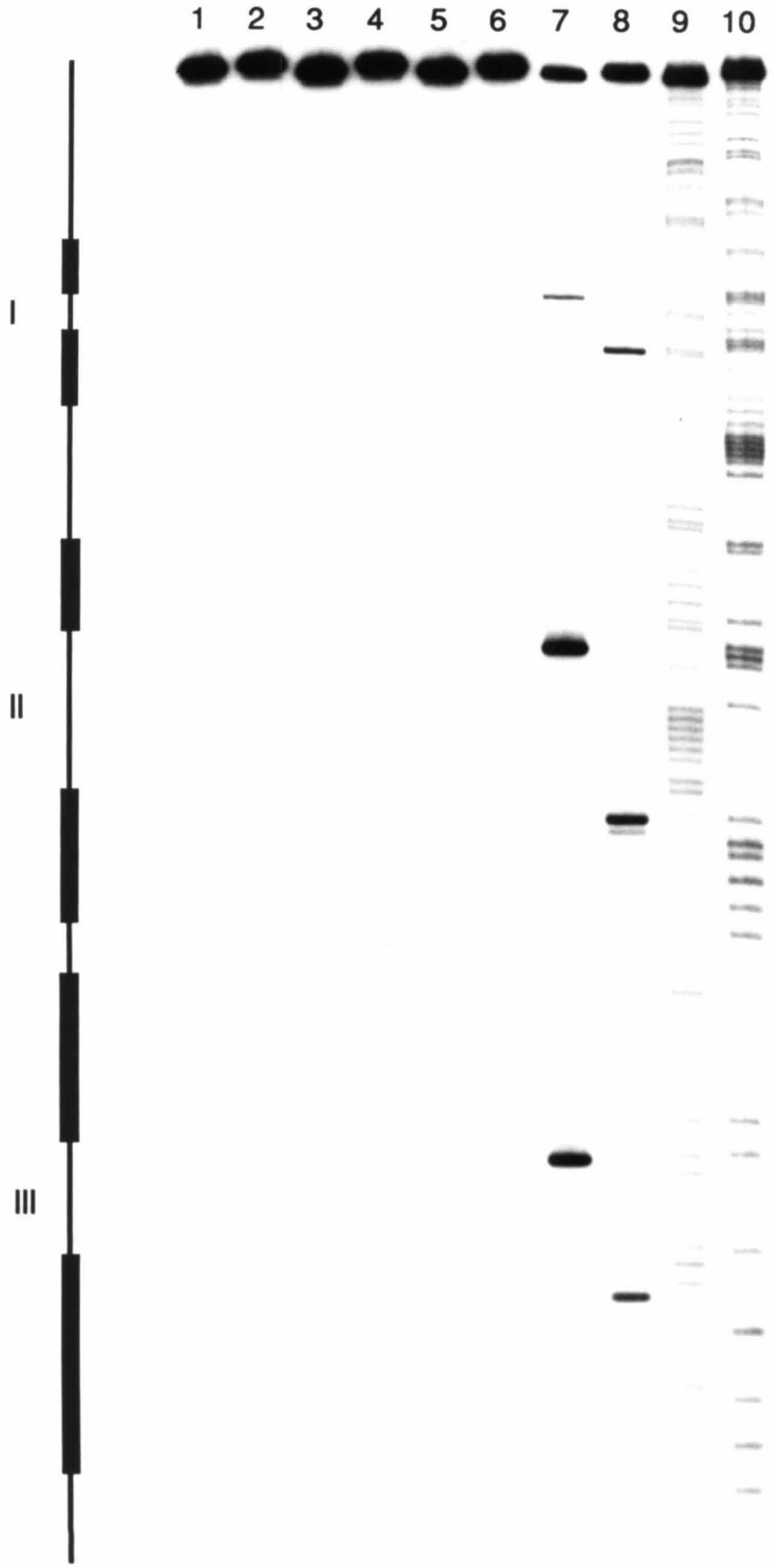

1
2
3
4
5
6
7
8
9
10
11
12
I
II
III

footprinting pattern from γ,δ (141-183) (lanes 5 and 6, Figure 3) and that of GGH-γ,δ (141-183) (lanes 7 and 8, Figure 3) reveal that the addition of GGH to the amino terminus of the DNA-binding domain from γ,δ-resolvase does not alter the peptide's binding specificity. In addition, comparison of the MPE footprint of GGH-γ,δ (141-183) in the absence and presence of Ni $(OAc)_2$, reveals that the presence of Ni (II) does not alter the binding specificity of the synthetic protein (lanes 9 and 10, Figure 3). A schematic representation of the footprint produced by all three of these proteins is found in Figure 5a.

Nickel-Mediated Cleavage Reactions. GGH-γ,δ (141-183) at 2.0 μM concentrations in the presence of one equivalent of Ni $(OAc)_2$ and one equivalent of monoperoxyphthalic acid cleaves the DNA restriction fragment containing *Res* at all six resolvase binding sites (Figure 4, lanes 7 and 6). Control lanes (lanes 1-6) demonstrate that cleavage requires the presence of nickel, GGHγδ (141-183) and monoperoxyphthalic acid. The cleavage patterns are strong and predominate with one-base specificity located 3' to each nine base pair consensus sequence bound by γδ resolvase (Figure 5b). The location of this cleavage is consistent with the location of the amino terminus of γδ (141-183) in the minor groove at 3'-end of each recognition sequence, as determined by Fe(II)-EDTA-mediated affinity cleaving (Graham and Dervan, 1990) (Figure 6).

Several items should be noted concerning cleavage by GGH γδ (141-183). The cleavage mediated by hydrogen peroxide occurs at a much lower efficiency by that produced by monoperoxyphthalic acid. Second, when the d-isomer of histidine is substituted for the *l*-isomer of histidine in the metal-binding region of the peptide, a peptide that produces the same cleavage pattern is created. However, the peptide GG_dH γδ (141-183) cleaves DNA with an approximately

Figure 4 An autoradiogram of a high-resolution denaturing gel containing ^{32}P-end labeled DNA fragments from the Hind III/Sal I/restriction fragment of PRW80 that contains *res*. The bars on the left-hand side of the gel indicate the location of the γδ resolvase-binding sites. Odd number lanes contain 5'-end labeled DNA, and even numbered lanes contain 3'-end labeled DNA. Lanes 1 and 2 contain 2.0 μM monoperoxyphthalic acid and 2.0 μM $Ni(OAc)^2$. Lanes 3 and 4 contain 2.0 μM GGH γδ (141-183) and 2.0 μM $Ni(OAc)^2$. Lanes 5 and 6 contain 2.0 μM GGH γδ (141-183) and 2.0 μM monoperoxyphthalic acid. Lanes 7 and 8 contain 2.0 μM GGH γδ (141-183), 2.0 μM $Ni(OAc)^2$. and 2.0 μM monoperoxyphthalic acid. The DNA in lanes 1 through 8 was subject to base treatment with 0.1 N-butylamine for 30 min at 90 °C. Lanes 9 and 10 are A-specific cleavage lanes (Iverson and Dervan, 1987).


1
2
3
4
5
6
7
8
9
10
I
II
III

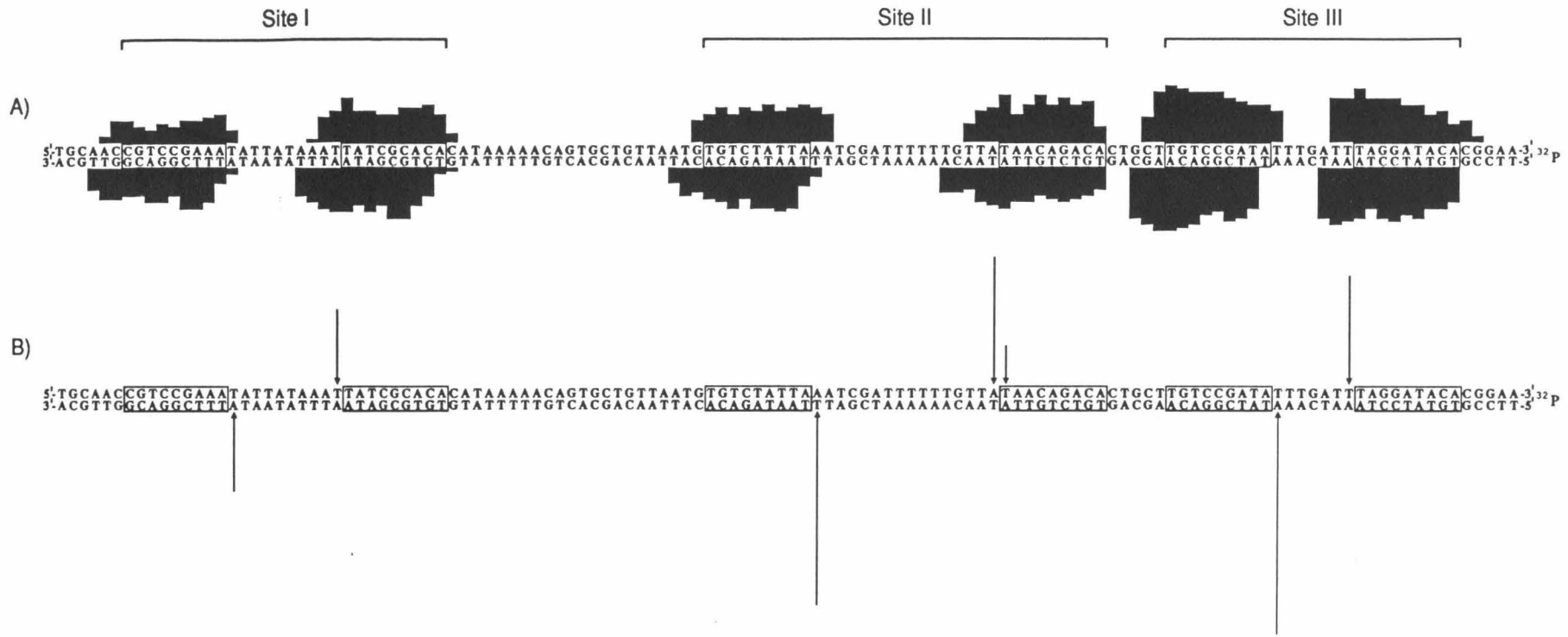
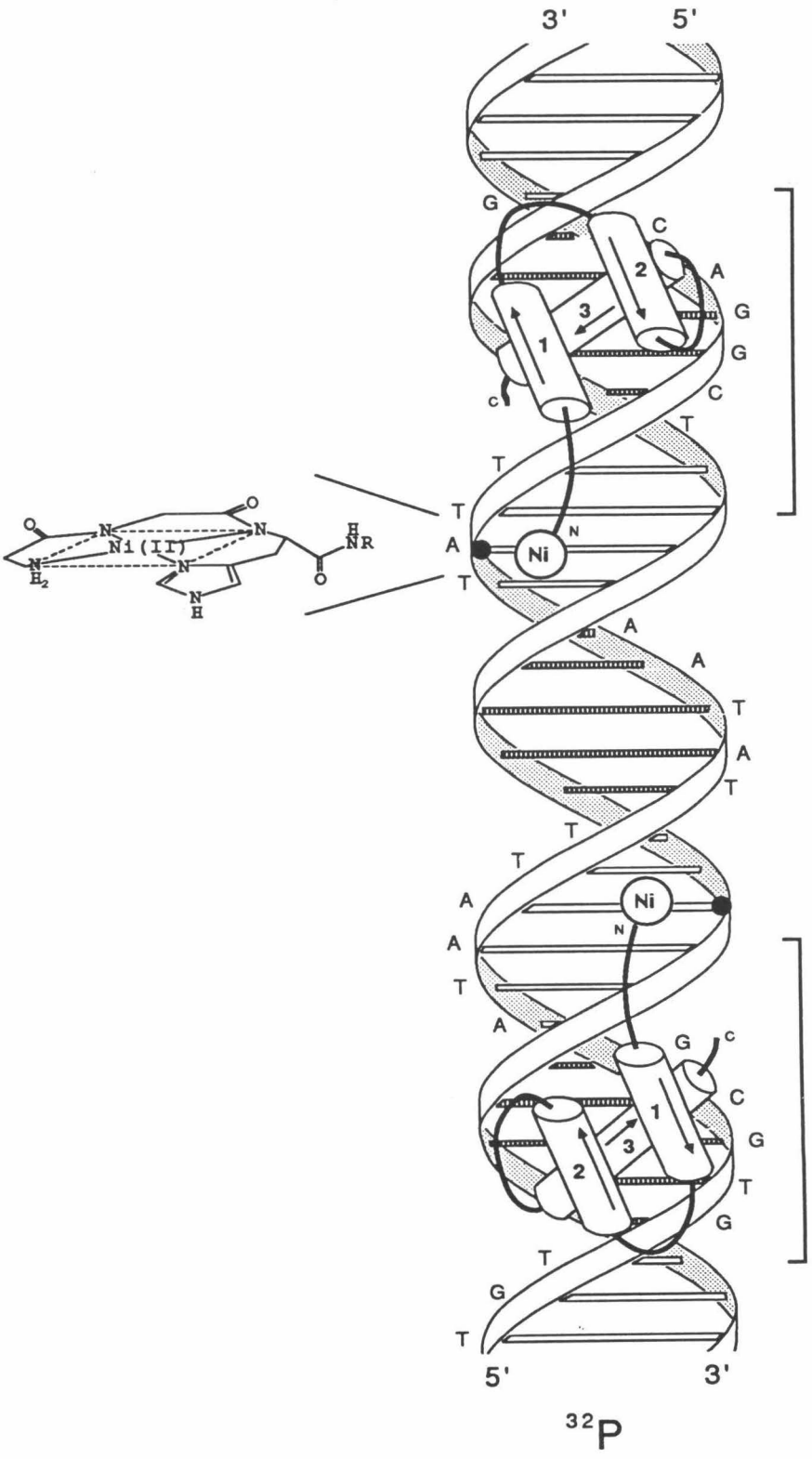


Figure 5 Histograms of footprinting and cleavage data from the autoradiograms in Figures 2 and 3. γδ resolvase-binding sites, I, II, and III are indicated by brackets. The DNA sequences within the boxes represent the binding sites assigned for γδ resolvase (Grindley, et al., 1982). A) Bars represent the extent of protection from MPE cleavage in the presence of Ni(II)•GGH γδ (141-183) at 2.0 μM. B) Arrows represent the extent and location of cleavage by Ni(II)•GGH γδ (141-183) at 2.0 μM concentration in the presence of 2.0 μM monoperoxyphthalic acid.

Figure 6 A schematic representation of the model for the designed metalloprotein Ni•GGH γδ (141-183) bound to *res* site I. Brackets indicate the location of the 9 bp γδ resolvase-recognition sequences. Dots indicate the location of DNA cleavage. Putative α-helices are shown as cylinders with arrows pointing from the amino to carboxyl terminus. The orientation of the helices is based upon affinity-cleavage data (Graham and Dervan, 1990).


3'
5'
G
C
A
G
G
C
T
T
T
A
T
Ni
N
A
A
T
A
T
T
T
A
A
T
Ni
N
G
C
C
G
T
G
T
G
T
5'
3'
1
2
3
Ni(II)
^{32}P

10-fold lower efficiency than the peptide containing the native *l*-isomer of histidine, GGH γδ (141-183) (data not shown). This indicates that the reaction may be susceptible to positional effects relative to the DNA. Finally, it is important to note that while the initial peptide synthesized by Mack, et al. (1988) was capable of cleaving DNA in the presence of Cu(II), cleavage by GGH γδ (141-183) was never observed in the presence of copper.

End Product Analysis. Knowledge of the type of DNA termini produced by a DNA-cleavage reaction can provide insight into the nature of the cleavage reaction. The type of DNA termini produced by the Ni(II), peroxide-mediated cleavage of DNA by GGH γδ (141-183) was examined by high-resolution denaturing gel electrophoresis. The electrophoretic mobility of DNA on a denaturing polyacrylamide gel is affected not only by the length of the DNA, but by the nature of its termini (Tapper and Clayton, 1981). In a DNA fragment 50 base pairs long, the removal of a phosphate from either the 5' or 3'-terminus causes a reduction in the electrophoretic mobility of the DNA strand on a denaturing polyacrylamide gel equivalent to approximately one half a nucleotide. Changes in gel mobility have been successfully used to investigate the nature of 5' and 3' DNA termini created by bleomycin- (Kross, et al., 1982) and MPE- (Hertzberg and Dervan, 1984) mediated cleavage of DNA. In the case of bleomycin-mediated cleavage of DNA, comparison of the electrophoretic mobility of 3'-end labeled DNA fragments with and without alkaline phosphatase treatment identified the 5'-terminal product as phosphate (Kross, et al., 1982). When the same type of analysis was performed using 3'-end labeled fragments of DNA generated by MPE cleavage, the 5'-termini was again found to be phosphate (Hertzberg and Dervan, 1984). The nature of the 3'-DNA termini producted by bleomycin- and MPE-mediated cleavage of DNA

was investigated, utilizing 5'-end labeled fragments and T4 kinase. T4 kinase possesses a 3' phophotase activity and can selectively remove 3' phosphates from DNA fragments (Cameron and Uhlenbeck, 1977). The results of this type of analysis demonstrated that DNA cleavage mediated by bleomycin produced fragments whose 3'-terminus was not a simple phosphate (Kross, et al., 1982) but rather a 3'-phosphoglyocolate (Giloni, et al., 1981). Analysis of the 3'-termini generated by MPE-mediated cleavage of DNA demonstrated that there were two products, 3'-phosphate and 3'-phosphoglycolate (Hertzberg and Dervan, 1984). In either case, the products of cleavage were consistent with the oxidative degradation of the deoxyribose ring of the DNA backbone (Hertzberg and Dervan, 1984; Kross, et al., 1982).

Figure 7 The sequence of the synthetic oligonucleotide duplex containing site I from *res*. Boxes define the location of the 9 bp recognition sequences for γδ resolvase binding.

To determine the nature of the DNA termini produced by Ni(II) peroxide mediate cleavage of DNA by GGH γδ (141-183) , the oligonucleotide shown in Figure 7 and its complementary strand were synthesized. To examine the nature of the 5'-termini produced by DNA cleavage, the top strand of the duplex was labeled at the 3'-end, using α ^{32}P-ddATP. In a separate experiment, the 3'-termini were examined, utilizing the same duplex, with the top strand labeled on the 5'-end. Since the same strand of DNA is utilized for both the 5' and the 3'-end product analysis, the termini on either side of the same lesion were examined.

5'-End Products. The DNA cleavage product produced when the 3'-end labeled oligonucleotide duplex was treated with Ni(II)•GGH γδ (141-183) in the presence of peroxide comigrates with the DNA cleavage product produced by MPE•Fe(II)-mediated cleavage of the duplex (Figure 8). When the cleavage product produced by Ni(II)•GGH γδ (141-183) and peroxide is treated with calf alkaline phosphotase, its electrophorectic mobility is reduced by approximately one-half base pair. These observations are consistent with the presence of a phosphate group at the 5'-termini of the DNA cleavage site.

3'-End Products. The DNA fragment produced when the 5'-end labeled oligonucleotide duplex is treated with Ni(II)•GGH γδ (141-183) and peroxide, comigrates with a 3'-phosphate cleavage product from the MPE•Fe(II) mediated cleavage of the 5'-labeled duplex (Figure 9). When the DNA fragment is treated with T4 kinase under conditions used to remove 3' phosphates from DNA (Cameron and Uhlenbeck, 1977), the electrophoretic mobility of a portion of the band on the gel is reduced. These results indicate that at least a portion of the 3'-termini generated by Ni(II)•GGH γδ (141-183) and peroxide are 3'-phosphate. The composition of the 3'-termini of the DNA fragments that comprise the portion of the band whose electrophoretic mobility was unaffected by kinase treatment is unknown. It is possible that these 3'-termini are phosphate and that the kinase reaction failed to go to completion; however, it is likely that these termini are not phosphate but are some other type of 3'-termini that are unaffected by kinase treatment, like 3'-phosphoglycolate.

Summary of End-Product Analysis. Analysis of the DNA termini produced by Ni(II) GGH γδ (141-183), peroxide-mediated cleavage of DNA has shown the

Figure 8 An autoradiogram of a high-resolution denaturing gel used for analysis of the 5'-end products produced by Ni(II) GGH γδ (141-183)-mediated cleavage of DNA in the presence of hydrogen peroxide. Lanes 1 and 2 contain G+A and G sequence-specific chemical marker lanes, respectively (Maxam and Gilbert, 1980). Lane 3 contains a DNA control. Lanes 4 and 5 were subject to treatment with Ni(II) GGH γδ (141-183) and hydrogen peroxide. Lane 4 was subsequently treated with calf akaline phosphotase to remove phosphates from the 5'-end of cleavage fragments. Lane 6 is an MPE cleavage lane.

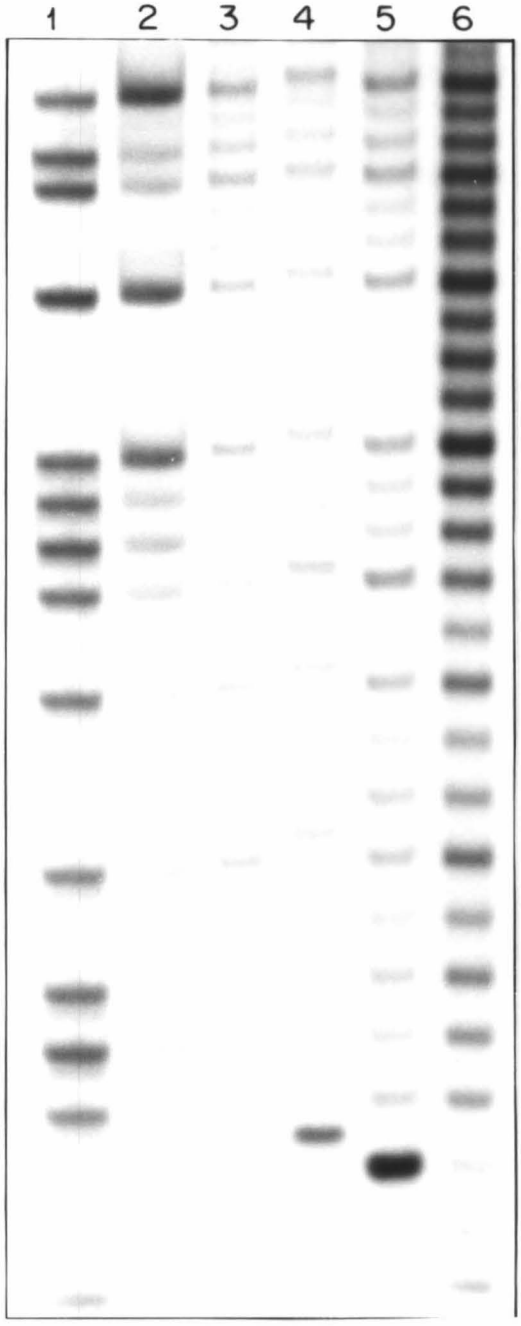
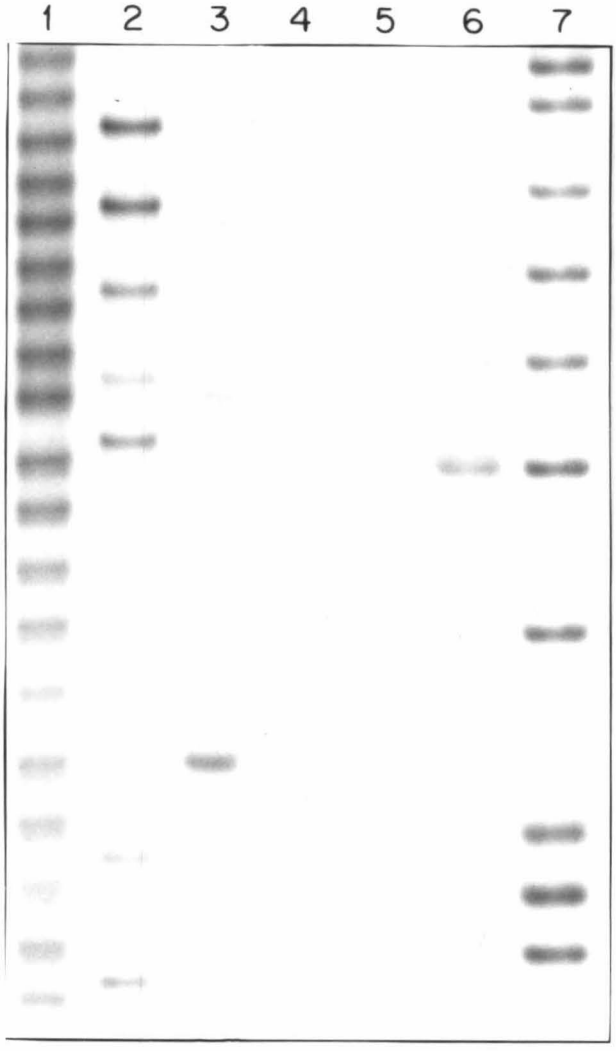

1
2
3
4
5
6

Figure 9 An autoradiogram of a high-resolution denaturing gel used for 3'-end product analysis of Ni(II)•GGH γδ (141-183)-mediated cleavage of DNA in the presence of hydrogen peroxide. Lane 1 is an MPE•Fe cleavage lane containing DNA fragments that end in 3' phosphate and 3' phosphoglycolate (Hertzberg and Dervan, 1984). Lane 2 is a DNAse I cleavage lane containing DNA fragments ending in 3' hydroxyl. Lanes 3 and 4 contain DNA fragments produced by Ni(II)•GGH γδ (141-183) cleavage of DNA. Fragments in lane 4 were treated with T4 kinase to remove the 3' phosphate from the DNA fragment. Lane 5 contains a DNA control. Lanes 6 and 7 are G and G+A sequence-specific chemical cleavage lanes, respectively (Maxam and Gilbert, 1980).


1
2
3
4
5
6
7

5'-end product and a fraction of the 3'-end products to be phosphate. The identity of a portion of the 3'-end products is uncertain; it could be either phosphate or a product like phosphoglycolate. This set of end products is consistent with DNA cleavage produced by oxidative degradation of the deoxyribose backbone. DNA termini consistent with oxidation of the deoxyribose were also observed for cleavage produced by the peptide GGH-Hin (139-190) (Mack, et al., 1988; Mack and Dervan, 1990). When DNA was cleaved using GGH-Hin (139-190) and copper, the resulting 3'- and 5'-termini were found to be phosphate. In the case of cleavage by this peptide in the presence of nickel, initially the 5'-end product was found to be phosphate, and the 3'-end product was found to be a mixture of phosphate and some other, unidentified product. However, when the DNA cleavage fragments produced by GGH-Hin (139-190) were treated with butylamine prior to end-product analysis, the only products identified were 5'-phosphate and 3'-phosphate (Mack and Dervan, 1990).

Cleavage Conditions. In order to more thoroughly characterize DNA cleavage by GGH γδ (141-183) in the presence of Ni(II) and peroxide, a series of studies were conducted to examine the effect of oxidant concentration, protein concentration, time, and pH on the efficiency of the reaction. These studies were conducted using 5'-end labeled oligonucleotides, and except where noted, analyzed using a Molecular Dynamics Phosphorimager. The studies were conducted using both hydrogen peroxide and monoperoxyphthalic acid as oxidants.

Concentration Studies. The cleavage of DNA as a function of monoperoxyphthalic acid concentration is graphically represented in Figure

10a-c. The first feature that becomes apparent from an examination of these graphs is that at approximately 10 μM concentration, maximal cleavage is reached, regardless of the concentration of Ni(II)-GGH-γδ (141-183). Significant cleavage above background occurs with as little as 0.01 μmole of monoperoxyphthalic acid. Comparison of cleavage with and without butylamine treatment shows that incubating the cleavage reactions for 30 min at 90 °C in the presence of 0.1 normal butylamine can produce a 20-25% increase in cleavage. By comparing graphs a through c directly it becomes apparent that the extent of cleavage is dependent on the concentration of Ni(II)-GGH-γδ (141-183). It is interesting to note that this relationship does not appear to be linear, at least between 0.5 and 2 μM.

