Design, Synthesis and Characterization of Sequence-Specific DNA-Cleaving Metalloproteins

Thesis by

Martha G. Oakley

In Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

California Institute of Technology

Pasadena, California

1994

(Submitted November 8, 1993)

To My Parents

ACKNOWLEDGMENTS

I would like to thank my research advisor, Professor Peter Dervan, for his guidance and for giving me the opportunity to work on two exciting and very different projects. I cannot imagine that there are very many research groups in which both of the projects described in this thesis could be done or very many advisors who would allow them to be done by the same person. I'd also like to thank Peter for his enthusiastic support when my research was going well and, more importantly, for his patient yet equally strong support during the slower periods. I have Peter's infectious enthusiasm to thank not only for its direct effect on me and my research but also for attracting the exceptional group of people with whom I've had the privilege to work during my time at Caltech.

I'd like to thank all the Dervan Group members, past and present, for their contributions to this work and to my understanding of science. I would especially like to thank the people who contributed directly to this thesis. The study described in Chapter Three was done in collaboration with Milan Mrksich, who provided me with 2-Imidazole-Netropsin and with whom I have enjoyed many helpful and stimulating scientific discussions on a variety of topics. I'd also like to thank Ken Turnbull, to whom I am deeply grateful both for his direct contributions to the bleomycin project and for his advice on matters synthetic and otherwise. In addition, I'd like to thank Jen Wales, Jim Kelly, and George Best for their careful reading of this thesis and helpful comments.

Special thanks also go to a number of people whose contributions were less direct. Warren Wade and John and Linda Griffin helped me get started in the Dervan group; Dave Mack taught me everything I needed to know

about peptide synthesis; and Laura Kiessling, Dave Horne, and Erick Carreira provided generous synthetic advice. Scott Singleton has been my friend, colleague and labmate for my entire time in the Dervan Group, and I simply could not have asked for a better person with whom to share the highs and lows of the graduate school experience. I was also very fortunate to have colleagues and friends like Laura Kiessling, Mark Distefano, Karen Draths, Jumi Shin, Kevin Luebke, Hogyu Han, and Jen Wales, without whom grad school would have been infinitely less enjoyable, entertaining and stimulating. In addition, I'd like to thank Jürg Hunziker and Guido Kurz for making my last few months go by more quickly and for making sure I didn't spend my last summer in Southern California without going to the beach.

I've also been fortunate in the friends I had outside the lab. Annmarie Eldering has been a wonderful person with whom to share an apartment (some of the time) and a friendship (all of the time). I'd like to thank her for making sure that I knew a *few* people who weren't chemists and for all the great times I had with her and with the friends I met through her, especially Lisa Maggiore. Amazingly, Erica Johnson, whom I've been lucky enough to have as a friend for over twenty years now, turned up at Caltech a little over a year ago; she deserves a great deal of the credit for keeping me sane as I was trying to finish up.

Finally, I'd like to thank my family, Mom, Dad, Susan, Robert, Stuart, Grandma Carrie, and Grandma Jack, for their love and support, without which this thesis would never have been written.

ABSTRACT

This thesis describes research pursued in two areas, both involving the design and synthesis of sequence specific DNA-cleaving proteins. The first involves the use of sequence-specific DNA-cleaving metalloproteins to probe the structure of a protein-DNA complex, and the second seeks to develop cleaving moieties capable of DNA cleavage through the generation of a non-diffusible oxidant under physiological conditions.

Chapter One provides a brief review of the literature concerning sequence-specific DNA-binding proteins. Chapter Two summarizes the results of affinity cleaving experiments using leucine zipper-basic region (bZip) DNA-binding proteins. Specifically, the NH_2-terminal locations of a dimer containing the DNA binding domain of the yeast transcriptional activator GCN4 were mapped on the binding sites 5′-CTGACTAAT-3′ and 5′-ATGACTCTT-3′ using affinity cleaving. Analysis of the DNA cleavage patterns from Fe•EDTA-GCN4(222-281) and (226-281) dimers reveals that the NH_2-termini are in the major groove nine to ten base pairs apart and symmetrically displaced four to five base pairs from the central C of the recognition site. These data are consistent with structural models put forward for this class of DNA binding proteins. The results of these experiments are evaluated in light of the recently published crystal structure for the GCN4-DNA complex. Preliminary investigations of affinity cleaving proteins based on the DNA-binding domains of the bZip proteins Jun and Fos are also described.

Chapter Three describes experiments demonstrating the simultaneous binding of GCN4(226-281) and 1-Methylimidazole-2-carboxamide-netropsin (2-ImN), a designed synthetic peptide which binds in the minor groove of DNA at 5′-TGACT-3′ sites as an antiparallel, side-by-side dimer. Through the

use of Fe•EDTA-GCN4(226-281) as a sequence-specific footprinting agent, it is shown that the dimeric protein GCN4(226-281) and the dimeric peptide 2-ImN can simultaneously occupy their common binding site in the major and minor grooves of DNA, respectively. The association constants for 2-ImN in the presence and in the absence of Fe•EDTA-GCN4(226-281) are found to be similar, suggesting that the binding of the two dimers is not cooperative.

Chapter Four describes the synthesis and characterization of PBA-β-OH-His-Hin(139-190), a hybrid protein containing the DNA-binding domain of Hin recombinase and the putative iron-binding and oxygen-activating domain of the antitumor antibiotic bleomycin. This 54-residue protein, comprising residues 139-190 of Hin recombinase with the dipeptide pyrimidoblamic acid-β-hydroxy-L-histidine (PBA-β-OH-His) at the NH_2 terminus, was synthesized by solid phase methods. PBA-β-OH-His-Hin(139-190) binds specifically to DNA at four distinct Hin binding sites with affinities comparable to those of the unmodified Hin(139-190). In the presence of dithiothreitol (DTT), Fe•PBA-β-OH-His-Hin(139-190) cleaves DNA with specificity remarkably similar to that of Fe•EDTA-Hin(139-190), although with lower efficiency. Analysis of the cleavage pattern suggests that DNA cleavage is mediated through a diffusible species, in contrast with cleavage by bleomycin, which occurs through a non-diffusible oxidant.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER ONE: PROTEIN-DNA RECOGNITION

CHAPTER TWO: STRUCTURAL INVESTIGATIONS OF LEUCINE ZIPPER PROTEINS BY AFFINITY CLEAVING

CHAPTER THREE: EVIDENCE THAT A DESIGNED SYNTHETIC PEPTIDE AND A DNA-BINDING PROTEIN CAN SIMULTANEOUSLY OCCUPY A COMMON BINDING SITE ON DOUBLE-HELICAL DNA....85

CHAPTER FOUR: DESIGN OF A HYBRID PROTEIN CONTAINING THE DNA-BINDING DOMAIN OF HIN RECOMBINASE AND THE IRON BINDING LIGAND OF BLEOMYCIN

CHAPTER ONE

PROTEIN-DNA RECOGNITION

Many key cellular functions, including DNA replication, transcription of active genes, and DNA repair are performed by DNA-binding proteins. Regulation of these functions is crucial to the proper control of differentiation, response to extracellular signals, and cell growth; in many cases, this regulation is achieved by sequence-specific DNA-binding proteins. Often, relatively small, discrete domains are sufficient for sequence-specific DNA recognition, allowing for extensive biochemical, genetic, and structural studies of these DNA-binding domains. Although the principles by which proteins bind to specific sequences of DNA remain incompletely understood, such studies have indicated that many DNA-binding proteins can be grouped into classes that use related structures for recognition.[1,2] A great deal of progress has been made in our understanding of protein-DNA interactions since the first three-dimensional structures of sequence-specific DNA-binding proteins were reported over a decade ago.[3-5] Much of our improved understanding results from X-ray crystallographic and multidimensional NMR studies of these proteins in complex with their DNA binding sites. The first cocrystal structure was reported in 1986;[6] at present, more than twenty such structures, representing members of most known classes of DNA-binding proteins, have been solved. Excellent reviews of DNA-binding proteins have been published recently,[1,2,7] and reviews of the discovery of DNA-binding proteins[8] and early structural work in the field[9] can be found in theses by former Dervan group members. The intent of this chapter is to provide a brief description of the classes of sequence-specific DNA-binding

proteins for which structural data are available, with an emphasis on structures reported in the last two years and their relevance to our understanding of the principles of protein-DNA recognition.

HELIX-TURN-HELIX PROTEINS

Prokaryotic Repressors.

The helix-turn-helix (HTH) proteins were the first class of sequence-specific DNA-binding proteins to be discovered.[10,11] One subclass of HTH proteins consists of a group of prokaryotic repressor proteins. The three dimensional structures of protein-DNA complexes for seven of these proteins, including the repressor and cro proteins of the bacteriophages 434[12-15] and λ[16-18], and the *E. coli* CAP protein[19] and *lac*[20] and *trp*[21] repressors have been determined. In addition, the structure of the *E. coli* Fis protein in the absence of DNA has been solved[22,23], and preliminary data have been reported for the protein-DNA complex of Hin recombinase from *Salmonella typhimuruim*.[24] These proteins bind to their C_2-symmetric DNA operator sites as symmetric homodimers. Each protein monomer contains an approximately 20 residue HTH element that forms two α-helices connected

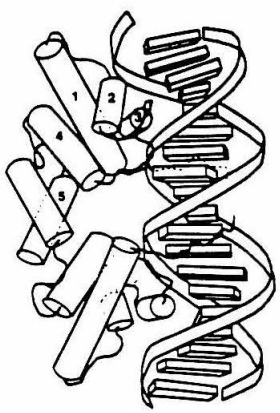


Figure 1. A schematic view of the λ repressor dimer bound to DNA.[17]

by a tight turn and crossing each other at an angle of approximately 120°.[10,11,25] The second of these helices is often described as the "recognition helix" due to its importance in determining sequence specficity,[26] and it lies in the major groove of DNA (Fig. 1).

The amino acid sequences for the HTH elements for the proteins listed above are shown in Figure 2. In addition to the examples listed, numerous proteins have been classified as HTH proteins based on sequence similarity.[25] Conserved features in the HTH sequence include a glycine at residue 9 to allow a sharp turn, hydrophobic residues at positions 4 and 15 that permit conserved packing of the helices against one another, and a small residue, usually glycine or alanine, and rarely a β-branched residue, at position 5, which is wedged between the two helices.[25]

		HELIX 1								Turn				HELIX 2								
Position	-2	1				5				10				15				20				
λ CRO(14)	F	G	Q	T	K	**T**	**A**	K	D	L	**G**	V	Y	Q	S	A	**I**	N	K	A	I	H (35)
434 CRO (16)	M	T	Q	T	E	**L**	**A**	T	K	A	**G**	V	K	Q	Q	S	**I**	Q	L	I	E	A (37)
λ REP (31)	L	S	Q	E	S	**V**	**A**	D	K	M	**G**	M	G	Q	S	G	**V**	G	A	L	F	N (52)
434 REP (16)	L	N	Q	A	E	**L**	**A**	Q	K	V	**G**	T	T	Q	Q	S	**I**	E	Q	L	E	N (37)
LAC REP (4)	V	T	L	Y	D	**V**	**A**	E	Y	A	**G**	V	S	Y	Q	T	**V**	S	R	V	V	N (25)
CAP (167)	I	T	R	Q	E	**I**	**G**	Q	I	V	**G**	C	S	R	E	T	**V**	G	R	I	L	K (188)
TRP REP (66)	M	S	Q	R	E	**L**	**K**	N	E	L	**G**	A	G	I	A	T	**I**	T	R	G	S	N (93)
FIS (72)	G	N	Q	T	R	**A**	**A**	L	M	M	**G**	I	N	R	G	T	**L**	R	K	K	L	K (93)
HIN (158)	H	P	R	Q	Q	**L**	**A**	I	I	F	**G**	I	G	V	S	T	**L**	Y	R	Y	F	P (179)

Figure 2. Sequences of a number of HTH proteins. Sequences can be found in references 9, 22, 23, and 25.

As an example of sequence-specific recognition of DNA by HTH proteins, the complex of λ repressor with its DNA operator is illustrated in Figure 3.[17] Importantly, residues both inside and outside the conserved HTH motif make crucial contacts to the DNA bases and backbone. The NH_2 termini of both HTH helices are in a position to allow their helix dipoles to interact favorably with the phosphate backbone. Sequence-specific contacts are made

exclusively in the major groove, which allows greater discrimination between base pairs than the minor groove.[27] The NH_2 terminus of the second HTH helix (helix 3) projects into the major groove allowing sequence specific hydrogen-bond contacts to the DNA bases to be made by Gln-44 and Ser-45.[17] Additional sequence-specific contacts are made by Asn-55, located in the loop after helix 3, and Lys 4 in the NH_2-terminal arm, a region required for efficient DNA binding (Fig. 3A).[28,29] (A higher resolution structure demonstrates that Lys-3 also makes a specific contact through its main-chain carbonyl group).[18] Hydrophobic contacts involving Ala-49, Ile-54, Gly-46, and Ser-45 are also made with thymine C5-methyl groups in base pairs 3 and 5.[17] Nonspecific contacts with the phosphodiester backbone are also made by residues within and without the HTH unit (Fig. 3B); these contacts are thought to be crucial for anchoring the protein to make the appropriate sequence-specific contacts.[2,11,17]

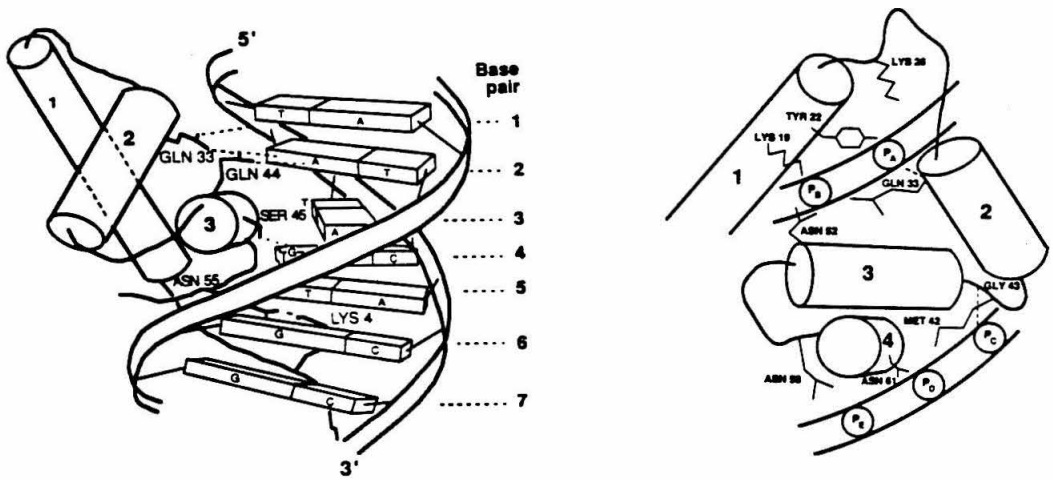


Figure 3. Schematic views of the interaction of a λ repressor monomer with its half site showing specific **(A)** and nonspecific contacts **(B)**.[17]

As is seen in other HTH proteins, some contacts, such as the hydrogen-bonding of the Gln-44 side chain amido group with N6 and N7 of adenine, had been anticipated as those likely to discriminate among base pairs.[27] In

other cases, such as the hydrogen bond between the side chain hydroxyl group of Ser-45 and the N7 of guanine, contacts were unanticipated and, acting alone, would not discriminate between base pairs. Another indication of the highly cooperative nature of specific interactions is that more than one side chain may be required to recognize a specific base pair; for example, the conformation of Gln-44 is stabilized by Gln-33, which also makes a contact with the phosphate backbone (Fig. 4). This arrangement of the Gln residues at the N-termini of both helices is also observed in the 434 repressor structure.[15] In spite of this conserved arrangement, the first two residues of the recognition helix in 434 repressor interact with adjacent base pairs, while the equivalent residues in λ repressor contact bases separated by one base pair, highlighting the difficulty of predicting the sequence specificity of closely related proteins.[15,17] Furthermore, interaction of multiple base pairs with one amino acid residue, as is seen with Ser-45 in the λ repressor structure, is also observed in many protein-DNA complexes.[2,11]

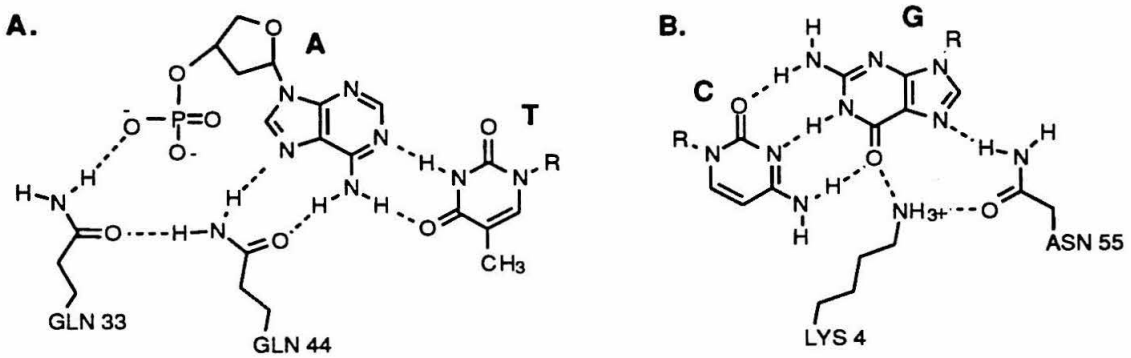


Figure 4. Sketches demonstrating multiple side-chain interactions in the major groove of **(A)** base pair 2 and **(B)** base pair 6 of the λ repressor operator half-site. Modified from reference 17.

The use of elements outside the HTH region for stabilizing contacts, such as those made by the NH_2-terminal arm of λ repressor,[29] is also seen in other

HTH proteins. The CO_2H-terminal domain of λ cro is critical for high affinity DNA binding.[30] In addition, the NH_2-terminal arm of the 434 repressor makes important contacts in the minor groove that are proposed to contribute to the sequence specificty of the protein.[15] Similary, the NH_2-terminal region of the Hin DNA-binding domain interacts with the minor groove in a manner that contributes significantly to its affinity and specificity.[8,31-34]

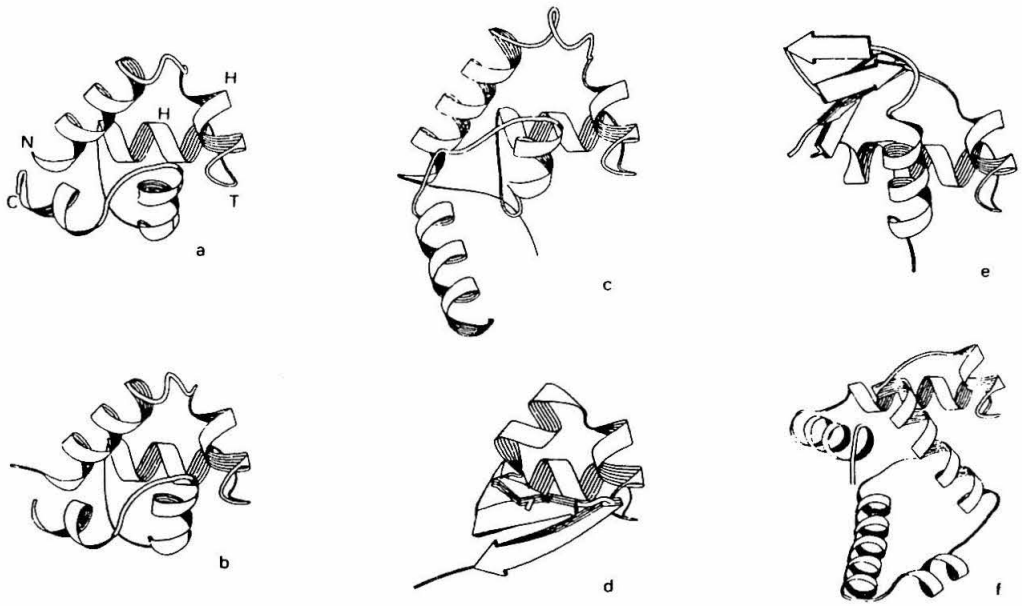


Figure 5. HTH-containing domains from a variety of proteins.[11] All HTHs are shown in the same orientation; the HTH helices and turn are labeled in part A. **(A)** 434 repessor DNA-binding domain. **(B)** Cro. **(C)** λ repressor DNA-binding domain. **(D)** λ Cro. **(E)** CAP DNA-binding domain. **(F)** *Trp* repressor.

It is important to note that the HTH element is not a stably folded domain; rather, it is a substructure that occurs in a variety of otherwise different, cooperatively folded domains (Fig. 5).[11] Although the three-dimensional structure of the HTH element is highly conserved across the known HTH proteins, the angle at which the second helix of the HTH interacts with the major groove varies significantly across this subclass in a

manner that depends on the structure of the rest of the domain.[11] For instance, the 434 proteins and λ repressor bind similarly, with their recognition helices approximately parallel to the major groove, pointing toward the center of the dimeric binding site.[11] In contrast, the recognition helix of CAP lies parallel to the bases[19] and the recognition helix of the *lac* repressor lies in the opposite orientation, projecting away from the center of the dimeric binding site.[20]

Another difference in the structures of HTH protein-DNA complexes is in the extent of DNA bending observed. The structure of the DNA in these complexes varies from the near canonical B-form seen in the λ repressor structure[17] to the 90° bend observed with CAP.[19] In the latter case, the severe bending, imposed largely by two 40° kinks allows the protein to contact 27 out of the 30 base pairs in its recognition site, rather than the 20 base pairs that could be contacted in the absence of bending.[19] The DNA bending in the CAP complex also contributes to the protein's sequence specificity. For example, both kinks occur at highly conserved 5'-TG-'3' steps thought to be more easily kinked than other sequence combination.[19] In addition, an A-T rich narrow minor groove in the confers sequence specificity on both the CAP complex and the 434 repressor complex.[15,19]

Sequence-dependent conformational flexibility of the DNA, or "indirect readout," has been proposed to account almost exclusively for the specific binding of the *trp* repressor to its operator site.[21] Remarkably, no direct hydrogen bonds or nonpolar contacts are seen between the protein and the DNA base pairs.[21] However, the DNA deformations observed are not far from B-form and are much less extensive than those discussed above, leading some investigators to suggest that the complex contains an incorrect binding site and is thus an example of non-specific DNA binding.[35] Later studies

have shown that the operator site used is in fact a high affinity site for trp repressor;[36,37] however, it appears that the DNA interaction is altered under the conditions of crystallization, resulting in a nonspecific complex.[25,37] Resolution of this controversy awaits further structural characterization of *trp*-operator complexes.

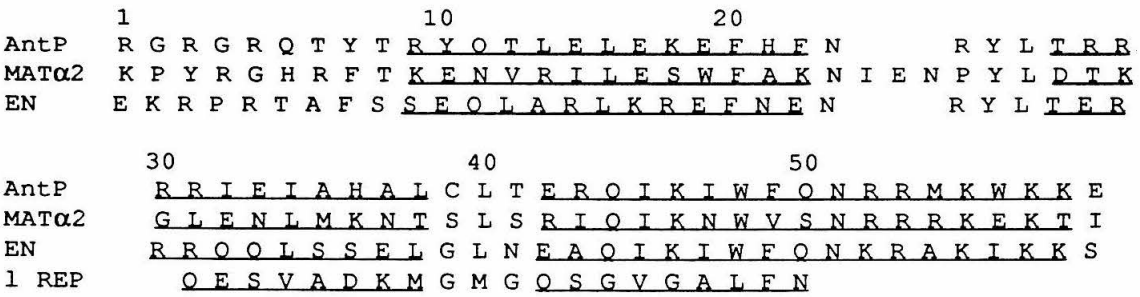


Figure 6. Sequences of the homeodomains of Antennepedia, MATα2, and engrailed. The sequence of the HTH from λ-repressor is shown for comparison. Underlined regions correspond to helices. Sequences are from references 8, 25, 42.

Homeodomain Proteins.

A number of eukaryotic transcriptional regulatory proteins contain a 60 amino acid sequence known as the homeodomain.[38] The structure of the Drosophila antennepedia protein and its complex with DNA have been determined by NMR,[39,40] and cocrystal structures of the Drosophila engrailed protein and the yeast repressor MAT α2 have been solved.[41,42] These structures have confirmed predictions based on weak sequence similarity that these proteins contained a HTH motif (Fig. 6).[43,44] Unlike the prokaryotic repressor HTH proteins, the homeodomain is a stably folded structure, and homeodomain proteins bind to their recognition sites as monomers.[1] The structure of the engrailed homeodomain is very similar to that of the other two homeodomains, consisting of an extended NH_2-terminal arm and three α-helices (Fig. 7A).[41] Helices 1 and 2 pack against each other in an

antiparallel arrangement, spanning the major groove approximately perpendicular to the DNA backbone.[41] Helix 3 is located in the major groove, perpendicular to helices 1 and 2, and makes numerous contacts with the DNA (Fig. 7).[41] The second HTH helix is markedly longer in homeodomain proteins than in the prokaryotic repressors; in the antennepedia structure, this long helix is kinked and has been described as two separate helices.[39] Sequence specfic contacts are made in both the major and minor grooves, the latter by the NH_2-terminal arm (Fig. 7).[41] Although the structure of the HTH element (helices 2 and 3) of the engrailed homeodomain superimposes well with the HTH from λ repressor, its position in relation to the DNA is markedly different. Specifically, the DNA is shifted toward the CO_2H terminus of the second HTH helix and rotated slightly away from the NH_2-terminus of the first HTH helix in comparison to the λ operator.[41]

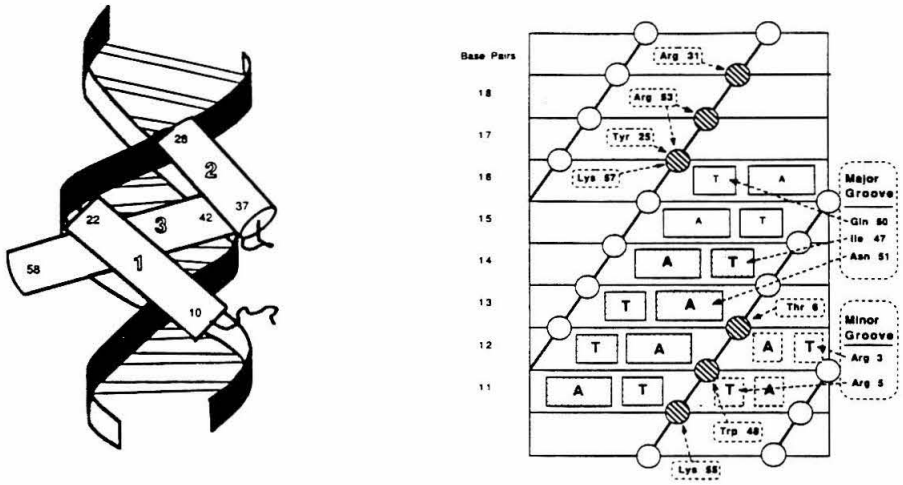


Figure 7. Schematic views of **(left)** the engrailed homeodomain-DNA complex and **(right)** the contacts made by engrailed in the major and minor grooves of DNA.[41]

Although MAT α2 is only 27% homologous to engrailed over its homeodomain (Fig. 6), the structure of its complex with its DNA binding site is remarkably similar to that of engrailed (compare Fig. 7 with Fig. 8).[42] Both

proteins bind to B form DNA with few distortions.[41,42] Six phosphate contacts and one base contact are conserved between the two protein complexes; these contacts appear to position the homeodomains properly with respect to the DNA binding sites.[42] Because the residues involved in these contacts are conserved in all homeodomain proteins, it is likely that, unlike prokaryotic repressors, all homeodomain proteins will be positioned similarly with respect to DNA.[42] The remaining base contacts are not conserved between the two complexes, allowing them to bind their different binding sites specifically, with base contacts including those predicted by genetic studies.[45] Finally, the NH_2-terminal arms of all three homeodomain proteins lie in the minor groove in similar conformations.[40-42]

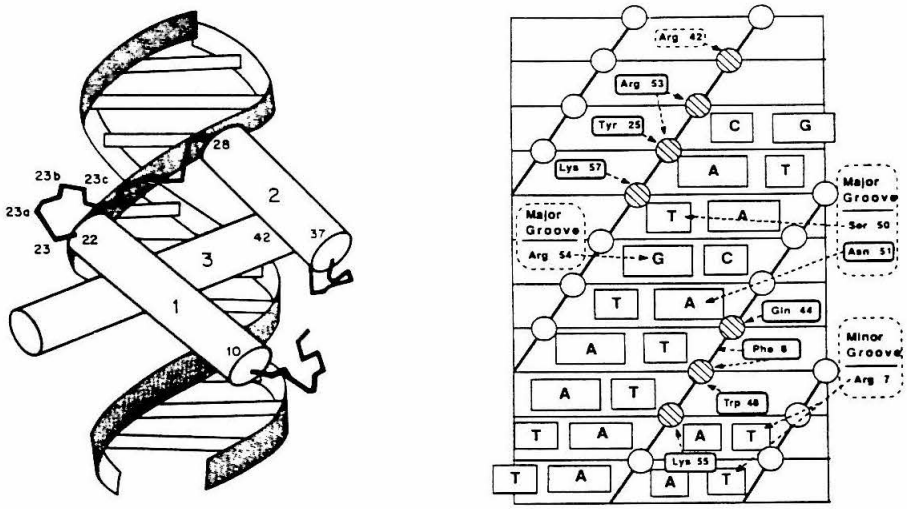


Figure 8. Schematic views of **(left)** the MATα2 homeodomain-DNA complex and **(right)** the contacts made by MATα2 in the major and minor grooves of DNA.[42]

Other Eukaryotic HTH Variants.

Several eukaryotic DNA-binding proteins unrelated to the homeodomain proteins have been characterized structurally and shown to contain elements similar to the HTH. Among these, a cocrystal structure has been reported for

only one protein-DNA complex, that of the HNF-3γ DNA-binding domain complexed to its recognition site.[46] The NH_2-terminal half of this domain forms a three-helix structure that places the third helix in the major groove of DNA.[46] Regions of β-sheet and extended structure also interact with the DNA in the minor groove.[46] Overall, the structure of the DNA-binding domain bears a marked resemblance to that of the globular domain of histone 5, which may interact with DNA in a similar fashion.[47] Another interesting feature of the HNF-DNA complex is a network of four well ordered water molecules, some of which mediate sequence specific hydrogen bonds between amino acid side chains and the DNA bases.[46]

Solution structures have also been reported for the protooncogene Myb[48] and for the POU-specific domain of the transcription factor Oct-1.[49,50] The POU domain structure resembles the DNA-binding domains of the λ and 434 repressors, with an extended first HTH helix and linker.[49,50] Similarly, helices 2 and 3 of Myb form a helix-turn-helix structure differing from the prototypical HTH in the length and makeup of its turn and in the relative orientation of the two helices.[48] Should these proteins recognize their binding sites as predicted, they will comprise new classes of eukaryotic HTH proteins.

ZINC BINDING PROTEINS

TFIIIA-like Zinc Finger Proteins.