Examination of Figure 10c and d shows the marked difference between the two oxidants, monoperoxyphthalic acid and hydrogen peroxide. Monoperoxyphthalic acid proves to be an effective oxidant at micromolar concentrations. However, hydrogen peroxide requires concentrations in the millimolar range even to produce detectable cleavage. On this basis alone it appears that monoperoxyphthalic acid is a 1,000 times more effective oxidant than hydrogen peroxide in this system. In other systems that use Ni(II)-GGH to cleave DNA, the same relative effectiveness of these two oxidants is also observed (Chen, et al., 1991; Mack, 1991). The effect of butylamine treatment on DNA cleavage mediated by hydrogen peroxide and Ni(II)-GGH-γδ (141-183) was not examined, because butylamine treatment of DNA following exposure to high levels of hydrogen peroxide results in significant amounts of non-specific cleavage.

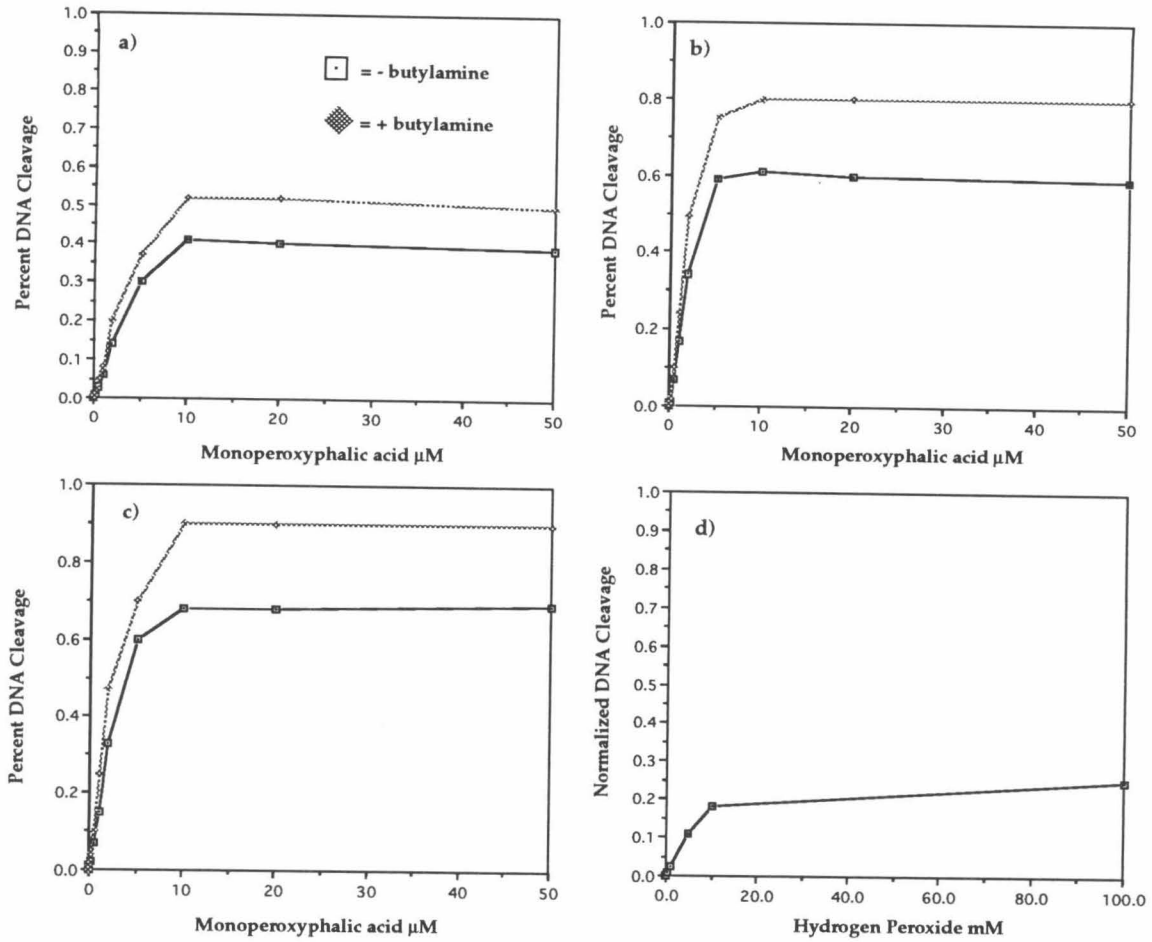


Figure 10 DNA cleavage produced by Ni(II)•GGH γδ (141-183) as a function of monoperoxyphthalic acid (A-C) or hydrogen peroxide concentration (D). The concentration of Ni(II)•GGH γδ (141-183) present in the reactions was 0.5 μM (A), 1.0 μM (B), and 2.0 μM (C and D). Reactions containing hydrogen peroxide were allowed to proceed for 90 min prior to precipitation; those containing monoperoxyphthalic acid were allowed to proceed for 10 min prior to ethanol precipitation. The reactions were treated with butylamine as noted. The extent of DNA cleavage in A-C was determined in phosphorimager analysis and in D by densitometric analysis.

Time Courses. Time courses of DNA cleavage produced by Ni(II)-GGH-γδ (141-183) in the presence of monoperoxyphthalic acid or hydrogen peroxide are

shown in Figures 11a and b. The time courses were run over a period of 120 min. The time point at 0 minutes represents the simultaneous addition of oxidant and quenching agent to the reactions. Examination of the time course for cleavage in the presence of monoperoxyphthalic acid shows that the reaction is complete within less than 10 minutes. In fact, even when the oxidant and quenching agent are simultaneously added to the reaction, cleavage approaches 40% of the maximum cleavage observed under these conditions.

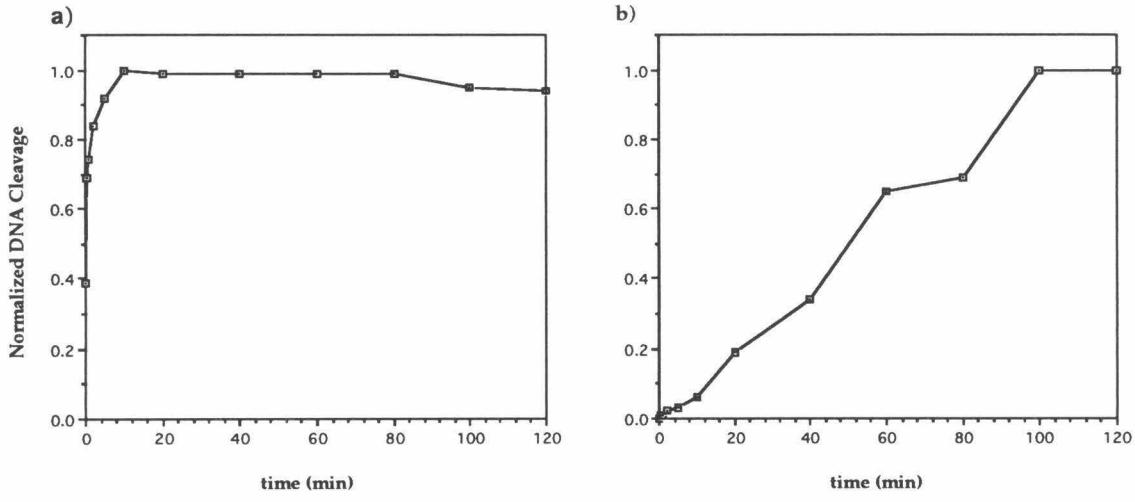


Figure 11 Time course of DNA cleavage by 2.0 μM Ni(II)•GGH γδ (141-183) in the presence of 1.0 μM monoperoxyphthalic acid (A) or 10.0 mM hydrogen peroxide (B). Reactions were stopped by ethanol precipitation. The zero time point represents simultaneous addition of oxidant and precipitation. Reactions containing monoperoxyphthalic acid were treated with 0.1 normal butylamine for 30 min at 90 °C prior to electrophoresis. The extent of DNA cleavage was determined by phosphorimager analysis.

Cleavage mediated by monoperoxyphthalic acid is so rapid that an accurate measurement cannot be obtained under the system that we are currently using. When the cleavage of DNA in the presence of hydrogen peroxide is examined as a function of time, a much different picture appears. When oxidant and

quenching agent are simultaneously added to the reaction, there is detectable cleavage above the background. The reactions must be incubated for at least five minutes before significant cleavage occurs. At this point, cleavage proceeds steadily until a maximum is reached at approximately 100 min (Figure 11b).

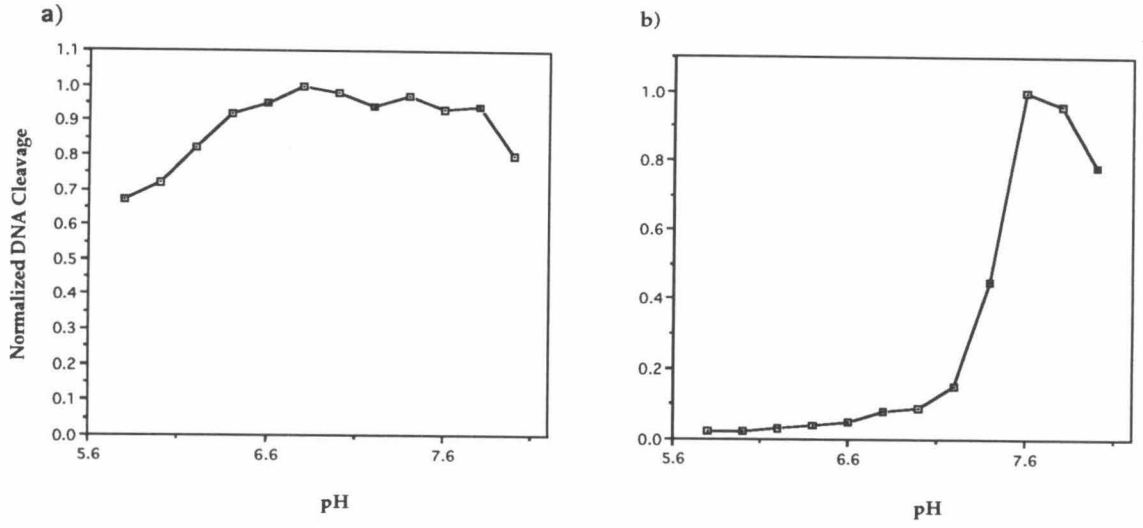


Figure 12 pH profiles of DNA-cleavage reaction mediated by Ni(II)•GGH γδ (141-183) in the presence of 1.0 μM monoperoxyphthalic acid (A) and 10.0 μM hydrogen peroxide (B). Reactions containing monoperoxyphthalic acid were allowed to proceed for 10 min prior to ethanol precipitation, and reactions containing hydrogen peroxide were allowed to proceed for 90 min prior to ethanol precipitation. Reactions containing monoperoxyphthalic acid were treated with 0.1 N-butylamine for 30 min at 90 °C prior to analysis. The extent of DNA cleavage was determined by phosphorimager analysis.

pH Profiles. The effect of pH on DNA cleavage by Ni(II)-GGH–γδ (141-183) in the presence of monoperoxyphthalic acid and hydrogen peroxide is shown in Figure 12 a and b. The pH of the phosphate reaction buffer was varied from 5.8 in 2/10ths of a pH unit up to 8.0. When DNA cleavage as a function of pH is examined with monoperoxyphthalic acid as the oxidant, a relatively flat profile

is obtained. Maximum cleavage seems to occur between the range of pH 6.8 and 7.4. Below pH 6.6 the amount of cleavage drops steadily to pH 5.8, where the last measurement was taken. At the high end of the curve, from pH 7.8 to 8.0, cleavage also appears to drop. When cleavage is examined in the presence of hydrogen peroxide, a different profile is obtained. From pH 5.8 to 6.6 very little cleavage is observed. From pH 6.6 to 7.2, a slight increase in cleavage is observed. Above pH 7.2 there is a sharp increase in DNA cleavage with a maximum occurring at about pH 7.6. From pH 7.6 to pH 8.0, the amount of cleavage declines. These pH profiles are unlike the sigmoidal curves obtained when Ni(II)-GGH-Hin (139-190) cleavage in the presence of monoperoxyphthalic acid and hydrogen peroxide was studied as a function of pH (Mack, 1991). The differences in the pH profiles obtained from Ni(II)-GGH-γδ (141-183) and those reported by Mack (1991) may be a result of several factors. The first may be differences in data collection; the studies conducted by Mack (1991) contained only five data points, while these contained twelve data points over a slightly wider pH range. A second factor is the change in the DNA-binding domain from Hin(139-190) to γδ (141-183). The ability of these two proteins to bind DNA as a function of pH may be somewhat different. It is possible that the differences in cleavage observed between these two different proteins may not be a function of reaction mechanism, but rather a function of DNA-binding. A somewhat more puzzling observation is the difference in the pH profiles for monoperoxyphthalic acid and hydrogen peroxide with the same protein Ni(II)-GGH-γδ (141-183). A possible explanation came when the pH of the stock solutions of oxidants used to initiate the reactions was examined. In the case of monoperoxyphthalic acid, a final concentration of 1 μM per acid was present in each of the reactions, and the pH of the stock solution used to achieve this was 6.2. However, in the case of hydrogen

peroxide, 10 millimolar peroxide was the final concentration in each of the reactions, and the pH of its stock solution was 5.2. It is possible that the low pH of the hydrogen peroxide stock solution in combination with the large amount of peroxide added to the reaction overloaded the 20 millimolar phosphate buffer that was used to maintain the pH of the reaction.

ESR Studies of Ni(II) GGH and Ni(III) GGH. The oxidation state of nickel peptide complexes has been studied extensively, using ESR spectroscopy (Sugiura and Mino, 1979; Bossu and Margerum, 1977; Bossu and Margerum, 1976). The Ni(II) complexes of tripetides are a low-spin, square-planar d^8 species, which are ESR-inactive. When the nickel in these complexes is oxidized to Ni(III), a tetragonal d^7 nickel complex is formed (Bossu and Margerum, 1976; Bossu and Margerum, 1977; Sugiura and Mino, 1979). The ESR spectrum of these trivalent nickel complexes obtained at liquid-nitrogen temperatures is similar to the spectra obtained for many Ni(III) tetraza-macrocycle complexes (Olson and Vasilevskis, 1969; Gore and Busch, 1973), with a G perpendicular greater than a G parallel. This type of ESR spectrum is consistent with an unpaired electron in the dz^2 orbital for d^7 Ni(III). If the conditions used to promote DNA cleavage in the presence of Ni(II)-GGH-γδ (141-183) generated a Ni(III)-GGH metal complex, we would expect ESR spectra with an anisotropic G tensor with values corresponding roughly to the values of G_X=2.256, G_Y=2.278, and G_Z=2.015, which are observed for GGH•Ni(III) (Lappin, et al., 1978). If the oxidizing conditions used created a ligand-centered radical, the paramagnetic product would be a nitrogen- or carbon-centered radical complex of Ni(II). Such a species would exhibit a dramatically different ESR spectrum; the G tensor would be isotropic with all three G values close to the spin-only value of 2.002 (Drago and Baucom, 1972; Barefield and Mocella, 1975).

The ESR spectrum of the nickel complex of GGH was examined in the presence and absence of oxidants at -50 °C. The peptide-nickel complex was dissolved in 20 millimolar phosphate buffer at pH 7.5, and the solutions were subject to several round of freeze/thaw degassing prior to measurement of the ESR spectrum. As expected, the Ni(II) complex of GGH showed no paramagnetic signal. When this complex was oxidized using $IrCl_6^{2-}$(IV), the characteristic ESR spectrum of Ni(III) was obtained (Figure 13). This spectrum showed the anisotropy characteristic of Ni(III) peptide complexes. Initially, when the EPR spectrum of the Ni(II)-GGH complex was measured in the presence of hydrogen peroxide, no paramagnetic signal was observed. To obtain a paramagnetic signal, a large excess of hydrogen peroxide had to be added to the complex; the final concentration in the ESR was 2.5 molar. This is approximately 500 times more hydrogen peroxide than is required to observe DNA cleavage in the presence of Ni(II)-GGH-γδ (141-183). When the ESR spectrum of a solution containing 5mM GGH-Ni(II) and 2.5 M H_2O_2 (the hydrogen peroxide was pH adjusted to 7.0) was measured, the spectrum shown in Figure 14 was obtained. This spectrum lacked the characteristic signals observed for a Ni(III) peptide complex. The spectrum showed three signals with G values of 2.1965, 2.0492, and 2.0059. While it is not possible to identify the radical species or specie responsible for this spectrum, it is apparent that Ni(III) is not formed under these conditions and is apparently not responsible for the Ni-GGH-γδ (141-183)-mediated cleavage of DNA. This observation was confirmed when Ni(II)-GGH-γδ (141-183) failed to cleave DNA in a sequence-specific manner in the presence of $IrCl_6^{2-}$(IV) (data not shown). Recently, an extensive ESR study was conducted, utilizing the Ni(II) complex of GGH and hydrogen peroxide (Inoue and Kawanishi, 1989). This study

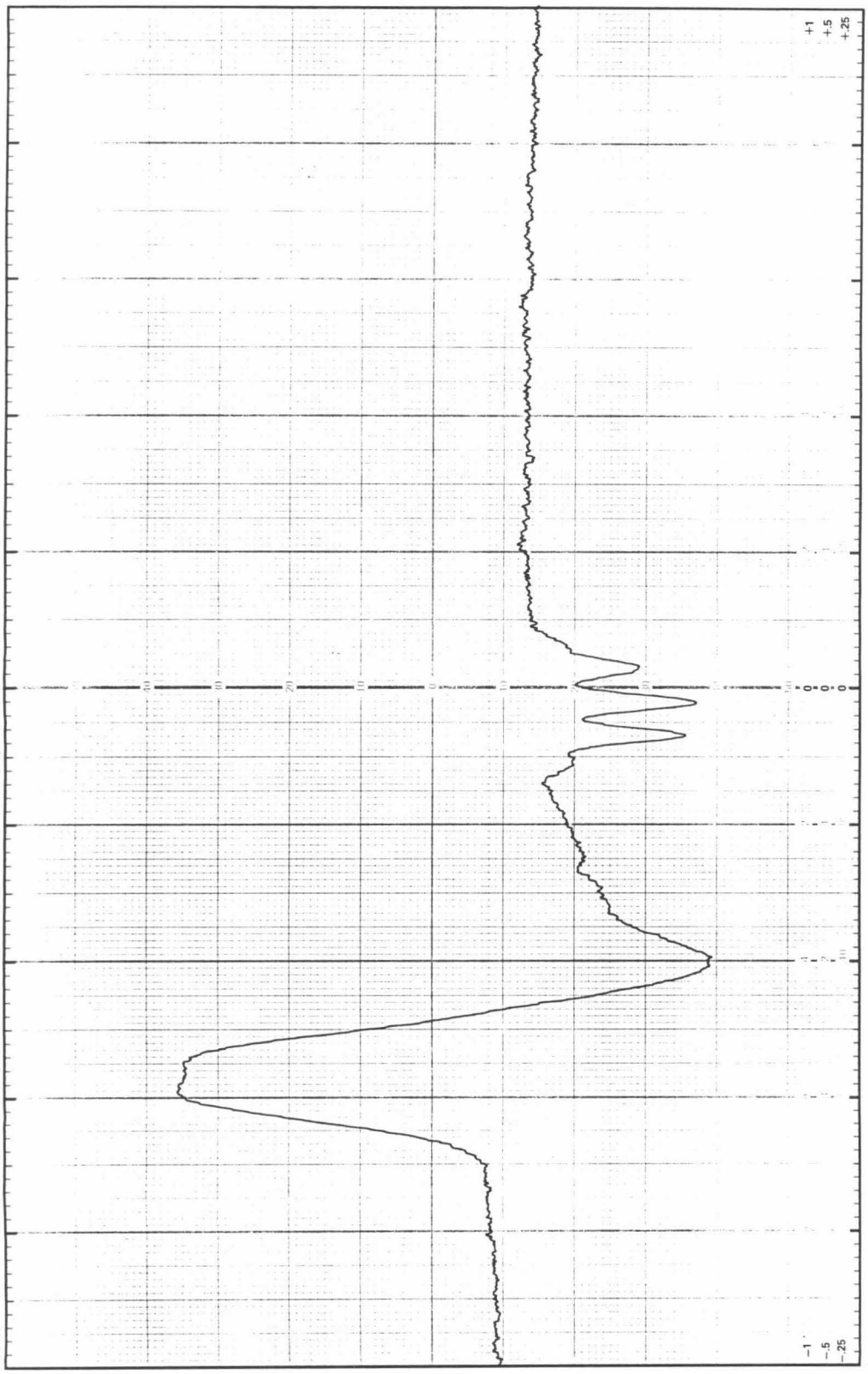


Figure 13 The ESR spectrum obtained at -150 °C from the oxidation of 5 mM Ni(II)•GGH γδ (141-183) with one equivalent of $IrCl_6^{2-}$ (IV). The spectrum shows the characteristic signal of Ni(III)•GGH (Sugiura and Mino, 1979). The metal-peptide complex was dissolved in a 20 mM phosphate buffer (pH 7.5), and the solution was subject to several rounds of freeze thaw degassing prior to spectroscopic measurement.

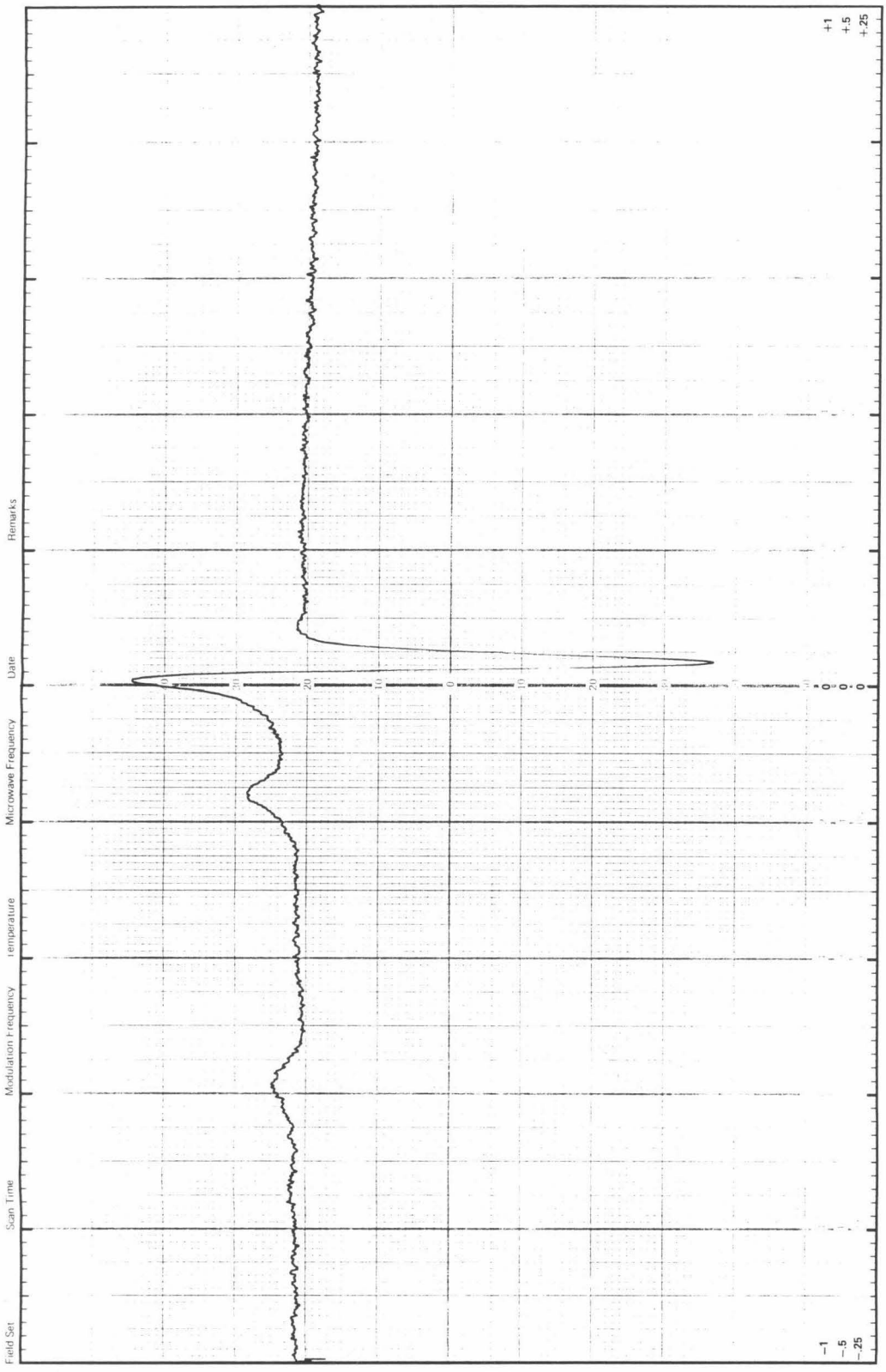


Figure 14 The ESR spectrum obtained at -150 °C from the oxidation of 5 mM Ni(II)•GGH in the presence of 2.5 M hydrogen peroxide. The metal peptide complex is dissolved in a phosphate buffer (pH 7.5), the hydogen peroxide added to the solution was adjusted to pH 7.5. The solution was subject to several rounds of freeze thaw degassing prior to spectroscopic measurement.

confirms our observation that Ni(III)-GGH is not formed in the presence of hydrogen peroxide. By using a series of spin traps, this ESR study revealed that hydroxyl radical adducts are produced by the decomposition of hydrogen peroxide in the presence of Ni(II)-GGH (Inoue and Kawanishi, 1989). Their results indicate that either an oxo-Ni(IV) complex or a Ni(III)-peroxide complex is formed under these conditions.

Kinetic Isotope Studies of the Mechanism of DNA Cleavage by Ni(II) GGH-γδ (141-183).

The type of DNA termini produced by Ni(II)-GGH γ δ (141-183) in the presence of oxidant are consistent with DNA cleavage via the oxidative degradation of the deoxyribose ring of the DNA backbone. Since DNA cleavage occurs exclusively at one nucleotide, this oxidizing species is highly localized, unlike the diffusable oxidizing species generated during EDTA•Fe-mediated cleavage of DNA. These observations are consistent with results obtained from a different sequence-specific, DNA-cleaving metalloprotein, Ni(II) Hin (139-190) (Mack and Dervan, 1990).

In order to gain insight into the nature of this highly localized, or more probably, bound oxidizing species, ESR studies were conducted utilizing Ni(II)-GGH. Under the conditions used to produce DNA strand scission, ESR provided no evidence for the formation of Ni(III), nor did they indicate the formation of a ligand radical. However, they did not preclude the formation of some type of activated metal-oxidant complex that would be capable of oxidatively cleaving DNA. ESR spin-trapping studies examining the decomposition of hydrogen peroxide in the presence of Ni(II)-GGH indicate

that in the presence of peroxide the tripeptide-nickel complex is capable of forming either a Ni(IV) oxo or a Ni(III)-peroxide complex (Inoue and Kawanishi, 1989). Further support for the generation of a GGH Ni(IV) oxo or GGH Ni(III)-peroxide complex is provided by the observation that in the presence of iodosylbenzene Ni(II)-GGH will epoxidize olefins (Mack, 1991; Mack and Dervan, 1990). Mechanistic studies of olefin epoxidation catalyzed by other Ni(II) complexes utilizing ^{18}O-labeling have demonstrated that the oxygen in the epoxide is entirely derived from the oxidant (Kinneary, et al., 1988). This indicates olefin epoxidation by Ni(II) complexes is mediated by some type of bound-metal oxidant species analogous with the Ni(IV)oxo or Ni(III)-peroxide complex detected by ESR spin-trapping studies. In order to improve our understanding of how the activated Ni-GGH γ δ (141-183) mediates DNA cleavage, a kinetic isotope experiment was conducted in which the effect of deuterium incorporation at discrete positions within the deoxyribose ring on the rate of DNA cleavage was examined.

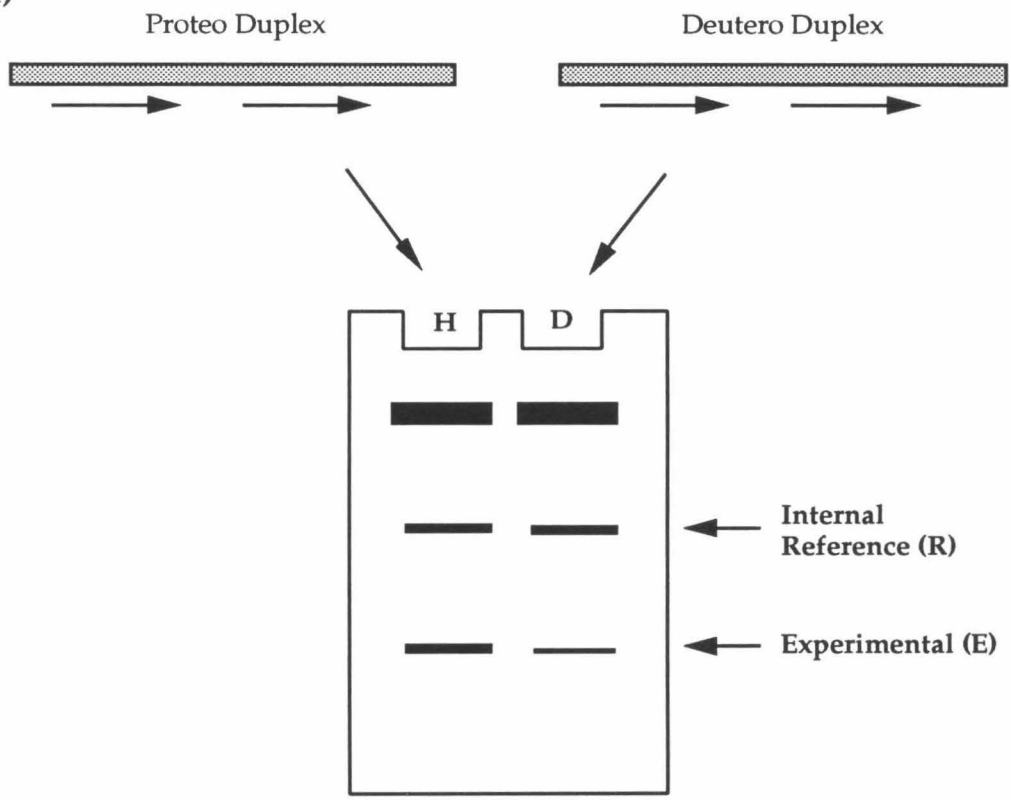
Kinetic isotope experiments of this type have provided useful insight into the mechanisms of DNA cleavage by the drugs neocarzinostatin and bleomycin (Kappen, et al., 1990; Frank, et al., 1991; Kappen, et al., 1991; Wu, et al., 1983; Kozarich, et al., 1989). In the case of neocarzinostatin, kinetic isotope effects have been useful in identifying the mechanism of cleavage at both major and minor sites (Frank, et al., 1991; Kappen, et al., 1991; Kappen, et al., 1990). The predominant type of DNA damage produced by neocarzinostatin is strand breaks at thymidilate and deoxyadenilate residues, resulting in DNA cleavage products with termini consisting of a 5'-phosphoaldehyde and a 3'-phosphate (Kappen, et al., 1982; Kappen and Goldberg, 1983). When the formation of these DNA strand breaks was investigated, using C4' deuterated and C5'

dideuterated thymidines, a kinetic isotope effect was observed at both positions with k_H/k_D = 2.4-5.5 and 1.0-2.6, respectively (Frank, et al., 1991). Close examination of the kinetic isotope data showed that at certain positions cleavage was strongly sensitive to deuterium substitution at the C4' position, and much less sensitive to deuteration at the C5' position. However, at other cleavage sites the situation was reversed, and sensitivity to 5'-deuteration was observed, while no evidence of sensitivity to deuteration at the 4' position was observed. Analysis of the end products of DNA cleavage at sites that were sensitive to deuteration at the 5' position showed that cleavage yielded 3'-phosphate termini, while at sites that were sensitive to 4' deuteration cleavage yielded 3' termini consisting of a modified abasic carbohydrate and phosphoglycolate. In a subsequent experiment, it was observed that the proportion of DNA cleavage resulting from hydrogen abstraction at the 4' or 5' position could be shifted as a result of isotope-selection effects (Kappen, et al., 1991). Kinetic analysis of neocarzinostatin cleavage at the minor sites, the deoxycytidylate residues in AGC sequences, demonstrated a third type of DNA cleavage chemistry for neocarzinostatin (Kappen, et al., 1990). At these sites, DNA cleavage appears to be mediated by C1' hydrogen bond cleavage. When the C1' position of deoxycysteine residues located at these cleavage sites was deuterated, a k_H/k_D of 3.3-4.2 was observed, confirming the location of hydrogen-atom abstraction as at the C1' position. Bleomycin, like neocarzinostatin, oxidatively cleaves DNA by a degradation of the sugar phosphate backbone. In the presence of metal ions such as Fe(II) or Cu(II) and oxygen, bleomycin cleavage of DNA yields free nucleic acid bases, base propenal, and DNA termini consisting of 5' phosphates and 3' phosphoglycolates. The first kinetic isotope study to examine the mechanism of bleomycin cleavage of DNA measured the effect of tritium selection on 4'-

hydrogen abstraction (Wu, et al., 1985). Under the conditions used for this study, k_H/k_D for 4' hydrogen abstraction was found to range from 7.2-12.5. This clearly indicated that 4'-hydrogen abstraction was rate-limiting and provided the first direct evidence that bleomycin cleavage of DNA is a consequence of 4'-hydrogen abstraction. In a subsequent experiment, a kinetic isotope study was conducted that measured the effect of deuterium incorporation at the C4' position on the rate of bleomycin cleavage (Kozarich, et al., 1989). In this study, k_H/k_D for 4'-hydrogen abstraction was found to range between 2 and 4.5. This was consistent with the results of the earlier experiment using tritium selection and confirmed that bleomycin cleavage of DNA involves rate-limiting abstraction at the 4' hydrogen. An interesting side note from this experiment was that the differences in the magnitude of the isotope effect at different sites were reproducible and suggests that local variability in the DNA sequence may be important in determining the magnitude of the kinetic isotope effect that is observed (Kozarich, et al., 1989).

Experimental Design. The experiment designed to determine the kinetic isotope effect for Ni(II)-GGH γ δ (141-183) is shown schematically in Figure 15. Like the kinetic isotope experiments examining neocarzinostatin and bleomycin, polyacrylamide gel electrophoresis is used in conjunction with sequence-specific incorporation of specifically deuterated nucleotides. Since Ni(II)-GGH γδ (141-183) cleaves at one nucleotide per binding site, a chemically synthesized oligonucleotide was designed that contained two binding sites arranged in a head-to-tail fashion. This provided two sites of DNA cleavage on the same oligonucleotide strand. Automated machine synthesis of this duplex allowed the insertion of a specifically deuterated nucleotide at a predetermined cleavage position. Two sets of oligonucleotides were prepared. The first set

A)



B)

$$KIE = \frac{\frac{H_E}{H_R}}{\frac{D_E}{D_R}}$$

Figure 15 A schematic representation of a kinetic isotope experiment using Ni(II)•GGH γδ (141-183). A) A representation of the cleavage products obtained from electrophoresis of a proteo and a deutero duplex. The bands used as internal references and experimentals are identified. B) The equation used to determine the kinetic isotope effect (KIE).

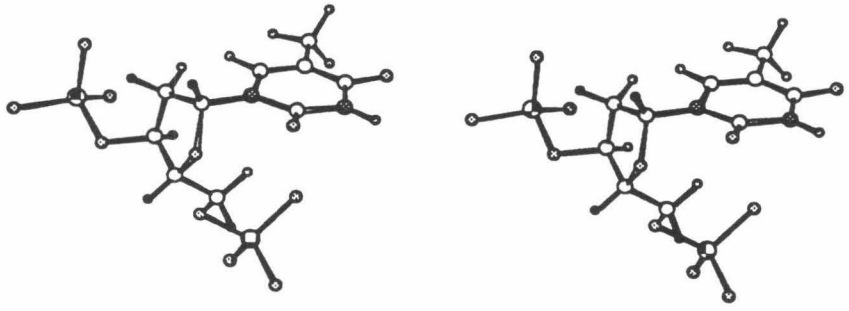
was an all proteo duplex, the second set contained the specifically deuterated nucleotide at a single cleavage site within the duplex. Using oligonucleotides that are 5'-end labeled with ^{32}P, two parallel sets of cleavage reactions are performed. The first one utilizes the all-proteo version of the target duplex, and the second one utilizes the target sequence containing a deuterated nucleotide. Following completion of the cleavage reactions, the DNA is treated with butylamine at 90 °C for 30 min to assure that all alkali labile products are converted to phosphate. This allows for the calculation of a kinetic isotope effect (KIE) based on one type of cleavage product. Following the butylamine treatment, the cleavage products are separated by gel electrophoresis using a 20% denaturing gel. This produces a gel with the appearance shown schematically in Figure 15. In each lane of the gel there are three bands. The first band corresponds to uncut DNA, while the second and third bands correspond to sequence-specific affinity cleavage by Ni(II)-GGH $\gamma\delta$ (141-183) adjacent to each of the protein-binding sites. The first of these affinity cleavage sites serves as an internal reference and always contains a proteo oligonucleotide. The second cleavage site is the experimental case; in the proteo duplex this site contains a proteo nucleotide, and in the deuterium duplexes this site contains a site-specific deuterated nucleotide. After electrophoresis, the intensity of the bands produced by affinity cleavage is measured, using a Molecular Dynamics Phosphorimager. These measurements are then used to calculate a KIE, using the formula found in Figure 15. Within each lane of the gel a ratio is first taken between the experimental case and the internal reference. This normalizes the value observed for the experimental band and serves to correct it for experimental error that might result from differential loading of DNA onto lanes of the electrophoretic gel, or for minor differences in the magnitude of affinity cleavage between individual reactions.

The kinetic isotope effect is then calculated by taking the ratio of this normalized value for the proteo case over the normalized value for the deutero case.

It is important to make note of the cleavage conditions used for the kinetic isotope experiment. The reaction conditions chosen produced approximately 1% total cleavage of the DNA after butylamine treatment. The choice of conditions that produce a low amount of cleavage is important for several reasons. First, it is important to maintain conditions where the DNA duplex is cleaved only once. At cleavage greater than 4-5%, a significant number of double cleavage events occur. It is important to avoid double-cleavage events because of the use of gel electrophoresis to isolate and quantitate the cleavage products. Gel electrophoresis separates DNA on the basis of size and charge. However, only the fragment of DNA that is ^{32}P-end labeled is detected. Thus, in the case of double-cleavage events, only the event closest to the label is detected. In the experimental design used to measure the isotope effect, double-cleavage events would lead to an error in reporting of the KIE.

Availability of Hydrogens for Abstraction. Affinity-cleaving studies have located the amino terminus of γ δ (141-183) in the minor groove at the center of each resolvase-binding site (Graham and Dervan, 1990). The location at which DNA cleavage occurs using Ni(II)-GGH-γ δ (141-183) and monoperoxyphthalic acid is consistent with this assignment for the position of the amino terminus of the peptide γ δ (141-183). As a consequence of this, one would expect the DNA-cleaving moiety Ni(II)-GGH to be positioned in close proximity with the minor groove, limiting the availability of hydrogens for abstraction to those that are found within the minor groove. By examining the major and minor grooves of

a) Stereo view from the minor groove



b) Stereo view from the major groove

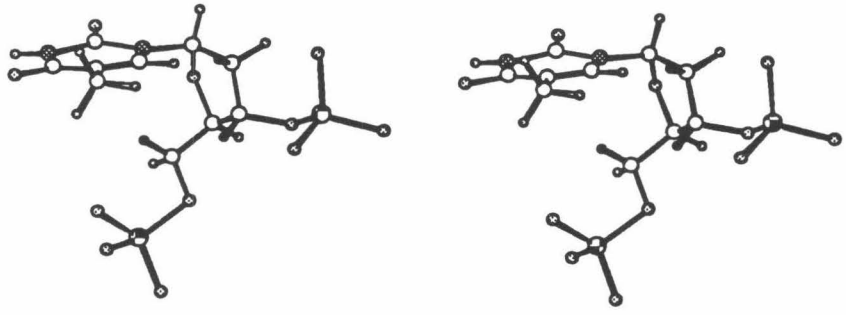


Figure 16 A stereo view of the sugar phosphate backbone of DNA from the minor (a) and the major (b) groove. Hydrogen atoms exposed to the corresponding grooves are shown in black.

duplex DNA (Figure 16), we see that only certain hydrogens are available for abstraction within each of these groups. Within the minor groove we see that hydrogens attached to carbons C1', C2', C4', and C5' of the deoxyribose are

available for abstraction; while within the major groove hydrogen atoms attached to the carbon atoms C2', C3', and C5' of the deoxyribose are available for abstraction. Since the Ni(II)-GGH is located in close proximity to the minor groove, it would be expect that DNA cleavage might involve hydrogen-atom abstraction from positions C1', C2', C4', and/or C5' of the deoxyribose ring. If the abstraction of hydrogen by the activated form of Ni(II)-GGH γ δ (141-183) is the rate-limiting step in DNA cleavage, then deuterium incorporation at one or more of these positions may produce an observable kinetic isotope effect and may thereby identify positions that are involved in the cleavage of DNA. An important assumption in this interpretation is that isotopic substitution does not affect the extent of strand breakage at any other step than the initial reaction with the activated Ni(II)-GGH γ δ (141-183).

Results of Kinetic Isotope Studies

Synthesis. Figure 17 shows the sequence of the synthetic oligonucleotide used for the kinetic isotope experiments, as well as the structures of the site-specific deuterated thymidines. The syntheses of the deuterated thymidines are shown in Schemes I-V.

Synthesis of the C1'-deuterated thymidine involved initial reduction of the tribenzoyl ribolactone, using deuterated disiamylborane. Following reduction of the lactone, the 1-O-acetyl compound (**2**) was synthesized by reaction with acetic anhydride. In the presence of the Lewis acid $SnCl_4$, (**2**) was reacted with the bis(O-trimethylsilylthymidine) forming the tribenzoylated methyluridine (**3**). The benzoyl groups were then removed and the 3' and 5'-alcohols were

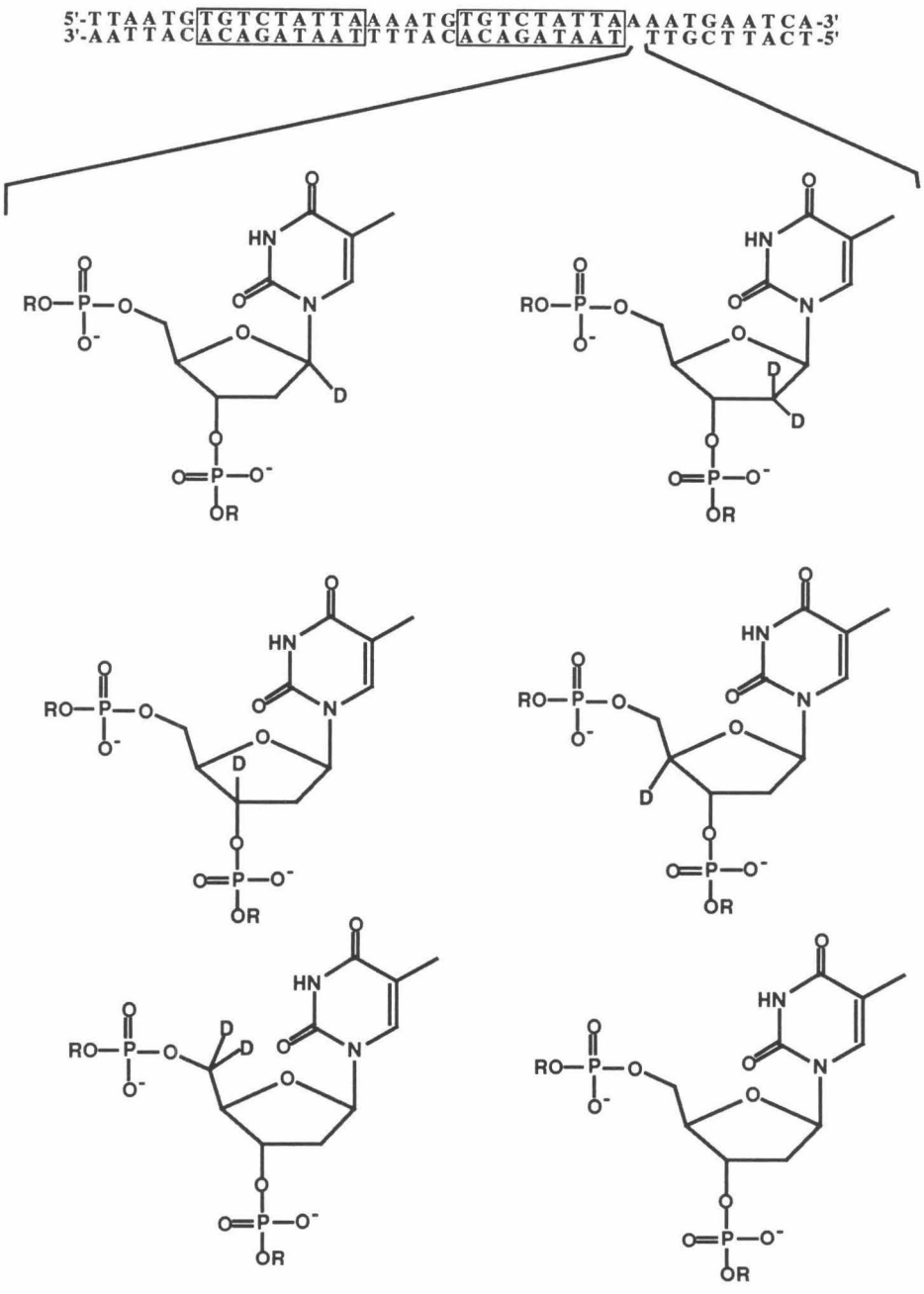
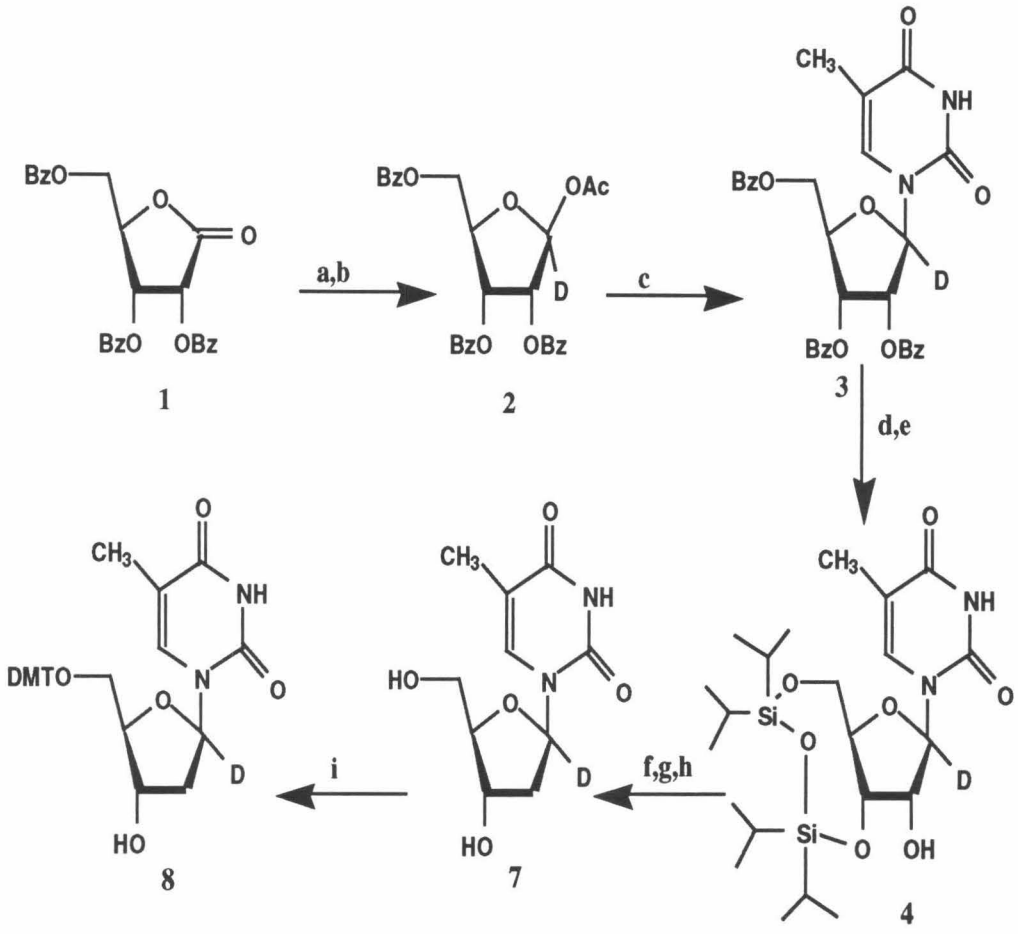


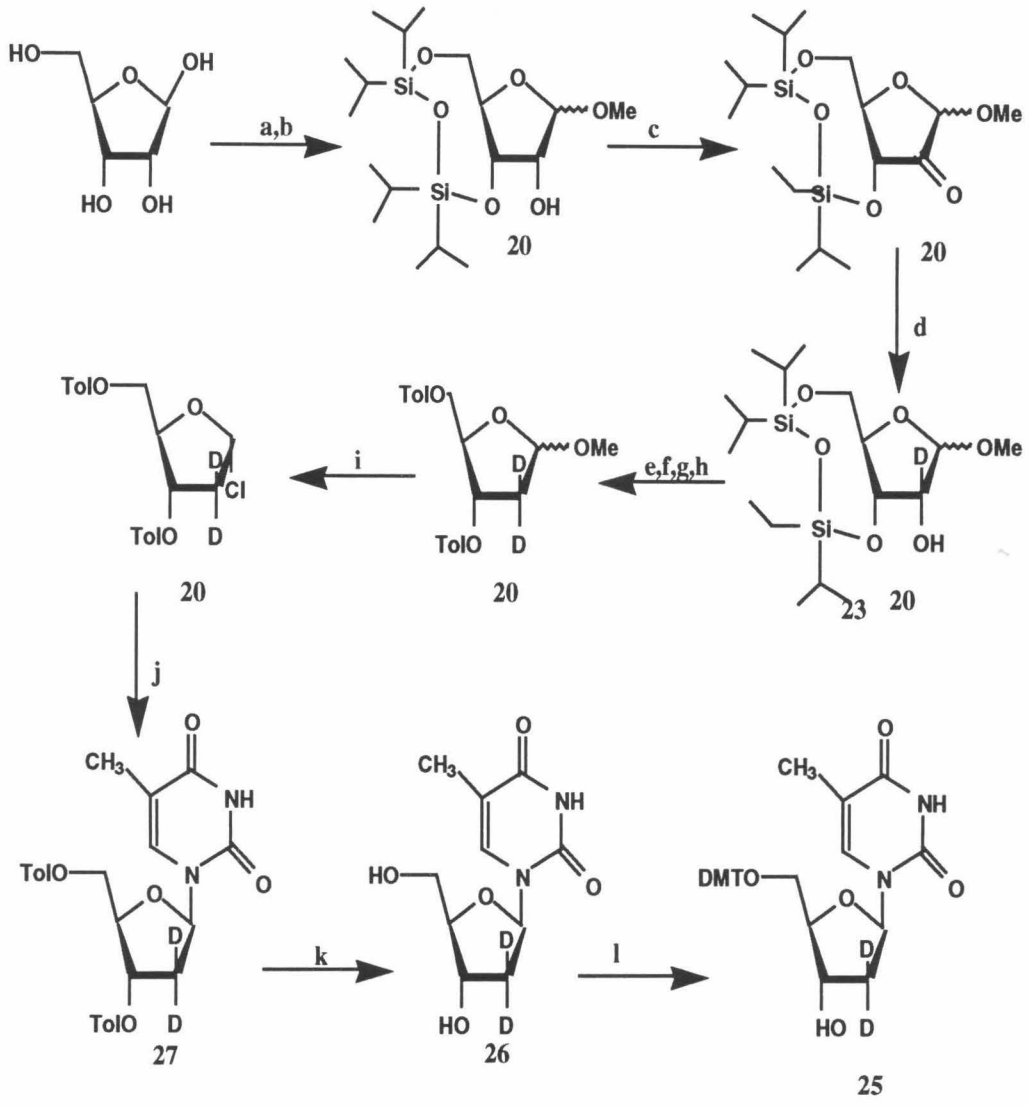
Figure 17 The sequence of a synthetic oligonucleotide duplex used for kinetic isotope experiments. Boxes indicate the 9 bp γδ resolvase binding sites. The structures of the site-specific deuterated thymines are shown below the point of incorporation.



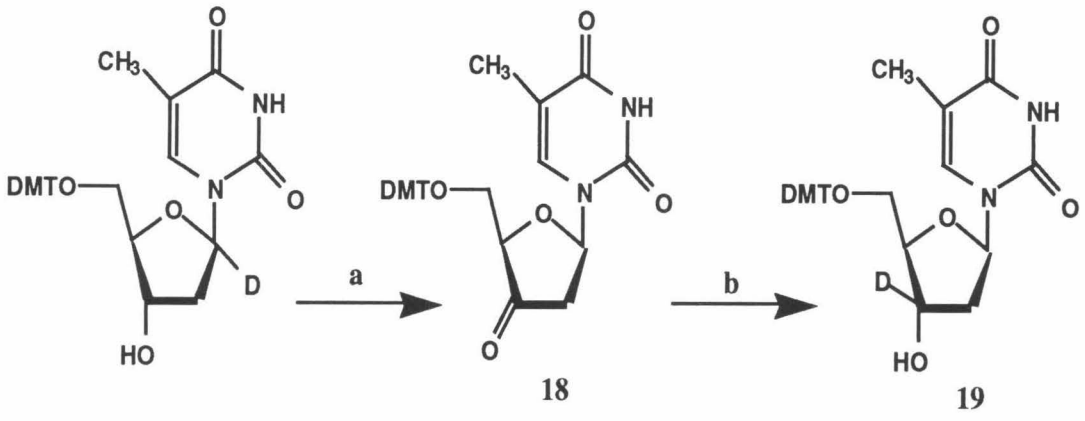
Scheme I. Synthesis of C1'Deuterated Thymidine. a) Disoamylborane, THF; b) Ac_2O, pyridine, 80 C; c) Bis(O-trimethylsilyl)thymine, $SnCl_4$. CH_2Cl_2, 0 C; d) NH_3 MeOH; e) 1,3-dichloro-1,1,3,3,-tetraisopropyldisiloxane, pyridine; f) Phenylchlorothionoformate, DMAP, CH_3CN; g) Bu_3SnH, toluene, 75 C; h) Bu_4NF, THF; i) DMTCl, pyridine.

protected as the tetraisopropyldisiloxane. The 2'-alcohol was then removed using the deoxygenation methodology of Robins, et al (1981). Following deoxygenation and removal of the silyl protecting group, a DMT group was attached to the 5' alcohol. The overall synthetic yield for seven sets was 12.2%, with a deuterium incorporation measured by mass spectrometry of 92.5%.