A number of otherwise unrelated proteins contain conserved cysteine and/or histidine residues and require zinc for proper folding and for the recognition of DNA.[51,52] The first class of zinc-containing DNA-binding proteins to be described were identified through sequence similarities to a

repeated motif in TFIIIA.[53] The consensus sequence of this repeat is (Phe, Tyr)-X-Cys-$X_{2\text{ or }4}$-Cys-X_3-Phe-X_5-Leu-X_2-His-X_3-His-X_5, where X is any amino acid, and most zinc finger proteins contain between two and nine repeats of this sequence.[54] Based on similarities with metalloproteins of known structure, a model was proposed for the structure of these zinc fingers.[54,55] In this model, a two-stranded antiparallel β-sheet, connected by a loop after the first conserved Cys residue, is followed by an α-helix beginning between the cysteine and histidine residues (Fig. 9).[54,55] The packing of the hairpin and helix was proposed to be stablized by conserved hydrophobic interactions and by tetrahedral coordination of the Zn^{2+} ion by the conserved cysteine and histidine residues.[54,55] Solution structures of fingers from various zinc finger proteins have been determined, and their structure is in general accord with the predicted structure, with stretches of 3_{10} helix in the helical region and, in some cases, minor deviations from β-haipin structure.[56-60]

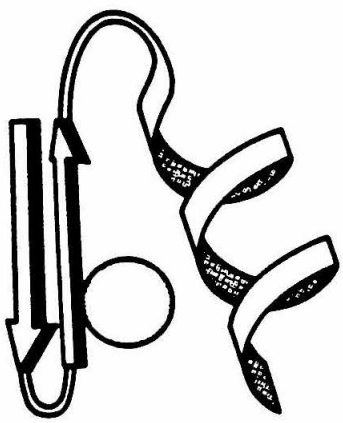


Figure 9. Schematic view of proposed zinc finger model.[54]

The structures of two protein-DNA complexes containing zinc finger proteins have been determined.[61,62] In the first complex, a three-finger fragment from Zif interacts with its DNA binding site in a very

straightforward manner: each finger contacts three base pairs of the nine base pair binding site with the amino-terminal end of its α-helix,[61] consistent with the model put forward by Berg.[54,63] The fingers are related to one another by a rotation of approximately 96° (3 x 32°) and a translation of roughly 10 Å (3 x 3.3Å) along the DNA helix (Fig. 10).[61] The DNA is essentially B-form.[61] Most of the DNA contacts, including all 11 hydrogen bonds made to the DNA bases, are with the G-rich strand of the DNA binding site (Fig. 10).[61]

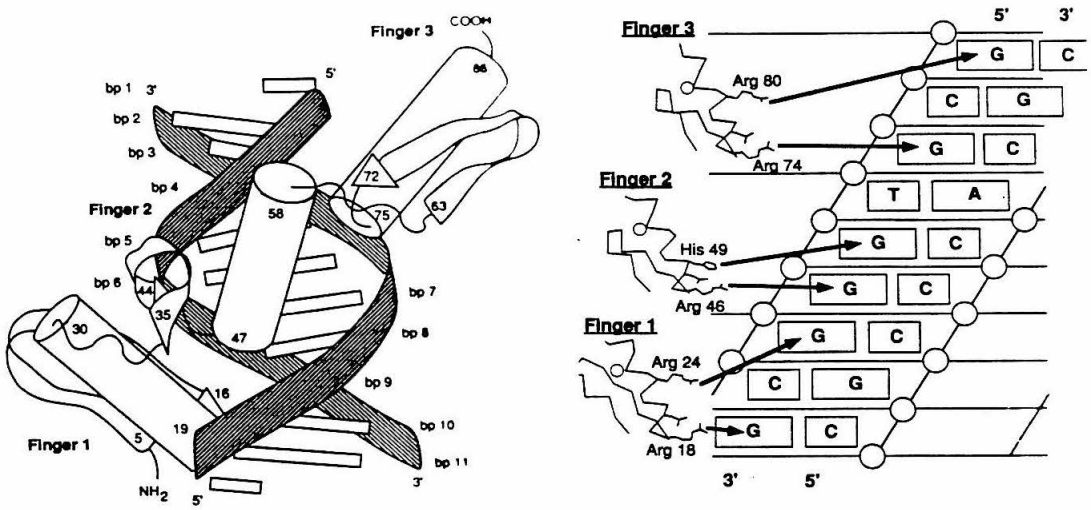


Figure 10. Schematic views of **(left)** the Zif-DNA complex and **(right)** the sequence-specific base contacts made by each Zif finger in the major groove of DNA.[61]

Contacts with the bases are made by the residue immediately preceding the α-helix, as well as the residues in the second, third, and sixth positions of the helix;[61] some of these interactions had been predicted from model building and genetic studies.[64] For fingers 1 and 3, the residues in those positions are Arg, Asp, Glu, and Arg, respectively; for finger 2, they are Arg. Asp, His, and Thr. In all cases the first Arg residue forms hydrogen bonds with N7 and O6 of the 3'-guanine in the recognition sequence; the

conformation of the Arg side chain is stabilized by the Asp residue.[61] In the first and third fingers, the second Arg interacts in a similar fashion with the 5'-guanine, and in the second finger, the histidine forms a hydrogen bond with N7 or O6 of the central guanine (Fig. 10).[61] Two conserved non-specific contacts are made by each finger, including a phosphate hydrogen bond by the first histidine coordinated by zinc, thereby involving the zinc atom in a DNA contact.[61] The other conserved phosphate contact is made by a conserved Arg on the second β-strand.[61] The protein makes an additional three contacts to the phosphate backbone, one of which involves the C-rich strand and constitutes the only direct contact with this strand. [61]

This straightforward, modular mode of recognition led to optimism that zinc fingers could be designed to bind each of the 64 possible 3 base sequences, providing a means for the construction of proteins with any desired specificity.[61] Although some efforts at achieving novel specificity through "finger swap" experiments have been successful,[64-66] the recently determined crysal structure of the DNA-binding domain of the GLI-DNA complex demonstrates that interactions by TFAIII-like zinc finger proteins cannot necessarily be predicted from our knowledge of the Zif-DNA complex (Fig. 11).[62]

Fingers 2-5 of the human glioblastoma DNA-binding domain fit into the major groove and wrap around the DNA for a helical turn, but the first finger does not contact the DNA.[62] Fingers 2 and 3 make several contacts with the phosphate backbone, but only a single hydrogen bond in finger 2 contacts a DNA base.[62] Most of the contacts to DNA bases occur in fingers 4 and 5, with a correlation between the positions of the residues contacting DNA bases and the positions of Zif residues serving a similar function.[62] However, only a weak correlation is observed between the position of these residues in the two

proteins that contact the phosphodiester backbone.[62] In addition, there are differences in the way that each finger docks against the DNA that preclude prediction of the GLI base contacts from the Zif contacts.[62] The structure of the DNA is also different in the two complexes; the region of the GLI DNA in which base contacts occur has a structure intermediate between B- and A-form DNA.[62] Thus, it appears that, like prokaryotic HTH element, the TFIIIA-like zinc finger is a DNA-binding element that may be presented to the DNA helix in a variety of different orientations. As the amino acid residues contacting the DNA will depend on the docking angle of the zinc finger, residues in different positions will make base contacts in different complexes, making these contacts difficult to predict.

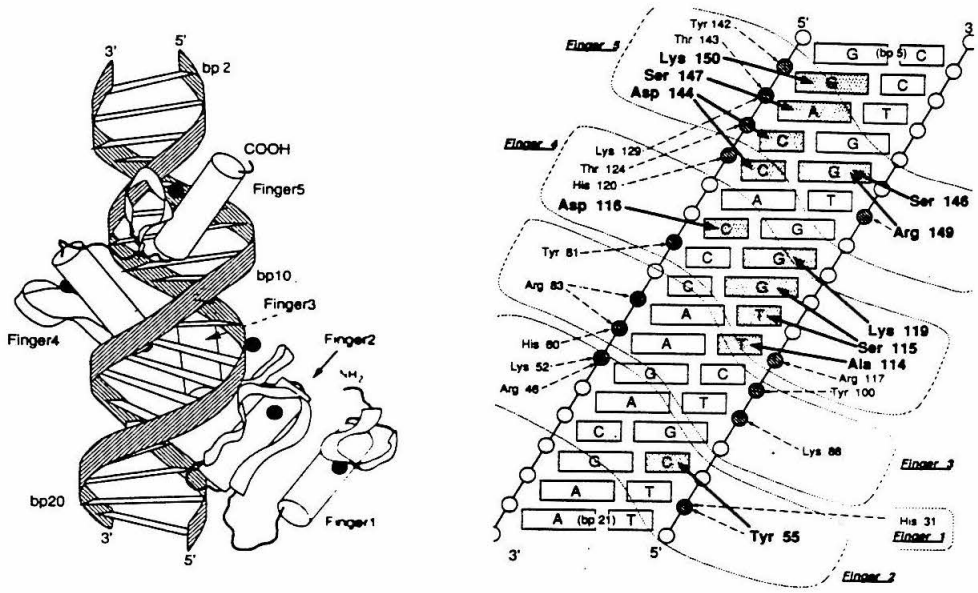


Figure 11. Schematic views of **(left)** the GLI-DNA complex and **(right)** the base and phosphate contacts made by each GLI finger with the DNA.[62]

Steroid Receptors.

The steroid receptors are a group of regulatory proteins including recpetors for various steroid hormones, vitamin D, retinoids, and thyroid

hormones. Each protein contains separate domains for ligand binding, transcriptional activation, and DNA binding.[67,68] Their DNA-binding domains consist of approximately 70 residues and contain eight conserved cysteine residues, with the conserved spacing Cys-X_2-Cys-X_{13}-Cys-X_2-Cys-X_{15-17}-Cys-X_5-Cys-X_9-Cys-X_2-Cys-X_4-Cys, where the last cysteine is not required for DNA binding and the first eight were proposed to form two zinc-binding modules (Fig. 12).[69] Steroid receptors bind to DNA as dimers, and their palindromic DNA binding sites can be classified into three groups: glucocoriticoid, estrogen, and thyroid response elements.[69] The glucocorticoid response elements (GREs) differ from the estrogen response elements (EREs) in sequence, while the EREs differ from the thryoid response elements (TREs) in the spacing between conserved inverted repeats (Fig. 12).[69] Three amino acid residues between and following the second pair of

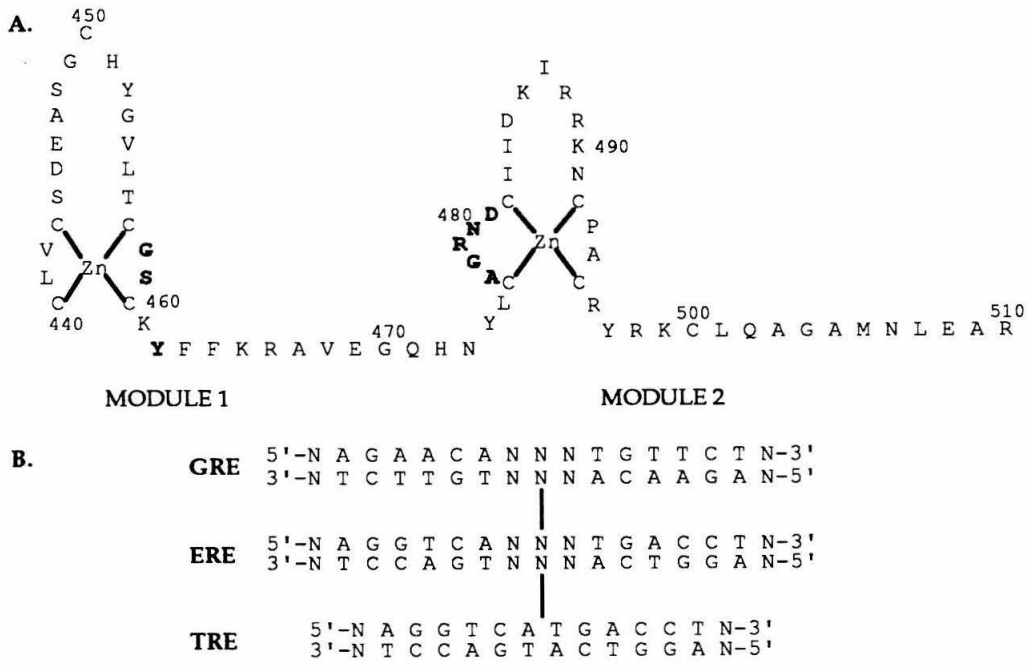


Figure 12. A. Sequence of the GR DNA-binding domain indicating cysteine residues serving as Zn^{2+} ligands. **B.** Three classes of steroid receptor binding sites. Adapted from reference 52.

cysteines (residues Gly-458, Ser-459, and Val-462 in the glucocorticoid receptor) have been idenified as responsible for distinguishing between the EREs and the GREs.[70,71] Similarly, residues in the second zinc-binding module (residues 478-481 in the glucocorticoid receptor) discriminate between TREs and EREs.[72]

Solution structures of the DNA-binding domains of the estrogen receptor (ER) and the glucocorticoid receptor (GR) have been reported, as has a cocrystal structure of the GR-DNA complex (Fig. 13).[73-75] In all three structures, the protein shows a similar globular fold, distinct from the largely independent folding of TFIIA-like zinc fingers. Each module contains a nucleated zinc coordination center followed by an amphipathic α-helix, the helices extending from residues Ser-459 to Glu 469 and from residue Pro-493 to Gly-504 in the GR.[74,75] The helices pack against one another and are stabilized by hydrophobic interactions between conserved residues.[73-75] Overall, the structure of the GR-DNA complex is very close to that predicted based on the structure of the GR;[74] the two modules serve different functions, with the first making contacts with the DNA and the second mediating dimerization (Fig. 13).[75] The DNA is largely B-form, and DNA contacts occur in the major groove. The region identified as discriminating between EREs and GREs makes contacts with the DNA bases; specifically, Lys-461 makes a water-mediated hydrogen bond with guanine, Val-462 is in van der Waals contact with thymine, and Arg-466 makes two hydrogen bonds to a second guanine (Fig. 13).[75]

One interesting feature of the GR-DNA complex results from an incorrect spacing of the two GR half-sites relative to one another. In the high-resolution complex, a spacing of four nucleotides, rather than the optimal three nucleotide gap, was used between the two conserved four base pair half

sites.[75] However, the spacing between the two monomers of the dimeric complex is more absolutely fixed than anticpated, resulting in the dimer binding to one optimal half site and a nonspecific half site three nucleotides away.[75] This preference underscores the influence of protein-protein interactions between monomers in selecting a binding site and explains why the residues found to discriminate between EREs and TREs are located in the dimer interface.[75]

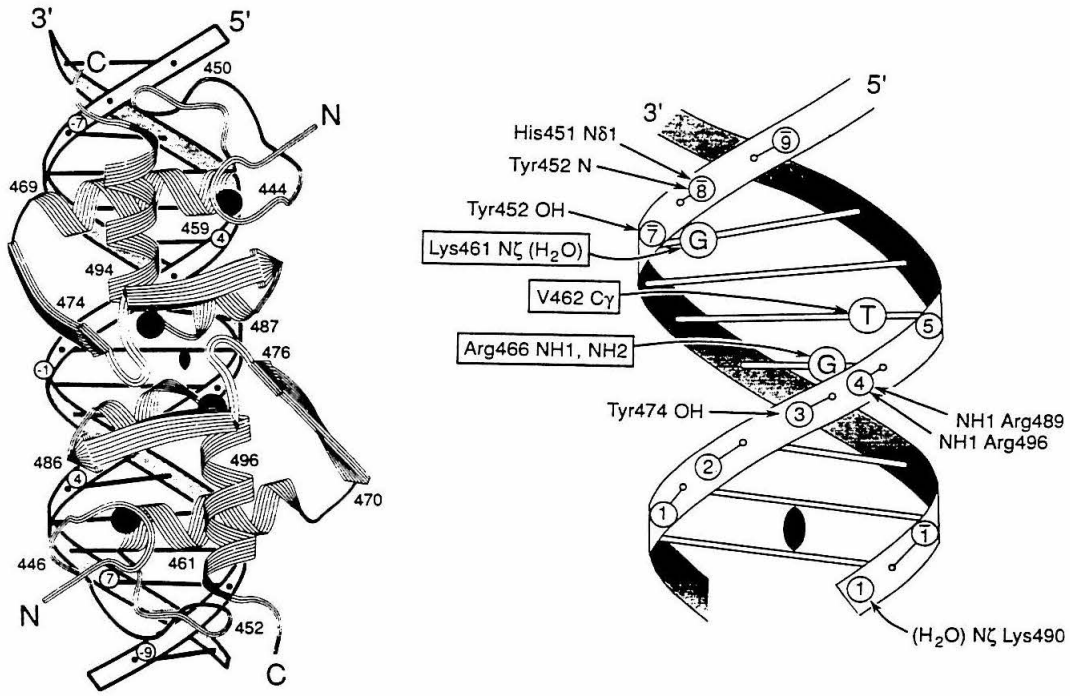


Figure 13. Schematic views of **(left)** the GR-DNA complex and **(right)** the contacts made by GR with its specific DNA binding site.[75]

GAL4.

Another class of sequence-specific DNA-binding proteins requiring zinc are related to the yeast transcriptional activator Gal4.[76] GAL4 is a transcriptional activator that recognizes a dyad symmetric 17-bp site as a dimer; the NH_2-terminal 65 amino acids contain its DNA-binding and

dimerization activities.[76,77] The DNA-binding domain contains six conserved cysteine residues that coordinate with two Zn^{2+} ions to form a binuclear metal cluster.[78,79] The structure of the DNA binding domain of GAL4 in complex with its DNA binding site has been determined by x-ray crystallographic techniques (Fig. 14),[80] and the solution structures of the zinc-binding domains have also been determined.[81,82] The compact, α-helical zinc-binding modules at the NH_2 terminus (residues 8-40) of each monomer lie in the major groove approximately a turn and a half of the DNA helix apart, making contacts to the conserved CCG triplets at the far ends of the palindromic binding site.[80] The zinc domain is connected by an extended linker region (residues 41-49) to short coiled-coil dimerization domain (residues 50-64);[80] this domain is monomeric and these regions are conformationally labile in the absence of DNA.[81] The DNA exhibits no significant deviations from B-form DNA.[80]

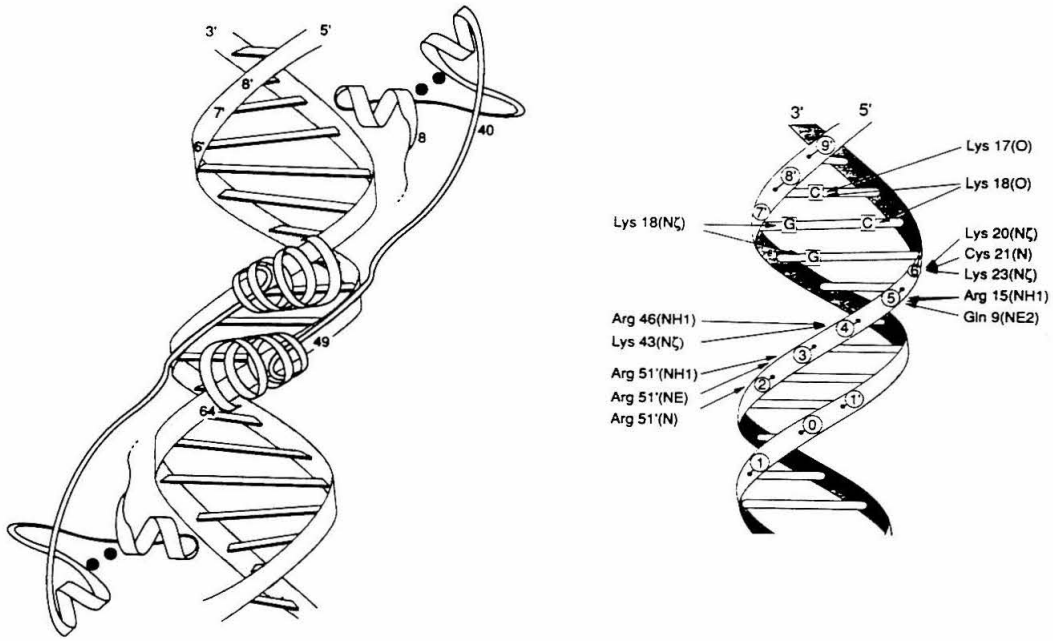


Figure 14. Schematic views of **(left)** the GAL4-DNA complex and **(right)** the contacts made by a GAL4 monomer with its half site.[7,80]

The coiled-coil region approaches the center of the DNA binding site in the minor groove, perpendicular to the double helical axis.[80] Its helical dipoles are directed toward the phophate backbone, and phosphate contacts are also made by the main-chain and side-chain NH groups of Arg-51.[80] Although the linker is less well ordered than other segments of the protein, several nonspecific contacts are made by positively charged side chains in this region.[80] Amino acid residues in the zinc-binding domain make base contacts with all three highly conserved base pairs (Fig. 14).[80] Interestingly, the interactions with two of the three conserved base pairs are through main-chain rather than side-chain atoms (the carbonyl groups of Lys-17 and Lys-18), indicating that altered specificity cannot be achieved by simple side-chain substitution of these residues.[80]

Other Zinc-Containing Nucleic Acid-Binding Proteins.

The three classes of zinc-containing, DNA-binding proteins discussed above almost certainly represent an incomplete list. Several other DNA transcription factors, such as the GATA factor, RAG-1 and RAD-18 contain cysteine-rich regions that may serve as metal binding domains.[2,51,52] In addition, a class of proteins related to the *gag* protein contain a small zinc-binding motif often referred to as the "retroviral knuckle."[2,51,52] The structure of the Cd^{2+} form of an 18 residue peptide from HIV *gag* has been determined by NMR; it is distinct from any of those described above.[83]

LEUCINE ZIPPER AND HELIX-LOOP-HELIX PROTEINS

Basic Region-Leucine Zipper Proteins.

A class of transcriptional activators and oncogenic proteins known as basic region-leucine zipper (bZIP) proteins were originally identified based on the sequence similarity of their dimerization domains.[84] As a thorough review of bZIP proteins is presented in Chapter Two, only a brief outline will be given here. The cocrystal structure of the bZIP domain of the yeast transcriptional activator GCN4 in complex with its DNA binding site has been solved.[85] Like GAL4, it contains a coiled-coil dimerization domain that is perpendicular to the DNA helix axis; however, in the GCN4 complex, this domain approaches the major groove.[85] Furthermore, no extended linker separates this domain from the basic region that makes specific and non-specific contacts with the DNA. Instead, the α-helices of the coiled-coil diverge smoothly and each monomer tracks along the major groove, resulting in a Y-shaped α-helical structure (Chapter Two, Fig. 8).[85] In contrast to GAL4, the dimerization domain of bZIP proteins forms a stable structure, but the basic region is largely disordered in the absence of DNA.[86-92]

Helix-Loop-Helix Proteins.

A number of transcriptional activators important in cell differentiation and growth control include a conserved element known as the helix-loop-helix (HLH) in their DNA-binding domains.[93] Two amphipathic α-helices linked by a loop of various lengths are found in this region, and it was proposed to serve as a dimerization domain.[93] Like the bZIP proteins, these HLH proteins contain a bipartite DNA binding domain consisting of a conserved basic region followed immediately by a dimerization domain, the

HLH.[94] One subclass of basic region-HLH (bHLH) proteins, including the c-Myc protein, also contain a leucine zipper region immediately C-terminal to the HLH element (Fig. 15).[95] For these proteins, both the HLH and zipper (Z) regions are required for dimerization and DNA binding.[95] The protein-DNA structure for the DNA binding domain of one of these b/HLH/Z proteins, Max, has recently been reported.[96] Max serves as the dimerization partner for Myc, but unlike Myc, it can form homodimers and recognize DNA sequence-specifically in the absence of its partner, making it an attractive candidate for structural studies.[95,97,98]

```
        <-----BASIC-----><- -HELIX 1---> <-LOOP->
MAX    ADKRAHHNALERKRRDHIKDSFHSLRDSVP SLQGEKAS
MYC    NVKRRTHNVLERQRRNDLRSRFLALRDQVPELENNEKAP
MYO D   ADRRKAATMRERRRLSKVNEAFETLKRCTS SNPNQRLP

       <---HELIX 2----><-----------ZIPPER---------->
MAX    RAQILDKATEYIQYMRRKNDTHQQDIDDLKRQNALLEQQVRALEKARS
MYC    KVVILKKATAYILSYQAEEQKLISEEDLLRKRREQLKHKLEQLRNSCA
MYO D   KVEILRNAIRYIEGLQALLRDQDAAP
```

Figure 15. Amino acid sequences of the b/HLH/Z proteins Myc and Max and the b/HLH protein MyoD. The basic, HLH, and zipper regions are indicated. Conserved residues are printed in bold type. Sequences and alignment were taken from reference 95.

The overall structure of the Max-DNA complex is reminiscent of the GCN4-DNA complex, approaching the major groove of DNA with a helical dimerization domain and positioning the basic region of each monomer appropriately for major groove contacts with a DNA binding site consisting of abutted inverted repeats (Fig. 16).[96] Unlike GCN4, Max contains a globular domain, the HLH dimerization domain, which forms a novel parallel four-helix bundle, a structure similar to those predicted by model building.[99-101] This structure is inconsistent with an antiparallel four-helix model based on NMR studies of an oxidized MyoD dimer. However, oxidized MyoD binds

DNA less actively than reduced MyoD and may not reflect the structure of the HLH upon DNA binding.[102]

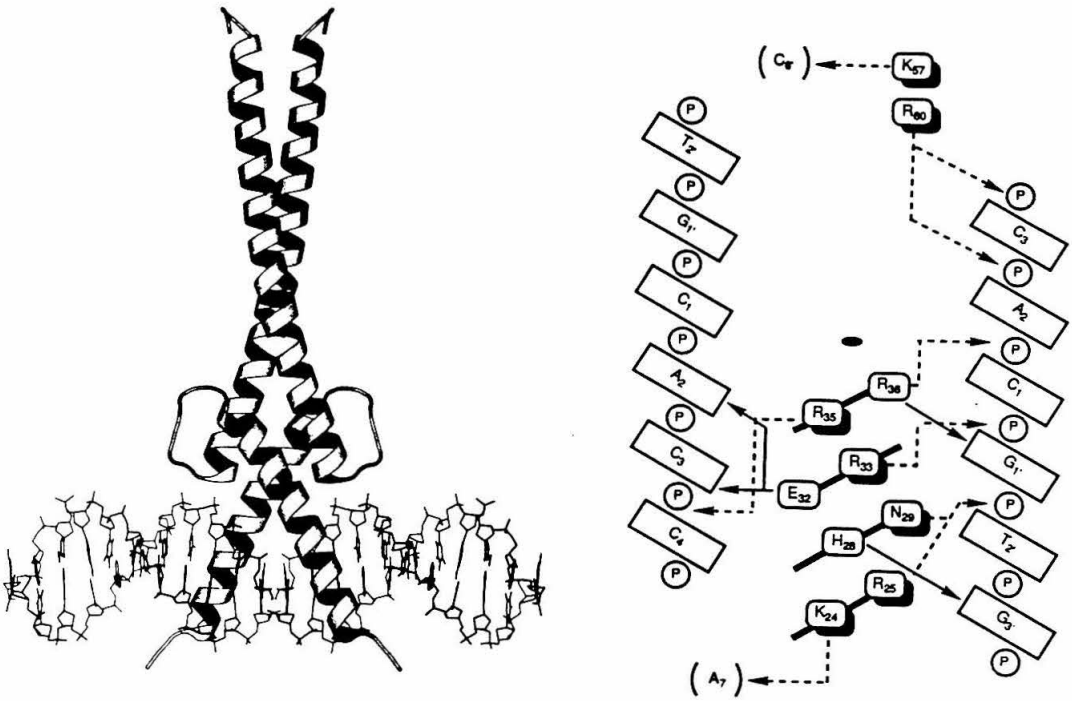


Figure 16. Schematic views of **(left)** the MAX-DNA complex and **(right)** the contacts made by a MAX monomer with its half site.[96]

Amino acid residues in the basic regions of Max make major groove contacts with each of the three bases in each half site as well as phosphate interactions spanning over six base pairs in each half-site in a manner similar to that seen with GCN4 (Fig 16).[96] Unlike GCN4, Max uses residues outside the basic region, including Arg-47 and Lys-57 in the loop region and Arg-60 at the start of the second HLH helix, to make phosphate contacts.[96] Lys-25 interacts with the minor groove, prompting the suggestion that proteins with long loop regions could make specific contacts with the adjacent minor groove, expanding the sequence selectivity of such proteins beyond the standard six base pair recognition element.[96] No contacts are made by the

zipper region, suggesting that bHLH proteins lacking a zipper domain will interact with the DNA very much as Max does.[96] As with GCN4, the DNA exhibits only minor deviations from B-form.[96]

BETA-RIBBON PROTEINS

Although α-helices provide the major source of protein-DNA contacts for the classes of DNA-binding proteins discussed above, other classes of proteins bind to their recognition sites through the use of β-sheet structure.[1,2,103] One such class of proteins includes the prokaryotic Met, Arc, and Mnt repressor proteins, and possibly the *E. coli* Tra Y proteins.[104,105] Mnt, Met and Arc repressors all bind as tetramers to 17 to 20 base pair operator sites.[106-109] The solution structure of Arc repressor has been determined, and the crystal structure of the Met repressor-operator complex is available.[104,108,110] In spite of only weak primary sequence homology (Fig. 17), the structures of the Arc and Met repressors are very similar; both form highly intertwined dimers in which residues near the NH_2-terminus of each monomer form an antiparallel β-sheet.[104,108]

```
              <SHEET-> <---HELIX----->          <--HELIX--->
Met  ...GKKSEQVKKITVSIPLKVLKILTDERTRRQVNNLRHATNSELLCEAFLHAFTGQPLPD...
Arc      MKGMSKMPQFNLRWPREVLDVLRKVAEEN      GRSVNSEIYQRVMESFKKEGRIGA
Mnt         ARDDPHFNFRMPMEVREKLKFRAEAN      GRSMNSELLQIVQDALSKPSPVTG
```

Figure 17. Sequences of the β-ribbon repressors. Regions of regular secondary structure are indicated. Sequences and alignment from reference 104.

In the Met repressor-DNA structure, two dimeric repressor molecules bind to adjacent direct repeats of eight base pairs with inter-dimer contacts suggestive of cooperative binding.[110] Each repressor dimer binds with its β-ribbon inserted into a compressed major groove making direct hydrogen bonds to the DNA bases, as predicted by a model based on the Arc solution structure (Fig. 18).[104,110] The β-ribbon is the primary source of sequence-

specific interactions, consistent with biochemical and genetic data for all three repressors.[106,111-114] However, many of the nonspecific phosphate contacts are made by residues in other segments of the protein, including the neighboring helices and loop regions (Fig. 18).[110]

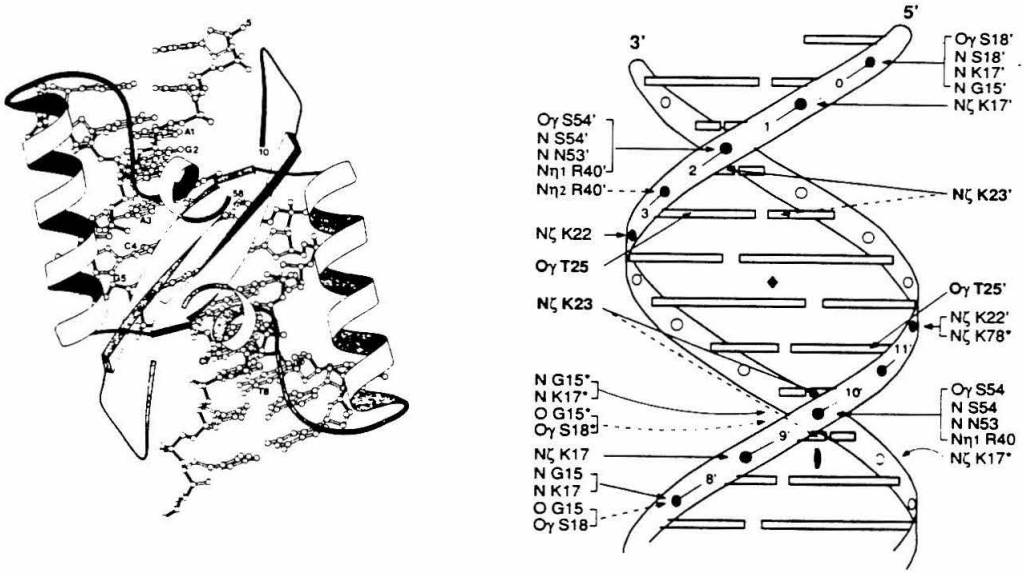


Figure 18. (Left) Ribbon diagram of the Met repressor-operator complex. **(Right)** Schematic view of the contacts made by a Met repressor dimer with its DNA site.[110]

It has been proposed that another group of DNA-binding proteins interacts with the minor groove of DNA through a β-ribbon motif.[1,2,103] The structure of a nonspecific histone-like protein, HU from *E. coli*, has been determined; the dimeric molecule has two symmetrically related β-arms positioned appropriately for DNA contacts.[115] The closely related DNA bindng protein IHF binds to DNA sequence-specifically in the minor groove, and a model has been proposed for the structure of its complex with DNA based on biochemical data and the HU structure.[116,117] Further structural studies will be required to test this model.

OTHER TRANSCRIPTIONAL ACTIVATORS

The cocrystal structures of the DNA-binding domains of the TATA-binding protein (TBP) and of the bovine papillomavirus E2 transcriptional activators complexed with DNA have been determined recently.[118-120] These DNA-binding domains bear no resemblance to those listed above, and although they are unlikely to represent general classes of DNA-binding proteins, their novel structural features are of interest.