The synthesis of the 2'-dideutero thymidine is shown in Scheme II. In a one pot synthesis, ribose is cyclized to form the 1-O-methyl compound and protect it at the 3- and 5'hydroxyls as the diisopropyltetrasilyl. This compound is then oxidized via Swern oxidation to form the ketone, **10**. The ketone is then reduced with $NaBD_4$, producing a mixture of the 2-deutero-arabino and ribo diestereomers in a 10:1 ratio. The arabino isomer was found by mass spectrometry to have 100% deuterium incorporation. This isomer was then deoxygenated, using the method of Robins, et al (1981) and the second deuterium was incorporated from tributyltindeuteride. This isotopic incorporation as determined by FAB mass spectrometry of the resulting dideuterodeoxyribose derivative (**13**) was 2% D_1 and 98% D_2. The α-chloro sugar was then synthesized, using the method of Hoffer (1960) and subsequently coupled with bis-O-trimethylsilylthymidine under the conditions of Hubbard, et al (1984). This afforded 76.4 mg of the desired β-isomer **15**. After ammonolysis, the deuterium incorporation of the resulting 2',2'-dideuterothymidine, as measured by DEI mass spectroscopy, was 2% D_1 and 98% D_2. Following attachment of the dimethoxytrityl group to the 5'-hydroxyl, 37 mg of the pure 2',2'-dideutero-5'-O-(4,4'-dimethoxytrityl)-thymidine was obtained in a 2.7% overall yield for nine steps.



Scheme II. Synthesis of C2'Dideuterated Thymidine. a) HCl, MeOH, 0C; b) 1,3-dichloro-1,1,3,3,-tetraisopropyldisiloxane, pyridine; c) oxalyl chloride, DMSO, Et_3N, CH_2Cl_2, -78 C; d) $NaBD_4$, EtOH; e) Phenylchlorothionoformate, DMAP, CH_3CN ; f) Bu_3SnD, toluene, 75 C; g) Bu_4NF, toluene; h) TolCl DMAP, CH_3CN; i) HCl Acetic Acid; j) Bis(O-trimethylsilyl)thymine, $CHCl_3$; k) NH_3 MeOH; l) DMTCl, pyridine.

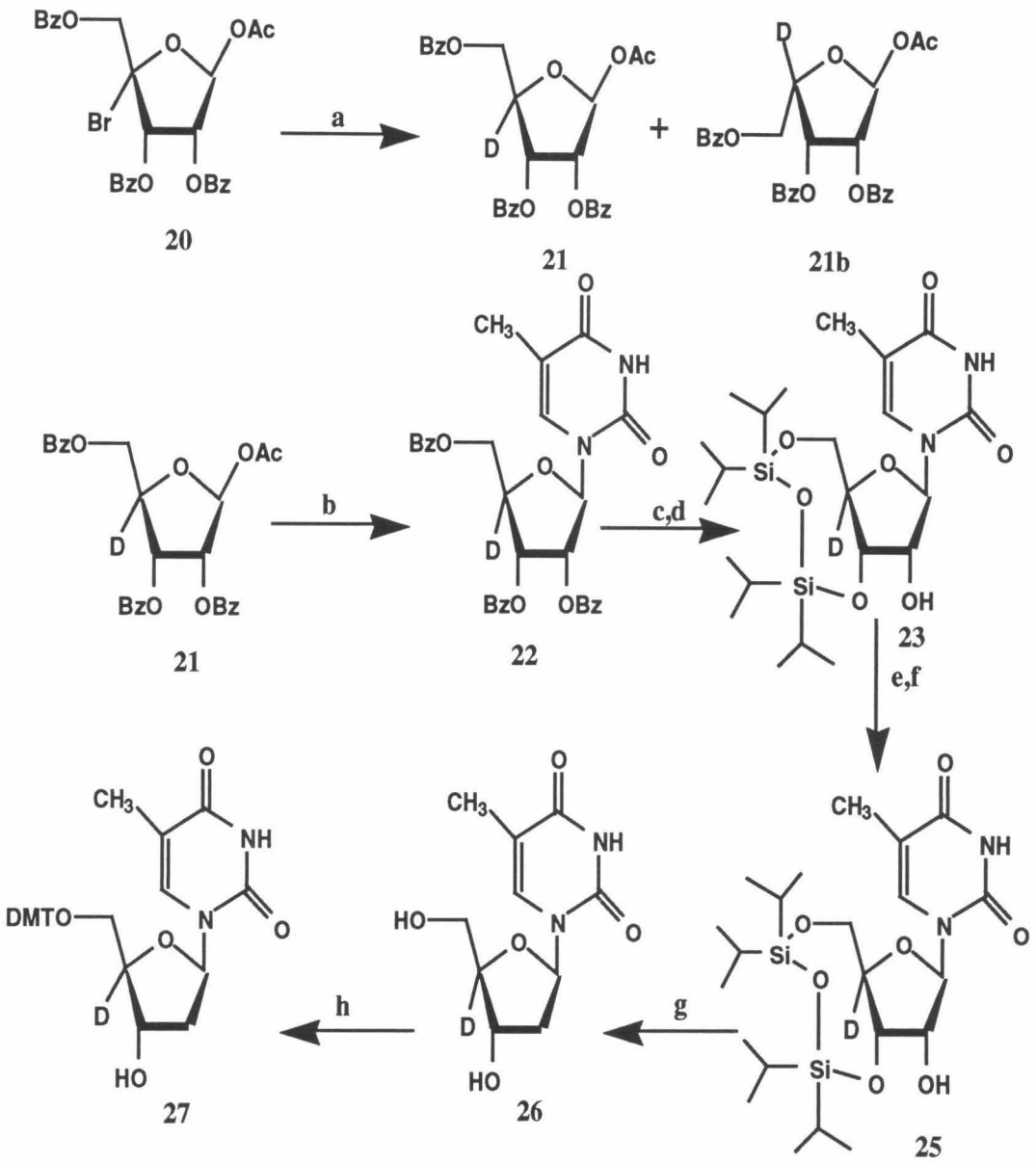


Scheme III. Synthesis of C3'Deuterated Thymidine. a) CrO_3, pyridine, acetianhydride CH_2Cl_2; b) $NaBD_4$, EtOH

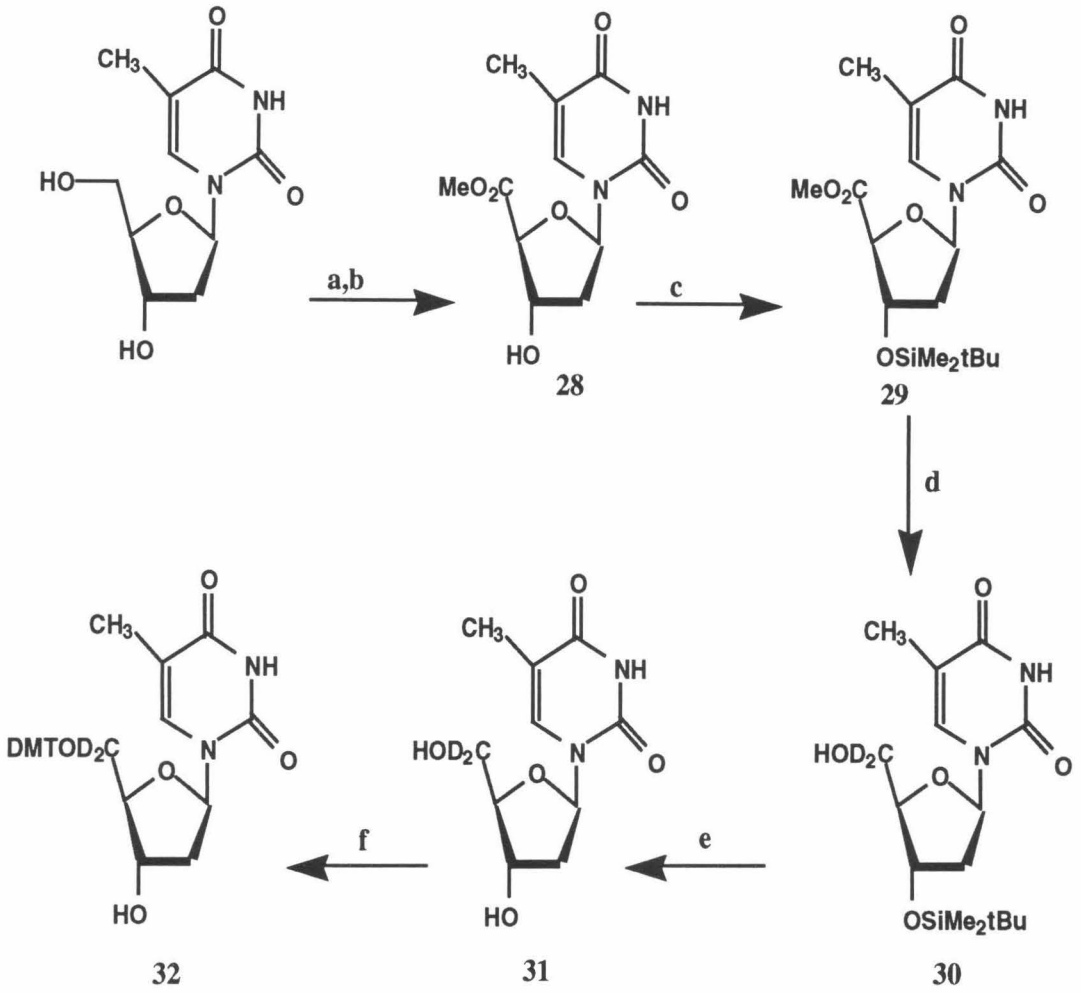
The synthesis of the 3'-deuterothymidine is shown in Scheme III. 5'-DMT protected thymidine is subject to chromium oxidation, forming the ketone **18**. The resulting ketone was then reduced, using sodiumborodeuteride, producing two isomers. The overall yield for two steps of the desired riboisomer was 8.5%. FAB mass spectroscopy showed the percent deuterium incorporation to be 97% D_1.

The synthesis of the C4'-deuterothymidine is shown in Scheme IV. 1-O-acetyl-4-bromo-2,3,5-tri-O-benzoyl-β-ribofuranose was prepared by the method of Ferrier, et al (1984). This compound was subsequently deuterated at position 4 by reaction with the tributyltindeuteride. After isolation of the desired isomer, the 1-O-acetyltribenzoylribose was subsequently converted to the corresponding tri-O-benzoyl-methyluridine by reaction with bis(O-trimethylsilyl)thymidine in the presence of $SnCl_4$ (**13**). Following ammonolysis and protection of the 3' and 5'-hydroxyls as the tetraisopropyldisiloxane, the 2' position was deoxygenated, using the method of Robins, et al (1981). The disiloxane-protecting group was subsequently removed, and the 5' position was protected as the dimethoxytrityl. The overall synthetic yield for seven steps was 2.3%, with a deuterium-isotope incorporation of 94.5% as measured by FAB mass spectroscopy.

The synthesis of the 5-dideuterated thymidine is shown in Scheme V. Initially, the 5'-hydroxyl is oxidized to the corresponding acid and then converted into the O-methylester. The 3' hydroxyl group is then protected as the dimethylisopropylsilanol to improve the solubility of the molecule. Following protection of the 3'-hydroxyl, the 5'-methylester is reduced to the 5'-alcohol, using $LiAlD_4$. The protecting group on the 3'-hydroxyl is subsequently



Scheme IV. Synthesis of C4'Deuterated Thymidine. a) Bu_3SnD, toluene, 75 C; b) Bis(O-trimethylsilyl)thymine, $SnCl_4$. CH_2Cl_2, 0 C; c) NH_3 MeOH; d) 1,3-dichloro-1,1,3,3,-tetraisopropyldisiloxane, pyridine 30 C; e) Phenylchloro-thionoformate, DMAP, CH_3CN; f) Bu_3SnD, toluene, 75 C; g) Bu_4NF, THF; h) DMTCl, pyridine.



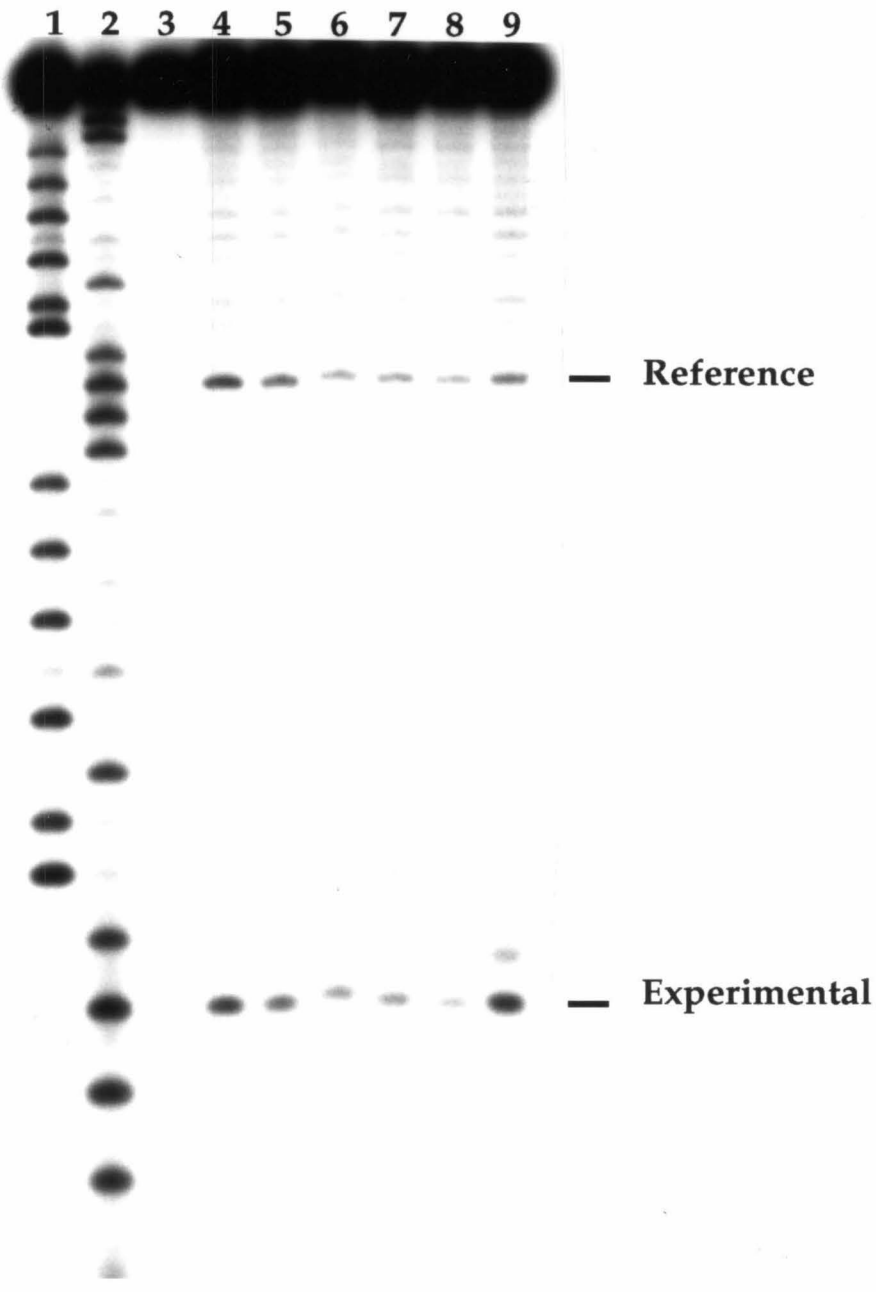
Scheme V. Synthesis of C5'Dideuterated Thymidine. a) PtO_2, O_2, H_2O, 75 °C; b) MeOH, H_2SO_4, reflux; c) tBuMe2SiCl imidazole, DMF, 25 C; d) $LiAlD_4$, THF, 0 C; e) Bu_4NF, THF; f) DMTCl, pyridine.

removed by reaction with tetrabutylammoniafluoride, and the 5'-hydroxyl is then protected, using DMT chloride. This produced the desired 5'-dideutero DMT-protected thymidine (**32**) in a 27% overall yield for five steps. Deuterium incorporation based on FAB mass spectroscopy was found to be 1.5% D_1 and 98% D_2.

Measurement of Kinetic Isotope Effects. Each of the six oligonucleotides (containing C1'-, C3'-, C4'-deuterated or C2'-, C5'-dideuterated, or all-proteo thymidine at the experimental position) was specifically labeled with ^{32}P at the 5'-end. The resulting labeled oligonucleotides were hybridized with the complementary strand and gel-purified, providing the duplex DNA used for the kinetic isotope experiments. Ni(II)-GGH γ δ (141-183) in the presence of monoperoxyphthalic acid was allowed to react with the DNA to produce a total of 1% strand scission after treatment with butylamine. Figure 18 contains a gel that shows cleavage with each one of the specific deuterated or the all-proteo nucleotide by Ni(II)-GGH γ δ (141-183) in the presence of monoperoxyphthalic acid. Examination of this gel shows that the protein cleaves the DNA specifically at only one nucleotide adjacent to the binding site. The site of cleavage at the experimental site occurs at the point in the DNA sequence where the specific mono- or dideuterated thymidines were incorporated.

A close examination of lane #9, which contains the 5'- dideuterated thymidine at the experimental site reveals the existence of an additional cleavage band adjacent to the site at which the 5'-dideuterated nucleotide was incorporated. This band migrates approximately one-half base pair slower than the cleavage band, placing its electrophoretic mobility in between that of DNA fragments ending in phosphate resulting from DNA cleavage at the site of deuterium

Figure 18 An autoradiogram of a high-resolution denaturing polyacrylamide gel containing ^{32}P-end labeled fragments from the oligonucleotide duplex in Figure 17. Lanes 1 and 2 are A and T specific chemical sequencing lanes. Lane 3 contains control DNA. Lanes 4 through 9 are affinity-cleaving lanes containing 1.0 μM Ni(II)•GGH γδ (141-183), and 0.05 μM monoperoxyphthalic acid. All reactions were subject to treatment with 0.1 N-butylamine for 30 min at 90 °C. The reference-cleavage position and experimental-cleavage position are indicated on the right-hand side. Lane 4 contains all-proteo duplex. Lane 5 contains a duplex with a C1'-deuterated thymine at the experimental position. Lane 6 contains a duplex with a C2'-deuterated thymine at the experimental position. Lane 7 contains an oligonucleotide duplex with a 3'-deuterated thymine at the experimental position. Lane 8 contains an oligonucleotide duplex with a 4'-deuterated thymine at the experimental position. Lane 9 contains an oligonucleotide duplex with a 5'-deuterated thymine at the experimental position.


1 2 3 4 5 6 7 8 9
Reference
Experimental

incorporation and the 3'-adjacent nucleotide. Out of 1,200 cleavage experiments that were performed during the course of the kinetic isotope studies, this slower-moving band was observed in only 50 of them, and it was always observed to occur adjacent to a site at which the 5'-dideuterated thymidine had been incorporated. The existence of a DNA fragment whose migration does not correspond to a phosphate termini is very surprising because all the reactions were treated with butylamine to eliminate any alkali labile end products. It is possible that this band may represent some as of yet unidentified "non-alkali labile" DNA end product resulting from oxidative cleavage of DNA. Additionally, since the occurrence of this uncharacterized band is always coincident with the incorporation of 5'-dideuterated thymidine, the possibility must be considered that deuterium incorporation at the 5' position alters the mechanism by which oxidative DNA cleavage occurs.

The relative intensities of the cleavage band produced by Ni(II)-GGH $\gamma\,\delta$ (141-183) in the presence of monoperoxyphthalic acid from the gel in Figure 18 are shown graphically in Figure 19. Panel A, which contains data from an oligonucleotide, which contains only proteo thymidine, shows that the intensity of cleavage at the experimental and control sites is approximately equal, as one would expect. Panel C, which contains data taken from cleavage of an oligonucleotide that has 2'-dideuterated thymidine incorporated at the experimental site, again shows no apparent difference between cleavage intensities at the experimental or reference site, which indicates that there should be no kinetic isotope effect with C2'-deuterated thymidine. Panels B and D, which contain information derived from duplexes that contain thymidine deuterated at the 1' and 3' positions, respectively, show a slightly greater cleavage intensity at the reference site relative to the experimental site.

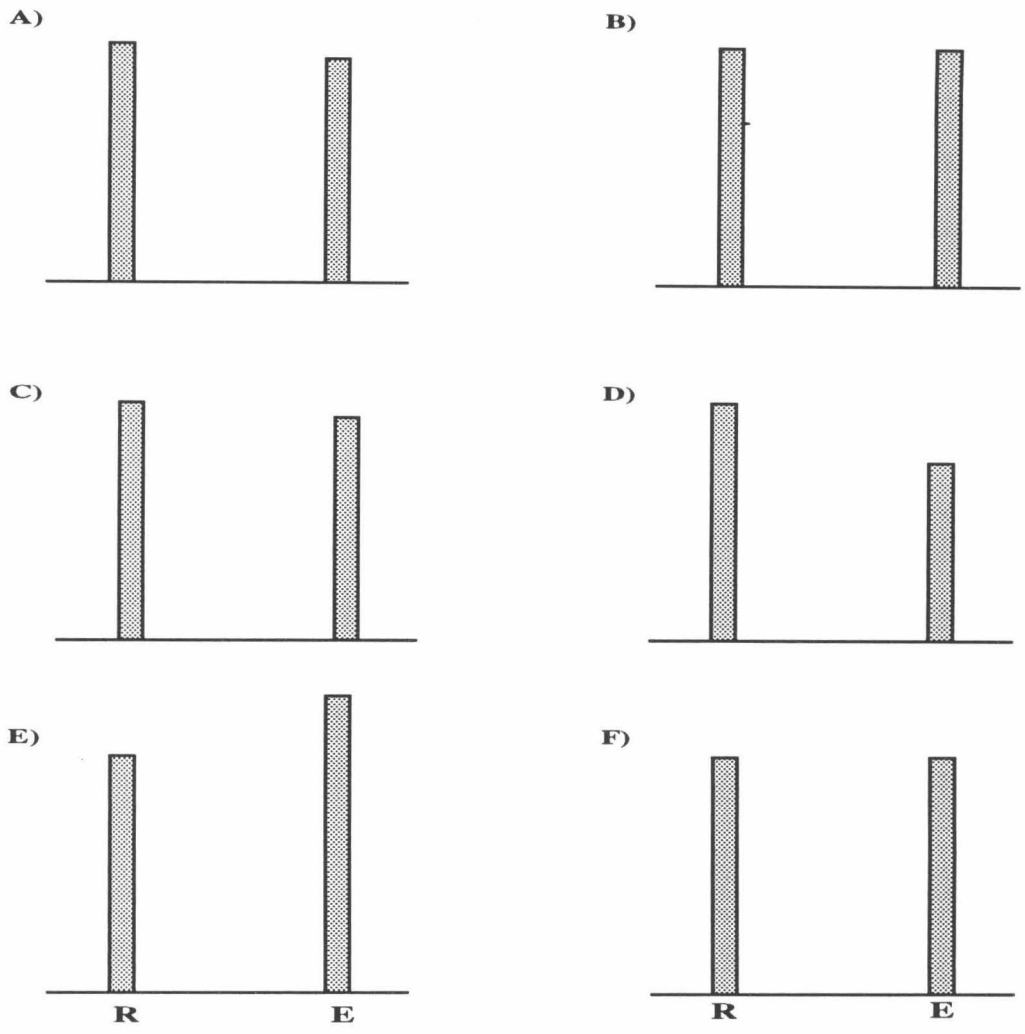


Figure 19 A graphic representation of the cleavage intensities at the reference (R) and experimental sites (E) obtained from phosphorimager analysis of the gel shown in Figure 18. A) Proteo thymine is incorporated at the experimental site. B) C1'-deuterated thymine is incorporated at the experimental site. C) C2'-dideuterated thymine is incorporated at the experimental site. D) C3'-deuterated thymine is incorporated at the experimental site. E) C4'-deuterated thymine is incorporated at the experimental site. F) C5'-dideuterated thymine is incorporated at the experimental site.

This would be consistent with the observation of a small kinetic isotope effect. Examination of Panel E, which contains data derived from an oligonucleotide

duplex with C4'-deuterothymidine incorporated at the experimental site, shows a much larger difference in cleavage intensity between the experimental and reference site. The higher intensity in cleavage observed at the reference site would indicate incorporation of deuterium at a C4' position should produce a significant kinetic isotope effect. The data in Panel F, which are derived from the cleavage of an oligonucleotide that contains C5'-dideuterated thymidine at the experimental cleavage position, show a significantly higher intensity of cleavage at the experimental site relative to the reference site. This type of result would indicate a negative, or inverse, kinetic isotope effect.

Quantitation and Statistical Analysis of Kinetic Isotope Effects. For each oligonucleotide duplex that contained thymidine with deuterium incorporated at a unique position, five gels containing ten experimental reactions and ten proteo control reactions were run. (Five gels containing ten experimental reactions using oligos in which no deuterium had been incorporated and ten proteo-control reactions were also run to serve as a control and to verify that $k_H/k_H = 1$.) Each gel produced one set of ten normalized-cleavage intensities for the experimental case (10 k_D's and 10 normalized-cleavage intensities for the control 10 k_H's). (The normalized-cleavage intensities for each gel are contained in Appendix II.) Since a total of five gels were run for each site of deuterium incorporation, we had collected fifty observations of k_D and fifty observations of k_H. Unfortunately, calculation of KIE (k_H/k_D) was not as simple as determining fifty ratios and taking their mean because of two complicating factors. First, analysis of variance showed that the values of k_D or k_H obtained from different gels could not be pooled. This meant that KIE, or the ratio of k_H/k_D, had to be determined prior to pooling the data from different gels. The second complication arose from a problem with the

experimental design. We had not conducted the experiment in a way that would produce paired values of k_H and k_D. This meant that values of KIE calculated by pairing the values of k_D with values of k_H from the same gel were not statistically meaningful. To overcome this problem, KIE was calculated, using values of k_H and k_D that were randomly generated and paired from the sets of ten observations using a resampling method called "Bootstrap Technique" (Fleiss, 1986; Dunn and Clark, 1987). With this technique, a value was selected from the sets of k_H's and the sets of k_D's. These values were used to calculate a KIE and then replaced into the respective data sets. This cycle was repeated until ten KIE's were calculated. The groups of ten KIE's from each of the five gels that contained the same deuterated nucleotide were pooled and used to calculate a mean value of KIE for the site of deuteration (Table 1). For the purposes of statistical analysis, 420 of these sets of fifty KIE's were generated for each site of deuteration and the control using the "Bootstrap Technique." Student's T tests were then performed to determine if there was a significant difference in the mean between the experimental group and the control group for each of the 420 data sets. Results of the statistical analysis showed that the KIE was significantly larger when deuterium had been incorporated at position C4' relative to the control in all 420 data sets (at the 95% C.I.). The value of KIE in the control was significantly higher than the value observed when deuterium had been incorporated at position C5' (with or without the inclusion of the extra, unidentified cleavage product in the calculation) in all 420 data sets (at the 95% C.I.). When the values of KIE obtained when deuterium were incorporated at positions C1' and C3' were compared with the value of KIE for the proteo control, they were found to be significantly higher than the control group in all 420 data sets. The P value for these two sites of incorporation is slightly larger, although the data

are still significant at the 95% confidence interval. When the value of KIE obtained with deuterium incorporated at the C2' position was compared with the value of KIE from the proteo control, significant differences were obtained with only 30 out of the 420 data sets examined. The value of KIE obtained with deuterium incorporation at position C2' was therefore deemed not statistically different from that of a value obtained with a proteo control.

Table 1. Kinetic Isotope effects (k_H/k_D) for Deuterated Thymidine. Value Represents the Mean of Fifty Rations ± the Standard Error of the Mean (SEM)

	Site of Deuterium Incorporation						
	C1'	C2'	C3'	C4'	C5'	C5'+Ex**	Control
Mean	1.07*	1.00	1.05*	1.27*	0.64*	0.48*	0.98
±SEM	±.01	±.01	±.01	±.01	±.01	±.01	±.01

* Statistical differences from the control at the 95% confidence interval.

** Represents k_H/k_D, including the unidentified cleavage product.

Interpretation of Kinetic Isotope Effects. Any interpretation of isotope effects must involve an understanding of their origin and magnitude. Isotope effects are usually divided into two categories, primary and secondary. A primary isotope effect results when the bond to the isotopically substituted atom is broken in the transition state (Richards, 1970). A secondary isotope effect occurs when the bond to the isotopically substituted atom is not broken, but the

presence of the isotopically substituted atom still produces a perturbation of the reaction rate.

Primary isotope effects arising from deuterium incorporation are frequently discussed as a result of differences in the zero-point energy between carbon-hydrogen bonds and carbon-deuterium bonds (Richards, 1970; Kresge, 1977; Melander, 1960). The difference in zero-point energy is a direct result of the difference in mass between hydrogen and deuterium. It is usually discussed in terms of the stretching frequency for a carbon-hydrogen bond and a carbon-deuterium bond, which are 2,900 cm^{-1} and 2,100 cm^{-1}, respectively. These vibrational frequencies correspond to zero-point energies of 4.15 kcal/mole for carbon-hydrogen bonds and 3.0 kcal/mole for carbon-deuterium bonds. If the primary isotope effects observed with deuterium incorporation were solely a result of the difference in zero-point energy, the maximum value of k_H/k_D that could be observed would be 6.4. However, there are many examples where significantly greater values of k_H/k_D have been observed, indicating that discussing deuterium isotope effects solely in terms of zero-point energy is a gross oversimplification (Kresge, 1977). The factors that account for the observation of the large primary isotope effects (isotope effects > 6.4) are tunneling and bending vibrations. For example, a deuterium-isotope effect corresponding to the complete loss of zero-point energy for one stretching vibration and two bending vibrations results in a k_H/k_D of 18 (Kresge, 1977). While the effects of tunneling and bending vibrations can combine with the normal stretching vibrations to produce extremely large kinetic isotope effects (k_H/k_D's as large as 45 have been observed by Caldin and Mateo, 1975). It is also possible to obtain primary kinetic isotope effects with magnitudes near unity (Kresge, 1977). An explanation of these small, primary isotope effects

was initially offered by Melander (1960) and Westheimer (1961). This explanation is based upon a three-centered model for the transition state of a hydrogen-transfer reaction and assumes that in the transition state the hydrogen that is being transferred is simultaneously bound by an acceptor and a donor. In simplest terms, this description predicts that hydrogen-isotope effects will be greatest when the partial bonds holding the hydrogen being transferred to the donor and to the acceptor are of equal strength, in other words, when the hydrogen is half-transferred at the center of the transition state (Kresge, 1977). The explanations put forth by Westheimer (1961) and Melander (1960) state that any deviation from this symmetric situation in either direction would produce a smaller isotope effect. In other words, if hydrogen transfer does not coincide with the energy maxima of the transition state, a small kinetic isotope effect will be observed. It should be noted that this explanation for small, primary isotope effects uses a linear model for the transition state. In recent years it has become apparent that bending frequencies and non-linear transition states can also play a role in producing small, primary kinetic isotope effects (Jencks, 1987).

As previously stated, secondary isotope effects are caused by isotopic substitution of atoms that are not "reacting." Secondary deuterium-isotope effects are caused by the changes in vibrational and bending frequencies and zero-point energies in ground and transition states that result from isotopic substitution. (Jencks, 1987; Melander, 1960; Richards, 1970) Secondary effects have been observed with deuterium substitution both, α and β, to the reacting atom. A classic example of a secondary isotope effect occurs as a result of the sp^3 sp^2 transition that occurs in carbonium ion formation. Substitution of deuterium α to the reacting carbon center has resulted in observed secondary

isotope effects in the range of 1.15 to 1.28 (Streitwieser, et al., 1958; Shiner, et al., 1968; Shiner, 1952). Secondary isotope effects of this type are probably a result of the change in one of the bending frequencies that occurs as a consequence in the change from sp^3 to sp^2 hybridization. Substitution of deuterium for hydrogen on a carbon atom adjacent (β substitution of deuterium) to the one that becomes cationic can also slow the rate of reaction (Streitwieser, 1958). This β substitution of deuterium is thought to produce a kinetic isotope effect because during the formation of a carbonium-ion, a partial breaking of the deuterium-carbon bond is thought to result from hyperconjugation with the vacant p-orbital of the carbonium ion carbon. This effect has been found to be maximal when the dihedral angle between the vacant p-orbital and the carbon-deuterium bond in question is zero (maximal overlap) and vanishes when this angle becomes 90°C (Streitwieser, et al., 1958; Shiner, 1960). While carbonium-ion formation may provide the classic examples of secondary isotope effects, it is important to realize that these effects occur with other types of reactions in which changes in hybridization from sp^3 to sp^2 are observed (Halevi, 1963; Brown and Cookson, 1965; Sunko and Borcic, 1970). It is interesting to note that in some circumstances secondary deuterium-isotope effects are additive. In other words, a greater effect is observed when two deuteriums are incorporated as compared with one. In other situations, particularly those involving hyperconjugation, this may not be so (Melander and Saunders, 1980).

When the products from reactions involving deuterated substrates can be isolated, it is easy to distinguish between primary and secondary isotope effects because it is possible to determine if product formation involved breaking of the carbon-deuterium bond. However, when experimental conditions necessitate that reactions are run on such a scale that product isolation and

characterization is not possible, distinguishing between primary and secondary kinetic isotope effects becomes somewhat more difficult. When this situation exists, the distinction is usually made on the basis of the magnitude of the observed isotope effect. It is generally accepted that an observed isotope effect of > 2 is a primary effect (Richards, 1970; Shiner, 1970; Harris and Wamser, 1976). It is also generally accepted that observed isotope effects of < 1.1 are typically secondary effects (Harris and Wamser, 1976). The problem comes with interpreting values that are intermediate, those between 1.1 and 2.0. This arises from the fact that there are many observed primary and secondary kinetic isotope effects with magnitudes that fall within this range. As a consequence of this, isotope effects with these intermediate values are typically analyzed on a case-specific basis by taking into account other factors that are known about the reaction mechanism in question. In light of the preceding discussions, an attempt will be made to interpret the kinetic isotope effects observed for Ni(II) GGH γ δ (141-183) cleavage of DNA in the presence of monoperoxyphthalic acid. At position C2' no isotope effect was observed. At positions C1' and C3', small kinetic isotope effects with values of 1.07 and 1.05, respectively, were observed. The small magnitude of these effects in the absence of contradictory information would lead to their interpretation as secondary isotope effects. It would be quite surprising indeed to observe a primary isotope effect at position C3', because the C3'-hydrogen is located in the major groove, and our cleaving moiety is located in the minor groove. These small isotope effects are likely a result of the change in carbon hybridization from sp^3 to sp^2 that would occur with base release from the C1' position and the loss of phosphate from the C3' position. This interpretation might raise the question as to why a β secondary isotope effect was not observed as a result of deuterium incorporation at position C2'. There are two

possible explanations for this. First, hyperconjugation may not have played an important role in the transition states for these reactions. Second, even if hyperconjugation could play a role, the conformational constraints of the deoxyribose-ring, 2'-endo sugar pucker would orient the deuteriums in such a manner that their contribution to an isotope effect was minimized (Saenger, 1984).

At the position C4', an isotope effect of 1.27 was observed. The magnitude of this effect makes its interpretation exceedingly difficult because values of k_H/k_D in this range have been observed for both primary and secondary isotope effects. Potential explanations for a secondary isotope effect with deuterium incorporation at this position would include an α effect that was due to opening of the deoxyribose ring, creating an sp^2 center at the C4' position, or β effects resulting from the loss of phosphate from either the 3' or 5' position, creating sp^2 centers at these locations. A primary isotope effect at this position would result from hydrogen abstraction from the C4' carbon by the activated species generated from Ni(II) GGH $\gamma\delta$ (141-183) in the presence of monoperoxyphthalic acid. Hydrogen-atom abstraction at this position would be consistent with the oxidative nature of the DNA cleavage observed. Neocarzinostatin and bleomycin, which cleave DNA in an oxidative manner, have been found to produce DNA cleavage by abstraction of the C4' hydrogen (Kappen et al., 1991; Frank, et al., 1991; Kozarich, et al., 1989). For nickel monoperoxyphthalic acid-mediated cleavage by the peptide GGH-Hin (139-190) Mack and Dervan (1990) have reported a kinetic isotope effect of 1.6 resulting from deuterium incorporation at the C4' position. If this is, indeed, a primary kinetic isotope effect, there are several explanations for its small magnitude. Two explanations that immediately come to mind, as previously

discussed, would be the effects resulting from either an asymmetric, or non-linear, transition state. While these classical explanations for small-magnitude, primary kinetic isotope effects would be more than adequate to account for our experimental observation, it should be pointed out that in other experiments that examined kinetic isotope effects on the cleavage of DNA, there was found to be a sequence variability in the magnitude of the isotope effect that was observed (Kozarich, et al., 1989; Kappen, et al., 1991; Frank, et al., 1991). In the kinetic isotope effect studies examining neocarzinostatin-mediated cleavage of DNA, primary isotope effects for 5' hydrogen abstraction in the range of 1.3 were observed (Frank, et al., 1991). Since the DNA cleavage observed is oxidative in nature and presumably involves hydrogen abstraction and no other larger kinetic isotope effects are observed, the isotope effect occurring at position C4' is probably a primary kinetic isotope effect with a small magnitude as a result of either characteristics of the reaction's transition state and/or effects of the DNA sequence that is cleaved.

The isotope effect observed at position C5' is probably the most diffult to explain rationally. The k_H/k_D observed for this position is 0.64. This is a negative, or inverse, kinetic isotope effect. In the system we are studying, there are three possible explanations for the occurrence of this type of effect. The first is that incorporation of deuterium at position C5' somehow alters the pathway by which the activated deoxyribose undergoes rearrangement, leading to strand scission. The second possibility is that we have observed a negative, primary isotope effect. And the third possibility is that we have observed a negative, secondary isotope effect.

The first one of these possibilities that we will deal with is that deuterium incorporation somehow alters the mechanism by which the deoxyribose rearranges prior to producing strand scission. This possibility is supported by the observation of an additional cleavage product when 5'-dideuterated thymidine is incorporated into the oligonucleotide duplex. This cleavage product is present even though all the cleavage reactions were subject to base workup using butylamine, indicating that it is some type of non-alkali, labile cleavage product. Inclusion of this addtional product in the calculation of the isotope effect for C5' deuteration leads to a significantly more negative effect of 0.48. The idea of an altered rearrangement mechanism, or change in partitioning of the rearrangement mechanism used by deoxyribose to lead to DNA strand scission, is highly attractive since a non-alkali, labile end product was not previously observed in DNA cleavage mediated by Ni(II) GGH γ δ (141-183) in the presence of monoperoxyphthalic acid. This hypothesis could be readily tested by examining the profile of DNA cleavage products that are produced both in the presence and absence of deuteration at the C5' position, with and without butylamine treatment. Careful characterization of these products and identification of the unknown, non-alkali labile product may provide support for this interpretation. This interpretation would still be consistent with the assignment of the effect at C4' as a primary isotope effect.

The two alternate explanations for the effect observed at C5' would require us to have observed a negative secondary, or primary, kinetic isotope effect. There are examples of both negative secondary and primary kinetic isotope effects in the literature. It has been found that a negative secondary kinetic isotope effect can occur as a result of the change in carbon hybridization from sp^2 to sp^3 (Brown and Cookson, 1965). This explanation for the observation of a negative

kinetic isotope effect as a result of deuterium incorporation at position C5' would be consistent with the mechanism for DNA cleavage, which first involved hydrogen abstraction from the C4' position to generate an sp^2 center, followed by a reaction of either a radical or cation at this position with oxygen or water, respectively. When the C4' position would react with oxygen or water, it would undergo a transition from sp^2 to sp^3 and the deuteriums present at the C5' position would be responsible for producing a negative β secondary isotope effect. O^{18} incorporation studies of bleomycin-mediated cleavage of DNA have shown that the radical generated at position C4' as a result of hydrogen abstraction can readily incorporate molecular oxygen and that a cation generated as a result of electron transfer from this radical can readily incorporate water at the C4' position (McGall and Stubbe, 1991; Rabow, et al., 1990; McGall, et al., 1992). These observations support our interpretation of the negative isotope effect at position C5' as being an inverse β secondary isotope effect resulting from a change in hybridization from sp^2 to sp^3.

The remaining possibility for the negative kinetic isotope effect observed with deuterium incorporation at position C5' would be a primary, negative kinetic isotope effect. It is important to point out that if a negative, primary kinetic isotope effect was observed at this position, our interpretation of the isotope effect at C4' should be changed to that of a secondary isotope effect to be mechanistically consistent. Inverse primary kinetic isotope effects have been observed in the reduction of sterically hindered ketones by LiAl (OBu-t)3D (Wigfield, et al., 1975), the formation of allylic-insertion products by carbenes (Baldwin and Andrist, 1971) and in the hydrogen-transfer reaction involved in the addition of hydrogen bromide to alkenes (Pasto, et al., 1974). In the hydrogen-transfer reactions involving hydrogen bromide and alkenes, kinetic

isotope effects in the range of 0.48 to 0.63 were observed. The magnitude of these observed inverse kinetic isotope effects is consistent with the transfer of a hydrogen ion from a species having a lower vibrational frequency than that of the hydrogen bond that is being formed (Pasto, et al., 1974). In simple terms, an inverse kinetic isotope effect of this type is thought to occur as the result of transferring hydrogen from a species having a weaker bond to the hydrogen to one that has a stronger bond to the hydrogen. Interestingly, our observed inverse kinetic isotope effects for deuterium incorporation at position C5', with or without the inclusion of the extra cleavage product, fall within the range observed for this type of effect. If the negative kinetic isotope effect observed as a result of deuterium incorporation at position C5' was a primary effect of this type, the mechanism of DNA cleavage would involve hydrogen abstraction, or transfer, from the C5' position and the subsequent formation of a bond between this hydrogen and the activated nickel species that is stronger than a carbon-hydrogen bond. Currently, there is no evidence to support the existence of such an intermediate.

Conclusions. The metallopeptide GGH γ δ (141-183) in the presence of nickel and monoperoxyphthalic acid produces highly specific cleavage of DNA adjacent to sites bound by γ δ (141-183). This DNA cleavage occurs at one nucleotide, indicating that it is mediated by some type of non-diffusible species. The nature of the cleavage is oxidative, producing phosphate termini on both the 5'- and 3'-ends. The results of ESR spectroscopy, using the tripeptide GGH and Ni(II) indicate that Ni(III) is not the active species that mediates this cleavage of DNA. Presumably then, DNA cleavage is mediated by some high-valent form of the nickel-peptide complex, presumably a nickel-peroxy or nickel-oxo species. Studies examining the effect of deuterium incorporation at

specific positions within the deoxyribose ring indicate that DNA cleavage may involve abstraction of the hydrogen at the C4' position. Currently, further studies are needed to elucidate the mechanistic details by which Ni(II) GGH γ δ (141-183) cleaves DNA in the presence of monoperoxyphthalic acid.

Experimental

General. UV vis spectra were recorded on a Perkin Elmer Lambda 4C spectrophotometer or a Hewlett Packard model 8451A diode array spectrophotometer. Proton, nuclear magnetic resonance spectra (^{1}H NMR) were taken on one of the following instruments: JEOL FX-90, or JEOL GX-400, Bruker 500 or a General Electric 300 MHz instrument. Chemical shifts (δ) were reported in parts per million (ppm) relative to TMS. Melting points were measured, using a Thomas Hoover Capillary Melting Point apparatus and were uncorrected. Exact mass and high resolution mass spectra were obtained from the NSF regional mass spectroscopy facility at the University of California, Riverside. DNA cleavage reactions were quantitated, using a Molecular Dynamics Phosphorimager model #400S. All chemicals were obtained from either Aldrich or Sigma, with the exception of sodium borodeuteride, which was purchased from Cambridge Isotopes. All enzymes were purchased from Boehringer Mannheim.

Peptide Synthesis and Purification. GGH-γδ(141-183) was synthesized by manual protocols on phenylacetamidomethyl (PAM) resin, substituted with tBoc-Asn. Standard tBoc synthetic protocols and previously described Boc protected amino acids were used. The overall yield for the synthesis was 72.8%. The peptide was deprotected, using the previously described HF protocol and purified via reverse phase chromatography, using a C8 column from Brownlee.

The sequence of the protein was confirmed via amino acid analysis and Edman degradation and mass spectral analysis agreed with the expected weight of the protein.

Labeling of the Hind III / Sal I Restriction Fragment from Plasmid PRW 80. Plasmid PRW 80 (Abdel-Maguid, et al., 1984) containing two copies of *Res* was digested with the restriction endonuclease Hind(III) to produce two fragments, 30 and 5 kilobase pairs in size. Labeling at the 3' end was accomplished using α-^{32}P-dATP with the Klenow fragment of DNA polymerase I. 5'-end labeling was accomplished by dephosphorylation of the 5'-DNA termini with calf alkaline phosphatase, followed by rephosphorylation of the hydroxyl ends using γ-^{32}P-ATP and T4 polynucleotide kinase. Following end labeling, the fragments were cleaved with the restriction endonuclease Sal (I). This yielded three different sized, end labeled fragments. The smallest fragment, 240 bp long, contained the intact *Res* site. This fragment was isolated by non-denaturing polyacrylamide gel electrophoresis, using a 15 cm X 2 mm 5% polyacrylamide gel.

Ni(II) GGH $\gamma\delta$ (141-183) Cleavage Reactions. Cleavage reactions were performed in a buffer containing 20 mM phosphate buffer (pH 7.5) 20 mM NaCl, 100 μM in base pair calf thymus DNA , approximately 30,000 cpm of ^{32}P-end labeled DNA. The reactions contained 0.5 μM, 1.0 μM, or 2.0 μM Ni(II) GGH $\gamma\delta$ (141-183) as noted in the text. The protein was allowed to equilibrate with the DNA in the buffer solution for 15 min prior to the addition of the oxidant. The reactions were initiated by the addition of either hydrogen peroxide or monoperoxyphthalic acid at the concentrations noted in the text. The times the reactions were allowed to proceed are also noted in the text. Reactions were

terminated by ethanol precipitation and dried by lyophilization. If butylamine treatment is indicated, they were resuspended in 50 µl of 0.1 normal N-butylamine and heated at 90 °C for 30 min, then frozen and lyophilized. All reactions were resuspended with 10 µl of 100 mM trisborate EDTA and 80% formamide solution. The resulting ^{32}P end labeled DNA fragments were analyzed by denaturing polyacrylamide gel electrophoresis. The resulting polyacrylamide gels were analyzed either by using autoradiography and densitometric analysis or by imaging with a Molecular Dynamics Phosporimager.

MPE Footprinting Reactions. MPE Footprinting reactions were performed in a buffer containing 20 mM NaCl, 100 µM in base pair, calf thymus DNA, 20 mM tris buffer (pH 8.0), and approximately 30,000 cpm ^{32}P-end labeled DNA. The reactions also contained 10 µM MPE•Fe(II), 5 mM DTT, and 2.0 µM protein when noted. For footprinting reactions, the protein was added to the buffer and allowed to equilibrate for the DNA for 15 min. MPE•Fe(II) was then added and then allowed to equilibrate for 5 min. The reaction was then initiated by the addition of DTT. After 10 min at 25 °C, the reaction was stopped by ethanol precipitation. Comparisons were run between MPE cleavage reactions in the presence and absence of protein.

Oligonucleotide Synthesis and Purification. Oligonucleotides used for end product analysis were synthesized on a Beckman System 1+ DNA synthesizer employing solid phase phosphotriester chemistry (Matteucci and Caruthers, 1981; Beaucage and Caruthers, 1984). Oligonucleotides were cleaved from the solid support and deprotected by treatment with concentrated ammonia hydroxide at 55 °C for 20h. Following removal of the ammonia hydroxide, the

oligonucleotides were purified by preparative polyacrylamide gel electrophoresis. Bands containing the desired product were visualized by UV illumination and excised from the gel. The oligonucleotide was eluted from the gel slice, using a solution that contained 0.2 M NaCl and 20 mM EDTA. After the polyacrylamide was removed by filtration, the resulting supernatant was desalted, using Sephadex G60. The concentration of the oligo in solution was then determined using UV spectroscopy.

5'-End Labeling of Oligonucleotides . 20 pmoles of single-stranded oligonucleotide was 5'-end labeled with γ ^{32}P ATP, using T4 polynucleotide kinase. After the labeling reaction was complete, the solution was ethanol-precipitated and reconstituted with 50 µl of STE buffer. To this solution 2 µl of solution containing 25 pmoles of the complementary strand of DNA were added. The mixture was heated to 90 °C and allowed to cool to room temperature overnight. The resulting labeled DNA duplex was purified using non-denaturing polyacrylamide gel electrophoresis. The labeled bands were visualized by autoradiography, and the desired band was excised from the gel. The labeled oligonucleotide was eluted from the gel slice using a buffer containing 0.2 M NaCl and 20 mM EDTA. After the polyacrylamide was removed by filtration, the solution was desalted, using Sephadex G60 and lyophilized. The DNA was then reconstituted with water to a concentration of 30,000 cpm per µl.

3'-End Labeling of Oligonucleotides. 25 pmoles of single-stranded oligonucleotide were 3'-end labeled with α-^{32}P ddATP and terminal transferase. After the labeling reaction was completed, the DNA was ethanol-precipitated and reconstituted with 50 µl of STE buffer. To this solution 2 µl of a solution

containing 25 pmoles of the complementary strand were added. The mixture was heated to 90 °C and allowed to cool to room temperature overnight. The resulting duplex was isolated, using the same procedures described above for 5'-labeled oligonucleotides.

Analysis of DNA Termini by Gel Electrophoresis. The analysis of DNA termini was conducted using 3'- and 5'-end labeled oligonucleotide duplexes. The conditions used for MPE•Fe(II) cleavage of the DNA were previously described. Ni(II) GGH γ δ (141-183) cleavage of DNA was performed in the previously described reaction buffer. These reactions contained 2.0 μM final concentration of peptide and 1.0 mM final concentration of the oxidant, hydrogen peroxide. Cleavage reactions were allowed to proceed for 90 min at 25 °C prior to ethanol precipitation. The resulting DNA cleavage fragments were analyzed by electrophoresis, using a 20% denaturing polyacrylamide gel. Analysis for the presence of 5'- and 3'-terminal phosphate groups was conducted using the methodology described by Hertzberg and Dervan (1984).

Electron Paramagnetic Resonance Spectroscopy (ESR). ESR spectroscopy was conducted using a Varian E-Line Century Series Spectrometer. The spectra were obtained at -150 °C, using degassed samples. All of the samples contained 5 mM tripeptide nickel complex (Ni(II) GGH) dissolved in a 20 mM phosphate buffer adjusted to pH 7.5. When noted, $IrCl_6^{2-}$ (IV) and hydrogen peroxide were present at final concentrations of 5 mM and 2.5 M, respectively. The pH of the hydrogen peroxide was adjusted to 7.0 prior to addition.

Sequence-Specific Marker Lanes. A-specific cleavage lanes were produced using the protocol of Iverson and Dervan (1987). G-specific and G+A cleavage lanes were produced using the protocol of Maxam and Gilbert (1980).

Protocols Modified for Kinetic Isotope Studies

Oligonucleotide Synthesis and Purification. Oligonucleotide synthesis was conducted on an Applied Biosystems model 490A DNA synthesizer, using Applied Biosystems reagents in all cases except for the specific deuterated thymidines. The standard Applied Biosystems cycle for 1.0 micromolar synthesis, using cyanoethyl phosphoramidites was used. The oligonucleotides were synthesized with the trityl group on. All couplings were done automatically with the exception of the couplings for the specific deuterated thymidines. In the case of the deuterated thymidines, couplings were performed manually, using the protocol described below. In all cases the thymidine was incorporated at base #29 on the oligo. The oligonucleotides differed only in the site of deuteration of the thymidine at this position. The DNA synthesis was allowed to proceed automatically through the completion of the coupling of base #28. After the coupling of base #28, the synthesis was allowed to continue through the detritylation and the wash steps. It was stopped at the eighth step for each column that was an aceytonitrile wash. The column was removed from the DNA synthesizer at this time. A 100 mmolar solution of the desired phosphoramidite was then made by adding 150 microliters of dry acetonitrile to 15 micromoles of phosphoramidite. To this solution, 99 microliters of 500 nmolar tetrazole were added. This mixture was flushed over the column for a period of not less than 4 min., using two oven-dried glass syringes. After this step, the column was returned to the machine,

and the synthetic cycle was allowed to complete normally with the exception that at the step where a phosphoramidite would have been added to the column, only acetonitrile was added. The synthesis was completed, using standard ABI coupling procedures. Upon completion of the synthesis, the columns were removed from the machine and dried on a lyophilizer overnight. The oligo cartridges were then opened and the glass beads placed in a vial to which 2 ml of fresh ammonia hydroxide were added. The vials were tightly capped and heated for 24h at 55 °C in a water bath. The vials were then removed from the water bath and allowed to cool to room temperature. The vials were then uncapped, and the remaining ammonia was removed by a stream of argon. The beads were removed from the remaining solution by a filtration, using a Centrex spin column containing a 0.45 micron filter. The beads were rinsed two times with 1 ml of H_2O, which was also filtered through the spin column. This filtrate was aliquoted into 4 eppendorf tubes, and the tubes were frozen and lyophilized. 1ml of water was added to each of these dry tubes, and the tubes were frozen and relyophilized. 1/4 of each synthesis was then detrytilated by the addition of 300 microliters of 80% acetic acid. The acetic acid was added to the tube and allowed to stand for 20 min., at which time 300 microliters of absolute ethanol were added, and the tubes were frozen and lyophilized. The samples were then reconstituted with 1000 microliters of water and relyophilized. This step was repeated a second time. The resulting deprotected oligos were then purified on a 30 cm X 2mm 20% denaturing acrylamide gel. The gel was run for 36h at 900 volts until the xylene cyanol had run off the bottom of the gel. Oligonucleotide bands were visualized by UV shadowing at 254 nm. The bands were excised from the gel and crushed. The oligonucleotides were eluted from the gel overnight with 4 mls of a solution containing 200 mM NaCl and 1 mM EDTA. The acrylamide was removed from

the solution, using filtration, and the solution was desalted, using Sep-pak C18 columns. The resulting desalted oligos were frozen and lyophilized. They were reconstituted with a known volume of milli Q H_2O and their concentrations were determined, using UV spectroscopy and a calculated extinction coefficient e_{260}=361,368 for the experimental strand and ε_{260}=385,056 for the complementary strand.

5´-End Labeling of Oligonucleotides. 120 picomoles of single-stranded oligonucleotide and buffer were incubated with 50 mCi $\gamma\,^{32}P$ ATP(Amersham) and 30 units of T4 polynucleotide kinase for 20 min. at 37 °C. At this point, 15 picomoles of cold ATP were added to the reaction, and the reaction was incubated for 30 additional min. at 37 °C. The reactions were cooled to room temperature and ethanol-precipitated. Following precipitation, 50 microliters of STE buffer and a 10% molar excess of the complementary strand of DNA were added to each tube. The tubes were heated to 90 °C and allowed to cool slowly to room temperature. The resulting oligonucleotide duplexes were purified on a 15% non-denaturing polyacrylamide gel 15 cm long by 2 mm thick. They were electrophoresed for 5 h at 230 volts. Labeled oligonucleotides were then visualized by autoradiography, and bands containing oligonucleotide duplexes were excised, crushed and eluted into 2 mls of buffer containing 200 millimolar sodium chloride and 1 millimolar EDTA at room temperature for 24h. The acrylamide buffer slurry was filtered through a Quick-Sep fritted filter into a sterile centrifuge tube. The acrylamide layer was rinsed two times with 500 ml portions of milli Q H_2O. The resulting solution was then filtered through a 0.45 micron centrex filter. The solution was then desalted using a NAP 25 column and the manufacturer's procedure. The

resulting desalted, labeled oligonucleotides were concentrated on a speed vac until dryness and reconstituted with water to 25,000 cpm per µl.