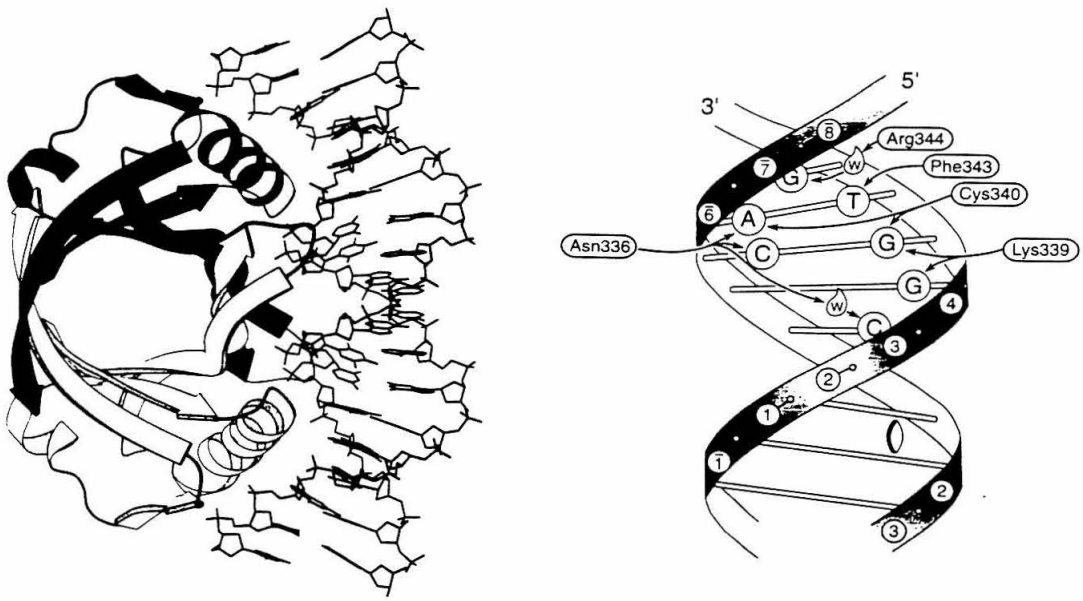


Figure 19. (Left) Ribbon diagram of the E2-DNA complex. **(Right)** Schematic view of the base contacts made by an E2 monomer at its DNA half site.[7,118]

The E2 protein is a transcriptional activator found in papillomaviruses; an 85 amino acid CO_2H-terminal domain from E2 is sufficient for both sequence-specific DNA binding and dimerization.[121-123] The crystal structure of the E2-DNA complex reveals an unprecedented dimeric antiparallel β-barrel with four strands contributed by each monomer. In contrast to the Met repressor, which uses a largely α-helical scaffold to present

a β-arm to the DNA, the β-barrel of E2 presents of pair of symmetrically related α-helices to the major grooves of each half site (Fig. 19).[118] Contacts to the bases are made by four conserved amino acid residues in each recognition helix, including an unprecedented hydrogen bond between the cysteine sulfhydryl group and O6 of guanine (Fig. 19).[118] In addition, a remarkable 10 direct and 14 water-mediated phosphate interactions are made with each half site, requiring significant DNA bending.[118] Unlike the bending observed in the prokaryotic HTH proteins, such as CAP, where the bending is localized mainly to two large kinks,[19] the DNA is smoothly bent with both major and minor grooves compressed on the concave side of the DNA that faces the protein.[19]

The TATA-binding protein, also known as TFIID, is required for transcription by all three eukaryotic RNA polymerases. The structure of TBP from *Arabidopsis thaliana* was reported in 1992; its structure is shown in Figure 20A.[124] It is a saddle-shaped structure with two symmetrically related and structurally similar domains of approximately 90 amino acids, each consisting of a five-stranded β-sheet and two α-helices.[124] On the basis of modeling studies, it was predicted that the DNA bound to the underside of TBP at a right angle to the long axis of the protein.[124]

The cocrystal structures of both the *A. thaliana* TBP and the DNA-binding domain of yeast TBP have been reported very recently.[119,120,125] Consistent with biochemical studies, the protein contacts the DNA in its minor groove, and the DNA helix is bent dramatically.[126-128] However, the DNA lies at an angle of 90° from the predicted orientation, following the long axis of TBP and bending to complement the concave face of the protein.[119,120] The entire eight base pair TATA box is bent severely toward the major groove,

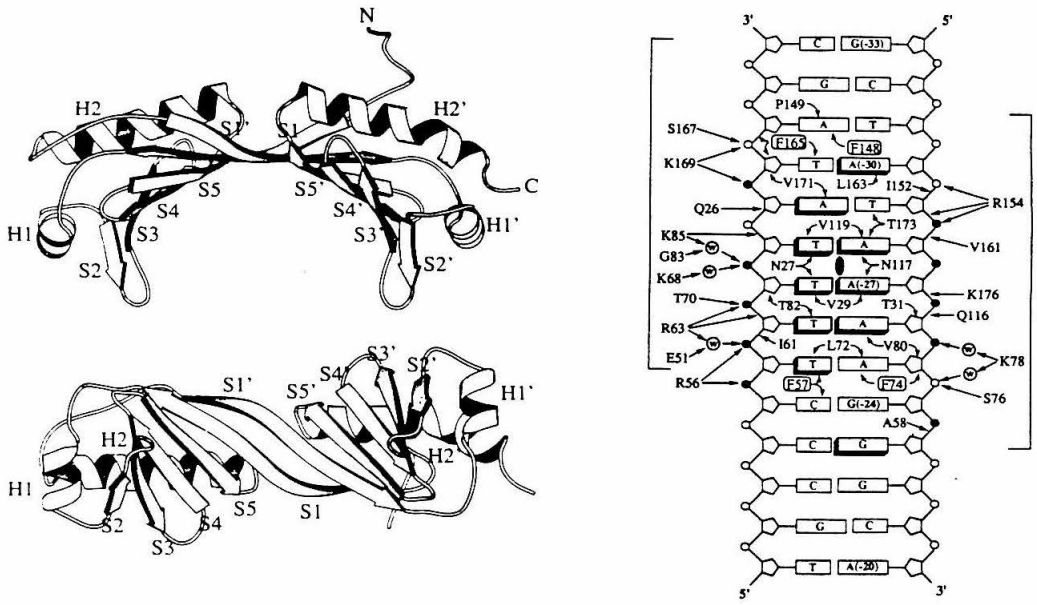


Figure 20. **(Left)** Ribbon diagrams showing two different views of the structure of TBP in the absence of DNA. **(Right)** Schematic view of the contacts made between TPB and the TATA box.[7,120]

allowing the protein to interact through its entire concave surface with its wide minor groove. Outside the TATA box, the DNA is suprisingly close to B-form; kinks between the first and second and seventh and eighth bases of the TATA box result from the wedging of two pairs of phenylalanine side chains between these base pairs.[119,120] The contact surface is much more hydrophobic than is usually found in protein-DNA interfaces, and most of the stabilization appears to result through van der Waals contacts with the DNA bases and the ribose backbone moieties.[119,120] There is an excellent correlation between residues contacting the DNA and mutations affecting DNA binding.[119,120] "Indirect readout" of the protein sequence through its sequence-dependent deformability is proposed to contribute not only to the DNA-binding specificity of TBP, but perhaps also to determine the polarity with which the pseudosymmetric protein binds, with important

consequences for the binding of RNA polymerases and the directionality of RNA transcription.[119]

CONCLUSION

A key lesson to be taken from studies of sequence-specific DNA-binding proteins to date is that there is no simple recognition code for protein-DNA binding; that is, sequence specific binding is not achieved through predictable contacts of side chains with the edges of the DNA bases. Instead, a complex and cooperative network of interactions between the protein side chain and main chain atoms with the DNA backbone and bases is used for high affinity and selective DNA binding. Studies of a number of protein structural classes demonstrate that amino acid residues may interact to recognize a single base pair, and, conversely, a single residue may make contacts with more than one base pair. From structural studies of more than one member of a single class of DNA-binding proteins, such as HTH and TFIIIA-like zinc finger proteins, it is clear that substructures may dock against the DNA at a variety of angles, rendering predictions and manipulation of sequence specificity difficult. Some groups of classes of proteins, such as the homeodomain or some subclasses of TFIIIA-like zinc fingers may provide a better starting point for the design or random selection of protein with altered sequence specificity.

The large and growing number of structural classes of DNA binding proteins illustrates the broad range of structural scaffolds that may be used to present elements of secondary structure to DNA. Although the majority of these classes present α-helices to the major groove of DNA, recent structures reveal the first examples of exclusive minor groove recognition[119,120] and the use of β-strand structure as a primary source of sequence specific

contacts.[110] In addition, the importance of both protein and DNA flexibility in forming sequence specific complexes has become clear over the last several years.[2,7,85] The recent explosion in the number of solved three-dimensional structures of protein-DNA complexes from a number of structural classes provides fertile ground for further chemical, genetic, and structural studies.

REFERENCES

1. S. C. Harrison, *Nature* **353**, 715 (1991).
2. C. O. Pabo and R. T. Sauer, *Annu. Rev. Biochem.* **61**, 1053 (1992).
3. W. F. Anderson, Y. Takeda, D. H. Ohlendorf, and B. W. Matthews, *Nature* **290**, 754 (1981).
4. C. O. Pabo and M. Lewis, *Nature* **298**, 443 (1982).
5. D. B. McKay and T. A. Steitz, *Nature* **290**, 744 (1981).
6. J. A. McClarin et al., *Science* **234**, 526 (1986).
7. C. Wolberger, *Curr. Opin. Struct. Biol.* **3**, 3 (1993).
8. J. A. Shin, Ph.D. Thesis, California Institute of Technology (1992).
9. J. P. Sluka, Ph.D. Thesis, California Institute of Technology (1988).
10. R. G. Brennan, *Curr. Opin. Struct. Biol.* **2**, 100 (1992).
11. S. C. Harrison and A. K. Aggarwal, *Annu. Rev. Biochem.* **59**, 933 (1990).
12. C. Wolberger, Y. Dong, M. Ptashne, and S. C. Harrison, *Nature* **335**, 789 (1988).
13. A. Mondragón and S. C. Harrison, *J. Mol. Biol.* **219**, 321 (1991).
14. J. E. Anderson, M. Ptashne, and S. C. Harrison, *Nature* **326**, 846 (1987).
15. A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, and S. C. Harrison, *Science* **242**, 899 (1988).

16. R. C. Brennan, S. L. Roderick, Y. Takeda, and B. W. Matthews, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8165 (1990).

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

18. L. J. Beamer and C. O. Pabo, *J. Mol. Biol.* **227**, 177 (1992).

19. S. C. Schultz, G. C. Shields, and T. A. Steitz, *Science* **253**, 1001 (1991).

20. R. Boelens, R. M. Scheek, J. H. van Boom, and R. Kaptein, *J. Mol. Biol.* **193**, 213 (1987).

21. Z. Otwinowski et al., *Nature* **335**, 321 (1988).

22. D. Kostrewa et al., *Nature* **349**, 178 (1991).

23. H. S. Yuan et al., *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9558 (1991).

24. J.-A. Feng et al., *J. Mol. Biol.* **232**, 982 (1993).

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

26. R. P. Wharton and M. Ptashne, *Trends Biochem. Sci.* **11**, 71 (1986).

27. N. C. Seeman, J. M. Rosenberg, and A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 804 (1976).

28. C. O. Pabo, W. Krovatin, A. Jeffrey, and R. T. Sauer, *Nature* **298**, 441 (1982).

29. N. D. Clarke, L. J. Beamer, H. R. Goldberg, C. Berkower, and C. O. Pabo, *Science* **254**, 267 (1991).

30. A. J. Hubbard et al., *Biochemistry* **29**, 9241 (1990).

31. J. Sluka, S. Horvath, M. Bruist, M. Simon, and P. Dervan, *Science* **238**, 1129 (1987).

32. J. P. Sluka, S. J. Horvath, A. C. Glasgow, M. I. Simon, and P. B. Dervan, *Biochemistry* **29**, 6561 (1990).

33. A. C. Glasgow, M. F. Bruist, and M. I. Simon, *J. Biol. Chem.* **264**, 10072 (1989).

34. K. T. Hughes, P. C. W. Gaines, J. E. Karlinsey, R. Vinayak, and M. I. Simon, *EMBO J.* **11**, 2695 (1992).

35. D. Staacke, B. Walter, B. Kisters-Woike, B. von Wilcken-Bergmann, and B. Müller-Hill, *EMBO J.* **9**, 1963 (1990).

36. T. E. Haran, A. Joachimiak, and P. B. Sigler, *EMBO J.* **11**, 3021 (1992).

37. J. Carey, D. E. A. Lewis, T. A. Lavoie, and J. Yang, *J. Biol. Chem.* **266**, 24509 (1991).

38. A. Laughon, *Biochemistry* **30**, 11357 (1991).

39. Y. Q. Qian et al., *Cell* **59**, 573 (1989).

40. G. Otting et al., *EMBO J.* **9**, 3085 (1990).

41. C. R. Kissinger, B. Liu, E. Martin-Blanco, T. B. Kornberg, and C. O. Pabo, *Cell* **63**, 579 (1990).

42. C. Wohlberger, A. K. Vershon, B. Liu, A. D. Johnson, and C. O. Pabo, *Cell* **67**, 517 (1991).

43. A. Laughon and M. P. Scott, *Nature* **310**, 25 (1984).

44. J. C. Shepherd, W. McGinnis, A. E. Carrasco, E. M. DeRobertis, and W. J. Gehring, *Nature* **310**, 70 (1984).

45. S. D. Hanes and R. Brent, *Science* **251**, 426 (1991).

46. K. L. Clark, E. D. Haley, E. Lai, and S. K. Burley, *Nature* **364**, 412 (1993).

47. V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, and R. M. Sweet, *Nature* **362**, 219 (1993).

48. K. Ogata et al., *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6428 (1992).

49. N. Assa-Munt, R. J. Mortishire-Smith, R. Aurora, W. Herr, and P. E. Wright, *Cell* **73**, 193 (1993).

50. N. Dekker et al., *Nature* **362**, 852 (1993).

51. J. M. Berg, *Current Opin. Struct. Biol.* **3**, 11 (1993).

52. R. Kaptein, *Curr. Opin. Struct. Biol.* **2**, 109 (1992).

53. J. Miller, A. D. McLachlan, and A. Klug, *EMBO J.* **4**, 1609 (1985).

54. J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 99 (1988).

55. T. J. Gibson, J. P. M. Postma, R. S. Brown, and P. Argos, *Protein Eng.* **2**, 209 (1988).

56. G. Párraga et al., *Science* **241**, 1489 (1988).

57. M. S. Lee, G. P. Gippert, K. V. Soman, D. A. Chase, and P. E. Wright, *Science* **245**, 635 (1989).

58. J. G. Omichinski, G. M. Clore, E. Appella, K. Sakaguchi, and A. M. Gronenborn, *Biochemistry* **29**, 9324 (1990).

59. J. G. Omichinski et al., *Biochemistry* **31**, 3907 (1992).

60. M. Kochoyan et al., *Biochemistry* **30**, 3371 (1991).

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

62. N. P. Pavletich and C. O. Pabo, *Science* **261**, 1701 (1993).

63. J. M. Berg, *J. Biol. Chem.* **265**, 6313 (1990).

64. J. Nardelli, T. J. Gibson, C. Vesque, and P. Charnay, *Nature* **349**, 175 (1991).

65. J. R. Desjarlais and J. M. Berg, *Proteins* **12**, 101 (1992).

66. J. R. Desjarlais and J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7345 (1992).

67. R. M. Evans, *Science* **240**, 889 (1988).

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

69. J. M. Berg, *Cell* **57**, 1065 (1989).

70. M. Danielsen, L. Hinck, and G. M. Ringold, *Cell* **57**, 1131 (1989).

71. S. Mader, V. Kumar, H. de Verneuil, and P. Chambon, *Nature* **338**, 271 (1989).

72. K. Umesono and R. M. Evans, *Cell* **57**, 1139 (1989).

73. J. W. R. Schwabe, D. Neuhaus, and B. Rhodes, *Nature* **348**, 458 (1990).

74. T. Härd et al., *Science* **249**, 157 (1990).

75. B. F. Luisi et al., *Nature* **352**, 497 (1991).

76. M. Johnston, *Microbiol. Rev.* **51**, 458 (1987).

77. M. Carey, H. Kakidani, J. Leatherwood, F. Mostashari, and M. Ptashne, *J. Mol. Biol.* **209**, 423 (1989).

78. J. F. Povey, G. P. Diakun, C. D. Garner, S. P. Wilson, and E. Laue, *FEBS Lett.* **266**, 142 (1990).

79. T. Pan and J. E. Coleman, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2077 (1990).

80. R. Marmorstein, M. Carey, M. Ptashne, and S. C. Harrison, *Nature* **356**, 408 (1992).

81. J. D. Baleja, R. Marmorstein, S. C. Harrison, and G. Wagner, *Nature* **356**, 450 (1992).

82. P. J. Kraulis, A. R. C. Raine, P. L. Gadhavi, and E. D. Laue, *Nature* **356**, 448 (1992).

83. M. F. Summers, T. L. South, B. Kim, and D. R. Hare, *Biochemistry* **29**, 329 (1990).

84. W. H. Landschulz, P. F. Johnson, and S. L. McKnight, *Science* **240**, 1759 (1988).

85. T. E. Ellenberger, C. J. Brandl, K. Struhl, and S. C. Harrison, *Cell* **71**, 1223 (1992).

86. K. T. O'Neil, R. H. Hoess, and W. F. DeGrado, *Science* **249**, 774 (1990).

87. K. T. O'Neil, J. D. Schuman, C. Ampe, and W. F. DeGrado, *Biochemistry* **30**, 9030 (1991).

88. L. Patel, C. Abate, and T. Curran, *Nature* **347**, 572 (1990).

89. J. D. Shuman, C. R. Vinson, and S. L. McKnight, *Science* **249**, 771 (1990).

90. R. V. Talanian, C. J. McKnight, and P. S. Kim, *Science* **249**, 769 (1990).

91. M. A. Weiss et al., *Nature* **347**, 575 (1990).

92. E. K. O'Shea, J. D. Klemm, P. S. Kim, and T. Alber, *Science* **254**, 539 (1991).

93. C. Murre, P. S. McCaw, and D. Baltimore, *Cell* **56**, 777 (1989).

94. X.-H. Sun and D. Baltimore, *Cell* **64**, 459 (1991).

95. G. C. Prendergast and E. B. Ziff, *Trends Genet.* **8**, 91 (1992).

96. A. R. Ferré-D'Amaré, G. C. Prendergast, E. B. Ziff, and S. K. Burley, *Nature* **363**, 38 (1993).

97. E. M. Blackwood and R. N. Eisenman, *Science* **251**, 1211 (1991).

98. G. J. Kato, W. M. F. Lee, L. Chen, and C. V. Dang, *Genes Dev.* **6**, 81 (1992).

99. T. D. Halazonetis and A. N. Kandil, *Science* **255**, 464 (1992).

100. S. J. Anthony-Cahill et al., *Science* **255**, 979 (1992).

101. C. R. Vinson and K. C. Garcia, *New Biol.* **4**, 396 (1992).

102. M. Starovasnik, T. K. Blackwell, T. M. Laue, H. Weintraub, and R. E. Klevit, *Biochemistry* **31**, 9891 (1992).

103. S.-H. Kim, *Science* **255**, 1217 (1992).

104. J. N. Breg, J. H. J. van Opheusden, M. J. M. Burgering, R. Boelens, and R. Kaptein, *Nature* **346**, 586 (1990).

105. J. U. Bowie and R. T. Sauer, *J. Mol. Biol.* **211**, 5 (1990).

106. A. K. Vershon, J. U. Bowie, T. M. Karplus, and R. T. Sauer, *Proteins* **1**, 302 (1986).

107. A. K. Vershon, S.-M. Liao, W. R. McClure, and R. T. Sauer, *J. Mol. Biol.* **195**, 311 (1987).

108. J. B. Rafferty, W. S. Somers, I. Saint-Girons, and S. E. V. Phillips, *Nature* **341**, 705 (1989).

109. B. M. Brown, J. U. Bowie, and R. T. Sauer, *Biochemistry* **29**, 11189 (1990).

110. W. S. Somers and S. E. V. Phillips, *Nature* **359**, 387 (1992).

111. P. Youderian, A. Vershon, S. Bouvier, R. T. Sauer, and M. M. Susskind, , (1983).

112. K. L. Knight and R. T. Sauer, *J. Biol. Chem.* **264**, 13706 (1989).

113. K. L. Knight and R. T. Sauer, *EMBO J.* **11**, 215 (1992).

114. Y.-Y. He et al., *Nature* **359**, 431 (1992).

115. I. Tanaka, K. Appelt, J. Dijk, S. W. White, and K. S. Wilson, *Nature* **310**, 376 (1984).

116. S. W. White, K. Appelt, K. S. Wilson, and I. Tanaka, *Proteins* **5**, 281 (1989).

117. C.-C. Yang and H. A. Nash, *Cell* **57**, 869 (1989).

118. R. S. Hegde, S. R. Grossman, L. A. Laimins, and P. B. Sigler, *Nature* , 505 (1992).

119. Y. Kim, J. H. Geiger, S. Hahn, and P. B. Sigler, *Nature* **365**, 512 (1993).

120. J. L. Kim, D. B. Nikolov, and S. K. Burley, *Nature* **365**, 520 (1993).

121. N. Dostani, F. Thiery, and M. Yaniv, *EMBO J.* **7**, 3807 (1988).

122. I. Giri and M. Yaniv, *EMBO J.* **7**, 2823 (1988).

123. A. A. McBride, J. C. Byrne, and H. P. M., *EMBO J.* **7**, 533 (1988).

124. D. B. Nikolov et al., *Nature* **360**, 40 (1992).

125. R. Kaptein, E. R. P. Zuiderweg, R. M. Scheek, R. Boelens, and W. F. V. Gunsteren, *J. Mol. Biol.* **182**, 179 (1985).

126. D. B. Starr and D. K. Hawley, *Cell* **67**, 1231 (1991).

127. D. K. Lee, M. Horikoshi, and R. G. Roeder, *Cell* **67**, 1241 (1991).

128. M. Horikoshi et al., *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1060 (1992).

CHAPTER TWO

STRUCTURAL INVESTIGATIONS OF LEUCINE ZIPPER PROTEINS BY AFFINITY CLEAVING

Affinity Cleaving.

One of the most common means of investigating the interaction of a DNA-binding molecule with its recognition site is through chemical or enzymatic footprinting (Fig. 1, left).[1-4] In this technique, a DNA sample is allowed to react with a non-specific DNA-cleaving agent under conditions that result in the random cleavage of each DNA molecule at one position. If the ^{32}P-end-labeled DNA products from such a reaction are subjected to polyacrylamide gel electrophoresis, a ladder of cleavage bands results. When the reaction is carried out in the presence of a specific DNA-binding molecule, cleavage is blocked at the positions to which the ligand binds, resulting in a "footprint" in the DNA cleavage ladder. Such experiments yield valuable information about the location and size of the DNA binding sites for a given DNA-binding molecule.

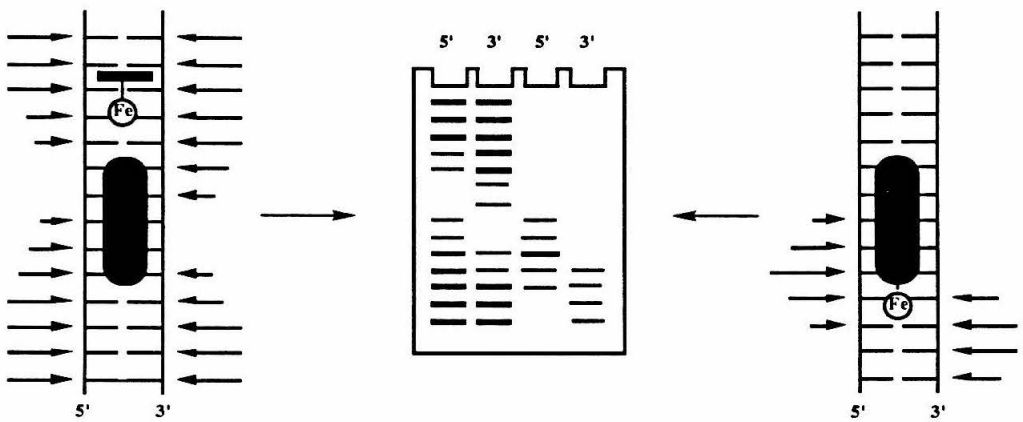


Figure. 1. Footprining and affinity cleaving assays.

A complementary technique known as affinity cleaving has been developed in the Dervan group (Fig. 1, right).[5] In this assay, the iron chelating agent EDTA is attached to a sequence-specific DNA-binding molecule to afford a sequence-specific DNA-cleaving molecule that functions at physiologically relevant pH, temperature, and salt conditions.[2,3,6,7] Following chemical activation with a reducing agent such as dithiothreitol (DTT), Fe•EDTA localized at a specific DNA binding site cleaves both DNA strands, typically over 4 to 6 base pairs, via a diffusible species, presumably hydroxyl radical (Fig. 1, right).[6,7] Because the cleaving moiety is not sequence specific, the cleavage specificity is derived from the binding specificity of the molecule under investigation. The location of the DNA cleavage provides information not only about the preferred binding sites of the DNA-binding molecule but also about its orientation at those sites.

In addition, due to the right-handed nature of double helical DNA, the groove in which the Fe•EDTA is located can be identified by analysis of the cleavage patterns (Fig.2). An Fe•EDTA located in the minor groove generates an asymmetric cleavage pattern with maximal cleavage loci shifted to the 3′ side on opposite strands (Fig. 2).[6,7] When the Fe•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 (Fig. 2).[8,9] This results in a pair of 3′-shifted asymmetric cleavage loci of unequal intensity on opposite strands (Fig.2).[8,9] These patterns can be explained if the diffusible radical generated from the localized Fe•EDTA reacts in the major and minor grooves of DNA with unequal rates, preferentially (though not necessarily exclusively) in the minor groove[10] or if the major groove is partially protected from cleavage by the presence of a bulky DNA-binding molecule.

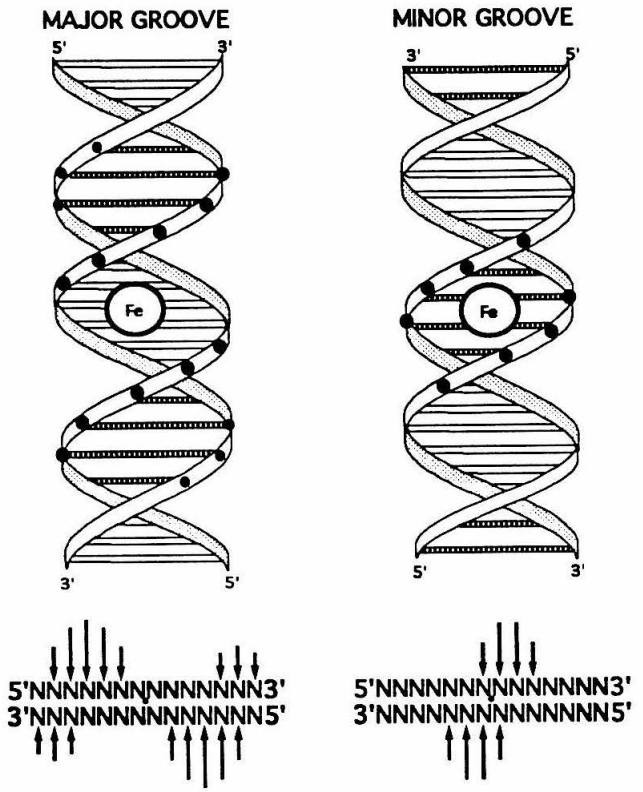


Figure 2. Cleavage patterns produced by a diffusible oxidant generated by Fe•EDTA located in the major and minor grooves of right handed DNA. Filled circles represent points of cleavage along the phosphodiester backbone. Sizes of circles represent extent of cleavage.

Affinity Cleaving Proteins.

Because the DNA-recognition domains of many DNA-binding proteins are of a reasonable size for solid phase peptide synthesis[11], chemical methods have been developed in the Dervan group to attach EDTA to DNA-binding proteins through solid-phase methods.[12-15] Incorporation of Fe•EDTA at discrete amino acid residues within a protein allows the positions of those residues in the protein-DNA complex relative to the DNA bases to be mapped. Such EDTA-equipped affinity cleaving proteins have been used to investigate the structure of the DNA binding domains of the putative helix-turn-helix proteins Hin recombinase and γδ resolvase.[12,14,16,17] By attaching

Fe•EDTA to the NH_2-termini of these proteins, the NH_2-terminal amino acid residues have been mapped to the minor groove near the center of the dimeric binding sites.[12,16,17] Peptides with Fe•EDTA attached near the CO_2H-terminus were used to determine the orientation of the putative recognition helices in the major groove of the binding sites[14,17]. In addition, the orientation of the recognition helix of the helix-turn-helix protein Lac repressor was confirmed using a protein modified with Fe•EDTA at its CO_2H-terminus.[18] These studies demonstrate the utility of affinity cleaving in the investigation of the structure of protein•DNA complexes. The studies described in this chapter extend the technique of affinity cleaving to another class of DNA bindng proteins, the basic region-leucine zipper (bZip) proteins.

Basic Region-Leucine Zipper Proteins.

A class of sequence-specific DNA binding proteins important in the regulation of gene expression is thought to bind DNA via a bipartite structural motif consisting of a DNA binding domain, termed the "basic region," and a dimerization domain, termed the "leucine zipper."[19,20] This class of proteins was first proposed by Landschulz *et al.*, who noted that several proteins, among them the transcriptional activators CEBP and GCN4, and the oncogenes Jun and Fos, contain a series of four or more heptad repeats of leucine.[21] These leucines occurred in an extremely α-helix permissive region. Recognizing that an α-helical structure would lead to the alignment of the leucines along one face of the helix, forming a very hydrophobic surface, these investigators suggested that this "leucine zipper" region constitutes a dimerization domain that also anchors the peptide on the DNA. In agreement with this model, both GCN4 and CEBP form homodimers.[21,22] Similarly, Jun and Fos have been shown to form

heterodimers[23-28], and Jun can also form homodimers.[26,28] Furthermore, deletion and site-directed mutagenesis analyses have demonstrated that the leucine zipper region is absolutely required for dimerization and for interaction with DNA.[23-33]

A basic region found immediately NH_2-terminal to and at a fixed distance from the leucine zipper region is also highly conserved among many of the leucine zipper proteins (Figure 3).[23-25,30,31] Compelling evidence of the importance of these basic residues was obtained in site-directed mutagenesis studies performed with several putative leucine zipper proteins. The mutation of groups of these basic residues to neutral residues severely affected the DNA binding affinities, but not the dimerization, of the Jun-Fos heterodimer, the Jun-Jun homodimer, and the CEBP homodimer.[23-25,30,31] These results supported the conclusion that the leucine zipper domain is necessary for dimerization, which is in turn required for DNA recognition, while the basic region is essential only for DNA recogntion.

```
GCN4    PESSD  PAALKR ARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER
yAP1    KQDLD RETKQKRTAQNRAAQRAFRQRKERKMKELEKKVQSLESIQQQNEVEATFLRDQLITL
v-jun   MESQERIKAERKR MRNRIAASKSRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQL
junB    MEDQERIKVERKR LRNRLAATKCRKRKLERIARLEDKVKTLKAENAGLSSAAGLLREQVAQL
junD    MDTQERIKAERKR LRNRIAASKCRKRKLERISRLEEKVKTLKSQNTELASTASLLRQQVAQL
v-fos   PEEEE  KRRIRR ERNKMAAAKCRNRRRELTDTLQAETDQLEDKKSALQTEIANLLKEKEKL
FRA1    PEEEE  RRRVRR ERNKLAAAKCRNRRKELTDFLQAETDKLEDEKSGLQREIEELQRQKERL
fos-B   PEEEE  KRRVRR ERNKLAAAKCRNRRRELTDRLQAETDQLEEKKAELESEIAELQKEKERL
CEBP    KNSNE  Y RVRR ERNNIAVRKSRDKAKQRNVETQQKVLELTSDNDRLRKRVEQLSTELDTL
CREB    QPAEE  RKREVR LRNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKALKDLYCHK
cpc-1   EDPSD  VVAMKR ARNTLAARKSRERKAQRLEELEAKIEELIAERDRWKNLALAHGASTE
cys-3             DKR KRNTAASARFRIKKKQREQALEKSAKEMSEKVTQLEGRIQALETENKWL
```

Figure 3. The peptide sequences of putative leucine zipper DNA binding proteins. Conserved residues are printed in bold. Sources of sequences: GCN4[34,35], yAP1,[36] jun,[37] junB,[38] junD,[39,40] fos[41,42], FRA1[43], fosB, [44]CEBP,[45] CREB,[46] cpc-1,[47] cys-3.[48]

Studies using peptides corresponding to the leucine zippers of GCN4, Jun, and Fos helped to define this region. First, Kim and co-workers demonstrated the GCN4 monomers were largely α- helical and aligned in a parallel fashion to form homodimers. They further noted the presence of a 4-3 repeat characteristic of classical coiled-coil proteins and suggested that the "leucine zipper" dimerization domain assumes this structure.[49] Similar investigations with the Jun and Fos leucine zippers confirmed that the dimerization specificity of the protein complexes is determined by this region.[50,51] Finally, the three-dimensional X-ray crystal structure of the leucine zipper region of GCN4 was determined and found to consist of a parallel two-stranded coiled-coil as predicted.[52]

The structure of the basic regions was probed through circular dichroism studies with intact bZip proteins, with peptides comprising the basic and leucine zipper regions, or with disulfide-linked basic region peptides. These studies demonstrated that this region is relatively disordered in the absence of DNA, but assumes an α-helical structure upon DNA binding.[53-58]

Two detailed structural models consistent with these results were proposed.[53,59] Although the models differ in their details, they have several major features in common. In both models, the leucine zippers form a parallel two-stranded coiled-coil running perpendicular to the axis of the DNA binidng site. The NH_2-terminal end of this region projects into the major groove of DNA, allowing the dipoles of the α–helices to interact favorably with the phosphate backbone. At the junction of the leucine zipper and basic regions, the α–helices of the dimer diverge and begin tracking in opposite directions along the major groove of the DNA binding site, resulting in a bifurcated, Y-shaped structure. As the DNA recognition elements for this

class of proteins consist of directly abutted half sites, this would allow one monomer to contact each half of the binding site in a symmetric fashion.

In their "scissors grip model," Vinson, Sigler, and McKnight proposed that the α-helical basic region forms a kink in order to allow all of the conserved basic residues to interact with DNA.[59] Specifically, they suggested the α–helices are broken by conserved asparagine residues, after which a second pair of α–helices begins at an angle of 75 degrees from the first. This structure would allow both halves of the DNA-binding domain to continue tracking along the major groove on the side opposite to that of their initial approach.

The "induced helical fork" model, O'Neil, Hoess, and DeGrado, on the other hand, contains no break in the helical structure of the DNA binding domain.[53] In addition, these investigators proposed that the conserved asparagine, as well as two conserved alanines and a conserved cysteine/serine were involved in specific contacts with the DNA.[53]

Both models proposed symmetrical contacts in the major groove of the palindromic DNA binding site.[53,59] Footprinting, methylation interference, and ethylation interference studies were consistent with this proposal. However, at the time our initial work was published,[60] relatively little other structural data were available. Recently, the three-dimensional structure of the GCN4-DNA complex has been reported.[61] In the next section, our studies with GCN4 affinity cleaving peptides will be presented. Our results and conclusions will then be evaluated in light of the known structure of the GCN4-DNA complex in the following section.