T-Specific Cleavage Reactions. 1,000,000 CPM of labeled single-stranded DNA were dissolved in 10 microliters of milli Q water. To this were added 3 microliters of calf thymus DNA (1.4 mM in base pair) was added. 20 µl of 20 µg/ml $KMnO_4$ were then added. The mixture was incubated for 9 min at room temperature, at which time 10 µl of allyl alcohol were added to the solution. This mixture was mixed vigorously with a vortex mixer for 45 sec. It was then frozen and lyophilized on a speed vac lyophilizer. 150 microliters of 1M piperidine were added and heated at 90 °C for 30 min. The solution was frozen and lyophilized. 150 microliters of milli Q water added to the residue, the residue was redissolved, and it was lyophilized an additional time. The reaction was resuspended with an 80% formamide loading buffer to 60,000 CPM/microliter.

Conditions for Ni(II) GGH γδ (141-183) Cleavage. Conditions were chosen to give approximately 1% total cleavage of the DNA duplexes. All reactions were run in a total volume of 20.0 µl. The reactions contained 20 mM NaCl, 20 mM sodium phosphate buffer, pH 7.5, 50 µm in base pair, calf thymus DNA, 75,000 CPM hot DNA duplex, 1.0 µM Ni(2)GGH (γ, δ 141-183) peptide-nickel complex, and 0.05 µM monoperoxyphthalic acid. To minimize the errors that were due to pipeting, for every ten reactions a stock solution was made up that contained the sodium chloride, sodium phosphate buffer, calf thymus DNA, and hot oligonucleotide duplex. This stock solution was then aliquoted into ten individual tubes. To each of these tubes, the protein-nickel complex was added. The protein-nickel complex was allowed to equilibrate with the DNA

in these tubes for 15 min, at which time the monoperoxyphthalic acid was added and the tubes were allowed to react for 10 min at room temperature. The reactions were then stopped by the addition of 2 µl of 20% sodium acetate, pH 5.2, and 70 microliters of absolute ethanol. The resulting mixture was frozen at -70 °C, and the DNA was precipitated by centrifugation. After removal of the supernatant, the tubes were dried on a speed vac lyophilizer for 5 min. 50 microliters of a 0.1 M aqueous solution of butylamine were then added to each tube. The tubes were then heated at 90 °C for 30 min. The tubes were then allowed to cool to room temperature, and the aqueous butyl amine was removed overnight on a speed vac lyophilizer . 15 µl of an 80% formamide loading buffer containing only the dye BPB were then added to each tube, and they were allowed to resuspend for 24 h at 4 °C.

Gel Analysis. Reactions for kinetic isotope experiments were analyzed, using a 25 cm X 0.2 mm 20% denaturing polyacrylamide gel. Each gel contained 24 lanes. Lanes 1 and 13 contained specific marker lanes, either an A or T lane. Lane 2 contained a control lane for the proteo reference duplex, and lanes 3 through 12 contained reaction lanes for the proteo duplex. Lane 14 contained a control lane for the deuterated duplex, and lanes 15 through 24 contained reaction lanes for the deuterated duplex. All gels were run such that the bromophenyl blue had traveled 21 cm.

Data Collection and Workup. Five gels were run for each of the six oligos to be tested. The six oligos were deuterated at either C1', C2', C3', C4', C5', or not deuterated at all (entirely proteo). After electrophoresis, the gels were exposed to a Molecular Dynamics Phosphorimager plate for 10-14 h. The plates were then scanned, using the Molecular Dynamics Phosphorimager. On each gel

there were 20 lanes containing cleavage reactions, and in each of these lanes containing cleavage reactions, there were two sites of cleavage. For ten of the reaction lanes, one of these sites of cleavage contained a thymidine in which one of the positions on the sugar had been deuterated (this was true in all cases except for the control, where an entirely proteo oligonucleotide was compared to an entirely proteo oligonucleotide). The other ten lanes contained an oligonucleotide, which at both sites of cleavage contained ribose that was entirely proteo. When looking at one of these gels, you will see two sites of cleavage; in all cases the upper site of cleavage served as a reference for a given lane, and the lower site of cleavage was used as the experimental case. These cleavage sites were analyzed and integrated by volume on the Phosphorimager; they were corrected for background, and used as the raw data to calculate the kinetic isotope effects. Calculation for kinetic isotope effects was done as follows: A ratio was taken between the lower cleavage site (or experimental case) and the upper cleavage site (or reference). This ratio represented an observation for one given reaction. This was done for each of the ten experimental lanes and for each of the ten that we will call control lanes. To calculate a kinetic isotope effect (KIE), one would then take a ratio of these two previously defined ratios. So, in effect, ten lanes on each gel provided a measurement of reactions, in which case deuterium was present and ten lanes on each gel provided a measurement of a case where there was no deuterium present and the substrate was entirely proteo. By determining the ratio defined by these two sets of ten observations, you would determine ten different values for the KIE or k_H/k_D.

Statistical Treatment and Analysis of Data. For each of the deuterated oligos plus the proteo control, five gels were run. Each of these gels provided ten

observations for the experimental case and ten observations for the control. This meant that there was a total of fifty experimental observations for each treatment and fifty associated control observations. The first issue which needed to be addressed was that of their gel-to-gel variability. If there was no gel-to-gel variability, it would mean that the observations from the five different gels could be pooled and could be treated as one set of fifty, as opposed to five sub-sets of ten with associated controls. Every set of ten experimental observations for each of the five deuterated oligos plus the proteo control was tested for normality, using Geary's test (Fleiss, 1986; Dunn and Clark, 1987). All of the subgroups of ten observations were found to be normally distributed. This meant that the data from each of the five gels for each deuterated oligonucleotide could be compared by an analysis of variance. Unfortunately, in every case comparison of the data from the five different gels showed that there was a gel-to-gel effect; in short, it meant that the data could not be pooled and had to be treated as five sets of ten observations as opposed to one set of fifty observations.

The next question that was addressed was one of intragel variability, and the question of whether there was any rationale to pair a specific individual from the ten control observations with a specific individual from the ten experimental. A comparison of data using different groupings of individual lanes within a gel demonstrated that there were no lane-to-lane positional effects on the kinetic isotope gels. This meant that while there is a statistical difference between gels, there is no statistical difference between lanes on the same gel that contain the same oligonucleotide or experiment. This means that no rationale exists for matching, or pairing, individual values from the experimental data sets on a given gel with individual values from the control

data sets on the same gel. Since a given control observation and experimental observation cannot be matched, the ratio of observations is not statistically meaningful (Fleiss, 1986; Dunn and Clark, 1987).

To obtain a statistically meaningful computation for the kinetic isotope effect ratio, a random sampling technique, called the Bootstrap Method, will be employed (Fleiss, 1986; Dunn and Clark, 1987). The Bootstrap Method for unpaired data involves the generation of a series of data sets based on the random pairing of the original observations. These data sets are, by design, the same size and composition as the originally collected group of observations. For example, for each deuterated oligonucleotide we ran five gels, and each gel contained ten measurements of the experimental case and ten measurements of the unpaired control. Our observed data set, therefore, consists of a total of fifty observations of KIE, ten observations from each of the five gels used to analyze a given deuterated oligonucleotide. Based on this, a data set generated using the Bootstrap Method would contain fifty ratios, or values, for KIE. This set of fifty ratios would comprise ten ratios from each of the five gels. The ten ratios from a given gel would be generated via a random pairing with replacement of observed values for the experimental and the control case. In general, several hundred data sets generated by this method would be compared in order to determine the statistical difference between two groups of data. It is generally accepted that the more random data sets that are compared, the more meaningful the values you get out. Typically, two hundred or more data sets are compared.

In our experiments examining the effect of deuteration on Ni(II) GGH γ, δ (141-183) mediated cleavage of DNA, the following comparisons were set up. In our

case we had five treatments and one control. The five treatments consisted of five different oligonucleotides containing a site-specifically deuterated thymidine at the point of DNA cleavage. The treatments are C1', C2', C3', C4', and C5'. The control case contains no deuteration at any point within the oligonucleotide. It will be referred to as proteo. We are interested in the statistical comparison between the deuterated oligonucleotides individually and the proteo, or control, case. Our comparison was done using the Bootstrap, random sampling, technique. 420 data sets of fifty ratios were computed, using Bootstrap sampling for each experimental and control group. Student's T tests were performed to see if there were significant differences in the ratio between the experimental group and the control for each data set.

Chemical Synthesis of Deuterated Thymidines. All glassware was oven-dried. Reactions were run under an argon atmosphere, unless otherwise noted. Diglyme was distilled from LiAlH4. Tetrahydroforan (THF) was distilled from potassium and benzophenone immediately before use. Triethyl amine, diisopropyl ethyl amine, acetonitrile (CH_3CN) and methylene chloride (CH_2CL_2) were distilled from CaH_2 before use. 1,1-dichloroethane and toluene were distilled from P_2O_5. Pyridine, N,N-dimethylformamide (DMF) and dimethyl sulfoxide (DMSO) were from Fluke over molecular sieves and used with no further purification. Oxalyl chloride was distilled immediately before use.

2,3,5-O-tribenzoylribolactone, (1). Benzoylchloride 7.8 g (55.6 mmoles) in chloroform, 6ml, was cooled in an ice bath. Pyridine, 8 ml, in chloroform was also cooled in an ice bath and then slowly added to the benzoylchloride. Solid ribolactone, 2g, 13.5 mM, was added over the course of 30 min. The solution

was allowed to warm to 25°C and stirred for 36h, at which time it was diluted with chloroform, 200 ml. Solid sodium bicarbonate, 8g, 0.1M, was added slowly to the mixture. When gas evolution ceased, the residue was dissolved in water, 50 ml, and the layers were separated. Organic layer was further extracted with water, 3 X 50 ml, washed with brine, 50 ml, and dried over sodium sulfate. Upon evaporation, the crystalline mass was recrystallized from ethanol chloroform 8:2. Melting point was 89-92 °C. Proton NMR 300 MHz in $CDCl_3$. δ 7.9-8.1 (m, 6H); 7.3 -7.65 (m, 9H); 6.2 (d, 1H; J=6 Hz), 5.96 (d, 1H; J=6 Hz); 5.08 (t, 1H; J=3 Hz); 4.7-4.8 (dd, 2H; J=3 and 14 Hz).

Preparation of deuterated diisoamylborane. The diborane generated was released through a side arm and bubbled through THF 40 ml at - 70°. Sodium boro deuteride 3.06 g (73.2 mmoles) in diglyme, 100 ml, were added over 1.5h, via an addition funnel to borontrifluorideetherate, 20.87 g, 146.4 mM. After the addition was complete, the diglyme solution was heated to 70 °C for 30 min. The diborane THF solution was warmed to 15 °C, and 2-methyl-2-butene in THF, 2 M, 40 ml, were added by a syringe pump over 1h. The solution was then stirred for 6h at this temperature.

1-O-acetyl-1-deutro-2,3,5-O-tribenzoyl-β-d-ribofuranose, (**2**). 2,3,5-O-tribenzoyl lactone, 3.65 g, 9 mM, in THF, 10 ml, were added to the THF solution of diisoamylborane prepared above and stirred for 60 h at 25 °C. The solution was cooled to 0 °C and carefully quenched with water. Additional water, 5 ml, was added and the solution was refluxed for 1h. After removing the THF on a rotary evaporator, the residue was taken up in 50 ml of acetone:water (8:2), and cooled to 0 °C. 30% hydrogen peroxide, 1 ml, was added and the pH was adjusted to 7.5 with 3 N sodium hydroxide, and the solution was stirred for 10

min. Dimethylsulfide, 5 ml, was added and the mixture was stirred for 30 min before extracting with chloroform (3 X 200 ml). The crude lactol solution was dried over sodium sulfate and concentrated *in vacuo*. The crude lactol was heated at 90 °C in pyridine (10 ml) and acetic anhydride (10 ml) for 4h. The solution was cooled to 25 °C, poured over ice, and extracted with methylenechloride (4 X 100 ml). The material was then dried over magnesium sulfate and concentrated on a rotary evaporator. Flash chromatography with 1:4 ethyl acetate hexane, followed by recrystallization from methanol yielded 1.79 g of **2**, 40% total yield. Melting point 124-125 °C, proton NMR 300 MHz, CD, CL 3. Delta 7.89 (d, 2H, J= 7 Hz); 7.86 (d, 1H, J=7Hz); 7.3-7.6 (m, 10 H, 5.89, dd, 1 H, J=5, 6Hz); 5.76 (d, 1H, J=5Hz); 4.73-4.78 (m, 2H); 4.48 (dd, 1 H, J= 5, 13 Hz); 1.98 (s, 3H). Mass Spec +ve DCI (NH_3) observed $(M+ MH_4)^+$ at M/Z = 523.18 (523.18 calculated for C_{28} H_{27} D_1 O_9 N_1).

1'-deutro-2',3',5'-O-tribenzylmethyluridine, (**3**). $SnCl_4$, 583 mg (2.25 mM) was added via a gas-tight syringe to 10 ml of dichloromethane containing 1.52 g (3.3 mM) of **2** and 1.22 g (4.5 mmoles) of bis(O-trimethylsilyl)thymidine. After stirring for 10h, the solution was poured into dichloromethane (100 ml) and washed with saturated sodium bicarbonate (50 ml) and filtered through celite. The organic layer was dried over sodium sulfate and concentrated *in vacuo*. Flash chromatography, ethyl acetate:methylenechloride (1:5) yielded 1.33 g, 76% of **3** as a white crystalline solid. Melting point 158-160 °C, proton NMR 300 MHz, $CDCL_3$, delta 7.92-8.4 (m, 6H); 7.34-7.64 (m, 9H); 7.14 (q, 1H, J=1.3 Hz); 5.89 (dd, 1H, J=5,6 Hz); 5.74 (d, 1H, J=6 Hz); 4.88 (dd, 1H, J=3, 13 Hz); 4.69 (dt, 1H, J=5,3 Hz); 4.63 (dd, 1H, J=3,13 Hz); 1.57 (d, 3H, J-1.3 Hz). Mass Spec

+ve FAB CH_2Cl_2/NBA observed MH^+ at M/Z = 572.18 (572.18 calculated for C_{31} H_{26} D_1 N_2 O_9).

1'-deutro-3',5'-O-1,1,3,3-tetraisopropyldisyloxanemethyluridine, (4). 1.01 g, 1.77 mM of **3** were dissolved in 10 ml of methanol. This solution was saturated with ammonia at 0 °C. The flask was stoppered and stirred for 12h at 25 °C, at which time the mixture was evaporated to dryness *in vacuo*. To remove any remaining traces of methanol, the residue was taken up in 15 ml of pyridine and evaporated to dryness *in vacuo*. Three such cycles were repeated. The residue was then taken up in 10 ml of dry pyridine and allowed to react with 1,3-dichloro-1,1,3,3-tetraisopropyldisyloxane, 613 mg, 1.94 mM for 2h at room temperature. The crude mixture was poured into water (10 ml) and extracted with diethyl ether (3 X 25 ml). The organic layer was washed two times with pH 7 buffer and dried once with brine, 20 ml. The organic layer was dried over sodium sulfate. Flash chromatography, ethyl acetate:methylenechloride (1:3) yielded 728 mg, 81% of the desired product, melting point 79-81 °C. ^1H NMR 300 MHz, $CDCL_3$. Delta 8.3, (bd, s, 1H); 7.34 (s, 1H); 4.41 (dd, 1H, J=5,8 Hz); 4.14 (m, 2H); 3.9804.05 (m, 2H); 2.95 (bd, s, 1H); 1.89 (s, 3H); 1.01-1.07 (m, 28H). Mass Spec +ve FAB CH_2Cl_2/NBA observed MH^+ at M/Z = 502.25 (502.25 calculated for C_{22} H_{40} D_1 N_2 O_7 Si_2).

1'-deutro-2'-O-phenylthiocarbonyl-3',5'-O-1,1,3,3-tetraisopropyldisyloxyl-methyluridine, (5). 221 mg, 0.44 mM of compound **4** and 70 mg, 0.57 mM of DMAP were dissolved in 4 ml of acetonitrile. To this solution, 91 Mg, 0.53 mM of phenylthiochloroformate were added. The solution was stirred for 4h at room temperature. The solvent was removed on a rotary evaporator, and the residue was purified by flash chromatography, ethyl acetate:methylenechloride

(1:9). The desired product was obtained in 80% yield, 225 Mg. Melting point 70-73 °C. ^{1}H NMR 300 MHz, CDCL3. δ 8.15 (bd, s, 1H); 7.36-7.42 (m, 3H); 7.26-7.3 (dd, 1H, J=7-8 Hz); 7.08-7.10 (d, 2H, J=8 Hz); 5.99 (d, 1H, J=6 Hz); 4.55 (dd, 1H, J=1,6 Hz); 4.21 (d, 1H, J=13 Hz); 4.01 (m, 2H); 1.90 (s, 3H); 1.00-1.09 (m, 28H). Mass Spec +ve FAB CH_2Cl_2/NBA observed MH^+ at M/Z = 638.25 (638.25 calculated for C_{29} H_{44} D_1 N_2 O_8 Si_2).

1'-deutro-3',5'-O-,1,1,3,3,-tetraisopropyldisyloxlythymidine, (**6**). 225 Mg, 0.35 mM of **5**, 123 Mg tributyltin hydride and 10 Mg AIBN were dissolved in 4 ml of toluene. The resulting solution was subject to three rounds of freeze/thaw degassing. The reaction mixture was then immersed in an oil bath at 75 °C for 45 min. After cooling to room temperature, the toluene was removed on a rotary evaporator equipped with a -78 °C cold finger and evacuated with a rough pump. Chromatography ethyl acetate:dichloromethane (1:6) yielded 140 Mg, 82% of **6**. Melting point 148-151 °C. ^{1}H NMR 300 MHz CDCL3. Delta 7.95 (bd, s, 1H); 7.40 (q, 1H, J=1 Hz); 4.46 (ddd, 1H, J=7,8,10 Hz); 4.08 (d, 1H, J=13 Hz); 4.00 (d, 1H, J13 Hz); 3.72 (d, 1H, J=7 Hz); 2.43-2.49 (m, 1H) 2.22-2.26 (m, 1H); 1.90 (d, 3H, J=11 Hz); 1.00-1.09 (m, 28H). Mass Spec +ve FAB CH_2Cl_2/NBA observed MH^+ at M/Z = 486.26 (486.26 calculated for C_{22} H_{40} D_1 N_2 O_6 Si_2).

1'-deutrothymidine, (**7**). 140 Mg, 0.29 mM of **6** were dissolved in 5 ml of THF. To this, 0.9 ml of a 1.0 M solution of $Bu_4N^+F^-$ in THF was added. After stirring at 30 min at room temperature, the THF was removed on a rotary evaporator. Chromatography in methanol:dichloromethane (1:5) yielded 64 Mg, 91% of the desired product, **7**. Proton NMR 90 MHz D_2O. Delta 7.72 (bd, s, 1H); 4.30 (d, 1H, J=6 Hz); 3.95 (m, 1H); 3.7-3.8 (m, 2H); 2.24 (dd, 2H, J=5,8 Hz); 1.76 (s, 3H).

Mass Spec +ve FAB CH_2Cl_2/NBA observed MH^+ at M/Z = 244.10 (244.1 calculated for C_{10} H_{14} N_2 D O_5). Percent deuterium incorporation as determined by FAB mass spectroscopy 92.5% D_1 and 7.5% D_0.

1'-deutro-5'-O-dimethoxytridylthymidine, (8). 63 mg (0.26 mmoles) deutrothymidine were taken up in pyridine, 5 ml. This was evaporated to dryness under vacuum three times to insure the dryness of the compound. The resulting residue was taken up in 2 ml of dry pyridine, blanketed under argon, and cooled in an ice bath. 97 mg, 0.28 mmoles of dimethoxytridylchloride were added and allowed to react for 6h. The reaction was quenched with methanol, 1 ml, poured into water, 10 ml, and extracted with diethyl ether. The organic layers were washed twice with pH 7 buffer and once with brine, 20 ml, and dried over sodium sulfate. After removing the ether *in vacuo*, the residue was purified by silica gel chromatography, using ethyl acetate:hexanes:methanol (40:55:5). The desired product was obtained in 83% yield, 120 Mg. Melting point 120-122 °C. Proton NMR 400 MHz $CDCL_3$. Delta 8.05 (bd, s, 1H); 7.18-7.54 (m, 10H); 6.81 (m, 6H); 4.55 (m, 1H); 4.01 (m, 1H); 3.78 (s, 6H); 3.46 (dd, 1H, J=1,10 Hz); 3.35 (dd, 1H, J=1, 10 Hz); 2.30-2.45 (m, 2H); 1.46 (s, 3H). Mass Spec +ve FAB CH_2Cl_2/NBA observed MH^+ at M/Z = 545.23 (545.23 calculated for C_{31} H_{31} D_1 N_2 O_7). Percent deuterium incorporation as determined by FAB mass spectroscopy 92.5% D_1 and 7.5% D_0.

1-O-methyl-3,5-O-tetraisopropyldisilyribofuranicide, (9). 0.8 gram (22 mmoles) of hydrogen chloride gas was dissolved in 285 g (225 ml) methanol. To this solution, 10 g (66 mmoles) of ribose were added. The mixture was stirred at room temperature for 4.5h, at which time 75 ml of pyridine were added to neutralize the solution. The methanol and pyridine were removed

under vacuum and 75 additional ml of pyridine were added to the residue, and it was also removed under vacuum. The residue was dissolved in 200 ml of dry pyridine. The flask was blanketed with argon and chilled to 0 °C. 21.8 g (69 mmoles) of tetraisopropyldisilychloride were added slowly via syringe. After 7h of stirring at 0 °C, the reaction was warmed slowly to room temperature, and the pyridine was subsequently removed under vacuum. The residue was taken up in 500 ml of dichloromethane and extracted one time with 500 ml of 10% citric acid. The aqueous layer was back-extracted with dichloromethane (2 X 250 ml). The organic layers were pooled and extracted with a 1:1 mixture of pH 7 buffer and saturated sodium chloride (2 X 200 ml), and one time with saturated sodium chloride (200 ml). The dichloromethane was dried with magnesium sulfate and removed under vacuum. The residue was purified by silica gel chromatography (ethyl acetate:hexane 1:10). This afforded 22.493 grams of pure material in an 84% overall yield for both steps. ^{1}H NMR (300 MHz CD CL_3) δ 4.83 (s, 1H), 4.52-4.48 (m, 1H), 4.06-4.00 (m, 4H), 3.75 (t, 1H, J = 9 Hz), 3.32 (s, 3H), 3.0 (s, 1H), 1.10-1.02 (m, 28H). Mass Spec +ve FAB Observed: $(M - H)^+$ at M/Z = 405.21 (405.21 calculated for C_{18} H_{37} O_6 Si_2).

1-O-methyl-2-keto-3,5-O-tetraisopropyldisylribofuranicide, (10). 1.29g (10.0 mmoles) of oxalochloride were dissolved in 30 ml of dry dichloromethane. The mixture was blanketed under argon and cooled to -78 °C. 1.66 g (21.3 mmoles) of DMSO were dissolved in 16 ml of dichloromethane. This was slowly added to the mixture at -78 °C via a syringe. This mixture was stirred for 10 min at -78 °C. 24 ml of dichloromethane containing 3.46 g (8.5 mmoles) of **9** were added slowly via syringe. Upon completion of the addition of the ribofurancide, the mixture was stirred for 20 min at -78 °C. A white precipitate was observed to

form. After the 20 min, 4.3g (42.5 mmoles) of triethylamine were added to the solution. The reaction was warmed to -40 °C and maintained at this temperature for 30 min. The mixture was subsequently added to 150 ml of cold diethylether and extracted with 150 ml of 10% aqueous citric acid. The citric acid was back-extracted with 3 portions of 50 ml of diethylether. The organic phases were pooled and were extracted with pH 7 buffer (3X 50 ml) and then extracted with brine (1 X 150 ml). The ether was then dried with sodium sulfate and removed under vacuum. The resulting residue was purified with silica gel chromatography (ethyl acetate:hexane:triethylamine 200:1000:1). This afforded 3.18 g of pure ketone in 92% overall yield. ^{1}H NMR (500 MHz $CDCl_3$) δ 4.69 (s, 1H), 4.59 (m, 1H), 4.09 (t, 1H, J = 8 Hz), 3.89 (m, 1H), 3.81 (m, 1H), 3.45 (s, 3H), 1.09-0.98 (m, 28H). Mass Spec CI/NH_3 Observed $(M+NH_4)^+$ at M/Z = 422.24 (422.24 calculated for $C_{18}H_{40}O_6N_1Si_2$).

1-O-methyl-2-deutero-3,5-O-tetraisopropyldisylarabanofuranicide, (11). 1.06g (25.6 mmoles) $NABD_4$ were placed in a dry flask and blanketed with argon. To this 68 ml of absolute ethanol were added and 1.53 ml (20.8 mmoles) of acetone were added. The mixture was stirred for 30 min at room temperature. After this time, 2.7 g (6.7 mmoles) of ketone, **10**, in 20 ml of absolute ethanol were added. This mixture was stirred for 14h at room temperature. At this time, the ethanol was removed under vacuum and the residue was taken up in 125 ml of ethyl acetate. The ethyl acetate was extracted with pH7 buffer (3 X 50 ml) and washed with brine (50 ml). The ethyl acetate was dried with sodium sulfate and removed under vacuum. The resulting crude material contained a mixture of the 2-deuteroarabino and ribodistereomers. The distereomers were separated and purified, using silica gel chromatography (ethyl acetate:hexane 7:100). After chromatography, 2.17 g of the arabino isomer were obtained (79.6

% yield) and 204 mg of the ribo isomer (7.5% yield). The overall yield for the reaction was 87%, and the ratio of arabao to ribo isomers was 10:1. Only the arabino isomer was used in subsequent synthetic steps because it was found that the ribo isomer contained proteal contamination. Arabino isomer ^1H NMR (300 MHz CD CL_3) δ 4.86 (s, 1H), 4.33 (d, 1H, J = 6Hz), 4.1-4.06 (dd, 1H, J = 10,3 Hz), 3.90-4.00 (m, 1H), 3.87 (t, 1H, J = 11 Hz), 3.51 (s, 3H), 2.39 (s, 1H), 1.22-1.09 (m, 28H). Mass Spec +ve FAB Observed $(M-H)^+$ at M/Z = 406.22 (406.22 calculated for C_{18} H_{36} D_1 O_6 Si_2). Isotopic incorporation as determined by +ve FAB is 100% D1.

1-O-methyl-2-deutero-2-O-phenythiocarbonate-3,5-O-tetraisopropyldisily arabinofuranose, (12). 1.05g (8.6 mmoles) dimethylaminopuridine and 1.46g (3.6 mmoles) of alcohol, **11,** were dissoved in 60 ml of dry acetonitrile and the flask was blanketed in argon. To this solution 865 μl (6.1 mmoles) of phenylchlorothiocarbonate were added via syringe. The mixture was stirred for 48h under an inert atmosphere. The acetonitrile was removed under vacuum. The residue was taken up in hexane, and the insoluble material was filtered off. The resulting filtrate was purified by silica gel chromatography (benzene:hexane 1:1). This afforded 1.42 g of the desired product in a 72.1% yield. ^1H NMR (400 MHz, CD, CL_3) δ 7.39-7.43 (m, 2H); 7.39-7.29 (m, 1H); 7.25-7.10 (m, 2H); 5.20 (s, 1H); 4.74 (d, 1H, J = 6 Hz); 3.94-4.02 (m, 2H); 3.85 (t, 1H, J = 9 Hz); 3.39 (s, 3H); 1.02-1.11 (m, 28H). Mass spectroscopy +ve FAB CH_2CL_2/MBA Observed: $(M-H)^+$ at M/Z = 542.22 (542.22 calculated for C_{25} H_{40} D_1 O_7 Si_2 S_1).

1-O-methyl-2,2-dideutor-3,5-O-tolyldeoxyribofuranose, (13). 1.28 g (4.3 mmoles) tributyltindeuteride, 1.3 g (2.4 mmoles) of compound **12** and 15 mg

AIBN were dissolved in 130 ml of dry benzene. The solution was subjected to three rounds of freeze/thaw degassing and placed under an argon atmosphere. The reaction was heated at 75 °C for 3h under argon, at which time it was removed from the heat and allowed to cool to room temperature. 10 ml of 1 molar $Bu_4N^+F^-$ in THF were added to the reaction mixture and the reaction was stirred for 12h at room temperature. The benzene and THF were removed under vacuum, and the brown residue was dissolved in 20 ml of dry acetonitrile and the flask was blanketed in argon. 1.5 g (11.8 mmoles) of dimethylaminopuridine were dissolved in the solution. 940µl (7.11 mmoles) of toluchloride were added to the mixture via syringe. The mixture was stirred for 18h at which time significant starting material remained. An additional 600 microliters (4.5 mmoles) of toluchloride were added, and the reaction was stirred for an additional 22h. At this time, no starting material was detected. The acetonitrile was removed under vacuum, and the residue was dissolved in 500 ml of diethylether. The diethylether was extracted with 250 ml of water. The water was back-extracted with diethylether (3 X 100 ml). The organic phases were pooled and extracted with saturated $NaHCO_3$ (2 X 100 ml). Subsequently, the organic phase was extracted with pH 7 buffer (2 X 100 ml) and saturated sodium chloride (150ml) and then dried with sodium sulfate. The ether was removed under vacuum, and the resulting residue was purified using silica gel chromatography (diethylether:CH_2Cl_2 1:100). After purification, 372 milligrams of the desired product was obtained. The overall yield for three steps was 41%. ^1H NMR (400 MHz CD CL_3) δ 7.97 (d, 2H, J = 8 Hz); 7.91 (d, 2H, J = 8 Hz); 7.23 (d, 2H, J = 8 Hz); 7.21 (d, 2H, J = 8 Hz); 5.59 (d, 1H, J = 1 Hz); 5.23 (s, 1H); 4.46-4.57 (m, 3H); 3.36 (s, 3H); 2.41 (s, 3H); 2.39 (s, 3H). Mass spec +ve FAB $CDCL_3$/NBA Observed: $(M-H)^+$ at M/Z = 385.16

(385.16 calculated for C_{22} H_{21} D_2 O). Isotopic incorporation determined by +ve FAB 2% D_1 and 98% D_2.

1-α–Cl-2,2-dideutero-3,5-O-tolyldeoxyribofuranose, (14). 278 mg (0.72 mmoles) of compound **13** were dissolved in 1.0 ml glacial acetic acid. This was added to 1ml of glacial acetic acid that was saturated with hydrogen chloride gas. Hydrogen chloride gas was bubbled through the resulting mixture until a precipitate formed. The solution was filtered through a medium glass frit, and the precipitate was washed with 50 ml of diethylether. The precipitate was dried under high vacuum. This yielded 131 mg of the chloro sugar (46.7% yield). Note: The chloro sugar is highly unstable and must be stored under vacuum with an acid trap. Ideally, it should be used immediately after synthesis. MP 120-121 °C DC. ^{1}H NMR (3 MHz CD CL_3) δ 8.02 (d, 2H, J = 7 Hz); 7.92 (d, 2H, J = 7 Hz); 7.25 (d, 2H, J = 7 Hz); 7.23 (d, 2H, J = 7 Hz); 6.49 (s, 1H); 5.58 (d, 1H, J = 3 Hz); 4.87 (m, 1H); 4.76 (dd, 1H, J = 3,12 Hz); 4.69 (dd, 1H, J = 3, 12 Hz); 2.44 (s, 3H); 2.42 (s, 3H). Mass spec DCI/NH_3 Observed: (M-Cl) at M/Z = 355.38 (355.37 calculated for C_{21} H_{19} D_2 O_5).

2',2'-dideutero-3',5'-O-tolylthymidine, (15). 228 mg thymine (1.8 mmoles) were refluxed overnight in a mixture of 4 ml chlorotrimethylsilene and 8 ml of hexamethyldisylizane. The solution was allowed to cool to room temperature, and the volatile material was removed on a rotovap under aspirator vacuum. 16 ml of chloroform, $CHCl_3$, freshly distilled from P_2O_5, were added to the residue. 125 mg of the chloro sugar (0.32 mmoles) were added as a solid to the reaction and the reaction was blanketed with argon. The mixture was stirred under argon for 12h. 3 ml of methanol were added to the reaction, and after 15 min it was filtered through a fine-fritted glass funnel to remove the solid

material. The liquid fraction was evaporated to dryness under vacuum, and the residue was purified by silica gel chromatography (CH_2Cl_2:MeOH 100:4). After chromatography, 132 mg of dideuterothymidine were obtained. NMR of this material showed it to be a mixture between the β and α isomers and the enatiomeric ratio to be 3.4:1, β to α. 125 mg of this material was crystallized from hot ethyl acetate. After this round of crystallization, 48 mg of pure 2',2'-dideutero-3',5'-O-tolyl-β-thymidine were obtained. The mother liquor from the crystallization was concentrated under vacuum. NMR of the resulting residue showed it to be 1:1 in β and α isomers. The residue was enriched for the desired β isomer, using a spinning chromatotron. The mixture was eluted from the chromatotron with ethyl acetate:dichloromethane 1:3, plus 0.5 ml of benzene for a 100 ml total of solvent. Purification using the chromatotron produced 35 mg of material, which contained 90% of the desired β isomer and 10% of the α isomer. This material was crystallized from hot ethyl acetate. Crystallization afforded 28.4 mg of pure β isomer. In all, 76.4 mg of pure 2',2'-dideutero-3',5'-O-tolylthymidine were obtained in a 50% overall yield. MP 196-197 °C. ^1H NMR (300 MHz $CDCl_3$) δ 8.20 (bd, s, 1H); 7.94 (m, 4H); 7.28 (m, 5H); 6.46 (s, 1H); 5.63 (d, 1H, J = 2 Hz); 4.78 (dd, 1H, J = 3,12 Hz); 4.66 (dd, 1H, J = 3,12 Hz); 4.53 (m, 1H); 2.44 (s, 3H); 2.43 (s, 3H); 1.61(s, 3H). Mass spec DEI Observed $[M\text{-}C_5H_5NO_2]^+$ mass minus thymidine at M/Z = 355. Percent deuterium incorporation determined by mass spectroscopy 2% D_1 and 98% D_2.

2',2'-dideuterothymidine, (16). 67 mg (0.14 mmoles) of **15** were added to a pressure tube containing 10 ml of methanol. The tube was blanketed with argon and cooled to 0 °C. Anydrous ammonia was bubbled through the solution for 15 min. The pressure tube was then sealed and allowed to stir at room temperature for 48 h. The tube was cooled back to 0 °C, opened, and

allowed to warm slowly to room temperature. When the evolution of ammonia had ceased, the methanol was removed under reduced pressure. The residue was tritirated with diethylether. This yielded 34.2 mg of solid material, which was dried under high vacuum. The material was crystallized from hot methanol. Crystallization yielded 31 mg of the desired product in a 91% overall yield. ^{1}H NMR (300 MHz D_2O) δ 7.62(m, 1H); 6.29 (s, 1H); 4.46 (d, 1H, J = 4 Hz); 4.03 (m, 1H); 3.82 (dd, 1H, J = 4,12 Hz); 3.76 (dd, 1H, J = 4, 12 Hz); 1.79 (s, 3H). Mass Spec DEI Observed: M^+ 244.10 (244.10 calculated for C_{10} H_{12} D_2 O_5 N_2). Deuterium incorporation by DEI mass spectroscopy 2% D_1 and 98% D_2.

2',2'-dideutero-5'-O-dimethoxytritylthymidine, (**17**). 25 mg (0.103 mmoles) 2',2'-dideuterothymidine was taken up in 5 ml of dry pyridine and evaporated to dryness under vacuum. This was repeated two more times. The residue was dissolved in 1 ml of dry pyridine and 38.5 mg (0.11 mmoles) of dimethoxytritylchloride were added. The mixture was stirred under argon for 12h. The pyridine was removed under vacuum, and the residue was taken up in 50 ml of dichloromethane. The dichloromethane was washed three times with pH 7 buffer (25 ml) and two times with a saturated sodium chloride solution (35 ml). The aqueous layers were layered and back-extracted with dichloromethane (2 X 30 ml). The organic layers were pooled and dried with sodium sulfate. The dichloromethane was removed under vacuum, and the residue was dried under high vacuum. The residue was crystallized from hot benzene. Crystallization afforded 37 mg (68%) yield of pure 2',2'-dideutero-5'-O-(4,4'-dimethoxytrityl)-thymidine. ^{1}H NMR (400 mHz, $CDCL_3$) δ 8.22 (bd, s, 1H); 7.58 (m, 1H); 7.21-7.38 (m, 9H); 6.82 (d, 4H, J = 9 Hz); 6.37 (s, 1H); 4.55 (bd, s, 1H); 4.05 (bd, s, 1H); 3.81 (s, 6H); 3.46 (dd, 1H, J = 1,10 Hz); 3.35 (dd, 1H, J =

1,10 Hz); 1.46 (s, 3H). Mass spec +ve FAB Ch_2Cl_2/MBA Observed M^+ at M/Z 546.23 (546.23 calculated for $C_{31}\ H_{30}\ D_2\ N_2\ O_5$).

3'-keto-5'-O-dimethoxytritylthymidine, (18). 1.50 g (15 mmoles) CrO_3 were dissolved in 70 ml dry dichloromethane and blanketed with an inert atmosphere. The reaction vessel was placed in a 20 °C water bath. 2.5 ml of dry pyridine were slowly added to the solution via syringe. Following the addition of the pyridine, 1.5 ml of acetic anhydride were added to the reaction mixture via syringe, and the mixture was stirred for 5 min. 2.73 g (5 mmoles) of 5'(4,4'-dimethyoxytrityl)thymidine were dissolved in 5 ml of dry dichloromethane and added slowly via syringe to the reaction mixture. The total reaction mixture was stirred at room temperature for 45 min. At this time the reaction was poured into a buchner funnel containing 2.5 ml celite and 75 ml cold toluene. The toluene was filtered through the celite, and the celite was rinsed with 25 additional ml of toluene. The filtrate was filtered a second time through celite, and the resulting filtrate was evaporated to dryness using rotary evaporator and high vacuum. The residue was taken up in 75 ml of cold toluene, and the resulting solution was evaporated to dryness again. This procedure was repeated three additional times. The residue was then treated with 50 ml of cold toluene plus 10 ml of hexane and filtered. The residue was washed with hexane (3 X 10 ml). The residue was then dried under high vacuum. The total yield of the reaction was 2.37 g of the desired ketone (87.4%).

3'-deutero-5'-O-4',4'-dimethoxytritylthymidine, (19). 300 mg (7 mmoles) of $NaBD_4$ were placed in a dry flask under argon. 30 ml of cold absolute ethanol were added and the entire mixture was cooled to 0 °C. Then 420 microliters

(5.7 mmoles) of acetone were added. This mixture was stirred at 0 °C for 2h. After this period of time, a suspension of 1.085 g (2 mmoles) of ketone, **18**, in 10 ml of absolute ethanol chilled to 0 °C were added to the reaction mixture. The reaction was stored at 4 °C for 22h. The resulting, green, homogeneous mixture was evaporated to approximately 10% of volume using a rotary evaporator and then taken up in 300 ml of ethyl acetate. The resulting mixture was extracted with 150 ml of a 1:1 mixture of pH 7 buffer and saturated sodium chloride. The aqueous phase was extracted with two more 300 ml portions of ethyl acetate. The resulting organic phases were pooled and washed with brine (2 X 150 ml) and then dried with sodium sulfate. The ethyl acetate was removed *in vacuo* yielding 769 mg of green foam. This foam contained a mixture of the desired riboisomer and a diastereomer. The foam was purified, using silica gel chromatography (CH_2Cl_2:MeOH 1,000:43). Purification resulted in 481 mg of an undesired isomer and 120 mg of the desired riboisomer. The riboisomer was recrystallized from a minimum volume of hexane. Recrystallization afforded 112.4 mg of the desired ribo product. Total yield for the reaction was 593 mg (54%). The diestereo selectivity was 1:4.3. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.12 (bd s, 1H); 7.58 (s, 1H); 7.2-7.5 (m, 12H); 6.84 (m, 4H); 6.39 (dd, 1H, J = 6,7 Hz); 4.05 (s, 1H); 3.76 (s, 6H); 3.48 (dd, 1H, J = 1,10 Hz); 3.35 (dd, 1H, J = 1,10 Hz); 2.29-2.44 (m, 2H); 1.44 (s, 3H).Mass spec -ve FAB CH_2Cl_2/NBA Observed $(M-H)^-$ at M/Z = 544.19, 545.26 calculated for C_{31} H_{31} D_1 N_2 O_7). Percent deuterium incorporation by -ve FAB CH2Cl2/NBA 96.8% D_1 and 3.2% D_0.

1-O-acetyl-4-deutero-2,3,5-tri-O-benzoyl-β-D-ribofuranose, (21). 1-O-acetyl-4-bromo-2,3,5-tri-O-benzoyl-β-D-ribofuranose **20** was prepared by the method of (Ferrier, et al., 1984). 7.8 g(13.4 m) of 1-O-acetyl-4-bromo-2,3,5-tri-O-benzoyl-β-D-ribofuranose, 7.81 g tributyltindeuteride (26.7 m) and AIBN (100 mg) were

dissolved in 70 ml of toluene. Argon was bubbled through the solution for 30 min at 25° C. The solution was then immersed in an oil bath at 75 ° C for 40 min. After removing the toluene *in vacuo*, the residue was crystallized from hot methanol. After two rounds of crystallization from methanol, 3.52g (52% yield) of the α-isomer were obtained. The mother liquors from the two crystallizations were combined and concentrated *in vacuo*. This afforded a residue now enriched in the desired, deuterated β-D-ribofuranose, **21**. The β-D-ribofuranose was crystallized from hot isopropyl alcohol, yield 954 mg (14%). MP 124-125 C. ^{1}H NMR (400 MHz, $CDCl_3$). δ 8.05 (d, 2H, J = 7 Hz); 7.98 (d, 2H, J = 7 Hz); 7.86 (d, 1H, J = 7 HZ); 7.30-7.60 (m, 10 H); 6.41 (d, 1H, J = 1Hz); 5.88 (dd, 1H, J = 5 Hz); 5.76 (dd, 1H, J = 1,5 Hz); 4.75 (d, 1H, J = 12 Hz); 4.48 (d, 1H, J = 12 Hz); 1.98 (s, 3H). Mass Spec: +ve FAB CH_2Cl_2/NBA. MH at M/Z 505.45 (calculated 505.4922 C_5 H_5 D_1 O_5. NMR for the α-isomer: ^{1}H (400 MHz, $CDCL_3$). d: 7.82-8.00 (m, 6H); 7.30-7.56 (m, 10 H); 6.50 (s, 1H); 6.08 (d, 1H, J = 5 Hz); 5.76 (d, 1H, J = 5 Hz); 4.64 (s, 2H); 2.08 (s, 3H).

4'-deutero-2',3',5'-tri-O-benzoyl-methyluridine, (**22**). 320 mg of $SnCL_4$ (1.24 m moles) were added via a gas-tight syringe to a 0° C solution of 425mg (0.84m mole) of 10-acetyl-4-deutero-2,3,5-tri-O-benzoyl-β-D-ribofuranose and 0.7 g bis(O-trimethylsylil)thymidine (2.58 m moles) in 6 m of dichloroethane. After stirring for 10h, the solution was poured into dichloroethane (50 ml) and washed with saturated Na (25 ml) and then filtered through celite. The organic layer was dried over Na_2SO_4 and concentrated *in vacuo*. Silica gel chromatography, ethyl acetate:dichloromethane (1:5) yielded 400 mg of **3** in an 83% yield. MP 158-160 °C 1H NMR (400 MHz, $CDCL_3$) d 7.92-8.14 (m, 6H); 7.34-7.64 (m, 9H); 7.15 (q, 1H, J = 1.3 Hz); 6.41 (d, 1H, J = 7 Hz); 5.89 (d, 1H, J = 6 Hz); 5.73 (dd, 1H, J = 6,7 Hz); 4.86 (d, 1H, J = 12 Hz); 4.63 (d, 1H, J = 12 Hz); 1.57

(d, 3H, J = 1.3 Hz). Mass Spec: +ve FAB CH_2Cl_2/NBA M^+ at M/Z 572.17 (calculated 572.18 for C_{31} H_{26} D_1 N_2 O_9).

4'-deutero-3',5'-O-tetraisopropyldisilymethyluridine, (23), A solution of 10 ml of methanol and 400 mg (0.70 mmoles) of **22** was saturated with ammonia 0 °C. The flask was stoppered and stirred at 25 °C for 12h, after which time the mixture was evaporated to dryness *in vacuo*. The residue was dried further by alternate cycles of dissolving in 15 ml of pyridine and evaporating to dryness *in vacuo*. After three such cycles, the residue was taken up in 10 ml of dry pyridine and allowed to react with 1,3-dichloro-1,1,3,3-tetratisopropyldisyloxane (237 mg, 0.75 mmoles) for 2h at 35 °C. The crude mixture was poured into water (10 ml) and extracted with diethylether (3 X 40 ml). The organic layers were washed two times with pH 7 buffer (20 ml) and once with brine (20 ml) and dried over Na_2SO_4. Flash chromatography (ethylacetate:CH_2Cl_2, 1:3) yielded 222 mg (63% yield). MP 78-80 °C. 1H NMR (400 MHz, $CDCl_3$) δ 8.3 (bd s, 1H); 7.33 (q, 1H, J = 1.3 Hz); 5.68 (s, 1H); 4.40 (d, 1H, J = 5 Hz); 4.09-4.17 (m, 2H); 3.98 (d, 1H, J = 13 Hz); 2.95 (bd s, 1H); 1.89 (d, 3H, J = 1.2Hz); 1.01-1.08 (m, 28H). Mass Spec: +ve/FAB CH_2Cl_2/NBA observed MH^+ at 502.24 (calcualted for C_{22} H_{40} D_1 N_2 O_7 Si_2 502.25).

4'-deutero-2'-O-phenylthiocarbonate-3',5'-O-tetraisopropyldisilymethyl-uridine, (24). 540 mg (1.08 m moles) of compound **23** and 172 mg (1.40 m moles) of DMAP were dissolved in 8 ml of dry acetonitrile. To this solution phenylthiochloroformate 223 mg (1.3 m moles) were added slowly via syringe. The soluiton was stirred for 4h at 25 °C. The solvent was then removed on a rotary evaporator, and the residue was purified by a flash chromatography using ethyl acetate:CH_2Cl_2, (1:9). 47 mg of the desired product (68% yield)

were obtained. Melting point 70-72 °C. ^{1}H NMR (400 MHz $CDCL_3$) δ 8.02 (bd s, 1H); 7.36-7.42 (m, 3H); 7.26-7.30 (t, 1H, J = 3 Hz); 7.09 (d, 2H, J = 8 Hz); 5.99 (d, 1H, J = 5 Hz); 5.89 (s, 1H); 4.58 (d, 1H, J = 5 Hz); 4.21 (d, 1H, J = 14 Hz); 4.01 (d, 1H, J = 14 Hz); 1.91 (s, 3H); 1.00-1.09 (m, 28H).

4'-deutero-3',5'-O-tetraisopropyldisilythymidine, (25). 470 mg of **24** (0.74 m moles), 260 mg tributyltinhydride (0.89 m moles) and 10 mg AIBN were dissolved in 8 ml of dry toluene. The solution was subjected to three rounds of freeze/thaw degassification. The reaction mixture was then placed under argon and immersed in an oil bath at 70 °C for 45 min. After cooling to room temperature, the toluene was removed on a rotary evaporator equipped with a -78 °C cold finger and evacuated through a rough pump. Flash chromatography ethyl acetate:CH_2Cl_2 (1:6) yielded 258 mg of the desired product (71% yield). MP 149-151 °C. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.95 (bd s, 1H); 7.4 (d, 1H, J = 1 Hz); 6.04 (dd, 1H, J = 2,7 Hz); 4.47 (dd, 1H, J = 7,9 Hz); 4.0 (d, 1H, J = 13 Hz); 3.99 (d, 1H, J = 13 Hz); 2.43-2.49 (m, 1H); 2.22-2.26 (m, 1H); 1.90 (d, 3H, J = 1 Hz); 1.00-1.10 (m, 28H). Mass Spec +ve/FAB CH_2Cl_2/NBA: Observed MH^+ at M/Z of 486.26 (calculated for C_{22}, H_{40}, D_1, N_2, Si_2, O_6, 486.26).

4'-deuterothymidine, (26). 255 mg of **25** (0.54 m mole) were dissolved in 8 ml of THF. To this solution, 1.6 ml of a 1M solution of $BU_4N^+F^-$ in THF were added. After stirring for 30 min, the THF was removed on a rotary evaporator. The residue was purified by flash silca gel chromatography methanol:CH_2Cl_2 (1:5). This afforded 115 mg of 4'-deuterothymidine (88% yield). ^{1}H NMR (90 MHz, D_2O) δ 7.70 (bd s), 1H); 6.15 (t, 1H, J = 7 Hz); 4.30 (m, 1H); 3.95 (m, 2H); 2.25 (m, 2H); 1.76 (s, 3H).

4'-deutero-5'-O-dimethoxytritylthymidine, (27). 115 mg of 4'-deuterothymidine(0.47 mmole) were taken up in 5 ml of pyridine and evaporated to dryness. This was repeated two times. The residue was taken up in 3 ml of dry pyridine and allowed to react with 175 mg of dimethoxytritylchloride (0.51 m moles) for 6h. The reaction was quenched with 1 ml of methanol and then poured into 10 ml of H_2O. The water was extracted with diethylether (3 x 50 ml). The organic phase was washed two times with pH 7 buffer (30 ml) and one time with brine (30 ml) and then dried over NA_2SL_4. After removing the ethylether *in vacuo*, the residue was purified by flash silica gel chromatography ethyl acetate:hexane:methanol (40:55:5). After chromatography, 190 mg of **27** were obatined in 74% yield. MP 120-122 °C. ^1H NMR (400 MHz, $CDCL_3$) δ 8.03 (bd s, 1H); 7.55 (s, 1H); 7.37 (d, 2H, J = 7 Hz); 7.25-7.30 (m, 5H); 6.81 (m, 6H); 6.37 (t, 1H, J = 6 Hz); 4.55 (t, 1H, J = 3 Hz); 3.77 (s, 6H); 3.46 (d, 1H, J = 11 Hz); 3.34 (d, 1H, J = 11 Hz); 2.33 (bd m, 2H); 1.46 (s, 3H). Mass Spec: +ve/FAB CH_2Cl_2/NBA Observed: M^+ at M/Z 545.23 (calculated for C_{31}, H_{31}, D_1, N_2, O_7 is 545.23). Percent deuterium incorporation by mass spec using - ve/FAB CH_2Cl_2/NBA 94.5% D_1,and 5.5% D_0.

Methylthymidine-5'-carboxolate, (28). 6 mg of crude thymidine (2.34 m moles) were refluxed in 150 ml of methanol and 0.5 ml of H_2SO_4 for 6h. The solution was cooled to room temperature and neutralized with solid Na_2CO_3. The solids were filtered off, and the methanol was removed *in vacuo*. The residue was purified by flash silica gel chromatography (methanol:CH_2Cl_2, 1:9). Purification yielded 586 mg of **28** (93% yield). ^1H NMR (90 MHz, D6-DMSO). δ 8.1 (bd s, 1H); 6.31 (m, 1H); 5.82 (d, 1H, J = 8 Hz); 4.40 (m, 1H); 3.72 (s, 3H);

2.10 (m, 2H); 1.72 (s, 3H). Mass Spec: +ve FAB CH_2Cl_2/NBA Observed: M^+ at M/Z 270.21 (270.24 calculated for C_{11}, H_{14}, N_2 O_6).

Methyl-3'-O-t-butyldimethylsilylthymidine-5'-carboxolate, (29). 0.59 g of **28** (2.17 m moles) and 0.37 g of imidizole (5.43 m moles) were dissolved in 7 ml of dry DMF. To this solution 0.39 g of t-butyldimethylsilylchloride (2.61 m moles) was added, and the solution was stirred at 37 °C for 10h. The mixture was taken up in 10 ml of water and extracted with diethylether (3 X 35 ml). The organic layers were combined, washed with pH 7 buffer (2X 20 ml) and brine (20 ml), and dried over $MgSO_4$. The ether was removed *in vacuo* and the residue was purified, using flash silica gel chromatography methanol:CH_2Cl_2 (1:19). This afforded 460 mg of the desired product (55% yield). MP 138-140 °C. ^1H NMR (90 MHz, $CDCl_3$) δ 8.1 (bd s, 1H); 6.5 (dd, 1H, J = 5,9 Hz); 4.43-4.53 (m, 2H); 3.8 (s, 3H); 2.20-2.35 (m, 2H); 1.96 (s, 3H); 0.89 (s, 9H); 0.10 (s, 6H). Mass spec: +ve FAB CH_2Cl_2/NBA Observed: MH^+ at M/Z = 385.48 (385.51 calculated for C_{17} H_{19} N_2 O_6 Si).

5',5'-dideutero-3'-O-t-butyldimethylsilylthymidine, (30). 35 mg of $LiAlD_4$ (1.2 m moles) were dissolved in 6 ml of THF and placed under an inert atmosphere and cooled to 0 °C. To this solution, 1 ml of THF containing 460 mg of **10** (1.2 m moles) was added via syringe. The reaction was stirred for 2h at 0 °C, after which time the reaction was quenched with the slow addition of 0.5 ml of water. The mixture was then concentrated on a rotary evaporator. 5 ml of water were added to the residue, and the mixture was extracted with diethylether (3 X 35 ml). The organic layers were pooled and washed with pH 7 buffer (2 X 20 ml), brine (20 ml) and dried with Na_2SO_4. The ether was removed on a rotary evaporator. The residue was purified, using flash silica

gel chromatography methanol:CH_2Cl_2 (1:19). This afforded 329 mg of the desired product (77% yield) as a clear oil. ^1H NMR (90 MHz, $CDCL_3$) δ 7.33 (s, 1H); 6.11 (t, 1H, J = 8 Hz); 3.92 (d, 1H, J = 6 Hz); 3.50 (m, 1H); 2.30 (m, 2H); 1.88 (s, 3H); 0.85 (s, 9H); 0.07 (s, 6H). Mass spec +ve FAB CH_2Cl_2/NBA Observed: MH^+ at M/Z = 359.49 (359.48 calculated for C_{16} H_{17} D_2 N_2 O_5 Si).

5',5'-dideuterothymidine, (**31**). 329 mg of **30** (0.92 m moles) were dissolved in 8 ml of THF, blanketed with an inert atmosphere and cooled to 0 °C. To this solution, 1.9 ml of a 1 M solution of $Bu_4N^+F^-$ in THF were added via syringe. The solution was stirred at 30 °C for 30 min, then allowed to warm to 25 °C. The reaction was then stirred for 3h at 25 °C. The THF was removed on a rotary evaporator, and the residue was purified by flash silica gel chromatography methanol:CH_2Cl_2 (1:9). This afforded 168 mg of the desired product in 75% yield. ^1H NMR (90 MHz, D_2O) δ 7.60 (s, 1H); 6.15 (t, 1H, J = 7 Hz); 4.3 (m, 1H); 3.89 (d, 1H, J = 4 Hz); 2.24 (dd, 1H, J = 6,5 Hz); 1.76 (s, 3H). Mass spec +ve FAB CH_2CL_2/NBA Observed: M^+ at M/Z 244.10 (244.10 calculated for C_{10} H_{12} D_2 O_5 N_2).