STRUCTURAL MOTIF OF THE GCN4 DNA BINDING DOMAIN CHARACTERIZED BY AFFINITY CLEAVING

This section describes the results of affinity cleaving studies designed to investigate the structure of the DNA binding domain of the yeast transcriptional activator GCN4, a putative basic region-leucine zipper (bZIP) protein. GCN4 is necessary for the coordinate induction of 30- 50 proteins involved in the biosynthesis of amino acids in response to amino acid starvation.[62] A functional dissection of GCN4 has demonstrated that the 60 amino acids at the CO_2H-terminus, residues 222 - 281, contain the specific DNA binding activity; however, 37 residues (245- 281) at the CO_2H-terminus were shown to be insufficient for DNA binding.[63] GCN4 exists as a dimer and the CO_2H-terminal 60 residues are sufficient for dimerization.[22] The optimum DNA binding site for GCN4 is 5'-rrTGAC<u>T</u>catt-3', where the underlined C corresponds to the center of pseudodyad symmetry.[64,65] Struhl and co-workers have proposed that this pseudosymmetric site behaves as two half sites, noting that mutation of the naturally occurring GCN4 binding site 5'-TGACTCT-3' to the perfectly palindromic 5'TGACTCA-3' increases the affinity of the protein for the DNA, while all other mutations in this region lead to unchanged or decreased affinity. However, mutation of the TGACT bases leads to the greatest reduction in binding affinity, and mutation of the central C to G abolishes activity, implying that the protein-DNA interactions at each half site are not equal.[64,65]

The DNA binding domain of the transcriptional activator GCN4, residues 222-281, contains the putative "basic" and "leucine zipper" regions[63]. Four different proteins based on the DNA binding domain of GCN4 were synthesized by solid phase methods (Fig. 4).[15] These proteins were incrementally shortened at the NH_2-terminus and differ in length from 45 to

60 amino acid residues. GCN4(237-281), (232-281), (226-281), and (222-281) were synthesized with and without EDTA attached to the NH_2-terminal residue for comparative affinity cleaving and footprinting studies of the corresponding protein-DNA complexes.

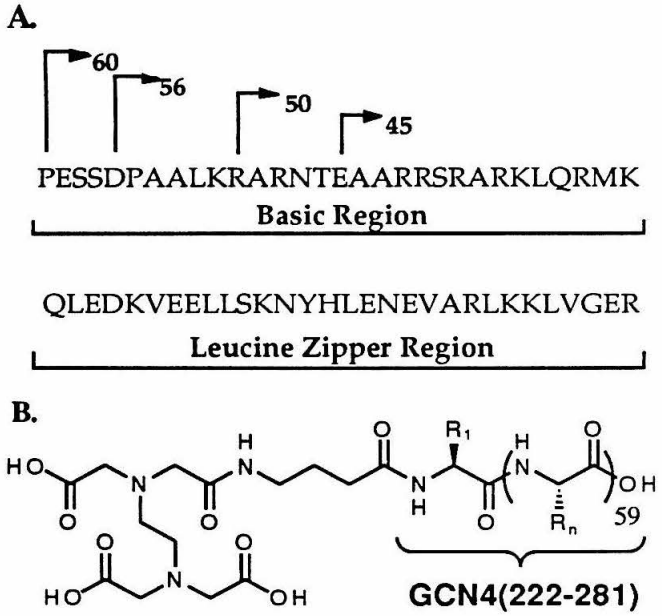


Figure 4. The sequence of 60 amino acids (222-281) corresponding to the CO_2H terminus of yeast transcriptional activator GCN4 listed from NH_2 to CO_2H terminus. **(B)** Synthetic protein EDTA-GCN4(222-281).

The GCN4 proteins were assayed for specific DNA binding and cleaving by analyzing the cleavage products from a restriction fragment containing two GCN4 recognition sites, termed ARE (5'-CTGACTAAT- 3') and GCRE (5'-ATGACTCTT-3').[66] DNase I and MPE footprinting demonstrate that GCN4(222-281), (226-281), and (232-281) (60, 56, 50mers, respectively) bind specifically to DNA at a concentration of 5 μM, whereas GCN4(237-281) (45mer) does not. Affinity cleaving experiments show that Fe•EDTA-GCN4 50, 56 and 60mers cleave DNA specifically at 5 μM concentrations. The GCN4 50mer binds DNA with reduced affinity compared to the longer proteins.

Figure 5. Autoradiogram of a high resolution denaturing polyacrylamide gel. Lanes 5-16 contain footprinting reactions with GCN4(222-281) and (226-281) and lanes 17-24 contain affinity cleaving reactions with Fe•EDTA-GCN4(222-281) and (226- 281). Odd numbered lanes are labelled with ^{32}P at the 5'-end and even numbered lanes are labelled with ^{32}P at the 3'- end. Lanes 1 and 2, intact DNA control; lanes 3 and 4, chemical sequencing A reaction; lanes 5 and 6, DNase I control; lanes 7 and 8, DNase I cleavage protection in the presence of GCN4(222-281) at 5 μM; lanes 9 and 10, DNase I cleavage protection in the presence of GCN4(226-281) at 5 μM; lanes 11 and 12, MPE•Fe control; lanes 13 and 14, MPE•Fe cleavage protection in the presence of GCN4(222-281) at 5 μM; lanes 15 and 16, MPE•Fe cleavage protection in the presence of GCN4(226-281) at 5 μM; lanes 17 and 18, Fe•EDTA-GCN4(222-281) at 5 μM; lanes 19 and 20, Fe•EDTA-GCN4(222-281) at 1 μM; lanes 21 and 22, Fe•EDTA-GCN4(226-281) at 5 μM; and lanes 23 and 24, Fe•EDTA-GCN4(226-281) at 1 μM.

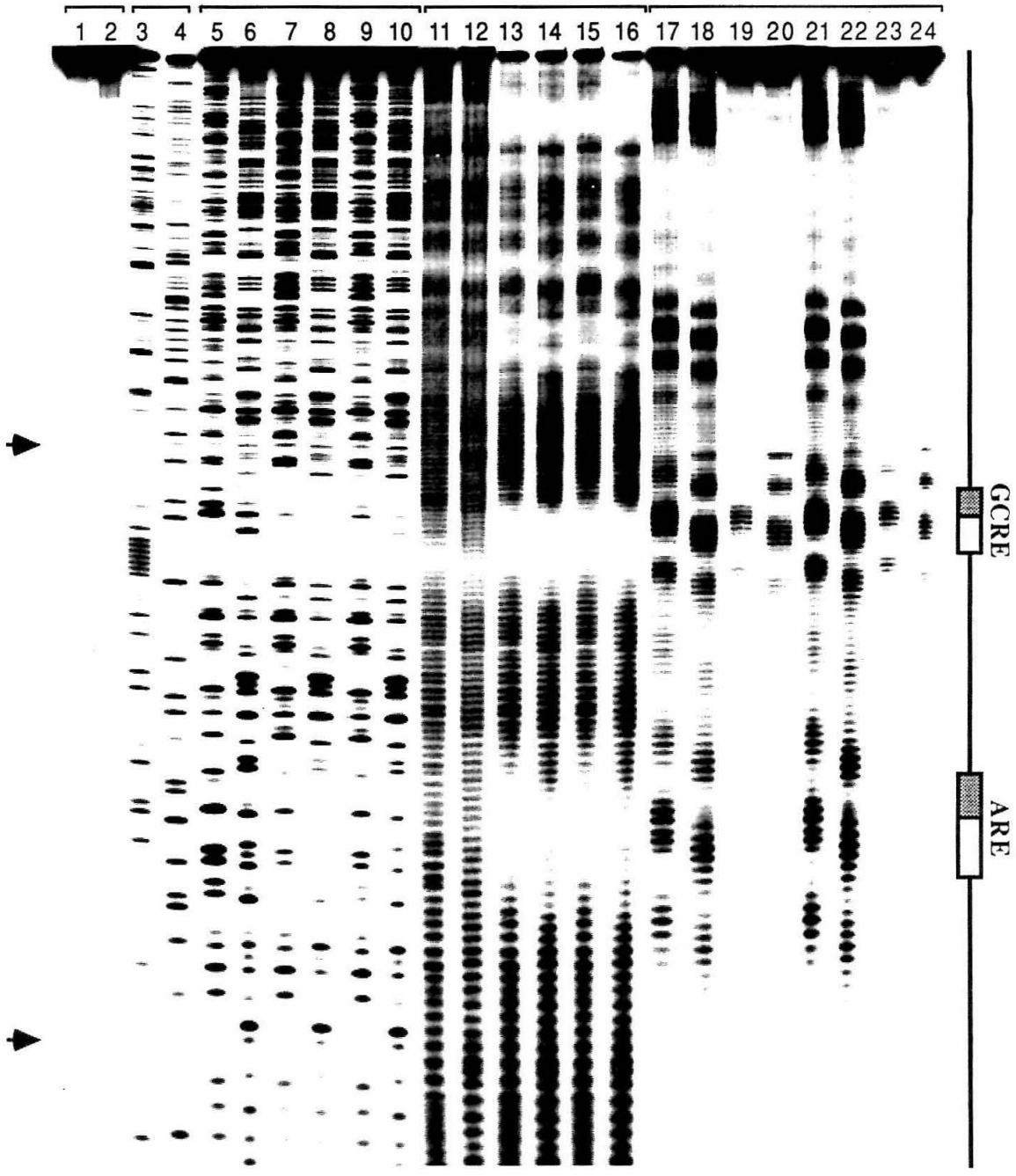
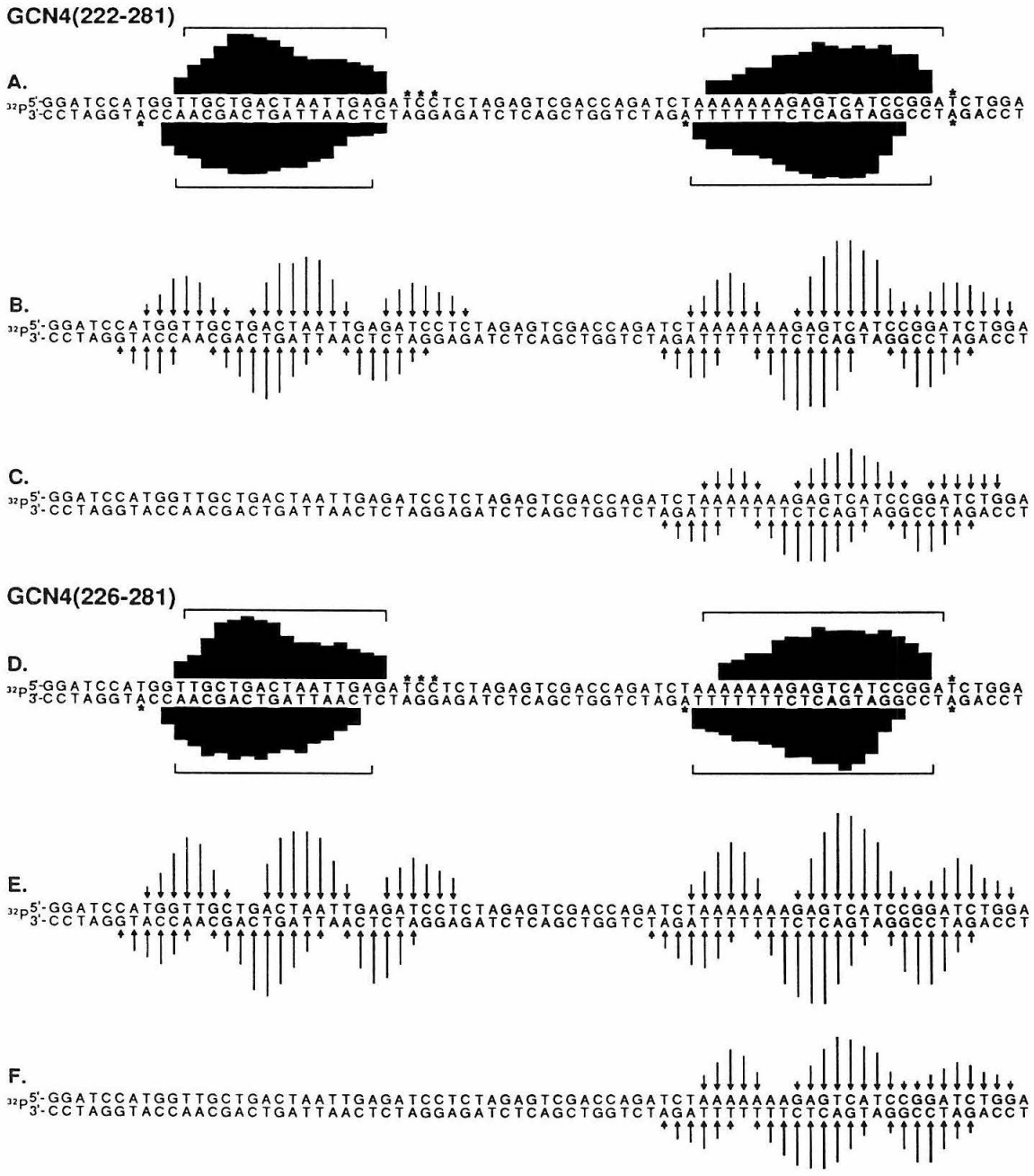

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
GCRE
ARE

Figure 6. The sequence from left to right represents the data between the two arrows from the bottom to the middle of the gel shown on the left in Fig. 5. **(A)** Bars represent the extent of protection from MPE•Fe cleavage in the presence of the GCN4(222- 281) (Fig. 5, lanes 13 and 14). Brackets represent bases protected from DNase I cleavage, and asterisks represent DNase I hypersensitivity sites (Fig. 5, lanes 7 and 8). **(B)** Arrows represent the extent of cleavage at the indicated base positions for Fe•EDTA-GCN4(222-281) at 5 μM (Fig. 5, lanes 17 and 18). **(C)** Cleavage pattern for Fe•EDTA-GCN4(222-281) at 1 μM (Fig. 5, lanes 19 and 20). **(D)** MPE•Fe and DNase I protection in the presence of GCN4(226-281) (Fig. 5, lanes 9, 10, 15, and 16). **(E)** Cleavage pattern for Fe•EDTA-GCN4(226-281) at 5 μM (Fig. 5, lanes 21 and 22). **(F)** Cleavage pattern for Fe•EDTA- GCN4(226-281) at 1 μM (Fig. 5, lanes 23 and 24).

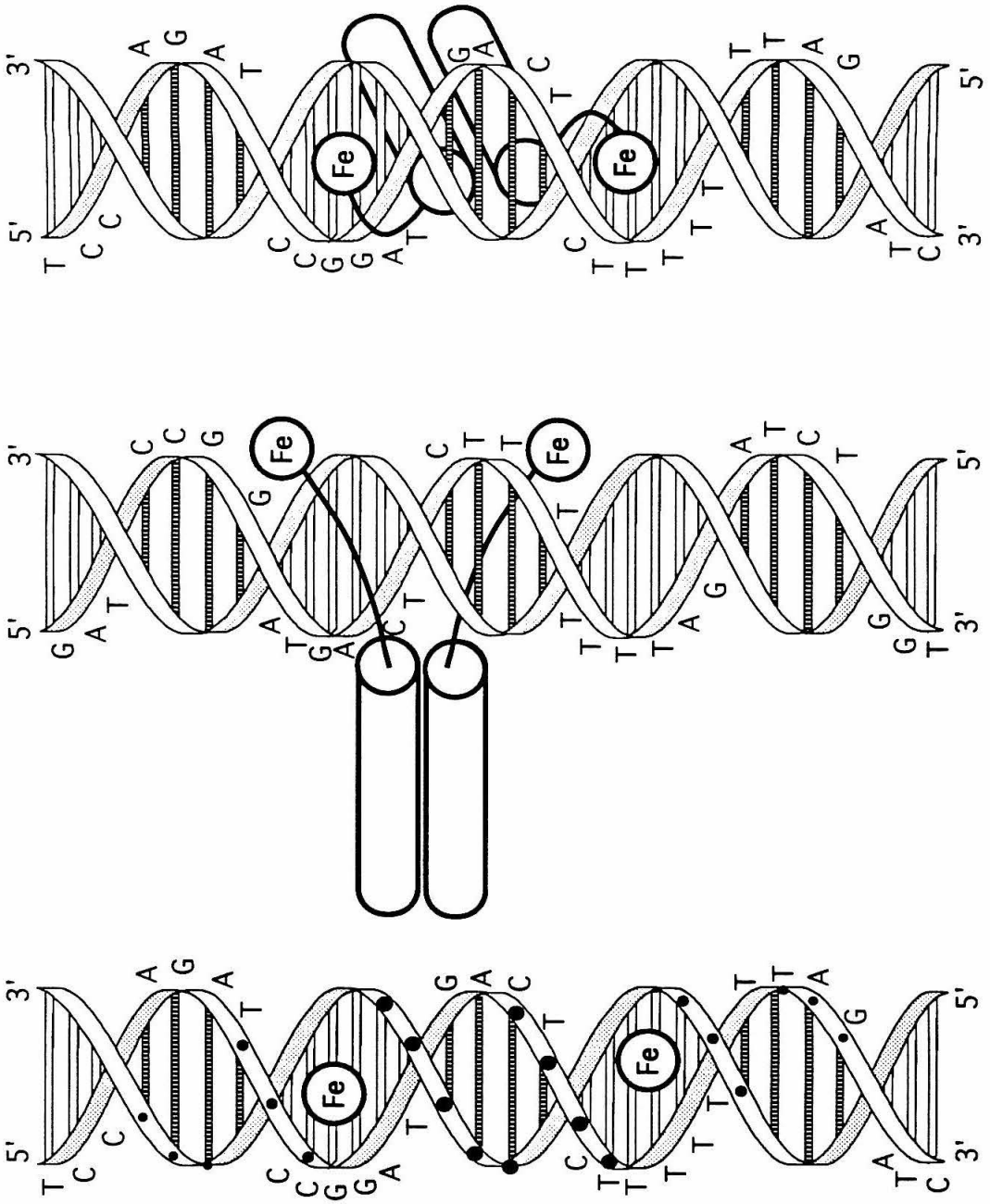

GCN4(222-281)
A.
B.
C.
GCN4(226-281)
D.
E.
F.
32P 5'-GGATCCATGGTTGCTGACTAATTGAGATCCTCTAGAGTCGACCAGATCTAAAAAAAAGAGTCATCCGGATCTGGA
3'-CCTAGGTACCAACGACTGATTAACTCTAGGAGATCTCAGCTGGTCTAGATTTTTTTTCTCAGTAGGCCTAGACCT

There are ten positively charged residues in the basic region extending from Lys-231 to Lys-251. GCN4 45mer, which contains seven basic residues, is not able to bind to DNA; GCN4 50mer, which contains nine basic residues, binds with reduced affinity, suggesting that the missing residues are important. GCN4 56mer and GCN4 60mer bind with approximately equal affinity.

The footprinting and affinity cleaving results for GCN4(222- 281) and GCN4(226-281) are shown in gel form in Figure 5; the data are represented by histograms in Figure 6. Both proteins protect regions of approximately 18 base pairs at the ARE and the GCRE binding sites from DNase I and MPE•Fe cleavage, demonstrating sequence specific binding (Fig. 5, lanes 5-16, Fig. 6A,D). Fe•EDTA-GCN4 56 and 60mers give cleavage patterns centered at those sites (Fig. 5, lanes 17-24; Fig. 6B,C,E,F). The GCN4 56 and 60mers have greater affinity for the GCRE site than for the ARE site. This is in contrast to results reported for the entire GCN4 protein which apparently binds to both sites with equal affinity.[66]

Because both halves of the Fe•EDTA-GCN4(222-281) and (226- 281) dimers contain Fe•EDTA at the NH_2-terminus, the affinity cleavage pattern should result from the sum of two localized Fe•EDTA moieties. The cleavage pattern observed consists of three cleavage loci of unequal intensity. With regard to the cleavage model in Figure 2, the simplest interpretation of this cleavage pattern is that it results from the superimposition of two adjacent major groove cleavage patterns along one face of the DNA (Fig. 7A). Cleavage occurs on both strands in three adjacent minor grooves, with the most efficient cleavage in the central minor groove proximal to both Fe•EDTA moieties. These data strongly suggest that the two Fe•EDTA moieties are located in adjacent major grooves. The positions of the Fe•EDTA moieties may be assigned by assuming that they lie in the center of

Figure 7. (A) Location of the two Fe•EDTA moieties at the NH_2- termini of a GCN4(222-281) dimer assigned from the cleavage patterns for Fe•EDTA-GCN4(222-281) on the GCN4 binding site, 5′- ATGACTCTT-3′. Filled circles represent the positions of cleavage along the phosphodiester backbone. Sizes of circles represent the extent of cleavage at the indicated base position. **(B)** Front view of a Y-shaped model (11) for the dimer of the DNA binding domain of GCN4 terminating with Fe•EDTA at the NH_2- termini. **(C)** Side view. The leucine zipper dimerization domains are represented by a pair of parallel cylinders pointing into the DNA.

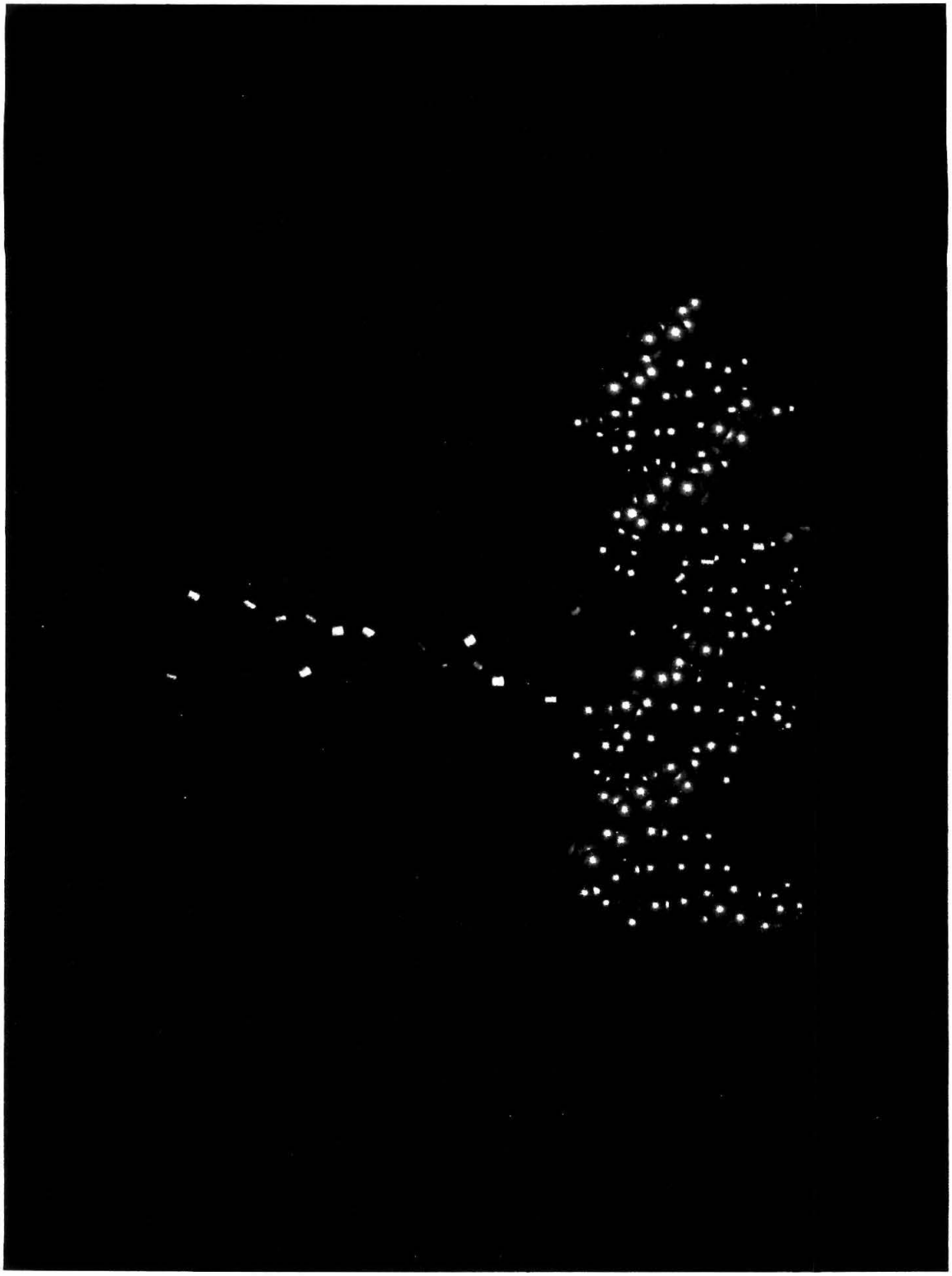

3'
5'
Fe
Fe
Fe
Fe
Fe
Fe

the two 5'-shifted patterns. The Fe•EDTA moieties, and hence the NH_2 termini of the dimer, are nine to ten base pairs apart and are located four to five base pairs on either side of the central C in the binding sites 5'-CTGACTAAT-3' and 5'-ATGACTCTT- 3' (Fig. 7). The cleavage patterns at the ARE and GCRE sites are similar and indicate that the protein dimer binds the two sites in much the same way.

Close scrutiny of the cleavage patterns reveals that the DNA cleavage by Fe•EDTA-GCN4(226-281) is more efficient than that by GCN4(222-281). This difference is especially noticeable at 1 μM concentrations, but it is found at all sites at all concentrations (Fig. 5, lanes, 19, 20, 23, 24). As the footprinting data indicate that the proteins bind with roughly equal affinity, this difference in cleavage intensity may indicate that the Fe•EDTA is closer to the DNA when it is attached to Asp-226. One curious aspect of our data is that the Fe•EDTA moiety appears to be located in approximately the same base pair position whether it is attached to Pro-222 or to Asp-226. A secondary structural element could bring these two residues in near enough proximity for the Fe•EDTA moiety to give similar cleavage patterns.[67]

Affinity cleaving indicates that the NH_2-termini of the GCN4(222-281) dimer are (1) in the major groove of DNA, (2) separated by nine to ten base pairs, and (3) symmetrically displaced four to five base pairs from the central C of the recognition site (Fig. 7B,C). This experimental result places the NH_2-termini in successive major grooves on one face of the DNA, and is consistent with the proposed Y-shaped models for the structure of bZIP proteins.[53,59]

Figure 8. Cocrystal structure of the GCN4-DNA complex.[61] The protein monomers are shown as ribbons and the DNA is shown as a CPK model. Coordinates of the complex were provided by T. E. Ellenberger and S. C. Harrison.



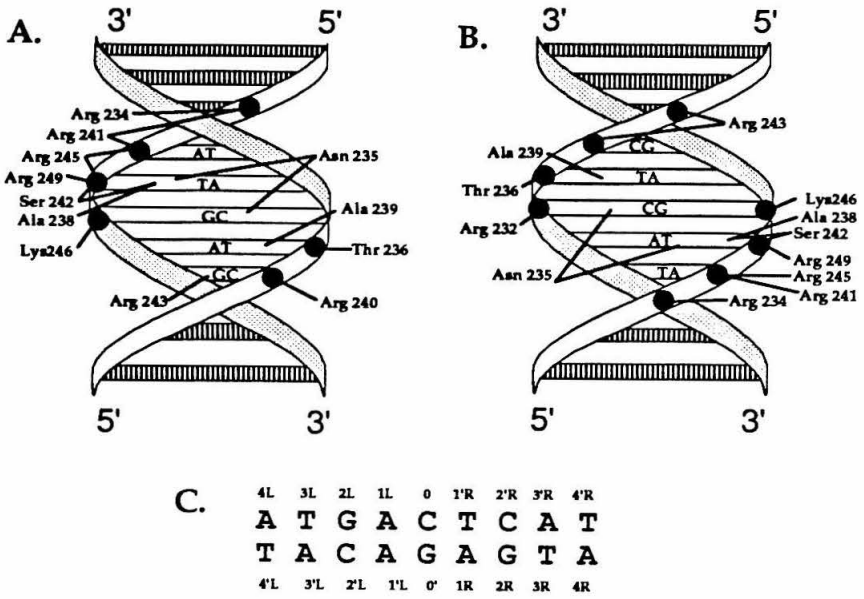


Figure 9. GCN4-DNA contacts. Phosphates contacted are represented as filled circles. **A.** Left half site monomer. **B.** Right half site monomer. **C.** DNA binding site numbering scheme. Adapted from Reference 61.

THREE-DIMENSIONAL STRUCTURE OF THE GCN4-DNA COMPLEX

The crystal structure of the GCN4 56mer, GCN4(226-281), in a complex with its DNA binding site has been determined recently by Harrison and co-workers (Fig. 8).[61] The protein does indeed have a Y-shaped structure with each monomer forming a smoothly curving α-helix.[61] The structure of the coiled-coil leucine zipper region is much the same as in the isolated leucine zipper,[52] and this region approaches the DNA helix at an angle of nearly 90 degrees.[61] The basic regions track through the major groove of each half site with no sharp bend or kink in this region, in general agreement with the predicitons of the induced helical fork model.[53]

Numerous amino acid residues, ranging from Arg-232 to Arg-249 (Fig. 4), make contacts with the DNA binding site. The contacts are similar, but not identical in the two half sites (Fig. 9).[61] Specific contacts include hydrogen bonds between the amido group of Asn-235 and N_4 of Cyt-2 and O_4 of Thy-3;

van der Waals contacts between the methyl side chains of Ala-238 and Ala-239 with the 5-methyl groups of Thy-3 and Thy-1, respectively; and a pair of hydrogen bonds between Arg-243 of the left half site monomer and O_6 and N_7 of Gua-0'.[61]

Of the GCN4 peptides synthesized in our affinity cleaving studies, only peptides of 50 amino acid residues or longer bound to DNA. The 45mer, GCN4(237-281), lacks Asn-235, which makes specific DNA contacts, as well as Arg-232 and Arg-234, which make non-specific contacts with the DNA backbone, explaining its inability to bind DNA. The 50mer, GCN4(232-281), on the other hand, contains all the residues involved in contacts to the DNA, but binds with reduced affinity compared to the 56mer and 60mer. The reduced binding affinity of the 50mer could be explained by a destabilization of the complex by unfavorable interactions between the negatively charged Fe•EDTA complex and the phosphodiester backbone.

The crystal structure shows that Asp-226 is three turns of the α-helix away from the residues that contact bases at the DNA binding site.[61] Although it is somewhat surprising that an interpretable cleavage pattern results from an Fe•EDTA moiety held so far away from the DNA, our experiments imply that it is not necessary to position an Fe•EDTA precisely in order to obtain structural information about a protein-DNA complex. On the other hand, the distance between Asp-226 and the DNA explains the lack of cleavage by Ni(II)•Gly-Gly-His-GCN4(226-281) (data not shown), as Gly-Gly-His cleaves DNA through a nondiffusible oxidant.[68,69] The fact that the protein extends away from the DNA helix also explains the cleavage by all three proteins at the same nucleotide positions, in contrast to the shifting cleavage patterns observed with Hin affinity cleaving proteins of varying lengths.[16] Interestingly, the 50mer, which should position the Fe•EDTA closest to the

Figure 10. Cocrystal structure of the GCN4-DNA complex showing the relative positions of Asn-226 and DNA backbone positions cleaved by Fe•EDTA-GCN4(226-281). The protein monomers are represented as ribbons, with the positon of Asn-226 highlighted in yellow. The nucleotides highlighted in yellow correspond to the sites of maximal cleavage by Fe•EDTA-GCN4(226-281) along the DNA backbone. The complex has been rotated 90° around the DNA double helical axis from the position seen in Figure 8.

DNA gives the weakest cleavage pattern. Quantitative affinity cleaving titration experiments[70-72] could be used to determine whether the poor cleavage is largely the result of poor binding or if the DNA is protected from cleavage in the major grooves by the protein.

The DNA base positions of maximal cleavage by Fe•EDTA-GCN4(226-281) are highlighted in the GCN4-DNA structure shown in Figure 10 as are the two Asn-226 residues. The Asn-226 residues are in the major groove and symmetrically displaced from the center of the DNA binding site, as predicted by our analysis. They are most closely situated to phosphates 6 and 2' of either half site,[61] in general agreement with our cleavage patterns, which are centered in each half site, 4-5 nucleotides away from the central C-G base pair. Interestingly, Ellenberger *et al.* point out that our data are more compatible with the induced helical fork model than with scissors grip model, which would predict that Asn-226 would be located farther from the center of the binding site.[61]

AFFINITY CLEAVING INVESTIGATIONS OF JUN AND FOS DIMERS

Introduction.

The basic region-leucine zipper proteins v-Jun and v-Fos were originally identified as the cellular counterparts of the transforming factors from the avian leukemia and FBJ murine sarcoma viruses, respectively.[37,41,42] Both Jun and Fos are nuclear phosphoproteins and are members of a group of factors known as AP1 which can act to stimulate transcription from genes that are responsive to tumor promoters.[73] The enhancers of these tumor promoter responsive genes contain the AP1 recognition site, 5'-TGAC/GTCA-3', a site which is identical to the consensus recognition site for GCN4.[64,73]

Because AP1 is a heterogeneous mixture of proteins, the presence of Jun or Fos in this mixture does not prove it activates transcription directly. However, Jun has been shown to activate transcription from phorbol ester responsive promoters *in vivo*[32,74] and its transcriptional activity is dramatically increased in the presence of Fos.[32]

As discussed above, it has been shown by many investigators that Jun and Fos can form heterodimers [23-28] and that the leucine zipper region is critical for dimerization and for DNA binding.[23-33] Jun can also bind to DNA specifically as a homodimer.[26,28] Although Fos appears to be incapable of binding to its recognition site in the absence of Jun,[25,27,75] it has been reported to stimulate the binding activity of Jun by factors ranging from 30- to 1000-fold.[23,25,27,31,32,38,74] As Fos contains a basic region very similar to that of Jun and GCN4, it is believed that its lack of DNA binding activity reflects its inability to form homodimers.[26]

Indirect evidence that the basic regions of Fos can interact with DNA comes from studies of chimeric leucine zipper proteins. A number of researchers have reported that chimeric proteins containing the basic region of Fos and the leucine zipper region of Jun are capable of forming homodimers or heterodimers with Jun. Both dimers contain two Fos binding regions and interact specifically with AP1 sites.[76-78] Similarly, a chimera (FosG) containing the leucine zipper of GCN4 and the DNA binding domain of Fos forms homodimers or heterodimers with GCN4 and binds to a GCN4/AP1 site.[79] Interestingly, in all cases, dimers with two Fos basic regions interact less strongly than those with a Jun and a Fos basic region, implying that the complex interacts asymmetrically with its palindromic binding site. As the dimerization domains are the same in both cases, it has been suggested that Fos has stronger affinity for one half site than the other,

leading to decreased binding efficiency when it is forced to interact with both half sites.[76] FosG homodimers also bind with lower affinity than GCN4 homodimers or heterodimers of GCN4 and FosG.[79] Again, the dimerization domains are identical in all three cases, indicating that the interaction of two basic regions from Fos leads to decreased binding affinity.

Crosslinking studies have provided direct evidence that Fos can associate with DNA.[80,81] Furthermore, crosslinking results with bromouridines substituted for specific thymidines in the binding site indicate that Fos binds preferentially to one half site. Asymmetry of binding by leucine zipper peptides has been suggested previously due to the asymmetry of the effects of mutations in the palindromic binding site.[64,82] The recent GCN4 crystal structure shows that contacts between the half sites are indeed asymmetric, with only one monomer making base-specific contacts to the central C-G base pair (Fig. 8).[61] As Jun, Fos, and GCN4 are highly homologous in their basic regions (Fig. 11), it has been proposed that the contacts made by the Jun-Fos heterodimer to the AP-1/GCN4 bindings site will be essentially the same as those made by GCN4.[61] A preferential interaction by the Fos or Jun basic region with the guanine of the central base pair would explain preferred binding by Fos or Jun to one half site.[61]

```
                        RR **  ** *** **  *
GCN4    PESSD  PAALKRARNTEAARRSRARKLQRMKQL
Jun     MESQERIKAERKRMRNRIAASKSRKRKLERIARL
Fos     PEEEE  KRRIRRERNKMAAAKCRNRRRELTDTL
```

Figure 11. Conservation of amino acid residues making DNA contacts in the GCN4-DNA complex.[61] Asterisks indicate residues making contact in both monomers; "R" indicates contacts made only by the right half site monomer.

Using affinity cleaving on a heterodimeric protein, it is possible to test this asymmetric binding model directly (Fig. 12). If Fos exhibits a preference

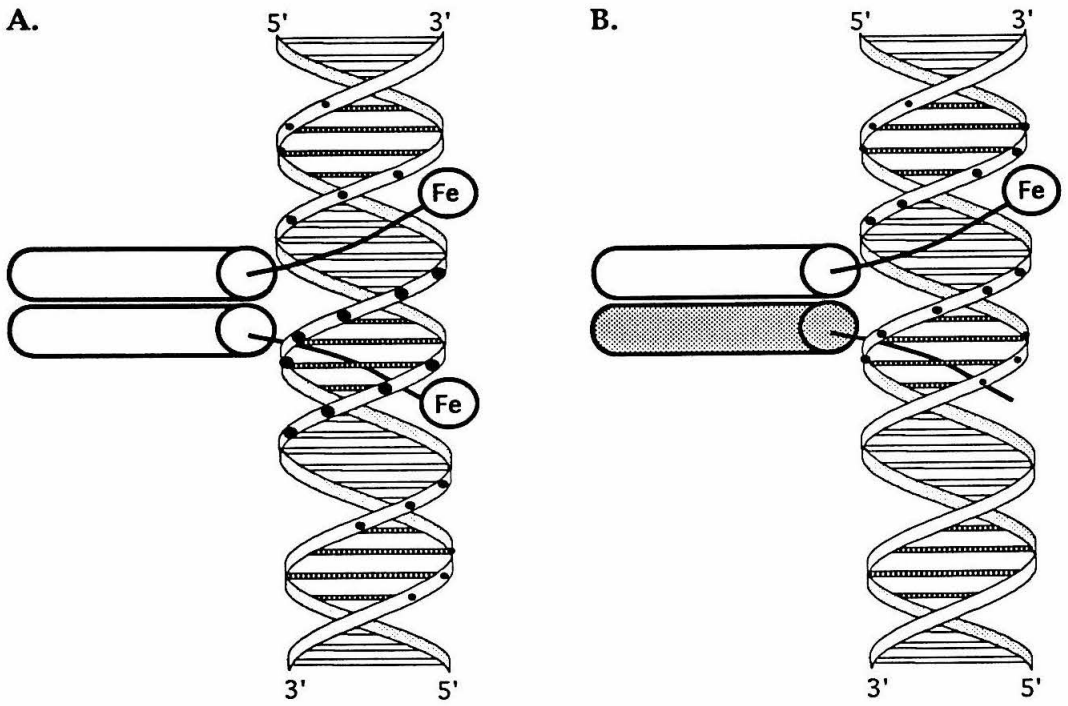


Figure 12. Difference between cleavage patterns expected for a homodimer with Fe•EDTA on each monomer **(A)** and for a heterodimer containing only one monomer with Fe•EDTA **(B)**. If the heterodimer binds only in the orientation shown, it should result in the affinity cleavage pattern seen on the right. If it binds equally well in either orientation, a cleavage pattern resulting from two major groove Fe•EDTA moieties (left) will be obtained.

for one half of the binding site over the other, then affinity cleaving by a dimer containing only one Fe•EDTA moiety should result in a standard major groove cleavage pattern at one half site (Fig. 12B). However, if there is no polarity in binding, then the Fe•EDTA moiety should be equally distributed in both half sites. In this case, a homodimer-like cleavage pattern resulting from the superimposition of two major groove Fe•EDTA moieties will be observed (Fig. 12A). These affinity cleaving studies will also investigate the relative DNA affinity and specificity of Jun-Jun homodimers, Fos-Fos homodimers, and Jun-Fos homodimers. In addition, if any of these

protein dimers interacts with the DNA in a different fashion than GCN4, these differences should be reflected in the cleavage patterns.

Preliminary Investigations.

In order to address these questions, three Fos peptides and three Jun peptides were synthesized (Fig 13). These peptides correspond to the 50, 56, and 60mers of GCN4 (Fig 4) with an extra seven residues at the CO_2H-terminus. Although these seven residues do not constitute a leucine repeat as such, they do contain hydophobic residues in the proper positions for formation of a coiled coil, and deletion of this region of Fos has been shown to reduce DNA binding.[23] Kim and co-workers have also included these additional residues in their studies of the Jun and Fos leucine zippers.[50] Fos(134-200), Fos(139-200) and Fos(144-200) were prepared from a single manual synthesis; the CO_2H-terminal histidine was replaced by tyrosine in order to facilitate determination of peptide concentrations. Jun(212-280), Jun(218-280), and Jun(224-280) were synthesized from a single automated synthesis with poorer than average yields as measured by quantitative

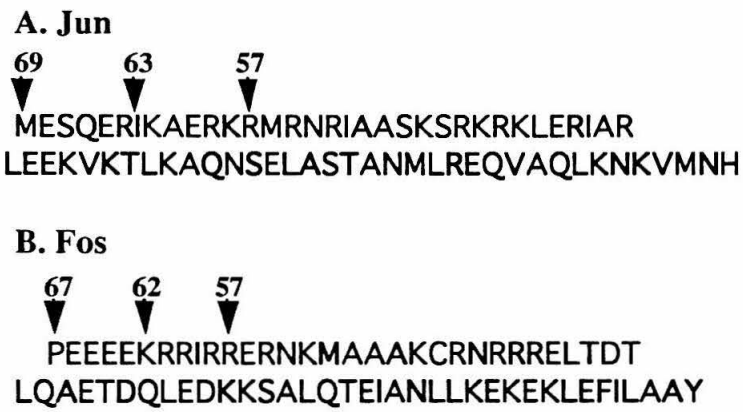


Figure 13. The sequences from NH_2 to CO_2H terminus of the Jun and Fos peptides synthesized. Arrows indicate the NH_2-termini of each peptide made; numbers indicate the length. **A.** Jun (212-280). **B.** Fos(134-200).

ninhydrin analysis.[83] All six peptides were prepared both with and without EDTA at the NH_2-termini. Because purification of the Jun peptides by preparative reverse phase HPLC did not result in pure fractions as assayed by analytical reverse phase HPLC, partially purified material was used in preliminary investigations.

Unexpectedly, only the two longest EDTA-containing proteins, EDTA-Jun(212-280) and EDTA-Fos(134-200), cleave DNA specifically as homodimers (data not shown). Because the homology with GCN4 extends through the regions corresponding to the smaller of these peptides (Fig. 3), it was expected that these peptides would be more likely to cleave DNA than would the longer peptides. Both Jun and Fos contain basic amino acid residues NH_2-terminal to the region of strongest homology with GCN4, and these residues could be required for binding. Attempts to measure binding of any of the 12 proteins by DNAse or MPE footprinting were unsuccessful.

DNA affinity cleaving results obtained with EDTA-Jun(212-280) and EDTA-Fos(134-200) are shown in gel form in Figure 14, and selected lanes are displayed in histogram form in Figure 15. Cleavage by EDTA-GCN4 56mer is shown for comparison. Both EDTA-Jun and EDTA-Fos homodimers are able to cleave DNA at the GCRE but not at the ARE. A much higher concentration of EDTA-Jun is required than of EDTA-Fos (5 μM). For the lanes in which heterodimers were allowed to form, Jun and Fos were incubated for one hour at 37° in the absence of DNA. No attempt was made to purify heterodimers. Instead, only one of the peptides in each heterodimer affinity cleaving reaction was equipped with EDTA and heterodimer formation was driven by an excess of the peptide without EDTA. DNA cleavage should therefore result primarily from the heterodimer. The use of high concentrations of the protein not equipped with EDTA led to decreased

Figure 14. Autoradiogram of a high resolution denaturing polyacrylamide gel. Lanes 1-14 are labeled with ^{32}P at the 5'-end and lanes 15-28 are labelled with ^{32}P at the 3'- end. Lanes 1 and 15, intact DNA control; lanes 2 and 16, chemical sequencing A reaction; lanes 3 and 17, EDTA-Jun(212-280) at 10 μM; lanes 4 and 18, EDTA-Jun(212-280) at 10 μM and Fos(134-200) at 10 μM ; lanes 5 and 19, EDTA-Jun(212-280) at 10 μM and Fos(134-200) at 20 μM; lanes 6 and 20, EDTA-Jun(212-280) at 20 μM; lanes 7 and 21, EDTA-Jun(212-280) at 20 μM and Fos(134-200) at 40 μM; lanes 8 and 22, Fe•EDTA-GCN4(226-281) at 5 μM; lanes 9 and 23, EDTA-Fos(134-200) at 2.5 μM; lanes 10 and 24, EDTA-Fos(134-200) at 2.5 μM and Jun(212-280) at 5μM ; and lanes 11 and 25, EDTA-Fos(134-200) at 2.5 μM and Jun(212-280) at 10μM; lanes 12 and 26, EDTA-Fos(134-200) at 2.5 μM and Jun(212-280) at 20 μM; lanes 13 and 27, EDTA-Fos(134-200) at 5 μM; lanes 14 and 28, EDTA-Fos(134-200) at 5 μM and Jun(212-280) at 20 μM.

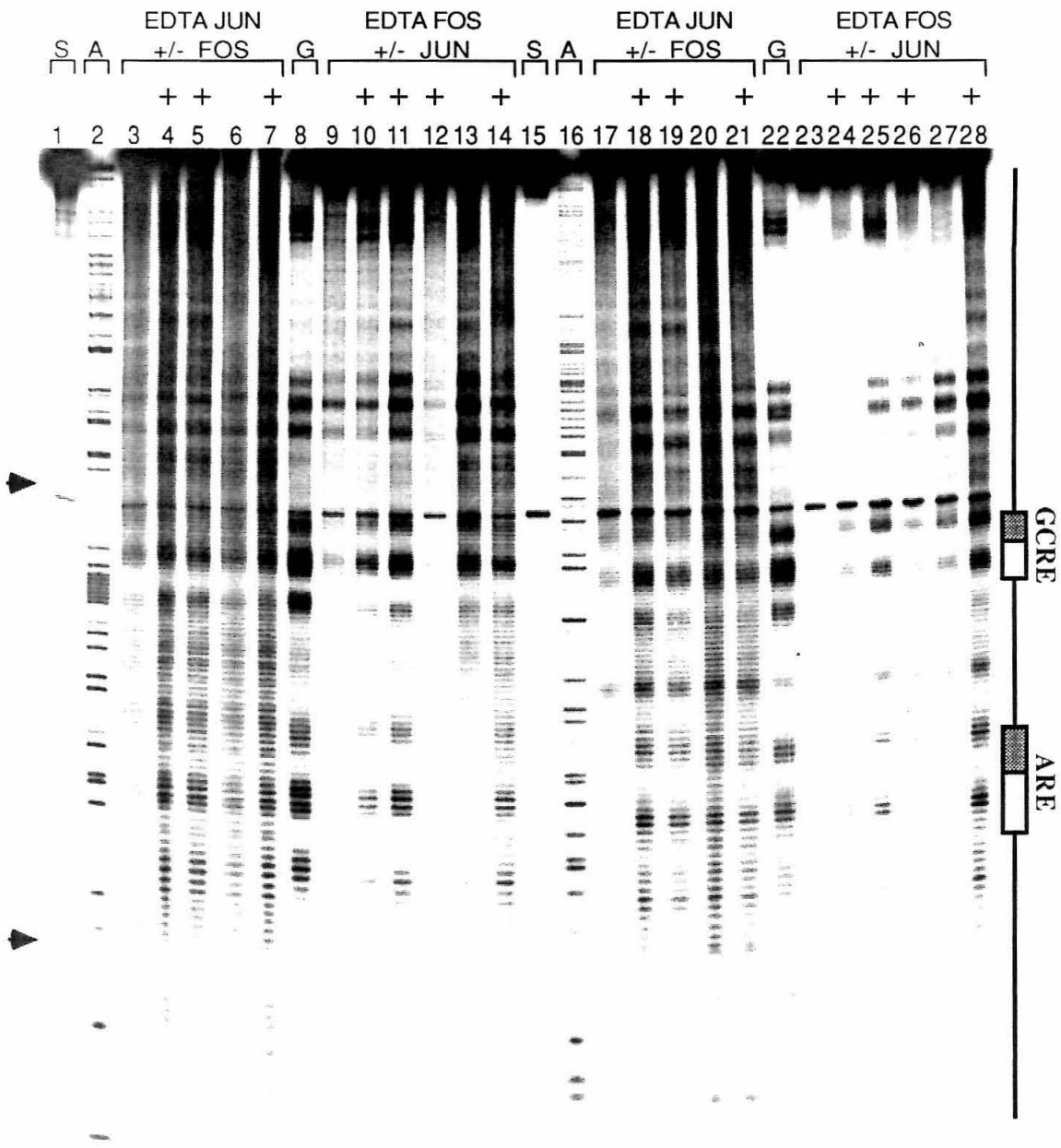

EDTA JUN
+/- FOS
EDTA FOS
+/- JUN
EDTA JUN
+/- FOS
EDTA FOS
+/- JUN
S
A
G
S
A
G
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
GCRE
ARE

Figure 15. The sequence from left to right represents the data between the two arrows from the bottom to the middle of the gel shown on the left in Fig. 8. Arrows represent the extent of cleavage at the indicated base position for **(A)** EDTA-Jun(212-280) at 10 μM and Fos(134-200) at 10 μM (Fig. 13, lanes 4 and 18) **(B)** EDTA-Jun(212-280) at 20 μM (Fig. 13, lanes 6 and 20) **(C)** Fe•EDTA-GCN4(226-281) at 5 μM (Fig. 13, lanes 8 and 22). **(D)** EDTA-Fos(134-200) at 2.5 μM and Jun(212-280) at 10 μM (Fig. 13, lanes 11 and 25). **(E)** EDTA-Fos(134-200) at 5 μM (Fig. 13, lanes 13 and 27).

Fe•EDTA-Jun(212-280) + Fos(134-200)

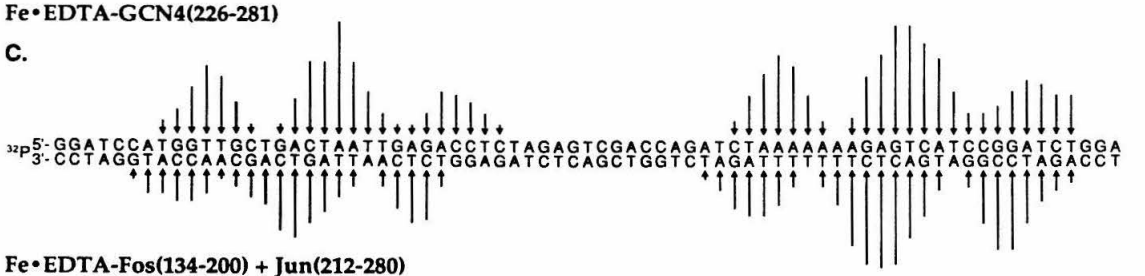
A.

Fe•EDTA-Jun(212-280)

B.

Fe•EDTA-GCN4(226-281)

C.



Fe•EDTA-Fos(134-200) + Jun(212-280)

D.

Fe•EDTA-Fos(134-200)

E.

cleavage (Fig. 14, lanes 12 and 26), presumably due to significant binding by the unmodified homodimer.

The high ratio of background cleavage to specific cleavage seen in many lanes in the gel is reproducible and precludes a quantitative assessment of these data. In addition, there is a concern about the purity of the Jun peptides. However, several qualitative conclusions may be drawn. First, leucine zipper and basic region domains of Jun and Fos are sufficient for specific DNA binding. Second, the DNA binding domains of both Jun and Fos are capable of binding to DNA in the absence of the other protein. While it is formally possible that Jun and Fos bind to DNA as monomers, this is unlikely. In either case, the result is at odds with previous studies concerning the inability of Fos to form dimers[26] and bind to specific sequences of DNA.[26-28,74] Kim and colleagues, however, have determined that peptides corresponding to the leucine zipper region of Fos are capable of forming both homo- and heterodimers.[50] The Fos-Fos homodimer is much weaker than the Jun-Jun and Jun-Fos dimers, but Fos can exist as a homodimer at room temperature with a dissociation constant of about 6 μM.[50] Such concentrations are much higher than those used in *in vitro* translation experiments, but approximately equivalent to the conditions used in these affinity cleaving studies.[50] Considering this data and the fact that Fos 67mer is able to interact with DNA at lower concentrations than the Jun 69mer, which is known to form homodimers, it is most reasonable to assume that Fos binds as a dimer.

A third conclusion to be drawn from the data shown in Figures 14 and 15 is that Jun-Fos heterodimers appear to form, as is evidenced by a clear change in DNA binding affinity. It is not possible to determine from this data whether the appearance of binding activity for the ARE is due to a change in sequence specificity or simply to increased affinity for DNA: GCN4 binds to

the GCRE and to the ARE at concentrations of 5 μM, while binding only to the GCRE at a concentration of 1 μM (Fig. 5). An interesting feature of the EDTA-Jun(212-280) + Fos(134-200) and EDTA-Fos(134-200) + Jun(212-280) cleavage patterns (Figure 15A and 15D) is that the interaction of the heterodimers with the two sites is markedly different. This is in contrast to GCN4, which interacts with both sites in a similar fashion. Although no polarity in binding is exhibited by Jun-Fos heterodimers at the ARE, it appears that EDTA-Fos(134-200) + Jun(212-280) cleaves with lower intensity at the 5'-CTCTT-3' half of the binding site than at the 5'-ATGAC-3' half site, as does the EDTA-Fos(134-200) homodimer. Conversely, the EDTA-Jun(212-280) + Fos(134-200) heterodimer leads to greater cleavage intensity at the 5'-CTCTT-3' half of the binding site than at the 5'-ATGAC-3' half site. The EDTA-Jun(212-280) affinity cleaving data is difficult to interpret due to the low cleavage intensity, but the intensity of cleavage appears to be approximately equal at each half site. Thus, it is possible that the apparent polarity of Jun-Fos heterodimer binding at the GCRE results from the preference of Fos to bind to the 5'-ATGAC-3' half site (or to avoid the 5'-CTCTT-3' half site).

A. Jun

69 67 65 63 60 58
▼ ▼ ▼ ▼ ▼ ▼
MESQERIKAERKRMRNRIAASKSRKRKLERIAR
LEEKVKTLKAQNSELASTANMLREQVAQLKNKVMNH

B. GBR-FZIP

PESSDPAALKRARNTEAARRSRARKLQRMKQ
LQAETDQLEDKKSALQTEIANLLKEKEKLEFILAAY

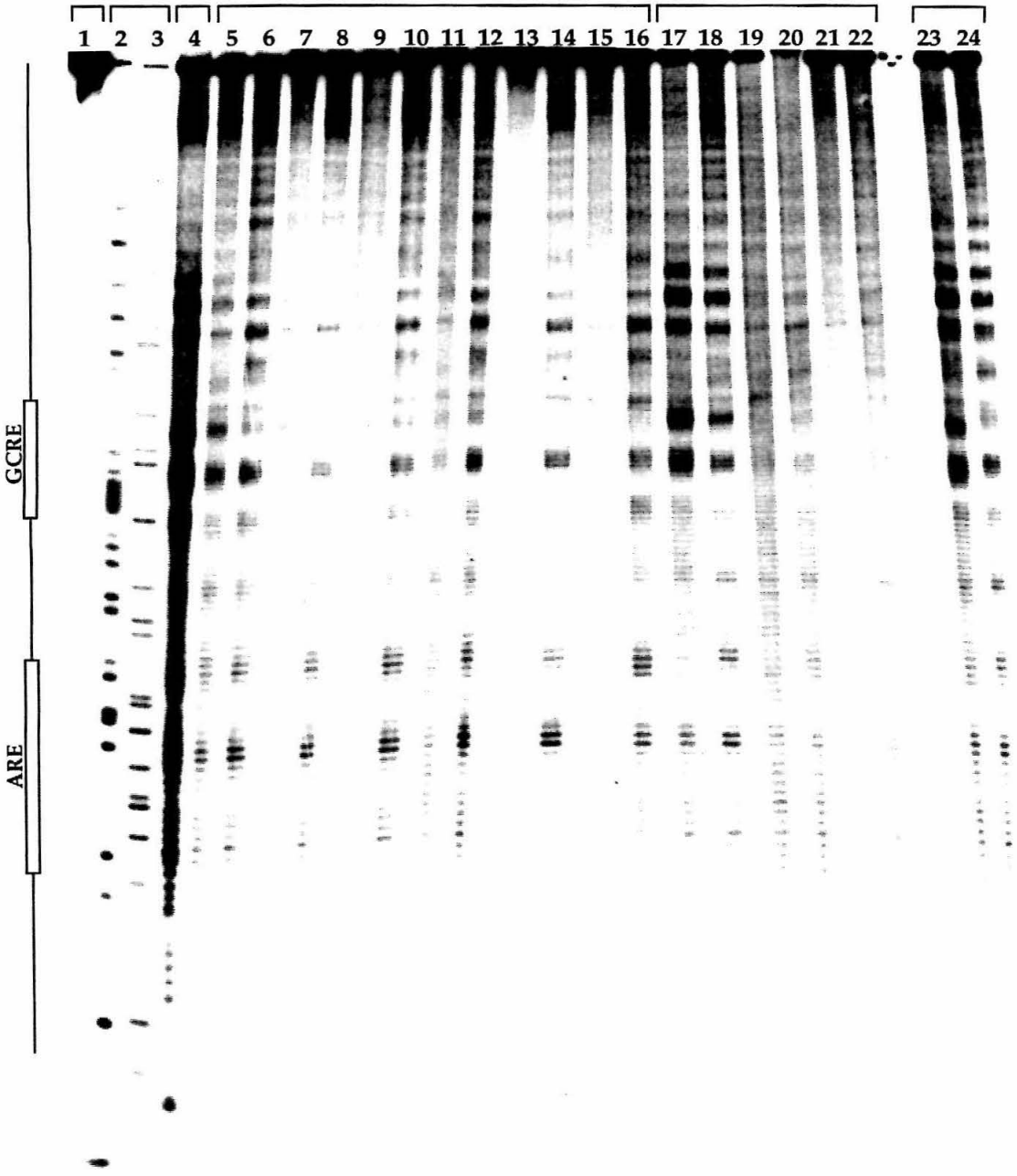
Figure 16. The sequences from NH_2 to CO_2H terminus of **(A)** Jun (212-280) and **(B)** GCN4(222-252)-Fos(165-200). Arrows indicate the NH_2-termini of each Jun peptide made; numbers indicate the length.

Further Studies.

Jun was resynthesized in hopes of obtaining more conclusive data. In order to explore the length dependence of affinity cleaving, a series of six Jun proteins was prepared, ranging in length from 58 to 69 residues (Fig. 16A). In addition, a chimeric protein, GCN4(222-252)-Fos(165-200) (GBR-FZIP), consisting of the leucine zipper of Fos and the basic region of GCN4 was synthesized (Fig. 16B). Each protein was sythesized with and without EDTA. Both syntheses gave acceptable yields and upon purification afforded a single peak by analytical HPLC. However, this sample of Fe•EDTA-Jun(212-280) (69mer) provided data at least qualitatively indistinguishable from that previously obtained with apparently less pure material. The Fos proteins were identical to those used in the last section. Again, the signal-to-noise is poor, and pure heterodimers were not isolated, allowing only for qualitative conclusions from the data.

The length dependence of Jun-Fos homo- and heterodimer formation and DNA cleavage was explored with the truncated Jun peptides. The Fe•EDTA-Jun(212-280) (69mer) cleaved as a homodimer at the GCRE at 10 μM, and cleavage fell off with length; no specific cleavage was seen for the 60- and 58mers (Fig. 17, lanes 5-16). Only the 69mer showed cleavage, though very weak, at the ARE. Although previous studies indicated that truncated Jun peptides could not form heterodimers with truncated Fos peptides, all six Fe•EDTA-Jun proteins cleaved DNA both at the ARE and at the GCRE in the presence of 20 μM Fos(134-200) (67mer; Fig 17, lanes 5-16). The cleavage patterns are similar in each case and show the same asymmetry at the GCRE noted above. The results for the Fos proteins were different (Fig. 17, lanes 17-22). As with Jun, at a concentration of 10 μM, the longest Fe•EDTA-Fos pro-

Figure 17. Autoradiogram of a high resolution polyacrylamide gel. All lanes contain 3′-^{32}P-end-labeled DNA unless otherwise noted. Lane 1, DNA standard. Lanes 2 and 3, 5′- and 3′- A-specific sequencing reactions. Lane 4, Fe•EDTA-GCN4(226-281), (5 μM). Lane 5, Fe•EDTA-Jun 69mer (10 μM). Lane 6, Fe•EDTA-Jun 69mer (10 μM) and Fos(134-200) (20 μM). Lane 7, Fe•EDTA-Jun 67mer (10 μM). Lane 8, Fe•EDTA-Jun 67mer (10 μM) and Fos(134-200) (20 μM). Lane 9, Fe•EDTA-Jun 65mer (10 μM). Lane 10, Fe•EDTA-Jun 65mer (10 μM) and Fos(134-200) (20μM). Lane 11, Fe•EDTA-Jun 63mer (10 μM). Lane 12, Fe•EDTA-Jun 63mer (10 μM) and Fos(134-200) (20 μM). Lane 13, Fe•EDTA-Jun 60mer (10 μM). Lane 14, Fe•EDTA-Jun 60mer (10 μM) and Fos(134-200) (20 μM). Lane 15, Fe•EDTA-Jun 58mer (10 μM). Lane 16, Fe•EDTA-Jun 58mer (10 μM) and Fos(134-200) (20 μM). Lane 17, Fe•EDTA Fos 67mer (10 μM). Lane 18, Fe•EDTA-Fos 67mer (10 μM) with Jun(212-280) (20 μM). Lane 19, Fe•EDTA Fos 62mer (10 μM). Lane 20, Fe•EDTA-Fos 62mer (10 μM) with Jun(212-280) (20 μM). Lane 21, Fe•EDTA Fos 57mer (10 μM). Lane 22, Fe•EDTA-Fos 57mer (10 μM) with Jun(212-280) (20 μM). Lane 23, Fe•EDTA-GBR-FZIP (10 μM). Lane 24, Fe•EDTA-GBR-FZIP (10 μM) with Jun(212-280) (20 μM).


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
GCRE
ARE

tein, the 67mer, cleaved DNA specifically as a homodimer. Cleavage was most notable at the GCRE, with only faint cleavage at the ARE. In the presence of 20 μM Jun, a large increase in cleavage by the 67mer at the ARE was observed. In contrast, neither the 62mer nor the 57mer cleaved DNA specifically in the presence or absence of Jun. These results imply that the deletion of NH_2-terminal segments in Jun has a greater effect on the stability of the homodimer than the heterodimer, suggesting that those segments are necessary only in one half of the heterodimer. It is unclear why the deletions in Fos affect the stability of the homo- and heterodimers so dramatically. None of the deleted residues is analagous to those making contacts in the GCN4-DNA cocrystal structure,[61] (Fig. 11) though the segment is highly basic.

Cleavage by the chimeric GBR-FZIP 67mer is also shown in Figure 17 (lanes 23, 24). At concentrations greater than 5 μM, it cleaves DNA as a homodimer. However, at a concentration of 10 μM, Fe•EDTA-GBR-FZIP cleaves much less efficiently than does Fe•EDTA-GCN4 at a concentration of 5 μM (compare lanes 4 and 23). Because the basic regions are the same in the two homodimers, this indicates that the decreased specific DNA cleavage seen with the Fos proteins is in large part due to weak dimerization. Fe•EDTA-GBR-FZIP also cleaves specifically in the presence of the Fos 67mer (data not shown) and more strongly in the presence of the Jun 69mer (lane 24), but not in the presence of GCN4, as expected. Interestingly, no polarity in the cleavage pattern is seen for the homodimer or for the Jun heterodimer, but the cleavage pattern for Fe•EDTA-GBR-FZIP + Fos appears to resemble that for Fe•EDTA-Jun + Fos, suggesting that Fos has a binding preference for the 5'-ATGAC-3' half site at the GCRE.

In crosslinking studies with Jun-Fos heterodimers at a 5'-CTGACTCAT-3' site (TRE), Fos formed crosslinks preferentially with a bromouridine at the

1R' position, while Jun formed crosslinks when a bromouridine was substituted at the 1L' position (Fig. 9C).[80] These results are in contrast to the polarity observed at the GCRE, in which Fos appeared to bind preferentially to the left half site, and to the lack of polarity observed at the ARE. Both the GCRE and the ARE differ from the optimal GCN4 and Jun-Fos binding site, the palindromic site 5'-ATGACTCAT-3'. The binding polarity of Jun-Fos heterodimers may therefore vary from sequence to sequence. While the TRE should allow all contacts seen in the GCN4 structure to be made, two bases at the GCRE and one base at the ARE that are involved in sequence-specific contacts are mutated.[61] It is somewhat surprising that the GCRE is a stronger GCN4 and Jun-Fos binding site than the less severely altered ARE, suggesting that alternate contacts can be made. Thus, the degree of binding polarity at each site could be determined not only by the residues flanking the Arg (243 in GCN4) making specific contacts at the central base pair,[61] but also by the degree to which each monomer can rearrange structurally to accomodate particular mutations in the binding site.

EXPERIMENTAL

Materials.

Protected amino-acid derivatives were purchased from Peninsula Laboratories or Applied Biosystems. Boc-L-His(DNP) was obtained from Fluka. Benzhydrylamine resin was obtained from U. S. Biochemical Corp. Arg-(Phenylacetamido)methyl (PAM) resin, *N, N* - dimethylformamide (DMF), diisopropylethylamine (DIEA), dicyclohexylcarbodiimide in dichloromethane, *N*-hydroxybenzotriazole (HOBt) in DMF, and trifluoroacetic acid (TFA) were purchased from Applied Biosystems. Dichloromethane and methanol (HPLC grade) were obtained from

Mallinckrodt, HOBt, *p* -cresol and *p* -thiocresol from Aldrich, and diethyl ether (low peroxide content) from Baker.

Doubly distilled water was further purified through the Milli Q filtration system from Millipore. Sonicated, deproteinized calf thymus DNA was purchased from Pharmacia and dissolved in H_2O to a concentration of 1 mM in base pairs. tRNA (*E. coli* strain W, Type XX) was obtained from Sigma and was dissolved in water and sterile-filtered. Enzymes were obtained from Boehringer-Mannheim or New England Biolabs and used with the buffers supplied. Deoxyadenosine- 5'- [α- ^{32}P]-triphosphate and adenosine- 5'-[γ -^{32}P]-triphosphate were obtained from Amersham. The plasmid pARE/GCRE was a generous gift from Keith Harshman and Prof. Carl Parker.

UV-vis spectra were recorded on a Hewlett-Packard Diode Array spectrophotometer. Cerenkov radioactivity was measured with a Beckman LS 3801 Scintillation Counter. Laser densitometry of gel autoradiograms was performed on a LKB Ultrascan XL densitometer. Manual peptide syntheses were carried out in 20 mL vessels fitted with coarse glass frits as described by Kent.[15] Automated peptide syntheses were performed in a 20 mL Teflon/KelF reaction vessel on an ABI 430A peptide synthesizer modified by the removal of the in-line filters to the top and bottom of the reaction vessel. Coordinates for the GCN4-DNA cocrystal structure were generously provided by Dr. T. E. Ellenberger and Prof. S. C. Harrison.

Methods.

Peptide Syntheses. All peptide syntheses were carried out using stepwise solid phase methods with N-α–Boc L-amino acids.[15] All four unmodified GCN4 peptides were made from a single manual synthesis starting with an Arg-PAM resin. [14,16] The average yield over 59 amino acid couplings in the synthesis of GCN4(222-281) was 99.4% as determined by Ninhydrin analysis.[83]

All three unmodified Fos peptides were synthesized from a single peptide synthesis from a BHA resin.[14,16] The average yield over 67 amino acid couplings in the synthesis of Fos(134-200) was 99.7%. The first synthesis of three unmodified Jun peptides was performed in a single automated synthesis from BHA resin[14,16] on an ABI 430A peptide synthesizer using protocols developed at the California Institute of Technology.[15] The average coupling yield for the synthesis of Jun(212-280) was 99.2%, with individual yields as low as 93.4%. The second set of unmodified Jun proteins was carried out as a single manual synthesis from BHA resin with an average yield per coupling of 99.6%. The chimeric protein GBR-FZIP 67mer was prepared from resin-bound Fos(165-200) (taken from the same synthesis described above) in an automated synthesis with an average coupling yield of 99.7%.