5',5'-dideutero-5'-O-'dimethoxytritylthymidine, (**3 2**). 168 mg 5',5'-dideuterothymidine (0.69 m moles) were taken up in 5 ml of pyridine and evaporated to dryness. This was repeated two times. The residue was taken up in 3 ml of dried pyridine and allowed to react with 260 mg of dimethoxytritylchloride (0.77 m moles) for 6h. The reaction was then quenched with 1 ml of methanol. The mixture was poured into 10 ml of water and extracted with diethylether (3 X 50 ml). The organic layers were cooled and washed with pH 7 buffer (2 X 30 ml) and with brine (30 ml) and dried over Na_2SO_4. The ether was removed under vacuum and the residue was purified

by flash silica gel chromatography ethyl acetate:hexanes:methanol (40:55:5). Chromatography yielded 343 mg of **13** (91% yield). MP 119-121 °C ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.05 (bd s, 1H); 7.55 (s, 1H); 7.36-7.38 (d, 2H, J = 7 Hz); 7.25-7.30 (m, 5H); 6.81 (m, 6H); 6.37 (t, 1H, J = 6 Hz); 4.55 (m, 1H); 4.02 (d, 1H, J = 3 Hz); 3.81 (S, 6H); 2.3-2.45 (m, 2H); 1.46 (s, 3H). Mass spec +ve FAB CH_2Cl_2/NBA Observed M^+ at M/Z = 546.23 (546.23 calculated for $C_{31}H_{30}D_2N_2O_7$). Isotopic incorporation based on mass spectroscopy, using -ve FAB CH_2Cl_2/NBA matrix 98% D_2, 1.5% and 0.5% D_0.

Phosphitylation of Deuterated Thymidines. Typically, 25 mg (46 μmoles) deuterated thymidine were dissolved in 0.75 ml of dry dichloromethane, 32 μl (230 μmoles) of freshly distilled triethylamine were added, and the flask was blanketed with argon. To this mixture, 16 μl (69 μmoles) of 2-cyanoethyl N,N-diisopropylchlorophosphoramidite were added via syringe. The mixture was stirred under argon for 80 min, at which time the dichloromethane was removed under vacuum. The residue was purified by silica gel chromatography (ethyl acetate:dichloromethane:triethylamine 30:70:1). Typical yields for the reaction range between 85 and 95%. The phosphoramidites were stored under vacuum and used as rapidly as possible.

References

Abdel-Meguid, S. S., N. D. F. Grindley, et al. (1984). Proc. Natl. Acad. Sci. U. S. A. **81**: 2001.

Baldwin, J. E. and A. H. Andrist (1971). Chem. Commun. : 1512.

Bannister, C. E., J. M. T. Raycheba, et al. (1982). Inorg. Chem. **21**: 1106.

Barefield, E. K. and M. T. Mocella (1975). J. Am. Chem. Soc. **97**: 4238.

Beaucage, S. L. and M. H. Caruthers (1984). Tetrahedron Lett. **22**: 1859.

Bossu, F. P. and D. W. Margerum (1977). Inorg. Chem. **16**: 1210.

Bossu, F. P. and D. W. Margerum (1976). J. Am. chem. Soc. **98**: 4004.

Brown, P. and R. C. Cookson (1965). Tetrahedron Lett. **21**: 1993.

Bryce, G. F., R. W. Roeske, et al. (1966). J. Biol. Chem. **241**: 1072.

Caldin, E. F. and S. Mateo (1975). J. Chem. Soc. Faraday Trans. **I:71**: 1876.

Camerman, N., A. Camerman, et al. (1976). Can. J. Chem. **54**: 1309.

Cameron, V. and O. C. Uhlenbeck (1977). Biochem. **16**: 5120.

Chen, X., S. E. Rokita, et al. (1991). J. Am. Chem. Soc. **113**: 5884.

Dervan, P. B. (1986). Science (Washington, D. C. **232**: 464.

Drago, R. S. and E. I. Baucom (1972). Inorg. Chem. **11**: 2064.

Dunn, O. and V. Clark (1987). Applied Statistics: Analysis of Variance and Regression. New York, John Wiley and Sons.

Ferrier, R., S. R. Haynes, et al. (1984). J. Chem. Soc. Perkin Trans. **1**: 1683.

Fleiss, J. (1986). The Design and Analysis of Clincial Experiments. New York, John Wiley and Sons.

Frank, B. L., L. Worth, et al. (1991). J. Am. Chem. Soc. **113**: 2271.

Frank, B. L., L. W. Worth, et al. (1991). Biochem. **30**: 2034.

Freeman, H. C., J. M. Guss, et al. (1968). Chem. Commun. : 485.

Freeman, H. C. and M. R. Taylor (1965). Acta Cryst. **18**: 939.

Garnier, J., J. Osguthorpe, et al. (1978) J. Mol. Biol. **120**: 97.

Giloni, L., M. Takeshita, et al. (1981). J. Biol. Chem. **256**: 8608.

Gore, E. S. and D. H. Busch (1973). Inorg. Chem. **12**: 1.

Graham, K. S. and P. B. Dervan (1990). J. Biol. Chem. **265**: 16534.

Grindley, N. D. F., M. R. Lauth, et al. (1982) Cell **30**:19.

Halevi, E. A. (1963). Prog. Phys. Org. Chem. **1**: 109.

Harris, J. M. and C. C. Wamser (1976). Fundamentals of Organic Reaction Mechanisms. New York, John Wiley and Sons.

Hertzberg, R. P. and P. B. Dervan (1984). Biochem. **23**: 3934.

Hoffer, M. (1960). Chem. Ber. **93**: 2777.

Hubbard, A. J., A. S. Jones, et al. (1984). Nucleic Acids Res. **12**: 6827.

Hurst, J. K. (1976). J. Am. Chem. Soc. **98**: 4003.

Inoue, S. and S. Kawanishi (1989). Biochem. and Biophys. Res. Commun. **159**: 445.

Iverson, B. L. and P. B. Dervan (1987). Nucleic Acids Res **15**: 7823.

Iverson, B. L. and P. B. Dervan (1987b). J. Am. Chem. Soc. **109**: 1241.

Jencks, W. P. (1969). Catalysis in Chemistry and Enzymology. New York, Dover Publications.

Kappen, L. S. and I. H. Goldberg (1983). Biochem. **22**: 4872.

Kappen, L. S., I. H. Goldberg, et al. (1991). Biochem. **30**: 2034.

Kappen, L. S., I. H. Goldberg, et al. (1982). Proc. Natl. Acad. Sci., USA **79**: 744.

Kappen, L. S., I. H. Goldberg, et al. (1990). J. Am. Chem. Soc. **112**: 2797.

Kent, S. B. H. (1988). Ann. Rev. Biochem. **57**: 957.

Kinneary, J. F., J. S. Albert, et al. (1988). J. Am. Chem. Soc. **110**: 6124.

Kinneary, J. F., T. R. Wagler, et al. (1988). Tetrahedron Letts. **29**: 877.

Kirsch, J. F. (1977). Secondary kinetic isotope effects. Isotope Effects on Enzyme-Catalyzed Reactions. Baltimore, University Park Press.

Koola, J. D. and J. K. Kochi (1987). Inorg. Chem. **26**: 908.

Kozarich, J. W., L. Worth, et al. (1989). Science **245**: 1396.

Kresge, A. J. (1977). Magnitude of primary hydrogen isotope effects. Isotope Effects on Enzyme Catalyzed Reactions. Baltimore, University Park Press.

Kross, J., W. D. Henner, et al. (1982). Biochem. **21**: 4310.

Lappin, A. G., C. K. Murray, et al. (1978). Inorg. Chem. **17**: 1630.

Lau, A.-J., T. P. A. Kruck, et al. (1974). J. Biol. Chem. **249**: 5878.

Mack, D. P. (1991). Design and Chemical Synthesis of Sequence-Specific DNA-Cleaving Metalloproteins. California Institute of Technology.

Mack, D. P. and P. B. Dervan (1990). J. Am. Chem. Soc. **112**: 4604.

Mack, D. P., B. L. Iverson, et al. (1988). J. Am. Chem. Soc. **110**: 7572.

Matteucci, M. D. and M. H. Caruthers (1981). J. Am. Chem. Soc. **103**: 3185.

Maxam, A. M. and W. Gilbert (1980). Methods Enzymol. **65**: 499.

McGall, G. H., L. E. Rabow, et al. (1992). J. Am. Chem. Soc. **114**: 1958.

McGall, G. H. and J. Stubbe, et al. (1991). J. Org. Chem.. **56**: 48.

Melander, L. (1960). Isotope Effects on Reaction Rates. New York, Ronald Press.

Melander, L. and W. H. Saunders (1980). Reaction Rates of Isotopic Molecules. New York, John Wiley and Sons.

Moser, H. E. and P. B. Dervan (1987). Science. **238**: 645.

Muller, J. G., X. Chen, et al. (1992). J. Am. Chem. Soc. **114**: 6407.

Olson, D. C. and J. Vasilevskis (1969). Inorg. Chem. **8**: 1611.

Paniago, E. B., D. C. Weatherburn, et al. (1971). Chem. Commun. : 1427.

Pasto, D. J., B. R. Meyer, et al. (1974). J. Am. Chem. Soc. **96**: 1858.

Rabow, L. E. f., G. H. McGall, et al. (1990). J. Am. Chem. Soc. **112**: 3203.

Richards, J. H. (1970). Kinetic isotope effects in enzyme reactions. The Enzymes. New York, Academic Press. Third, ed.

Robins, M. J., J. S. Wilson, et al. (1977). J. Medical Chem. **22**: 518.

Robins, M. J. and J. S. Wilson (1981). J. Am. Chem. Soc. **103**: 932.

Sarin, V. K., S. B. H. Kent, et al. (1981). Anal. Biochem. **117**: 147.

Saenger, W. (1984). Principles of Nucleic Acid Structure. New York, Springer-Verlag.

Sakurai, T. and A. Nakahara (1979). Inorg. Chimica Acta **34**: L243.

Shiner, V. J. (1952). J. Am. Chem. Soc. **74**: 2925.

Shiner, V. J. (1960). J. Am. Chem. Soc. **82**: 2655.

Shiner, V. J. (1970). Isotope Effects in Chemical Reactions. New York, Van Nostrand Reinhold Co.

Shiner, V. J., W. E. Buddenbaum, et al. (1968). J. Am. Chem. Soc. **90**: 418

Sigman, D. S. and C. B. Chen (1990). Ann. Rev. Biochem. **59**: 207.

Sluka, J. P., S. J. Horvath, et al. (1987). Science **238**: 1129.

Streitwieser, A., R. H. Jagow, et al. (1958). J. Phys. Chem. **80**: 2326.

Sugiura, Y. and Y. Mino (1979). Inorg. Chem. **18**: 1336.

Sunko, D. E. and S. Borcic (1970). Isotope Effects in Chemical Reactions. New York, Van Nostrand Reinhold Co.

Tapper, D. P. and D. A. Clayton (1981). Nucleic Acids Res. **9**: 6787.

Wells, R. G. and N. D. F. Grindley (1984). J. Mol. Biol. **179**: 667.

Westheimer, F. H. (1961). Chem. Rev. **61**: 265.

Wigfield, D. C., D. J. Phelps, et al. (1975). J. Am. Chem. Soc. **97**: 898.

Wu, J. C., J. W. Kozarich, et al. (1985). Biochem. **24**: 7562.

Wu, J. C., J. W. Kozarich, et al. (1983). J. Biol. Chem. **258**: 4694.

Youngquist, R. S. and P. B. Dervan (1985). Proc. Natl. Acad. Sci. USA **82**: 2565.

Appendix A: Phosphorimager Data for the Ni(II) GGH γδ (141-183) Cleavage-Condition Studies.

Table I. Phosphorimager Data from the Concentration Studies of DNA Cleavage in the Presence of 2.0 mM Ni(II) GGH γ δ (141-183) and Monoperoxyphthalic Acid with and without Butylamine Treatment

Monoperoxyphthalic Acid Concentration (mM)	% DNA Cleavage w/o Butylamine Treatment	% DNA Cleavage with Butylamine Treatment
0.00	0.00%	0.00%
0.01	0.58%	1.05%
0.05	0.88%	1.57%
0.10	1.65%	3.66%
0.50	6.76%	9.83%
1.00	15.22%	24.97%
2.00	33.30%	46.97%
5.00	59.54%	70.35%
10.00	67.54%	89.81%
20.00	68.48%	89.77%
50.00	68.78%	89.65%

Table II. Phosphorimager Data from the Concentration Studies of DNA Cleavage in the Presence of 1.0 mM Ni(II) GGH γ δ (141-183) and Monoperoxyphthalic Acid with and without Butylamine Treatment

Monoperoxyphthalic Acid Concentration (mM)	% DNA Cleavage w/o Butylamine Treatment	% DNA Cleavage with Butylamine Treatment
0.00	0.00%	0.00%
0.01	0.44%	0.50%
0.05	0.75%	1.04%
0.10	1.02%	1.75%
0.50	7.12%	10.23%
1.00	16.85%	24.08%
2.00	34.10%	48.55%
5.00	58.69%	75.44%
10.00	60.77%	79.85%
20.00	59.93%	80.47%
50.00	59.21%	79.96%

Table III. Phosphorimager Data from the Concentration Studies of DNA Cleavage in the Presence of 0.5 mM Ni(II) GGH γδ (141-183) and Monoperoxyphthalic Acid with and without Butylamine Treatment

Monoperoxyphthalic Acid Concentration (mM)	% DNA Cleavage w/o Butylamine Treatment	% DNA Cleavage with Butylamine Treatment
0.00	0.00%	0.00%
0.01	0.22%	0.31%
0.05	0.35%	0.66%
0.10	0.58%	0.88%
0.50	2.63%	4.95%
1.00	6.05%	7.87%
2.00	13.93%	19.97%
5.00	29.93%	36.88%
10.00	40.87%	51.58%
20.00	40.33%	51.86%
50.00	39.10%	50.20%

Table IV. Densitometer Data from the Concentration Studies of DNA Cleavage in the Presence of 2.0 mM Ni(II) GGH γ δ (141-183) and Hydrogen Peroxide

Hydrogen Peroxide Concentration (mM)	Normalized DNA Cleavage as a Function of Measured Optical Density
0.00	0.00
0.50	0.02
1.00	0.05
5.00	0.21
10.00	0.35
100.00	0.50

Table V. Phosphorimager Data from the Time Course of DNA Cleavage by Ni(II) GGH γδ (141-183) in the Presence of Monoperoxyphthalic Acid or Hydrogen Peroxide

	Monoperoxyphthalic acid		Hydrogen Peroxide	
time (min)	% cleavage	Normalized cleavage	% cleavage	Normalized Cleavage
background	0.05%	0.01	0.05%	0.00
0	3.86%	0.39	0.13%	0.00
0.5	6.90%	0.69	0.34%	0.01
1	7.37%	0.74	0.44%	0.01
2	8.54%	0.84	0.91%	0.02
5	9.24%	0.92	1.42%	0.03
10	9.98%	1.00	2.71%	0.06
20	9.93%	0.99	8.34%	0.19
40	9.92%	0.99	15.08%	0.34
60	9.89%	0.99	28.85%	0.65
80	9.92%	0.99	30.41%	0.69
100	9.52%	0.95	44.11%	1.00
120	9.45%	0.94	44.19%	1.00

Table VI. Phosphorimager Data from the pH Profiles of DNA Cleavage by Ni(II) GGH γ δ (141-183) in the Presence of Monoperoxyphthalic Acid or Hydrogen Peroxide

	Monoperoxyphthalic acid		Hydrogen Peroxide	
pH	% cleavage	Normalized cleavage	% cleavage	Normalized Cleavage
background	0.59%	0.40	0.06%	0.03
5.8	0.98%	0.67	0.07%	0.02
6.0	1.05%	0.72	0.08%	0.02
6.2	1.20%	0.82	0.10%	0.03
6.4	1.34%	0.92	0.16%	0.04
6.6	1.39%	0.95	0.21%	0.05
6.8	1.46%	1.00	0.30%	0.08
7.0	1.43%	0.98	0.36%	0.09
7.2	1.37%	0.94	0.60%	0.15
7.4	1.41%	0.97	1.77%	0.45
7.6	1.36%	0.93	3.92%	1.00
7.8	1.37%	0.94	3.77%	0.96
8.0	1.17%	0.80	3.07%	0.78

Appendix B: Normalized Cleavage Values Used to Calculate KIE

Table I. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) in the Presence of Monoperoxyphthalic Acid when the Thymidine at the Experimental Cleavage Site in the "Deutero Duplex" is Deuterated at Position C1'

	Normalized Cleavage of the "Deutero Duplex" Containing C1' Deutero Thymidine				
Lane #	Gel #1	Gel #2	Gel #3	Gel #4	Gel #5
1	1.01	1.05	1.05	0.87	0.99
2	1.08	1.06	1.03	0.87	1.01
3	1.08	1.04	1.04	0.90	0.94
4	1.01	1.02	1.04	0.89	0.91
5	1.06	1.00	0.99	0.92	0.90
6	1.03	0.97	1.03	0.90	0.87
7	1.04	0.93	1.04	0.90	1.86
8	1.03	0.97	1.05	0.90	0.86
9	1.05	0.97	1.04	0.93	0.87
10	1.04	0.93	1.01	0.87	0.86

	Normalized Cleavage of the All-Proteo Control Duplex				
Lane #	Gel #1	Gel #2	Gel #3	Gel #4	Gel #5
11	1.14	1.01	1.20	0.98	0.95
12	1.12	1.04	1.14	1.03	0.98
13	1.09	1.02	1.13	1.11	1.00
14	1.11	1.01	1.16	1.00	1.03
15	1.09	1.02	1.08	0.99	1.05
16	1.10	0.99	1.08	0.93	0.98
17	1.10	1.01	1.09	0.96	0.93
18	1.08	0.98	1.09	0.98	0.93
19	1.16	1.03	1.08	0.99	0.93
20	1.10	0.97	1.08	0.97	0.98

Table II. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) in the Presence of Monoperoxyphthalic Acid when the Thymidine at the Experimental Cleavage Site in the "Deutero Duplex" is Dideuterated at Position C2'

	Normalized Cleavage of the "Deutero Duplex" Containing C2' Dideutero Thymidine				
Lane #	Gel #6	Gel #7	Gel #8	Gel #9	Gel #10
1	1.05	0.96	1.10	0.92	1.15
2	0.97	1.01	1.07	0.95	1.12
3	1.01	1.00	1.13	0.95	1.15
4	0.97	1.00	1.06	0.91	1.10
5	0.96	1.03	1.08	0.93	1.08
6	1.04	1.17	1.08	0.96	1.15
7	0.99	1.04	1.11	0.93	1.11
8	0.99	1.04	1.08	0.90	1.14
9	1.05	1.03	1.08	0.88	1.14
10	1.00	0.97	1.07	0.94	1.13

	Normalized Cleavage of the All-Proteo Control Duplex				
Lane #	Gel #6	Gel #7	Gel #8	Gel #9	Gel #10
11	0.97	0.96	1.11	0.96	1.20
12	1.01	1.10	1.10	0.98	1.16
13	1.01	0.98	1.09	1.01	1.28
14	0.98	0.93	1.08	1.00	1.18
15	0.99	0.89	1.11	1.01	1.16
16	1.01	0.94	1.08	0.98	1.08
17	1.04	0.97	1.09	0.93	1.13
18	0.97	0.96	1.15	0.93	1.11
19	0.98	0.95	1.06	0.92	1.14
20	0.98	0.89	1.04	0.97	1.07

Table III. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) in the Presence of Monoperoxyphthalic Acid when the Thymidine at the Experimental Cleavage Site in the "Deutero Duplex" is Deuterated at Position C3'

	Normalized Cleavage of the "Deutero Duplex" Containing C3' Deutero Thymidine				
Lane #	Gel #11	Gel #12	Gel #13	Gel #14	Gel #15
1	0.99	1.01	0.90	1.09	1.04
2	1.02	0.98	0.90	1.01	1.04
3	0.98	1.02	0.89	1.07	1.06
4	1.00	1.02	0.85	1.03	0.98
5	0.97	0.98	0.83	1.02	0.98
6	1.16	1.01	0.87	1.05	1.00
7	1.02	0.97	0.88	0.94	0.95
8	0.96	1.00	0.85	1.02	0.96
9	0.90	1.02	0.86	1.01	1.02
10	0.95	0.98	0.88	0.99	0.95

	Normalized Cleavage of the All-Proteo Control Duplex				
Lane #	Gel #11	Gel #12	Gel #13	Gel #14	Gel #15
11	1.02	1.00	0.96	1.15	1.05
12	1.02	0.99	0.98	1.03	1.07
13	1.03	1.01	0.97	1.09	1.02
14	1.02	1.00	1.00	1.15	1.04
15	1.06	1.00	0.96	1.05	1.02
16	1.02	1.01	1.05	1.06	1.01
17	1.03	0.98	1.01	1.05	1.00
18	1.01	1.00	0.96	1.06	0.98
19	1.01	1.03	1.04	1.05	1.05
20	0.99	1.00	0.96	1.05	1.01

Table IV. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) in the Presence of Monoperoxyphthalic Acid when the Thymidine at the Experimental Cleavage Site in the "Deutero Duplex" is Deuterated at Position C4'

	Normalized Cleavage of the "Deutero Duplex" Containing C4' Deutero Thymidine				
Lane #	Gel #16	Gel #17	Gel #18	Gel #19	Gel #20
1	0.87	0.90	0.83	0.86	0.88
2	0.81	0.91	0.77	0.88	0.81
3	0.80	0.88	0.80	0.86	0.83
4	0.81	0.87	0.80	0.85	0.84
5	0.82	0.92	0.80	0.88	0.82
6	0.84	0.88	0.81	0.83	0.80
7	0.82	0.80	0.87	0.84	0.89
8	0.91	0.86	0.84	0.86	0.88
9	0.83	0.93	0.80	0.85	0.82
10	0.82	0.85	0.79	0.86	0.79

	Normalized Cleavage of the All-Proteo Control Duplex				
Lane #	Gel #16	Gel #17	Gel #18	Gel #19	Gel #20
11	1.07	1.13	0.98	1.20	1.03
12	1.05	1.12	0.96	1.13	1.03
13	1.04	1.13	1.04	1.18	1.00
14	1.04	1.10	1.04	1.09	0.99
15	1.03	1.20	1.01	1.05	1.01
16	1.05	1.21	1.01	1.03	1.01
17	1.06	1.22	1.00	1.05	0.99
18	1.06	1.22	1.07	1.03	1.03
19	1.05	1.22	1.02	1.08	1.01
20	1.07	1.22	1.05	1.18	1.04

Table V. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) in the Presence of Monoperoxyphthalic Acid when the Thymidine at the Experimental Cleavage Site in the "Deutero Duplex" is Dideuterated at Position C5'

	Normalized Cleavage of the "Deutero Duplex" Containing C5' Dideutero Thymidine				
Lane #	Gel #21	Gel #22	Gel #23	Gel #24	Gel #25
1	1.83	1.56	1.75	1.80	1.77
2	1.70	1.44	1.71	1.73	1.69
3	1.86	1.41	1.74	1.70	1.70
4	1.90	1.52	1.73	1.79	1.73
5	1.80	1.38	1.64	1.77	1.67
6	1.87	1.49	1.72	1.78	1.78
7	1.86	1.45	1.68	1.61	1.76
8	1.75	1.37	1.68	1.71	1.68
9	1.84	1.59	1.80	1.63	1.81
10	1.80	1.59	1.68	1.66	1.73

	Normalized Cleavage of the All-Proteo Control Duplex				
Lane #	Gel #21	Gel #22	Gel #23	Gel #24	Gel #25
11	1.10	1.08	1.06	1.15	1.01
12	1.11	1.05	1.06	1.15	1.16
13	1.10	1.12	1.06	1.19	1.05
14	1.09	1.06	1.05	1.13	1.01
15	1.07	1.06	1.03	1.15	0.98
16	1.11	1.11	1.05	1.15	1.04
17	1.11	1.12	1.14	1.07	1.07
18	1.12	1.11	1.08	1.04	1.04
19	1.09	1.14	1.08	1.09	1.01
20	1.11	1.10	1.04	1.06	0.98

Table VI. Phosphorimager Data for the Kinetic Isotope Studies Examining DNA Cleavage by Ni(II) GGH γ δ (141-183) in the Presence of Monoperoxyphthalic Acid when Proteo Thymidine is Present at Both Cleavage Positions in the Experimental and Control Duplexes

	Normalized Cleavage of the All-Proteo Experimental Duplex				
Lane #	Gel #26	Gel #27	Gel #28	Gel #29	Gel #30
1	1.17	1.02	1.03	1.02	1.02
2	1.10	0.99	1.08	1.05	1.00
3	1.08	1.03	1.05	1.10	0.99
4	1.11	1.02	1.06	1.10	0.97
5	1.05	1.00	1.03	1.08	0.98
6	1.06	1.05	1.16	1.07	0.96
7	1.04	0.99	1.12	1.05	1.09
8	1.11	1.05	1.07	1.14	1.01
9	1.08	0.97	1.11	1.06	1.01
10	1.02	1.01	1.07	1.06	0.98

	Normalized Cleavage of the All-Proteo Control Duplex				
Lane #	Gel #26	Gel #27	Gel #28	Gel #29	Gel #30
11	1.09	1.03	1.07	1.02	0.96
12	1.10	1.05	1.02	1.04	0.95
13	1.04	1.10	1.05	1.02	0.94
14	1.09	1.05	1.06	1.04	0.94
15	1.08	1.03	1.11	1.02	0.91
16	1.10	10.6	1.05	1.03	0.93
17	1.09	1.06	1.06	1.05	0.97
18	1.07	1.02	1.03	1.04	0.94
19	1.09	1.02	1.04	1.05	0.96
20	1.07	1.02	1.01	1.05	0.94

Appendix C: Mass Spectrometry of Peptides

Mass Spectrometry. The synthetic proteins used in the previously described studies were examined by time of flight mass spectrometry. This technique in addition to amino acid analysis and sequencing of the peptides confirmed their identity as the desired γδ resolvase analogs. Mass spectroscopy was performed at the micro sequencing facility at the California Institute of Technology. Analysis was performed on an Applied Biosystems Bio Ion 20R mass spectrometer using plasma desorption and ^{252}Cf as the radiation source. The expected and observed masses for the purified proteins is shown in Table 1.

Table I. Mass Spectroscopy of γδ resolvase Affinity Cleaving Peptides. Values are in atomic mass units and expected mass is calculated based on normal isotopic distributions

Protein	Expected Mass	Observed Mass
γδ(141-183)	4888.6	4891.4
EDTA-γδ(141-183)	5248.0	5259.7
γδ(141-183)-EDTA (Asn^{183}- Lys^{183})	5245.9	5238.3
EDTA-γδ(141-183)-EDTA (Asn^{183}- Lys^{183})	5605.2	5620.1
GGH-γδ(141-183)	5139.9	5145.9
Ni(II) GGH-γδ(141-183)	5198.6	5190.7

As can be seen from Table I all of the observed masses match fairly well to the expected masses. The accuracy of the measurement is approximately 1% (50 mass units for a protein of molecular weight 5000). The differences between the observed and expected masses fall with in this range. The observed masses of the various EDTA labeled compounds are consistent with EDTA incorporation.

It should be noted that mass spectroscopy of γδ resolvase analogs was problematic. The proteins were not readily desorbed. They contain a high

number of lysines and arginines which leads to a high rate of fragmentation. The proteins also contain two methonines. During the course of analysis these methonines tended to become oxidized to their corresponding sulfoxides and sulfones. These factors combine to produce a low signal and distribution of masses that makes identification of the molecular ion peak difficult.