γ-Aminobutyric acid (GABA) and EDTA, protected as the tricyclohexyl ester[13], were coupled to the NH_2-terminus of each peptide. All peptides were deprotected with anhydrous HF and purified by reverse phase HPLC (C4 or C8).[14,16] All were single peaks by analytical RP-HPLC (C8) except in the case of the first Jun synthesis.

DNA Binding and Cleaving Assays. All assays were performed on a ^{32}P-end-labelled restriction fragment (EcoRI-Pvu II) from the plasmid pARE/GCRE,[66] which was prepared by standard procedures.[84] A-specific sequencing reactions were performed according to Iverson and Dervan.[85] DNase I footprinting[1] reaction conditions were 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 200 μM dithiothreitol (DTT), 0.33 μg/mL DNase I, 200 μM calf thymus DNA (bp), and 30,000 cpm end-labelled DNA, pH 7.0. Reactions were initiated with the addition of DNase and DTT, then allowed to proceed at room temperature for 3.5 minutes. MPE footprinting[2,3] reaction conditions were 20 mM phosphate, 20 mM NaCl, 5 mM DTT, 100 μM

calf thymus DNA and 30,000 cpm labelled DNA, pH 7.5. Reactions were initiated with the addition of MPE•Fe(II) and DTT, then allowed to proceed at room temperature for 10 minutes. Affinity cleaving reaction conditions were 30mM Tris•HCl, 3mM sodium acetate, 20 mM NaCl, 5 mM DTT, 100 μM calf thymus DNA and 20,000 cpm labelled DNA, pH 7.9. Reactions were initiated with the addition of DTT, then allowed to proceed at room temperature for 30 minutes. For all types of reactions, the peptides and the DNA were allowed to equilibrate for 30 minutes at room temperature before the reactions were initiated. For heterodimer experiments, proteins were allowed to equilibrate for one hour at 37° before DNA was added.

REFERENCES

1. D. Galas and A. Schmitz, *Nucleic Acids Res.* **5**, 3157 (1978).
2. R. P. Hertzberg and P. B. Dervan, *J. Am. Chem. Soc.* **104**, 313 (1982).
3. R. P. Hertzberg and P. B. Dervan, *Biochemistry* **23**, 3934 (1984).
4. T. D. Tullius, B. A. Dombroski, M. E. A. Churchill, and L. Kam, *Methods Enzymol.* **155**, 537 (1987).
5. P. B. Dervan, *Science* **232**, 464 (1986).
6. P. G. Schultz, J. S. Taylor, and P. B. Dervan, *J. Am. Chem. Soc.* **104**, 6861 (1982).
7. J. Taylor, P. Schultz, and P. Dervan, *Tetrahedron* **40**, 457 (1984).
8. H. E. Moser and P. B. Dervan, *Science* **238**, 645 (1987).
9. L. C. Griffin and P. B. Dervan, *Science* **245**, 976 (1989).
10. Several lines of evidence support this view. Tullius has shown that a sinusoidal cleavage pattern is obtained when a DNA is bound to a precipitate of calcium phosphate and allowed to react with Fe•EDTA, demonstrating that the two grooves have a different reactivity towards

Fe•EDTA.[86,87] Fe•EDTA attached to minor groove binding molecules affords cleavage patterns with single 3′- shifted asymmetric cleavage loci per Fe(II)•EDTA position.[5-7] However, when Fe(II)•EDTA is bound in the major groove of DNA by oligonucleotide triple helix formation, cleavage occurs along both strands of the adjacent minor grooves to afford a pair of cleavage loci of unequal intensity.[8,9] Cleavage is more intense along the strands proximal to the Fe(II)•EDTA than along the distal strands of the adjacent minor grooves.

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

12. J. Sluka, S. Horvath, M. Bruist, M. Simon, and P. Dervan, *Science* **238**, 1129 (1987).

13. J. P. Sluka, J. H. Griffin, D. P. Mack, and P. B. Dervan, *J. Am. Chem. Soc.* **112**, 6369 (1990).

14. D. P. Mack et al., *Biochemistry* **29**, 6561 (1990).

15. S. B. H. Kent, *Annu. Rev. Biochem.* **57**, 957 (1988).

16. J. P. Sluka, S. J. Horvath, A. C. Glasgow, M. I. Simon, and P. B. Dervan, *Biochemistry* **29**, 6561 (1990).

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

18. J. A. Shin, R. H. Ebright, and P. B. Dervan, *Nucleic Acids Res.* **19**, 5233 (1991).

19. T. K. Kerppola and T. Curran, *Curr. Opin. Struct. Biol.* **1**, 71 (1991).

20. D. Pathak and P. B. Sigler, *Curr. Opin. Struct. Biol.* **2**, 116 (1992).

21. W. H. Landschulz, P. F. Johnson, and S. L. McKnight, *Science* **240**, 1759 (1988).

22. I. A. Hope and K. Struhl, *EMBO J.* **6**, 2781 (1987).

23. R. Gentz, F. J. Rauscher, C. Abate, and T. Curran, *Science* **243**, 1695 (1989).

24. R. Turner and R. Tjian, *Science* **243**, 1689 (1989).

25. T. Kouzarides and E. Ziff, *Nature* **336**, 646 (1988).

26. T. D. Halazonetis, K. Georgopoulos, M. E. Greenberg, and P. Leder, *Cell* **55**, 917 (1988).

27. Y. Nakabeppu, K. Ryder, and D. Nathans, *Cell* **55**, 907 (1988).

28. L. J. Ransone, J. Visvader, P. Sassone-Corsi, and I. M. Verma, *Genes and Dev.* **3**, 770 (1989).

29. T. J. Bos, F. J. I. Rauscher, T. Curran, and P. K. Vogt, *Oncogene* **4**, 123 (1989).

30. W. H. Landschulz, P. F. Johnson, and S. L. McKnight, *Science* **243**, 1681 (1989).

31. M. Neuberg, J. Adamkiewicz, J. B. Hunter, and R. Müller, *Nature* **338**, 589 (1989).

32. P. Sassone-Corsi, W. W. Lamph, M. Kamps, and I. Verma, *Cell* **54**, 553 (1988).

33. M. Schuermann et al., *Cell* **56**, 507 (1989).

34. A. G. Hinnebusch, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6442 (1984).

35. G. Thireos, M. D. Penn, and H. Greer, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5096 (1984).

36. W. S. Moye-Rowley, K. D. Harshman, and C. S. Parker, *Genes Dev.* **3**, (1989).

37. Y. Maki, T. J. Bos, C. Davis, M. Starbuck, and P. K. Vogt, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2848 (1987).

38. J. Schutte et al., *Cell* **59**, 987 (1989).

39. S.-I. Hirai, R.-P. Ryseck, F. Mechta, R. Bravo, and M. Yaniv, *EMBO J.* **8**, 1433 (1989).

40. K. Ryder, A. Lanahan, E. Perez-Albuerne, and D. Nathans, *Proc. Natl. Acad. Sci.* **86**, 1500 (1989).

41. T. Curran et al., *Cell* **36**, 259 (1984).

42. C. Van Beveren, F. V. Straaten, T. Curran, R. Muller, and I. M. Verma, *Cell* **32**, 1241 (1983).

43. D. R. Cohen and T. Curran, *Mol. Cell. Biol.* **8**, 2063 (1988).

44. M. Zerial et al., *EMBO J.* **8**, 805 (1989).

45. W. H. Landschulz, P. F. Johnson, E. Y. Adashi, B. J. Graves, and S. L. McKnight, *Genes Dev.* **2**, (1988).

46. J. P. Hoeffler, T. E. Meyer, Y. Yun, J. L. Jameson, and J. F. Habener, *Science* **242**, 1420 (1988).

47. J. L. Paluh, M. J. Orbach, T. L. Legerton, and C. Yanofsky, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3728 (1988).

48. Y.-H. Fu, J. V. Paietta, D. G. Mannix, and G. A. Marzluf, *Mol. Cell. Biol.* **9**, 1120 (1989).

49. E. K. O'Shea, R. Rutkowski, and P. S. Kim, *Science* **243**, 538 (1989).

50. E. K. O'Shea, R. Rutkowski, W. F. Stafford, and P. S. Kim, *Science* **245**, 646 (1989).

51. E. K. O'Shea, R. Rutkowski, and P. S. Kim, *Cell* **68**, 699 (1992).

52. E. K. O'Shea, J. D. Klemm, P. S. Kim, and T. Alber, *Science* **254**, 539 (1991).

53. K. T. O'Neil, R. H. Hoess, and W. F. DeGrado, *Science* **249**, 774 (1990).

54. K. T. O'Neil, J. D. Schuman, C. Ampe, and W. F. DeGrado, *Biochemistry* **30**, 9030 (1991).

55. L. Patel, C. Abate, and T. Curran, *Nature* **347**, 572 (1990).

56. J. D. Shuman, C. R. Vinson, and S. L. McKnight, *Science* **249**, 771 (1990).

57. R. V. Talanian, C. J. McKnight, and P. S. Kim, *Science* **249**, 769 (1990).

58. M. A. Weiss et al., *Nature* **347**, 575 (1990).

59. C. R. Vinson, P. B. Sigler, and S. L. McKnight, *Science* **246**, 911 (1989).

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

61. T. E. Ellenberger, C. J. Brandl, K. Struhl, and S. C. Harrison, *Cell* **71**, 1223 (1992).

62. E. W. Jones and G. R. Fink, in *The Molecular Biology of the Yeast Saccharmomyces: Metabolism and Gene Expression* J. N. Strathern, E. W. Jones, J. R. Broach, Eds. Cold Spring Harbor, NY, 1982) pp. 181.

63. I. A. Hope and K. Struhl, *Cell* **46**, 885 (1986).

64. D. E. Hill, I. A. Hope, J. P. Macke, and K. Struhl, *Science* **224**, 451 (1986).

65. I. A. Hope and K. Struhl, *Cell* **43**, 177 (1985).

66. K. D. Harshman, W. S. Moye-Rowley, and C. S. Parker, *Cell* **53**, 321 (1988).

67. Based on sequence similarity (PESSD), one candidate for this structural element would be an Ω loop.[88]

68. D. P. Mack and P. B. Dervan, *Biochemistry* **31**, 9399 (1992).

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

70. S. F. Singleton and P. B. Dervan, *Biochemistry* **31**, 10995 (1992).

71. S. F. Singleton and P. B. Dervan, *J. Am. Chem. Soc.* **114**, 6957 (1992).

72. J. A. Shin, Ph.D. Thesis, California Institute of Technology (1992).

73. T. Curran and B. R. Franza, Jr., *Cell* **55**, 395 (1988).

74. R. Chui et al., *Cell* **54**, 541 (1988).

75. D. R. Cohen, P. C. P. Ferreira, R. Gentz, B. R. J. Franza, and T. Curran, *Genes Dev.* **3**, 173 (1989).

76. M. Neuberg, J. Adamkiewicz, J. B. Hunter, and R. Müller, *Nature* **341**, 243 (1989).

77. Y. Nakabeppu and D. Nathans, *The EMBO Journal* **8**, 3833 (1989).

78. T. Kouzarides and E. Ziff, *Nature* **340**, 568 (1989).

79. J. W. Sellars and K. Struhl, *Nature* **341**, 74 (1989).

80. G. Risse, K. Jooss, M. Neuberg, H.-J. Brüller, and R. Müller, *EMBO J.* **8**, 3825 (1989).

81. C. Abate, D. Luk, R. Gentz, F. J. I. Rauscher, and T. Curran, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1032 (1990).

82. A. R. Oliphant, C. Brandl, and K. Struhl, *Mol. Cell. Biol.* **9**, 2944 (1989).

83. V. K. Sarin, S. B. H. Kent, J. P. Tam, and R. B. Merrifield, *Anal. Biochem.* **117**, 147 (1981).

84. J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratory Manual, 2nd Edition* (Cold Spring Harbor Press, New York, 1989).

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

86. T. D. Tullius and B. A. Dombroski, *Science* **230**, 679-681 (1985).

87. T. D. Tullius and B. A. Dombroski, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5469 (1986).

88. J. F. Leszczynski and G. D. Rose, *Science* **234**, 849-855 (1986).

CHAPTER THREE

EVIDENCE THAT A DESIGNED SYNTHETIC PEPTIDE AND A DNA-BINDING PROTEIN CAN SIMULTANEOUSLY OCCUPY A COMMON BINDING SITE ON DOUBLE -HELICAL DNA

GCN4 is a yeast transcriptional activator responsible for the coordinate induction of 30 - 50 genes involved in amino acid biosynthesis in response to amino acid starvation.[1] This protein is a member of a class of sequence-specific DNA-binding proteins known as the leucine zipper-basic region proteins, which are proposed to bind to DNA through a conserved structural motif consisting of a coiled coil dimerization domain, often termed the leucine zipper, and a DNA-binding domain termed the basic region.[2-13] These proteins recognize, as dimers, binding sites on DNA consisting of abutted inverted repeats. For GCN4, the consensus DNA binding site is 5'-rrTGAC<u>T</u>catt-3', where the underlined C corresponds to the center of pseudodyad symmetry.[14,15]

The C-terminal 60 amino acids of GCN4 (residues 222-281) contain both the leucine zipper and basic regions and have been shown to be necessary and sufficient for both dimerization and sequence-specific DNA binding.[16,17] Two detailed modeling studies of leucine zipper-basic region proteins have been reported: the scissors-grip model of Vinson et al.[13] and the induced helical fork model of O'Neil et al.[10] Although these structural models differ slightly, both propose exclusively major groove binding. Footprinting, methylation interference, and ethylation interference studies are also consistent with largely major groove binding by these proteins.[13,18,19] Affinity cleaving proteins based on the DNA-binding domain of GCN4 with the DNA cleaving moiety Fe•EDTA at the NH_2-terminus have been synthesized (Fig 1, A and B).[20] Cleavage patterns generated by Fe•EDTA-GCN4(226-281) bound to the DNA sites 5'-

CTGACTAAT-3′ and 5′-ATGACTCTT-3′ reveal that the NH_2-termini of the GCN4 DNA binding domain are located in the major groove of DNA, nine to ten base pairs apart, consistent with a Y-shaped dimeric structure.[20] However, at the time this work was published, the structure of the DNA binding region of this class of proteins had not been reported.[21]

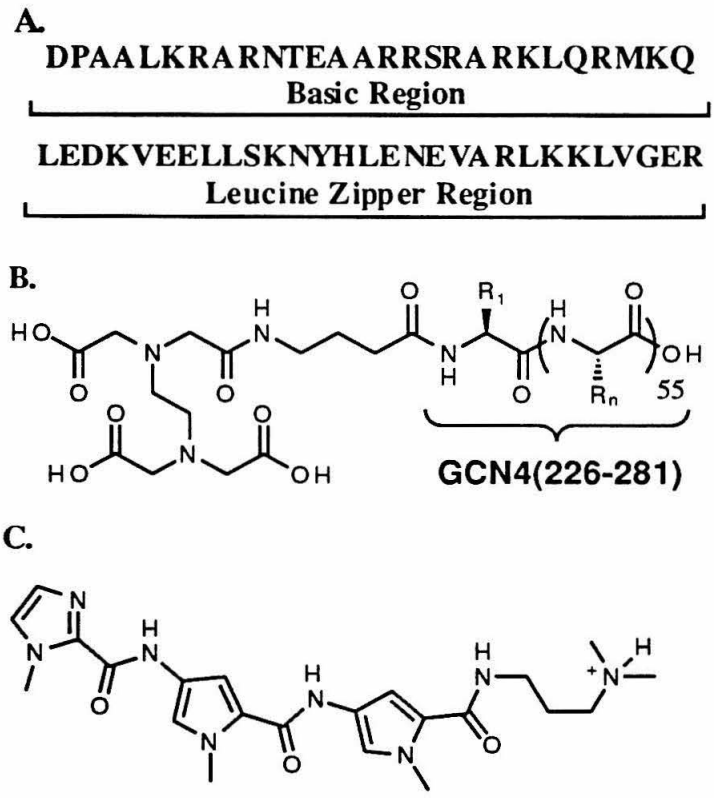


Figure 1. (A) The sequence of the carboxy terminal 56 amino acids (226-281) of GCN4. The sequence is shown from the NH_2-terminus of the peptide to its CO_2H-terminus. **(B)** Synthetic peptide EDTA-GCN4(226-281). **(C)** Synthetic minor groove binding peptide 2-ImN.

1-Methylimidazole-2-carboxamide-netropsin (2- ImN) is a synthetic analog of the naturally occurring minor groove binding peptide distamycin A, wherein the NH_2-terminal pyrrole has been replaced by 1-methylimidazole in order to allow for the binding of sites containing G, C base pairs (Figure 1C).

Footprinting and affinity cleaving studies reveal that 2-ImN binds to the five base pair sequence 5'- TGACT -3',[22] the conserved element of the GCN4 recognition site. Two-dimensional NMR studies have recently shown that 2-ImN binds to its recognition site in the minor groove of DNA as a side-by-side, antiparallel dimer,[23] forming a complex similar to that observed for distamycin binding to a 5'-AAATT-3' sequence at high concentrations.[24]

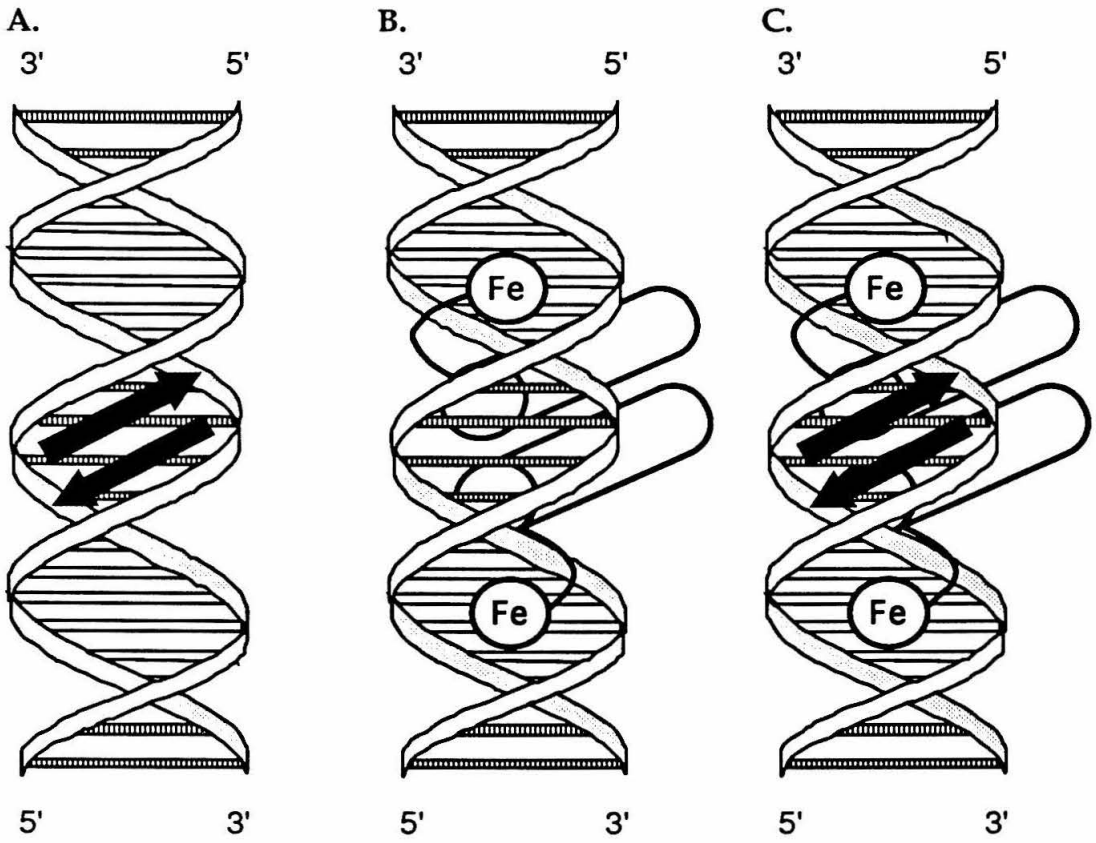


Figure 2. Schematic models for **(A)** the binding of the 2-ImN antiparallel dimer in the minor groove of DNA in which 2-ImN monomers are represented as arrows, **(B)** the leucine zipper DNA-binding motif in which the leucine zipper domains are represented by a pair of cylinders pointing into the DNA, and **(C)** both dimers binding to the same DNA molecule.

The complexes formed by 2-ImN and the GCN4 DNA binding domain with their common DNA binding site are shown schematically in Figures 2A and 2B,

respectively. The question arises as to whether these molecules can bind to their common DNA binding site simultaneously (Fig. 2C). This question of mutual recognition is addressed here by the use of Fe•EDTA-GCN4(226-281) as a sequence-specific footprinting reagent. Following chemical activation with a reducing agent such as dithiothreitol (DTT), Fe•EDTA localized in the major groove at a specific DNA binding site cleaves the proximal and distal strands of the adjacent minor grooves via a diffusible, non-sequence-specific radical, most likely hydroxyl[20,25-27]. The cleavage pattern generated by the dimeric Fe•EDTA-GCN4(226-281) extends beyond the five base pair recognition site for 2-ImN. Therefore, the peptide dimer bound in the minor groove should be capable of protecting its binding site from cleavage by diffusible radicals generated locally by Fe•EDTA-GCN4(226-281) bound at the same sequence in the major groove. This protection will be observed if, and only if, the major groove binding protein GCN4 and the minor groove binding synthetic peptide 2-ImN share their sites simultaneously on the same molecule of DNA.

RESULTS AND DISCUSSION

Simultaneous Binding of 2-ImN and GCN4. Footprinting and affinity cleaving assays were performed on a restriction fragment (*Eco* RI -*Pvu* II) from the plasmid pARE/GCRE,[28] which was separately 5'- and 3'-^{32}P-end-labeled at the *Eco* RI site. This DNA fragment contains two binding sites for GCN4 and 2-ImN, 5′-CTGACTAAT-3′ and 5′-ATGACTCTT-3′, previously designated ARE and GCRE, respectively.[28] The DNA cleavage products were separated by polyacrylamide gel electrophoresis and visualized by autoradiography (Fig. 3). From MPE•Fe footprinting , 2-ImN protects a region of approximately five base pairs at the site 5'-TGACT-3' (Fig. 3, lanes 7 and 8, Fig. 4B), consistent with previous footprinting and affinity cleaving data[22] and direct NMR studies.[23]

Figure 3. Autoradiogram of a high resolution denaturing polyacrylamide gel of 2-ImN and GCN4(226-281) footprinting and of Fe•EDTA-GCN4(226-281) affinity cleaving reactions in the presence and absence of 2-ImN. Odd numbered lanes are 5'-end-labeled and even numbered lanes are 3'-end-labeled. Lanes 1 and 2, DNA control lanes; lanes 3 and 4, A-specific chemical sequencing lanes; lanes 5 and 6, MPE•Fe control lanes; lanes 7 and 8, MPE•Fe cleavage protection by GCN4(226-281) (5 μM); lanes 9 and 10, MPE•Fe cleavage protection by 2-ImN (50 μM); lanes 11 and 12, Fe•EDTA-GCN4(226-281) at 5 μM; lanes 13 and 14, Fe•EDTA-GCN4(226-281) at 5 μM in the presence of 2-ImN at 50 μM.

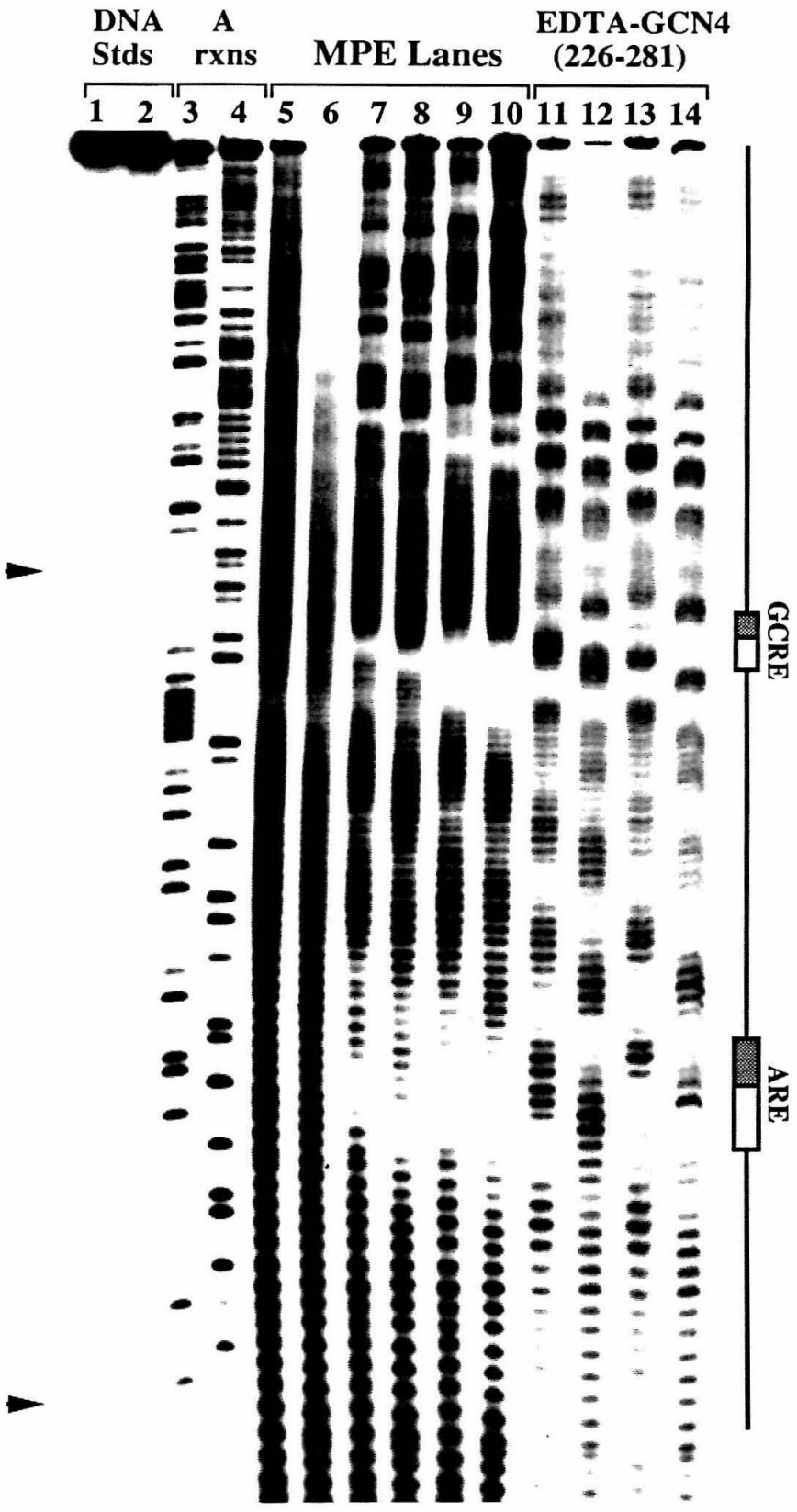
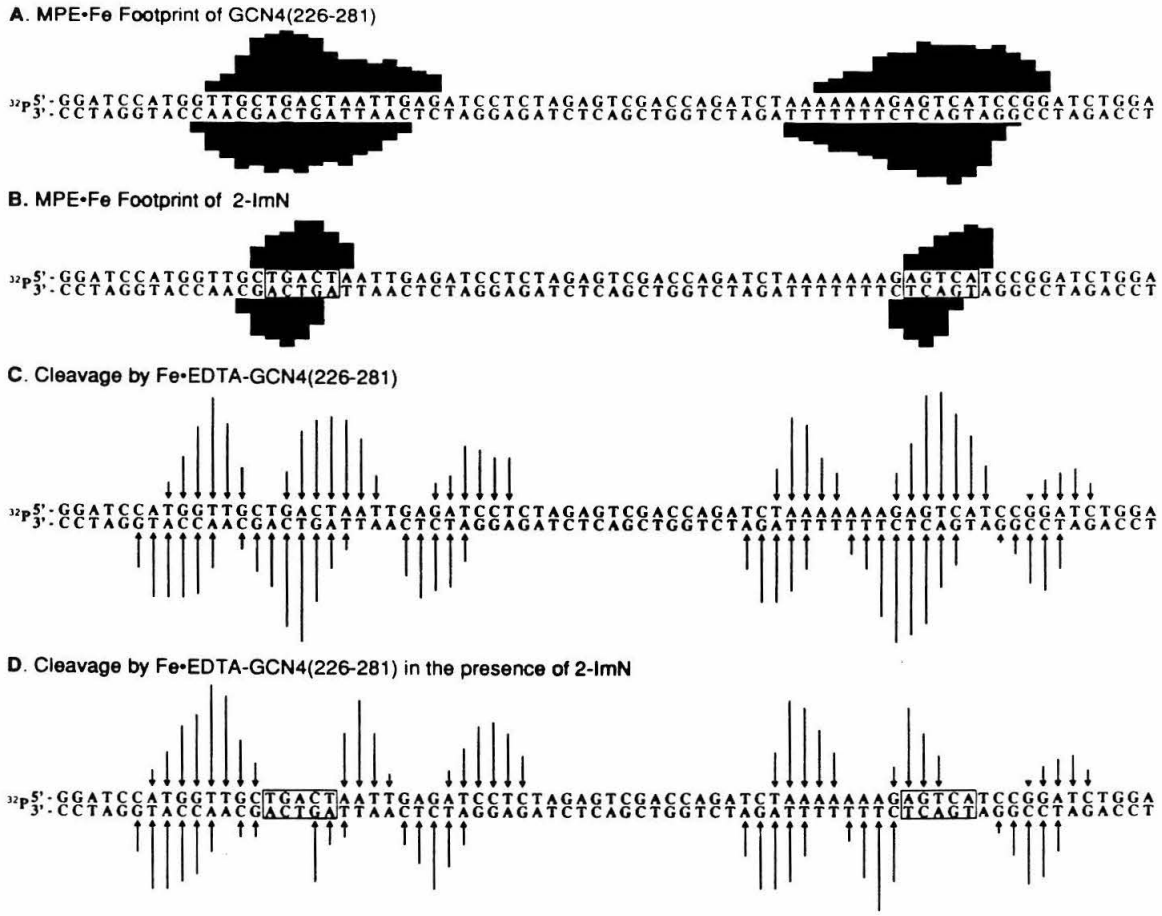

DNA Stds
A rxns
MPE Lanes
EDTA-GCN4 (226-281)
1 2 3 4 5 6 7 8 9 10 11 12 13 14
GCRE
ARE

Figure 4. The sequence from left to right represents the data between the markers on the left-hand side of the gel shown in Fig. 3. (**A**) Bars represent the extent of protection from MPE•Fe cleavage at the ARE and GCRE sites in the presence of 5 μM GCN4(226-281) (Fig. 3, lanes 7 and 8). (**B**) Bars represent the extent of protection from MPE•Fe cleavage at the ARE and GCRE sites in the presence of 2-ImN (50 μM); boxes denote 2-ImN binding sites (Fig. 3, lanes 9 and 10). (**C**) Arrows represent the extent of cleavage at the ARE and GCRE sites for Fe•EDTA-GCN4(226-281) at 5 μM (Fig. 3, lanes 11 and 12). (**D**) Arrows represent the extent of cleavage at the ARE and GCRE sites for Fe•EDTA-GCN4(226-281) at 5 μM in the presence of 2-ImN (50 μM); boxes denote 2-ImN binding sites (Fig. 3, lanes 13 and 14).


A. MPE•Fe Footprint of GCN4(226-281)
B. MPE•Fe Footprint of 2-ImN
C. Cleavage by Fe•EDTA-GCN4(226-281)
D. Cleavage by Fe•EDTA-GCN4(226-281) in the presence of 2-ImN

GCN4(226-281) protects an 18 bp region which contains the same five base pair site, 5'-TGACT-3'. Importantly, the products of cleavage by Fe•EDTA-GCN4(226-281) in the absence and in the presence of 2- ImN differ (Fig. 3, lanes 11 -14, Fig. 4C,D). The tripartite cleavage pattern for Fe•EDTA-GCN4(226-281) results from cleavage of the DNA binding site by two Fe•EDTA moieties localized in successive major grooves (Fig. 6A).[20] The cleavage pattern extends well beyond the five base pair binding site for 2- ImN in both directions. Comparisons of lane 11 with lane 13 and lane 12 with lane 14 demonstrate that the intensity of cleavage is reduced dramatically in the central region of the cleavage pattern in the presence of 2- ImN. This cleavage pattern is unaltered upon changing the order of addition of the two DNA-binding molecules (data not shown). The histograms in Figures 4C and D and the densitometry traces in Figure 5 compare the cleavage patterns observed in the presence and absence of 2-ImN. The base positions at which cleavage intensity is reduced correspond to those bases protected by 2-ImN from cleavage by MPE•Fe, demonstrating that 2-

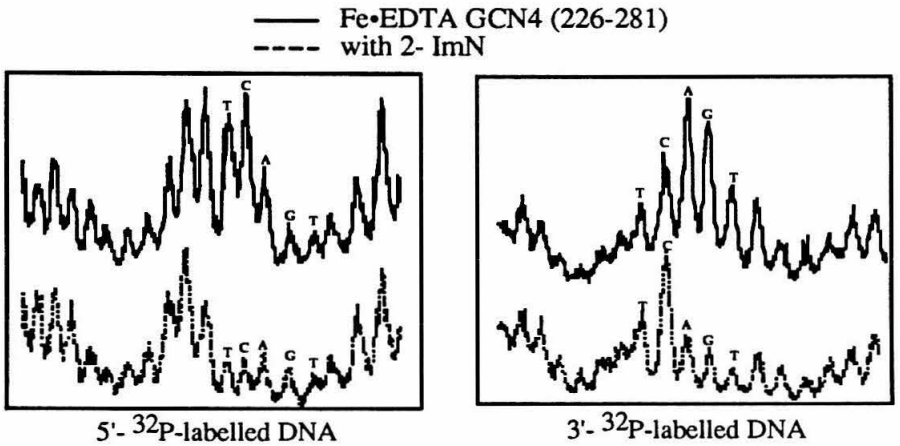


Figure 5. Phosphor storage densitometry traces of the 5'-TGACT-3' sequence from the lower binding site (ARE) in lanes 11-14 of Figure 3. The solid lines show the data obtained from cleavage of the DNA with Fe•EDTA-GCN4(226-281) in the absence of 2-ImN; the dashed lines show the data obtained in the presence of 2-ImN. The peaks corresponding to the sequence 5'-TGACT-3' are indicated.

ImN is capable of binding in the presence of Fe•EDTA-GCN4(226-281).

It is important to note that the cleavage pattern observed in Figure 4D can only occur if the two molecules occupy their common binding site *simultaneously*. If binding by 2-ImN in the minor groove prevents Fe•EDTA-GCN4(226-281) from binding, no specific cleavage will be observed. Similarly, if binding by Fe•EDTA-GCN4(226-281) in the major groove prevents 2-ImN from binding in the minor groove, an unaltered cleavage pattern will be observed (Fig. 3D). Furthermore, the cleavage intensity in the upper and lower sections of the tripartite Fe•EDTA-GCN4(226-281) cleavage pattern, which do not overlap a TGACT sequence, serves as an internal control for a decrease in the binding affinity of Fe•EDTA-GCN4(226-281) in the presence of 2-ImN. As can be seen from the data (Figure 3; Figure 4C and D), these bands are of similar intensity in the presence and absence of 2- ImN. The cleavage pattern for Fe•EDTA-GCN4(226-281) in the presence and absence of 2-ImN is shown mapped to a DNA double helix in Figure 6.

DNA Structure. This example of simultaneous sequence-specific complexation of a shared DNA binding site by a protein and a small molecule is in contrast to results seen with related systems. The minor groove binder distamycin inhibits cleavage by restriction endonucleases[29] and topoisomerase II[30] at A,T-rich sites. In addition, distamycin has been shown to inhibit the binding of regulatory proteins such as OTF-1, NFE-1[31] and the homeodomain proteins Ftz and Antp.[32] Surprisingly, distamycin is able to disrupt binding of a Ftz homeodomain peptide lacking the NH_2-terminal arm involved in minor groove specific contacts, implying that direct competition for the minor groove binding site is not required for inhibition.[32] It is proposed that the protein binding affinity is also reduced by a distamycin-induced conformational change in the local DNA structure.[32]

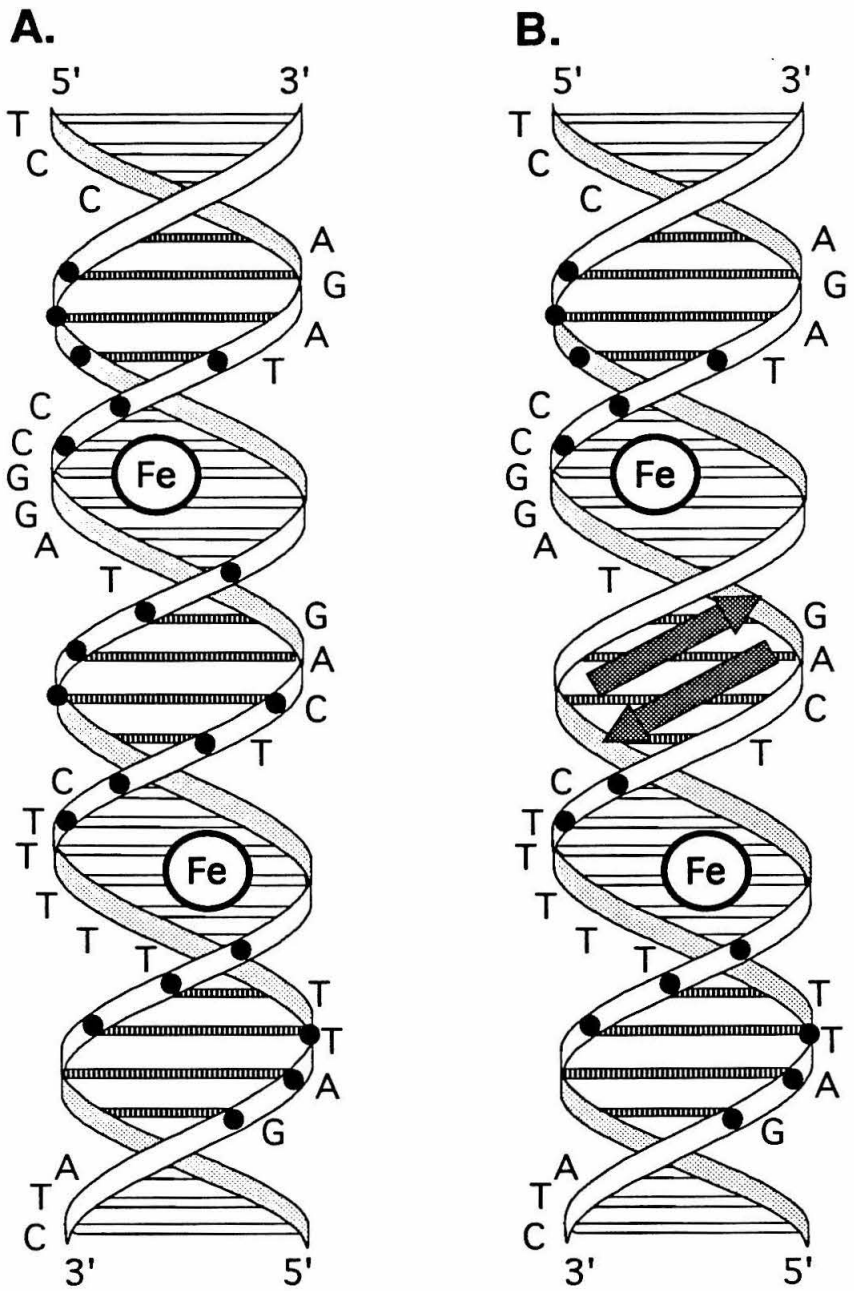


Figure 6. (**A**) Schematic for B-form DNA indicating the most intense cleavage sites for Fe•EDTA-GCN4(226-281) at the GCRE site. Filled circles represent the points of cleavage along the phosphodiester backbone. **(B)** A map of the most intense cleavage sites for Fe•EDTA-GCN4(226-281) in the presence of 2-ImN.

It is of interest to understand how the DNA is able to accommodate structurally the two dimeric species, GCN4 and 2-ImN. The simplest assumption is that different faces of the same local DNA structure are recognized by both dimers. Although some members of the leucine zipper-basic region class of DNA-binding proteins have been shown to bend their DNA recognition sites either toward (Jun-Jun homodimers) or away from (Jun-Fos heterodimers) the minor groove,[33,34] similar studies with GCN4 gave no evidence of protein-induced DNA bending.[18] Similarly, NMR studies with 2-ImN are consistent with no major distortion of the DNA double helix as a result of ligand binding.[23]

Table I. Binding constants for 2-ImN at two DNA binding sites measured by footprint titrations using Fe•EDTA-GCN4(226-281) and MPE•Fe.

Site	Footprinting Agent	K_{app} $(M^{-1})^{a,b}$
5'- CTGACTCTT-3'	Fe•EDTA-GCN4(226-281)	2.2 (0.3) x 10^5
	MPE	1.5 (0.4) x 10^5
5'-ATGACTAAT-3'	Fe•EDTA-GCN4(226-281)	2.3 (0.6) x 10^5
	MPE	2.2 (0.5) x 10^5

[a] Values reported are the mean values measured from three or four footprint titration experiments. Numbers in parentheses indicate the standard deviation for the data set.[b] The assays were performed at room temperature, pH 7.9, in the presence of 30 mM tris-hydrochloride, 3 mM sodium acetate, 20 mM sodium chloride.

Nevertheless, modeling studies indicate that the minor groove width in the 2-ImN•5'-TGACT-3' complex may be larger than that observed for canonical B-form DNA in order to accommodate the 2-ImN dimer.[23] If even modest structural changes are induced upon binding of either the GCN4 or the 2-ImN dimer, positive or negative cooperativity between GCN4 and 2-ImN might be

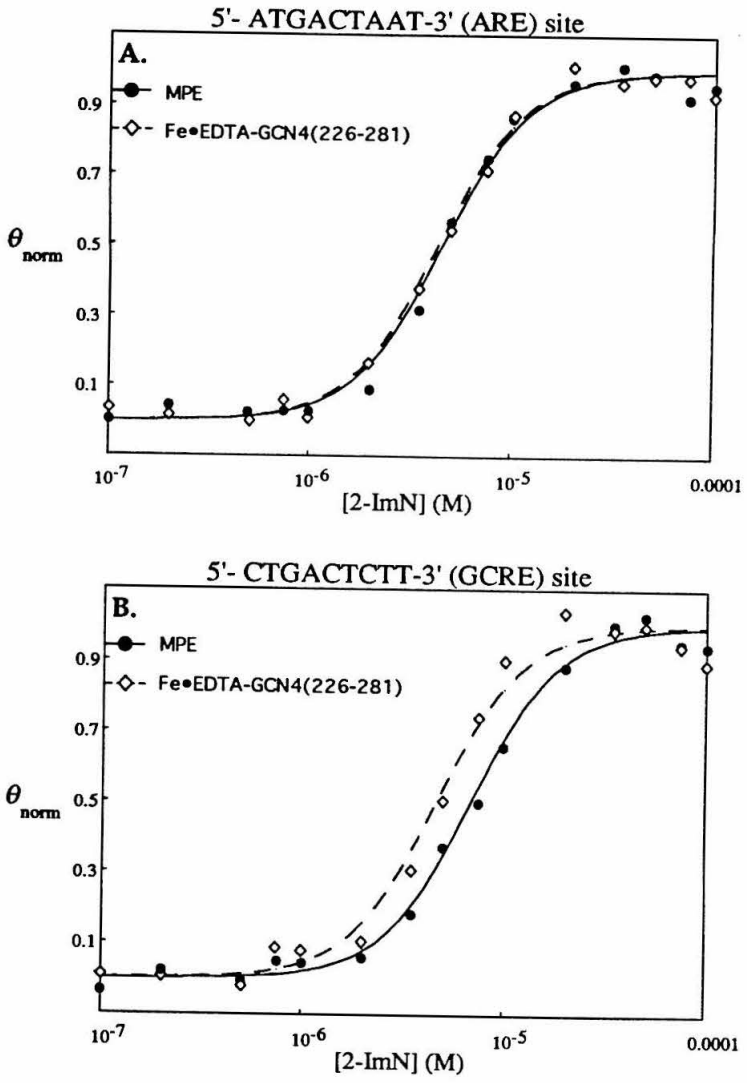


Figure 7. Data for the MPE•Fe and Fe•EDTA-GCN4(226-281) quantitative footprinting experiments at (**A**) the ARE site and (**B**) the GCRE site. The sigmoidal curves show the titration binding isotherms plotted with eq 2 using $\theta_{min} = 0$, $\theta_{max} = 1$, and the mean K_a value for each set (Table I). The solid curves represent the MPE•Fe footprinting data, the dotted curves, the Fe•EDTA-GCN4(226-281) data. The data points are taken from the average of the normalized θ_{app} values at each concentration of 2-ImN; filled circles correspond to MPE•Fe data, open diamonds to Fe•EDTA-GCN4(226-281) data.

Figure 8. Gray-scale representation of a storage phosphor autoradiogram of a denaturing 8% polyacrylamide gel showing quantitative MPE footprinting titration reactions for 2-ImN. All lanes except the sequencing lane contain 5'-^{32}P-end-labeled DNA. Lane 1, intact DNA; lane 2, 3'-A-specific sequencing reaction; lane 3, MPE control lane; lanes 4-18 show MPE cleavage protection by 2-ImN at concentrations of 100, 75, 50, 35, 10, 7.5, 5, 3.5, 2, 1, 0.75, 0.5, 0.35, 0.2, and 0.1 μM, respectively.

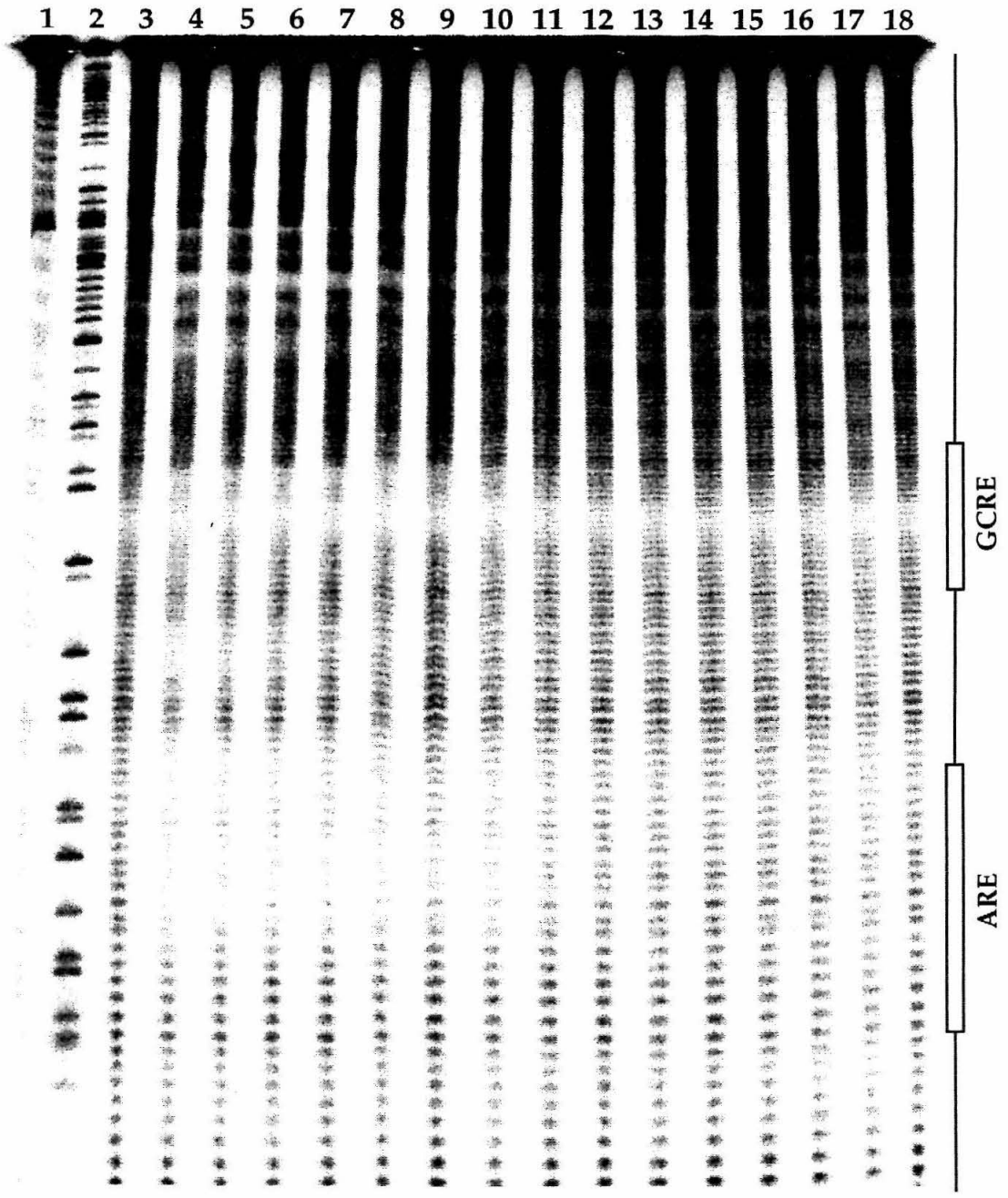
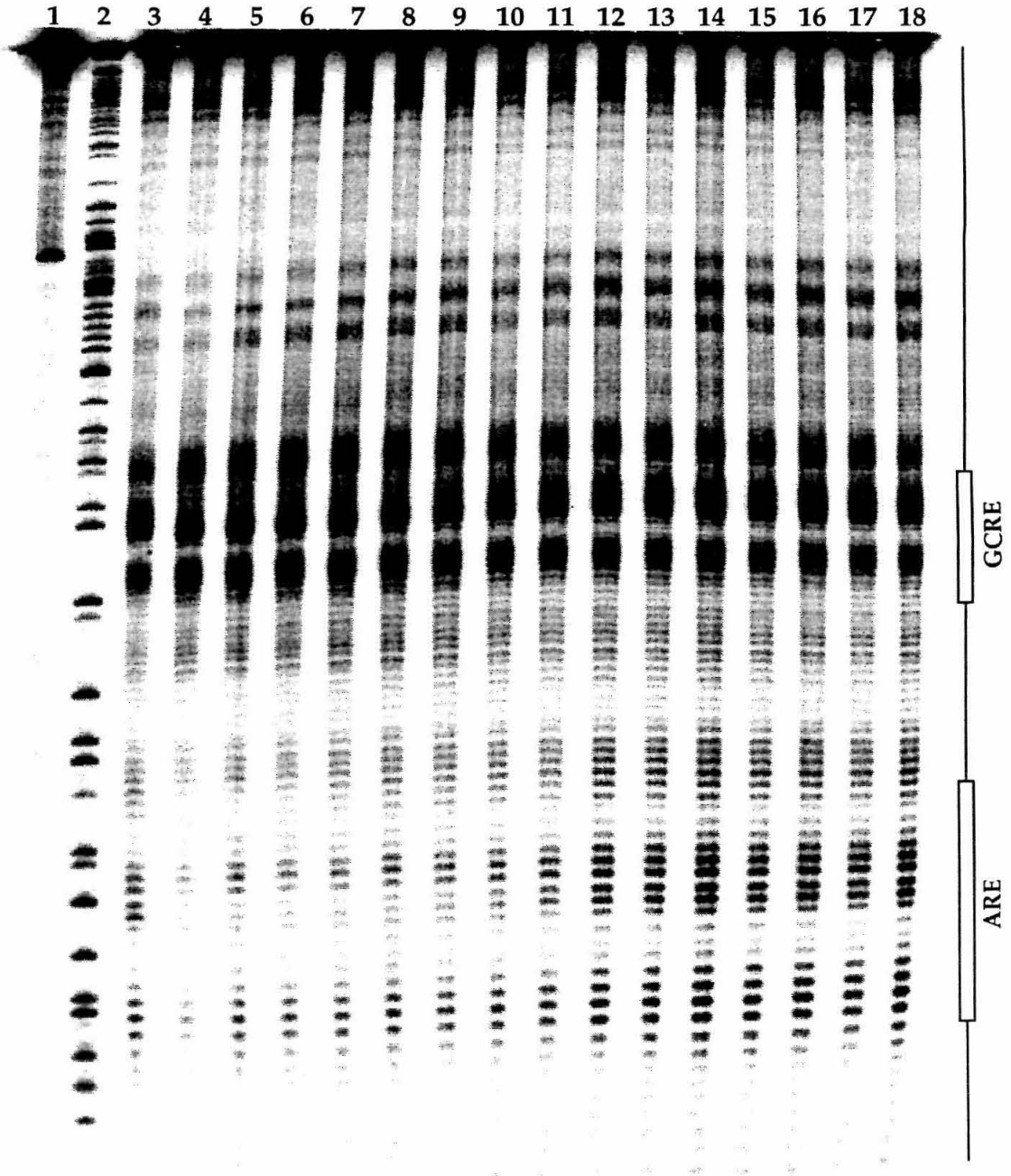

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
GCRE
ARE

Figure 9. Gray-scale representation of a storage phosphor autoradiogram of a denaturing 8% polyacrylamide gel showing quantitative EDTA-GCN4(226-281) footprinting titration reactions for 2-ImN. All lanes except the sequencing lane contain 5'-^{32}P-end-labeled DNA. Lane 1, intact DNA; lane 2, 3'-A-specific sequencing reaction; lane 3, EDTA-GCN4(226-281) control lane; lanes 4-18 show EDTA-GCN4(226-281) cleavage protection by 2-ImN at concentrations of 100, 75, 50, 35, 10, 7.5, 5, 3.5, 2, 1, 0.75, 0.5, 0.35, 0.2, and 0.1 μM, respectively.


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
GCRE
ARE

expected. In order to address this possibility, we have measured the binding affinity constants for 2-ImN in the presence and in the absence of Fe•EDTA-GCN4(226-281). To measure the apparent association constant of 2-ImN for each site in the absence of Fe•EDTA-GCN4(226-281), we used a quantitative MPE•Fe footprint titration procedure,[35] adapted from the quantitative DNase footprint titration procedure of Ackers and co-workers.[36] A typical MPE gel is shown in Figure 8. To measure the apparent association constant for each binding site in the presence of Fe•EDTA-GCN4(226-281), we used the protein itself as a site-specific footprinting agent; a typical Fe•EDTA-GCN4(226-281) gel is shown in Figure 9.

Because NMR[23] and affinity cleaving[22] experiments support the assumption of all-or-none binding, apparent fractional saturation values (θ_{app}, eq 1) were fit to a modified Hill equation (eq 2).[37,38] The means of the best-fit apparent monomeric association constants are listed in Table I; the corresponding binding isotherms are shown in Figure 7. The association constants for 2-ImN in the presence and absence of Fe•EDTA-GCN4(226-281) differ by less than a factor of two at both sites and probably are not different within the error of the experiment. These data indicate that there is little, if any, cooperativity between the two molecules, supporting a model in which neither the 2-ImN dimer nor the truncated GCN4 dimer perturbs the common DNA binding site significantly.

CONCLUSION

The DNA-binding domain of the leucine zipper-basic region protein GCN4 and the synthetic minor groove-binding peptide 2-ImN can bind to their common DNA binding site simultaneously. This result supports models in which leucine zipper-basic region proteins bind to their DNA binding sites

exclusively in the major groove, leaving the minor groove unoccupied. Furthermore, binding of Fe•EDTA-GCN4(226-281) in the major groove has little or no effect on the equilibrium binding constant for 2-ImN. This demonstration that both grooves of the DNA may be occupied without significantly weakening the binding affinity of the major or minor groove binding molecules suggests that tight-binding synthetic molecules may be designed with components interacting with both grooves of DNA over a limited sequence space.

EXTENSION TO OTHER SYSTEMS

Jun and Fos Homodimers and Heterodimers. Preliminary investigations were undertaken to examine the effect of 2-ImN binding on Jun-Jun and Fos-Fos homodimers, as well as Jun-Fos heterodimers. Unlike GCN4, Jun-Jun homodimers and Jun-Fos heterodimers have been shown to bend their DNA recognition sites.[33,34] Because Jun-Jun homodimers and Jun-Fos heterodimers bend DNA in opposite directions,[33,34] it was of interest to determine whether the two dimers exhibited different behavior in the presence of a minor groove binder at the same DNA sequence. Qualitative experiments indicated that 2-ImN can protect its binding site from cleavage by Fe•EDTA-derivatives of Jun-Jun homodimers, Fos-Fos homodimers, and Jun-Fos heterodimers (data not shown; MGO III-151). The Jun-Fos homodimers would be expected to widen the minor groove, allowing 2-ImN to bind. However, Jun-Jun homodimers bend DNA toward the minor groove and would be expected to have an inhibitory effect on 2-ImN binding. No such effect was observed. It is known that truncated Jun and Fos proteins such as those used in these experiments bend DNA to a lesser degree than the full proteins,[33,34] and this bending may not be significant enough to affect 2-ImN binding. Alternatively, the Jun and Jun-Fos dimers may bend DNA outside the optimal 2-ImN binding site of 5′-TGACT-3′. Quantitative

footprinting titrations might reveal small effects not seen in qualitative experiments.

Triple Helices. Michelle Parks performed experiments to determine whether or not the minor groove binders 2-ImN and distamycin could complex their DNA binding sites in a pyrimidine motif triple helix.[39] Optimal conditions for binding of a third strand to form a triple helix,[27] especially the inclusion of spermine in the reaction mixture, were found to reduce the binding efficiency of the minor groove binders such that very high concentrations of these molecules were required.[39] Under conditions in which quantitative footprinting and affinity cleaving titrations indicate that both the third strand and the minor groove binder should bind their sites, even in the presence of high concentrations of the other species, neither 2-ImN nor distamycin protected its binding site from cleavage by a pyrimidine-rich oligonucleotide complexed with Fe•EDTA.[39] Although the cleavage seen by pyrimidine triple helices does spill over into the adjacent minor grooves, generating the cleavage pattern shown in Chapter Two (Fig. 2),[27] the percentage of cleavage in the minor groove is much less than for Fe•EDTA-GCN4(226-281),[40] rendering a small reduction in the cleavage intensity more difficult to detect.

The crystal structure of GCN4 provides an explanation for the largely minor groove cleavage effected by Fe•EDTA-GCN4(226-281).[41] Asp-226, the residue to which Fe•EDTA is attached, is located three α-helical turns away from the amino acid residues making contacts with the DNA bases (approximately 13Å from the nearest phosphate group).[41] The generation of diffusible oxidant from an Fe•EDTA moiety held away from the DNA helix allows more even access to the major and minor grooves than would an Fe•EDTA moiety held close to the DNA in the major groove. Because GCN4 binds exclusively in the major groove, protecting it from cleavage, cleavage occurs largely from the minor grooves.

MATERIALS AND METHODS

Materials

Protected amino-acid derivatives were purchased from Peninsula Laboratories; Boc-L-His(DNP) was obtained from Fluka. Arg-(Phenylacetamido)methyl (PAM) resin, *N, N* - dimethylformamide (DMF), diisopropylethylamine, dicyclohexylcarbodiimide in dichloromethane, *N*-hydroxybenzotriazole (HOBt) in DMF, and trifluoroacetic acid (TFA) were purchased from Applied Biosystems. Dichloromethane and methanol (HPLC grade) were obtained from Mallinckrodt, HOBt, *p* -cresol and *p* -thiocresol from Aldrich, and diethyl ether (low peroxide content) from Baker. Mass spectrometry was performed at the New York University Electrospray Mass Spectrometry Regional Center by Dr. Steven Wilson and colleagues.

Doubly distilled water was further purified through the Milli Q filtration system from Millipore. Sonicated, deproteinized calf thymus DNA was purchased from Pharmacia and dissolved in H_2O to a concentration of 1 mM in base pairs. tRNA (*E. coli* strain W, Type XX) was obtained from Sigma and was dissolved in water and sterile-filtered. Enzymes were obtained from Boehringer-Mannheim or New England Biolabs and used with the buffers supplied. Glycogen was obtained from Boehringer-Mannheim. Deoxyadenosine- 5'- [α-^{32}P]-triphosphate and adenosine- 5'-[γ-^{32}P]-triphosphate were obtained from Amersham. UV-vis spectra were recorded on a Hewlett-Packard Diode Array spectrophotometer. Cerenkov radioactivity was measured with a Beckman LS 3801 Scintillation Counter. Storage phosphor technology autoradiography was accomplished using a Molecular Dynamics 400S Phosphorimager and ImageQuant software. Storage phosphor screens (Kodak #S0230) were purchased from Molecular Dynamics.

Methods

Peptide and Protein Syntheses. EDTA-GCN4(226-281) and GCN4(226-281) were synthesized, deprotected and purified as previously described (Chapter Two, same synthesis).[20,42-45] NH_2-terminal sequencing of the purified GCN4(226-281) protein confirmed the desired sequence. Molecular masses were verified by electrospray mass spectrometry [GCN4(226-281): calculated mass, 6615.7, observed, 6615.1; EDTA-GCN4(226-281): calculated mass, 6975.0, Na complex, 6998.0, Fe complex, 7030.9; observed 6998.2, 7031.7]. 2-ImN was prepared and purified as previously described.[22] Fe•EDTA-GCN4(226-281) was generated immediately prior to use by equilibration of 1 mM EDTA-GCN4(226-281) with an equal volume of 1 mM ferrous ammonium sulfate for approximately 10 min. Fe•EDTA-GCN4(226-281) was then diluted to a concentration of 25 μM.

DNA Manipulations. The 270 bp *Pvu* II -*Eco* RI restriction fragment of the plasmid pARE/GCRE[28] was isolated and labeled at the 5'-or 3'- end by standard procedures.[46] Chemical sequencing adenine-specific reactions were carried out as previously described.[47] MPE•Fe reaction conditions were 20 mM phosphate, 20 mM NaCl, 5 mM DTT, 5 μM MPE•Fe, 100 μM calf thymus DNA, and 20,000 cpm labeled DNA, pH 7.5. The DNA binding molecules were allowed to equilibrate with the DNA for 30 min. Reactions were then initiated by the addition of MPE•Fe, followed immediately by DTT, and allowed to proceed 10 min at room temperature. Affinity cleaving reaction conditions were 30 mM Tris•HCl, 3 mM sodium acetate, 20 mM NaCl, 5 mM DTT, 100 μM calf thymus DNA, and 20,000 cpm labeled DNA, pH 7.9. 2- ImN was allowed to equilibrate with the DNA at room temperature for 30 min, at which point Fe•EDTA-GCN4(226-281) was added and allowed to equilibrate with the DNA for a further 30 min. The reactions were initiated with the addition of DTT and allowed to proceed at room temperature for 30 min. All reactions were

terminated by ethanol precipitation; the residues were dried and resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer. Reaction products were analyzed by electrophoresis on 8% polyacrylamide denaturing gels (5% crosslink, 7M urea). After electrophoresis, gels were dried and autoradiographed. Gels were analyzed using storage phosphor technology. Storage phosphor screens were pressed against dried gels and exposed. A Molecular Dynamics 400S PhosphorImager was used to obtain data from storage screens, and data were analyzed by performing area integrations along lines drawn down the center of each lane using the ImageQuant v.3.0 software running on an AST Premium 386/33 computer.

Quantitative Footprint Titrations. The above reaction conditions were modified for quantitative footprinting as follows. tRNA was used in place of calf thymus DNA in order to minimize the binding of 2-ImN to carrier nucleic acid. Stock solutions of 2-ImN were diluted serially to give 15 solutions of concentrations ranging from 500 nM to 500 μM, leading to final reaction concentrations of 100 nM to 100 μM. Final reaction conditions were 30 mM Tris•HCl, 3 mM sodium acetate, 20 mM NaCl, 2 mM DTT, 100 μg/mL tRNA , 5 μM Fe•EDTA-GCN4(226-281) or 2.5 μM MPE•Fe, and 20,000 cpm 5′-^{32}P end-labeled DNA, pH 7.9. For MPE•Fe footprinting, 2-ImN was allowed to equilibrate at room temperature with the DNA for 1 h, followed by the addition of MPE•Fe and DTT. Reactions were quenched by ethanol precipitation after 30 min. For Fe•EDTA-GCN4(226-281) footprinting, 2-ImN was allowed to equilibrate with the DNA at room temperature for 45 min, followed by the addition of Fe•EDTA-GCN4(226-281). After 45 min, the reactions were initiated with DTT. Reactions were quenched by ethanol precipitation after a further 30 min.

Footprint Titration Fitting Procedure. The footprint titration gels were quantitated using storage phosphor technology, and data were analyzed by performing volume integrations of the target and reference sites using the ImageQuant v.3.0 software running on an AST Premium 386/33 computer. The target sites chosen were the two 5′-TGACT-3′ sites found in the restriction fragment. For the MPE•Fe footprint titrations, reference sites were chosen at which little 2-ImN binding was observed even at the highest concentrations, generally G, C-rich sites. For the Fe•EDTA-GCN4(226-281) footprint titration experiments, reference sites were taken from regions at each Fe•EDTA-GCN4(226-281) cleavage site at which the intensity of cleavage was unaltered. Specifically, the lower set of cleavage bands at each site was used: 5′-GGTT-3′ at the ARE site, and approximately 5′-TAAAAA-3′ at the GCRE site.

The data from MPE•Fe and Fe•EDTA-GCN4(226-281) footprint titrations were analyzed using a method analogous to that described for analysis of DNase I footprint titrations.[22,36] The apparent DNA target site saturation, θ_{app}, was calculated using the following equation:

$$\theta_{app} = 1 - \frac{I_{tot}/I_{ref}}{I_{tot}^{\circ}/I_{ref}^{\circ}} \qquad (1)$$

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values for an MPE•Fe or Fe•EDTA-GCN4(226-281) control lane to which no 2-ImN has been added. The ($[L]_{tot}$, θ_{app}) data points were fit by minimizing the difference between θ_{app} and θ_{fit}, using the modified Hill equation:

$$\theta_{fit} = \theta_{min} + (\theta_{max} - \theta_{min}) \cdot \frac{K_a^2[L]_{tot}^2}{1 + K_a^2[L]_{tot}^2} \qquad (2)$$

where $[L]_{tot}$ corresponds to the total 2-ImN concentration, K_a corresponds to the apparent monomeric association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively.

Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) running on a Macintosh IIfx computer with K_a, θ_{max}, and θ_{min} as the adjustable parameters. All lanes from each gel were used unless visual inspection of the computer image from a storage phosphor screen revealed a flaw at the target or reference site, or the θ_{app} value was greater than two standard errors away from values in both neighboring lanes and greater than two standard errors from the initial θ_{fit}. Data for experiments in which fewer than 80% of the lanes were usable were discarded. The data were normalized using the following equation:

$$\theta_{norm} = \frac{\theta_{app} - \theta_{min}}{\theta_{max} - \theta_{min}} \qquad (3)$$

The uncertainty in θ_{norm} was estimated from the scatter in the normalized data for the four lowest and four highest concentrations of 2-ImN. A standard uncertainty in θ_{norm} of 0.07 was used in a χ^2 analysis of normalized data, with $\chi^2 \leq 1.0$ as the criterion for an acceptable fit. Three or four sets of acceptable data were used in determining each association constant.

REFERENCES

1. E. W. Jones and G. R. Fink, in *The Molecular Biology of the Yeast Saccharmomyces: Metabolism and Gene Expression* J. N. Strathern, E. W. Jones, J. R. Broach, Eds. Cold Spring Harbor, NY, 1982) p. 181.

2. W. H. Landschulz, P. F. Johnson, and S. L. McKnight, *Science* **240**, 1759 (1988).

3. W. H. Landschulz, P. F. Johnson, and S. L. McKnight, *Science* **243**, 1681 (1989).

4. T. Kouzarides and E. Ziff, *Nature* **336**, 646 (1988).

5. P. Sassone-Corsi, L. J. Ransone, W. W. Lamph, and I. Verma, *Nature* **336**, 692 (1988).

6. M. Neuberg, M. Schuermann, and R. Müller, *Oncogene* **6**, 1325 (1991).

7. R. Gentz, F. J. Rauscher, C. Abate, and T. Curran, *Science* **243**, 1695 (1989).

8. E. K. O'Shea, R. Rutkowski, and P. S. Kim, *Science* **243**, 538 (1989).

9. E. K. O'Shea, J. D. Klemm, P. S. Kim, and T. Alber, *Science* **254**, 539 (1991).

10. K. T. O'Neil, R. H. Hoess, and W. F. DeGrado, *Science* **249**, 774 (1990).

11. R. Turner and R. Tjian, *Science* **243**, 1689 (1989).

12. R. V. Talanian, C. J. McKnight, and P. S. Kim, *Science* **249**, 769 (1990).

13. C. R. Vinson, P. B. Sigler, and S. L. McKnight, *Science* **246**, 911 (1989).

14. I. A. Hope and K. Struhl, *Cell* **43**, 177 (1985).

15. D. E. Hill, I. A. Hope, J. P. Macke, and K. Struhl, *Science* **224**, 451 (1986).

16. I. A. Hope and K. Struhl, *Cell* **46**, 885 (1986).

17. I. A. Hope and K. Struhl, *EMBO J.* **6**, 2781 (1987).

18. M. R. Gartenberg, C. Ampe, T. A. Steitz, and D. M. Crothers, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6034 (1990).

19. J. A. Nye and B. J. Graves, *Proc. Natl. Acad. Sci. USA* **87**, 3992 (1990).

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

21. M. G. Oakley and P. B. Dervan, *Biochemistry* **31**, 10969 (1992).

22. W. S. Wade, M. Mrksich, and P. B. Dervan, *J. Am. Chem. Soc.* **114**, 8783 (1992).

23. M. Mrksich et al., *Proc. Natl. Acad. Sci. USA* **89**, 7586 (1992).

24. J. G. Pelton and D. E. Wemmer, *J. Am. Chem. Soc.* **112**, 1393 (1990).

25. J. Taylor, P. Schultz, and P. Dervan, *Tetrahedron* **40**, 457 (1984).

26. R. P. Hertzberg and P. B. Dervan, *Biochemistry* **23**, 3934 (1984).

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

28. K. D. Harshman, W. S. Moye-Rowley, and C. S. Parker, *Cell* **53**, 321 (1988).

29. V. V. Nosikov, E. A. Braga, A. V. Karlishev, A. L. Zhuze, and O. L. Polyanovsky, *Nucleic Acids Res.* **3**, 2293 (1976).

30. M. Fesen and Y. Pommier, *J. Biol. Chem.* **264**, 11354 (1989).

31. M. Broggini et al., *Nucleic Acids Res.* **17**, 1051 (1989).

32. A. Dorn, M. Affolter, M. Müller, W. J. Gehring, and W. Leupin, *EMBO J.* **11**, 279 (1992).

33. T. K. Kerppola and T. Curran, *Cell* **66**, 317 (1991).

34. T. K. Kerppola and T. Curran, *Science* **254**, 1210 (1991).

35. W. S. Wade, Ph.D. Thesis, California Institute of Technology (1989).

36. M. Brenowitz, D. F. Senear, M. A. Shea, and G. K. Ackers, *Methods Enzymol.* **130**, 132 (1986).

37. C. R. Cantor and P. R. Schimmel, in *Biophysical Chemistry*(W. H. Freedman and Company, New York, 1980), vol. Part III: The behavior of biological macromolecules, pp. 863.

38. Data were also fit to the following equation:

$$\theta_{fit} = \theta_{min} + (\theta_{max} - \theta_{min}) \cdot \frac{K_1[L]_{tot} + K_1K_2[L]_{tot}^2}{1 + K_1[L]_{tot} + K_1K_2[L]_{tot}^2} \qquad (4)$$

where K_1 corresponds to the association constant for binding of the first 2-ImN monomer and K_2 to the association constant for binding of the second monomer to a singly occupied site. Non-linear least squares analysis using

eq 4 with K_1, K_2, θ_{min}, and θ_{max} as adjustable parameters resulted in fits of similar quality to those obtained using eq 2. Moreover, these results gave $K_2 \geq 1000 \bullet K_1$, and $K_1K_2 \equiv K_a^2$, justifying our use of an all-or-none approximation.

39. M. E. Parks and P. B. Dervan, personal communication.

40. A comparison of the cleavage intensity at positions of the distal minor groove strand (Chapter Two, Fig. 2) for which cleavage must result from a diffusible oxidant in the minor groove, with that on the proximal strand, which is cleaved from both grooves, indicates that ~80-90% of the proximal strand cleavage by Fe•EDTA-GCN4(226-281) comes from the minor groove.

41. T. E. Ellenberger, C. J. Brandl, K. Struhl, and S. C. Harrison, *Cell* **71**, 1223 (1992).

42. S. B. H. Kent, *Annu. Rev. Biochem.* **57**, 957 (1988).

43. J. P. Sluka, J. H. Griffin, D. P. Mack, and P. B. Dervan, *J. Am. Chem. Soc.* **112**, 6369 (1990).

44. J. P. Sluka, S. J. Horvath, A. C. Glasgow, M. I. Simon, and P. B. Dervan, *Biochemistry* **29**, 6561 (1990).

45. D. P. Mack et al., *Biochemistry* **29**, 6561 (1990).

46. J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratory Manual, 2nd Edition* (Cold Spring Harbor Press, New York, 1989).

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

CHAPTER FOUR

DESIGN OF A HYBRID PROTEIN CONTAINING THE DNA-BINDING DOMAIN OF HIN RECOMBINASE AND THE IRON-BINDING LIGAND OF BLEOMYCIN

Bleomycin.

The bleomycins (BLMs, Fig. 1) are a group of glycopeptide antibiotics with clinically useful antitumor activities.[1] Bleomycin has been shown to cleave double-stranded DNA efficiently and specifically in the presence of oxygen and certain metal ions, most notably Fe(II) (for reviews, see references 2-4). Oxidative cleavage occurs primarily at a subset of 5'-GC-3' and 5'-GT-3' sites[5,6] and results from degradation of the deoxyribose backbone of the 3' nucleotide of each site by an activated Fe•Bleomycin species, leading to a single strand break at each DNA binding site.[7-9]

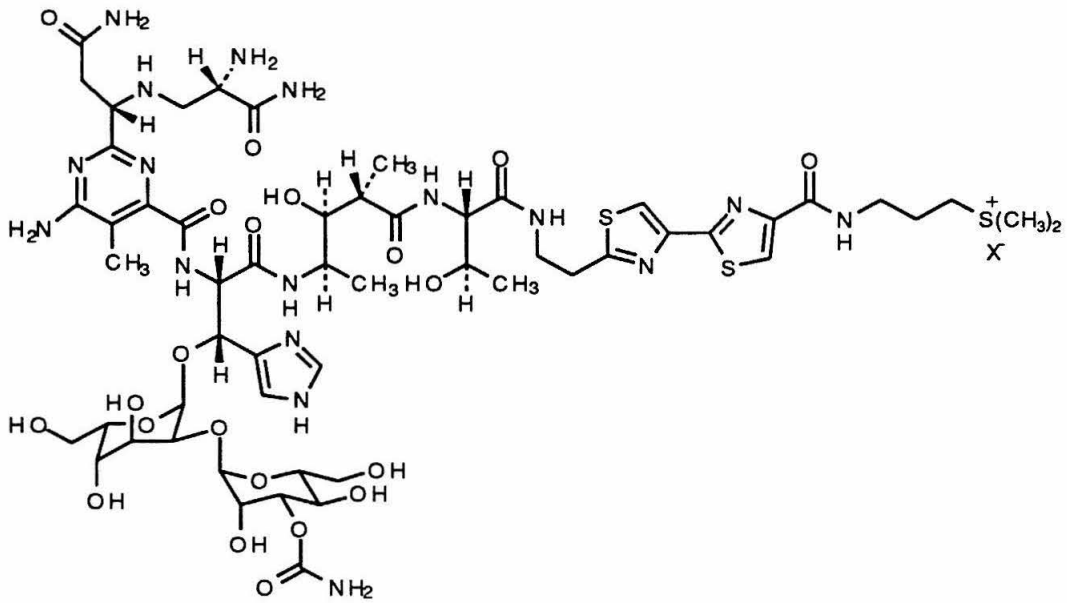


Figure 1. Bleomycin A_2.

Mechanism. An activated iron-bleomycin complex competent to cleave DNA may be formed from bleomycin in the presence of: Fe(II) and O_2; Fe(II), O_2, and a reducing agent, such as DTT or β-mercaptoethanol; Fe(III) and hydrogen peroxide;[9] or Fe(III) and single oxygen donors such as iodosyl benzene or oxone.[10] A proposed mechanism for the generation of "activated bleomycin" is shown in Figure 1.[9] Upon addition of oxygen to Fe(II)•BLM, a ternary complex is formed which must undergo one electron reduction in order to form "activated bleomycin," a complex distinguishable from its precursors by the appearance of an EPR spectrum.[9] If no external reductant has been added, Fe(II) or Fe(II)•BLM can function as an electron source.[11,12] "Activated bleomycin" may also be formed from Fe(III) and H_2O_2 in the absence of O_2. Although the structure of "activated bleomycin" remains controversial, it is known to have two more oxidizing equivalents than Fe(III)•BLM,[13] and the similarity between the spectral data for BLM and the iron-porphyrins has led some investigators to propose the intermediacy of a high-valent iron-oxo species in the degradation of DNA by BLM.[10]

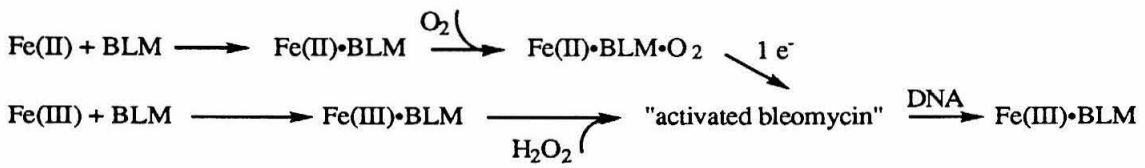


Figure 2. Proposed mechanism for the formation of "activated bleomycin." Adapted from reference 9.

The activated bleomycin complex initiates DNA cleavage by abstraction of the C-4'-hydrogen of the deoxyribose backbone, as shown by primary tritium and deuterium isotope effects at the 4'-position.[7,8,14] This leads to the release of two classes of products, free bases and base propenals.[2-4] The

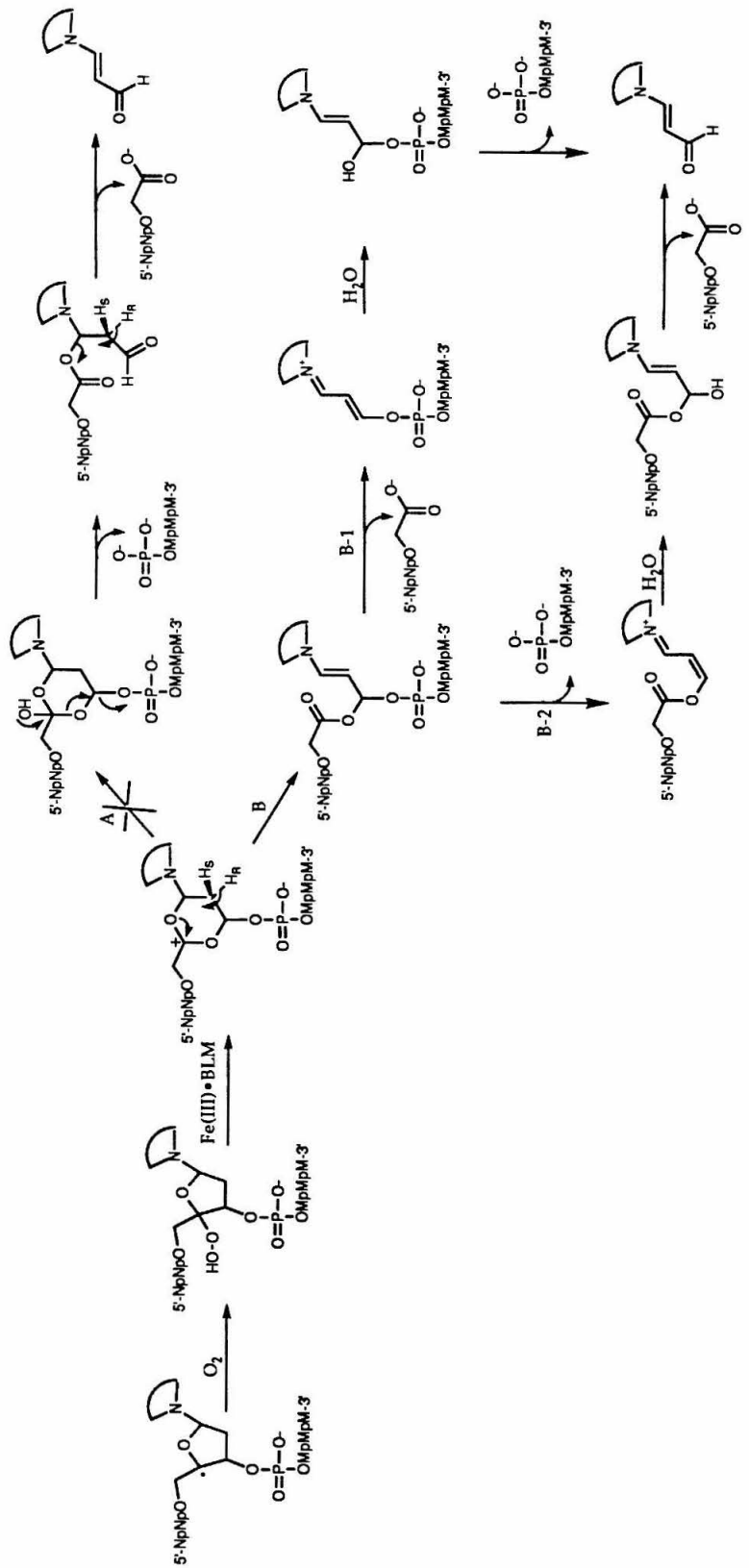


Fig. 3. Bleomycin-catalyzed oxygen-dependent degradation of the DNA deoxyribose backbone.

ratio of these two sets of products is dependent on oxygen concentration.[2-4]

The release of base propenals requires an additional equivalent of oxygen; a proposed mechanism for this oxygen-dependent production of base propenals from the 4'-deoxyribose radical intermediate is shown in Figure 3. The additional equivalent of O_2 leads to a 4'-hydroperoxide species which is thought to undergo a Crieggee rearrangement in which the Fe•BLM complex acts as a Lewis acid catalyst.[15] It was originally proposed that the resulting carbocation was trapped by H_2O or by the BLM•FeOH species resulting from the Fe•BLM catalysis of the Creigee rearrangement (Figure 3, pathway A).[2,16] However, this pathway is inconsistent with more recent ^{18}O labeling studies.[17] The oxygens in the 3'-phosphoglycolate moiety are derived from O_2 and from the O4′ of the original deoxyribose species. In addition, the oxygen of the base propenal is derived exclusively from water. A mechanism consistent with the results has been proposed (Fig. 3, pathway B).[17] In this model, abstraction of the 2'-*pro-R*-proton[17] with concomitant anti-elimination of O4'-occurs immediately following the Criegee rearrangement. The order in which the 5'-phosphate and 3'-phosphoglycolate moieties are eliminated is unclear (Fig. 3, pathways B-1 and B-2). However, in both models, the first elimination reaction is base-assisted and yields a conjugated species that can undergo Michael addition of water to form a reactive hemiacetal. This species then undergoes decomposition to the base propenal and 5'-phosphate or 3'-phosphoglycolate. No further studies to test this model have been reported.

The pathway for the formation of free base and an alkali-labile lesion has also been investigated extensively; a proposed mechanism is shown in Figure 4. The 4'-hydroxylated intermediate[18] is thought to form by electron transfer to the 4'-position followed by trapping of the carbocation with sol-

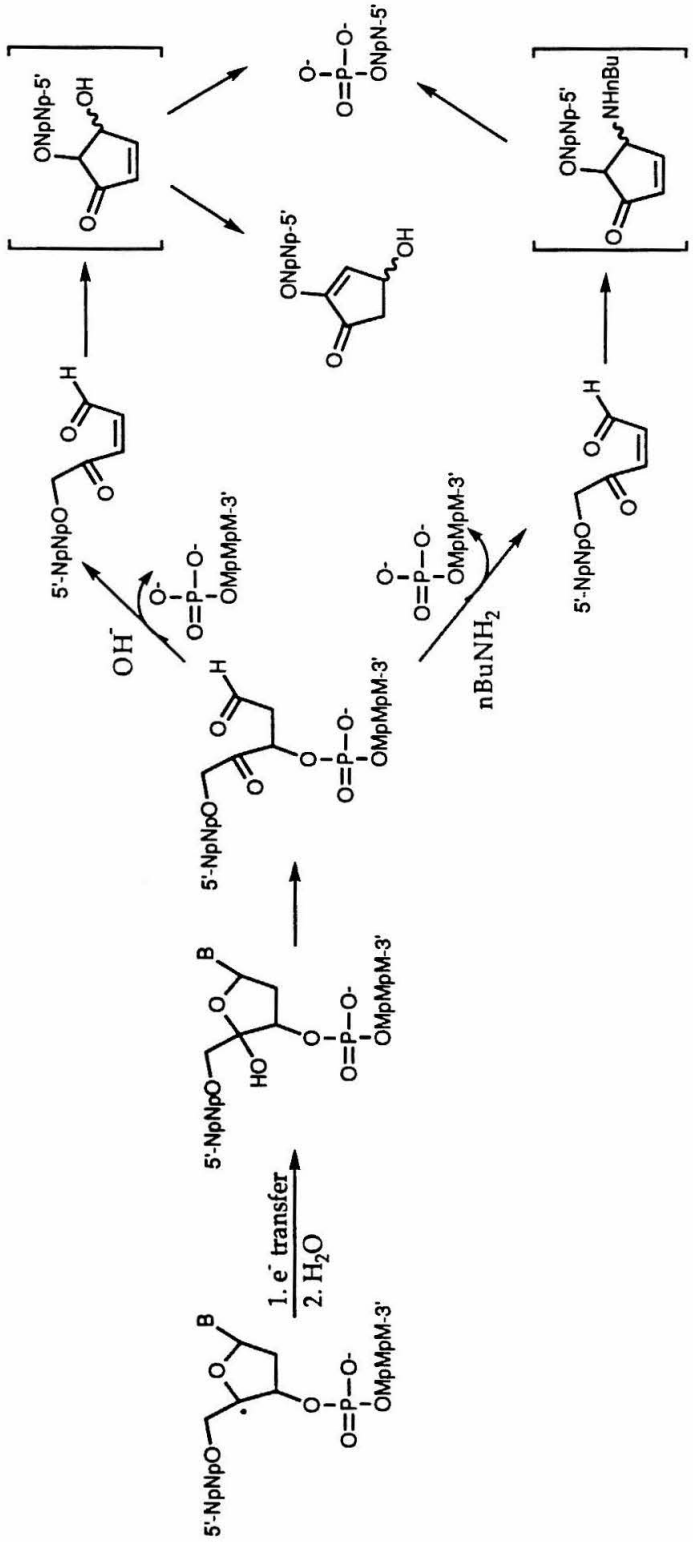


Figure 4. Bleomycin-catalyzed oxygen-independent degradation of the DNA deoxyribose backbone.

vent.[19] The alternative proposal of a "radical rebound" mechanism in which bleomycin catalyzes a recombination between the carbon-centered radical and hydroxyl radical is inconsistent with oxygen labeling experiments.[19,20] This intermediate then decomposes, leading to the release of free base and to an alkali-labile species that has been identified by trapping with sodium borohydride.[21] Upon treatment with base, this lesion is degraded further, resulting in the release of 5'-phosphate and the two 3'-products, phosphate and the cyclopentenone product shown in Figure 4. The identity of these products has been confirmed by HPLC analysis against authentic standards.[22] When butylamine is used instead of hydroxide in the base treatment, 3'- and 5'-phosphates are produced exclusively.[22]

Bleomycin Structure. Bleomycin has traditionally been divided into several structural and functional domains (Fig. 1).[2-4] The pyrimidine, β-aminoalanine, and β-hydroxyhistidine moieties are necessary for iron binding and O_2 activation.[23-26] Although the carbamoyl disaccharide moiety increases the efficiency of DNA cleavage, deglycobleomycin cleaves DNA with similar sequence specificity to that of bleomycin and produces the same DNA degradation products.[24,27-29] A polypeptide linker region separates the iron-binding domain from the bithiazole moiety and the positively charged tail, which are necessary for sequence-specific recognition of DNA.[26,30-32] That bleomycin models and analogs, some of which are deficient in sequence-specific DNA-cleaving activity, are capable of oxygen activation[25,33,34] suggests the possibility of constructing novel hybrid sequence-specific DNA-cleaving molecules in which the iron-binding domain of bleomycin is attached to a heterologous binding domain.

Bifunctional Design.

In an effort to harness the DNA cleaving activity of bleomycin to DNA-binding molecules with greater sequence selectivity, Zarytova and colleagues have devised a method to couple intact Bleomycin A5 to oligonucleotides.[35,36] These oligonucleotide-BLM conjugates cleave complementary strands of DNA through a non-diffusible species with a specificity reflective of both the oligonucleotide and bleomycin moieties of the conjugate. In a complementary approach, Ohno and co-workers have synthesized a novel DNA-cleaving molecule in which a model bleomycin ligand, termed PYML-6, is attached through an appropriate linker to distamycin.[37] The resulting ligand displays distamycin-like specificity, binding to and cleaving DNA at A,T-rich sites. However, instead of the single-nucleotide cleavage patterns seen with bleomycin, the cleavage patterns obtained are broad and reminiscent of the cleavage patterns produced by Fe•EDTA-distamycin, which cleaves DNA through a diffusible species, presumably hydroxyl radical.[38-40] Although these studies constitute promising leads for the design of novel bifunctional molecules, it remains to be shown that the iron-binding domain is capable of mechanistically similar cleavage to that effected by bleomycin when it is attached to a heterologous DNA binding domain.

DNA-Binding Domain of Hin Recombinase.

Hin Recombinase is a 190 amino acid enzyme that inverts a segment of DNA that controls the expression of the flagellin genes in *Salmonella typhimurium* .[41] Hin binds as a dimer to a number of similar DNA sites with the consensus half-site sequence 5'-TTNTCNNAAACCA-3'.[41,42] On the basis of sequence similarity with repressor proteins, Hin is thought to recognize DNA through a helix-turn-helix motif .[43] A synthetic 52-amino

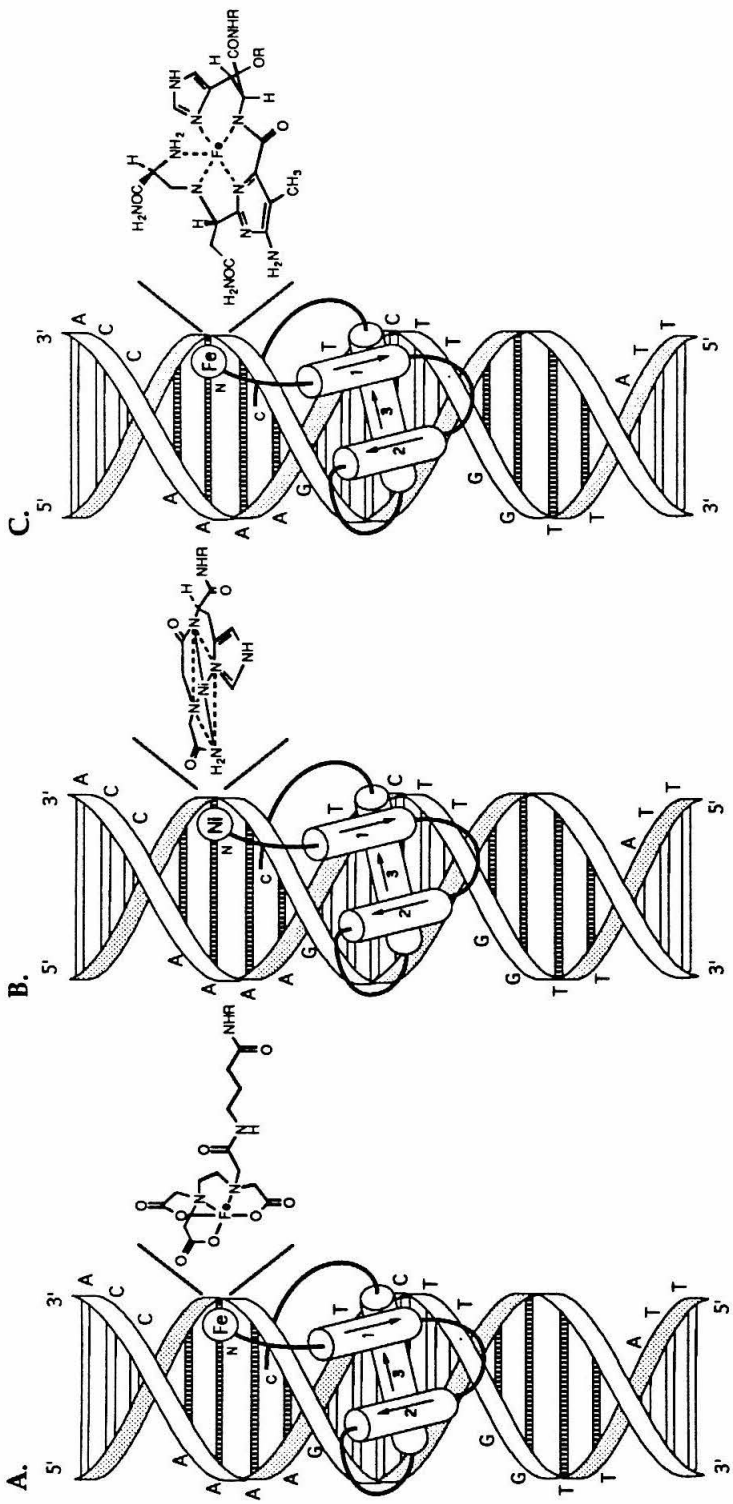


Figure 5. Schematic representations of models for affinity cleaving proteins containing the Hin recombinase DNA-binding domain bound to one *hix*L half-site (IRR). The DNA sequence is shown along the 5' strand. Putative α-helices are shown as cylinders with an arrow pointing from the NH2 to the CO2H terminus. **(A)** Fe(II)•EDTA-GABA-Hin(139-190). **(B)** Ni(II)•GGH-Hin(139-190). **(C)** Fe(II)•PBA–β-OH-His-Hin(139-190).

acid peptide identical to the CO_2H-terminus of Hin (residues 139-190) and containing the putative helix-turn-helix region has been shown to constitute the sequence specific DNA-binding activity of Hin.[44] Attachment of EDTA•Fe to the DNA-binding peptide Hin(139-190) affords a DNA-cleaving peptide that mediates oxidative degradation of the DNA at Hin binding sites via a diffusible oxidant.[45-47] Affinity cleaving studies using Hin(139-190) equipped with the EDTA•Fe at the NH_2 terminus have revealed that the NH_2 terminus of Hin(139-190) lies proximal to the minor groove near the symmetry axis of Hin recombination sites.[47] Affinity cleaving studies with Fe•EDTA attached near the CO_2H terminus of Hin(139-190) have demonstrated that the recognition helix of the helix-turn-helix element is oriented toward the symmetry axis of the DNA binding site.[48] A binding model put forward for Hin(139-190) includes a helix-turn-helix-turn-helix structure binding in the major groove with residues at the NH_2 terminus extending across the phosphodiester backbone making specific contacts in the minor groove (Fig. 5A).[47-49] The importance of these additional NH_2-terminal minor groove contacts has been demonstrated through methylation interference assays,[42] mutation of Hin recombination sites,[50] and affinity cleaving studies with proteins truncated at the NH_2-terminus.[49]

Ni(II)•GGH-Hin(130-190).

A second type of hybrid protein, consisting of all naturally-occurring α-amino acids has been synthesized. This protein, GGH-Hin(139-190), contains the tripeptide consensus sequence for the copper-binding domain of serum albumin, Gly-Gly-His, attached to the NH_2 terminus of the DNA-binding domain of Hin recombinase.[51] GGH-Hin(139-190) cleaves DNA at Hin binding sites in the presence of Cu(II), hydrogen peroxide, and sodium

ascorbate[51] or in the presence of Ni(II) and monoperoxyphthalic acid (Fig. 5B).[52] Cleavage by Ni•GGH-Hin(139-190) is rapid and efficient and occurs predominantly at a single deoxyribose position at each Hin binding site, indicating cleavage by a highly localized, nondiffusible oxidizing species.[52,53] A significant deuterium isotope effect at H4' has been measured for DNA cleavage by Ni•GGH-Hin(139-190)/monoperoxyphthalic acid, implying that oxidative DNA degradation by this protein results from hydrogen abstraction at the 4'-position of the deoxyribose backbone.[54] Thus, the DNA-binding domain of Hin recombinase can support DNA cleavage in the adjacent minor groove by a reactive metal complex that reacts with DNA through a non-diffusible oxidant.

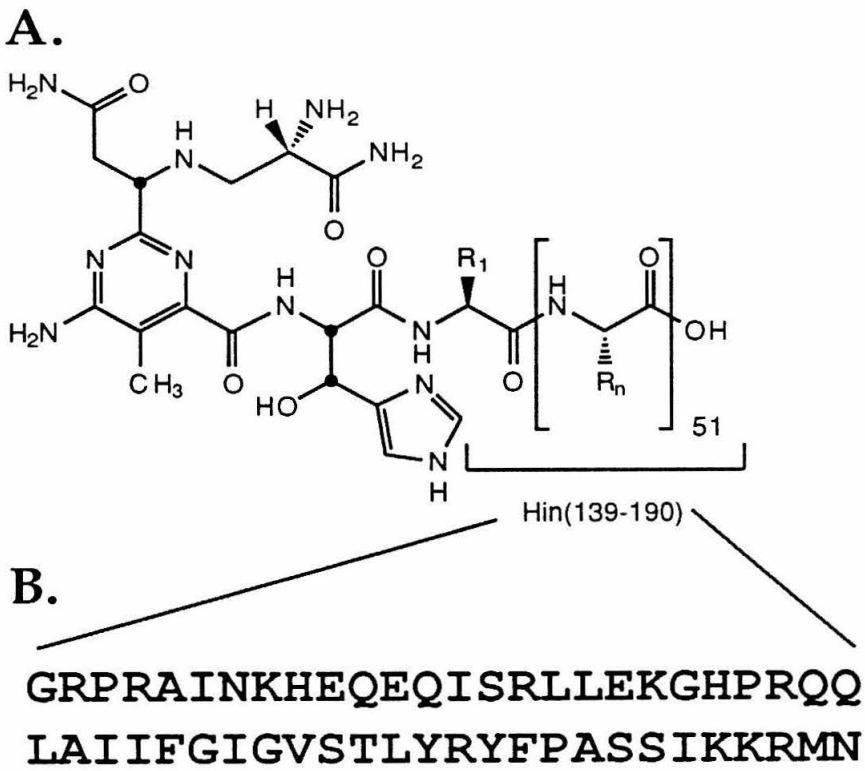


Fig. 6. **(A)** Synthetic protein PBA-β-OH-His-Hin(139-190). **(B)** Sequence of the CO_2H-terminal 52 amino acids(139-190) of Hin recombinase. Sequence is shown from the NH_2 terminus of the peptide to its CO_2H terminus.

Strategy for the Synthesis of the Bleomyin-Hin Conjugate Protein.

This chapter describes the synthesis and DNA-binding and -cleaving properties of a 54-residue protein hybrid protein, PBA-β-OH-His-Hin(139-190) (Fig. 5C,6). This protein comprises the DNA-binding domain of Hin recombinase, residues 139-190, and the putative iron-binding and oxygen-activating domain of bleomycin, pyrimidoblamic acid-erythro-β-hydroxy-L-histidine (PBA-β-OH-His). The chemical synthesis of Hin(139-190) by solid phase methods is well known.[44,47-49,55] Syntheses of pyrimidoblamic acid[56-60] and erythro-β-hydroxy-L-histidine[61-64] have been reported by several groups. Because these synthetic efforts were part of a larger program toward the total synthesis of bleomycin[65,66] or deglycobleomycin,[58] PBA and β-OH-His were synthesized in most cases in forms suitably protected for peptide synthesis.

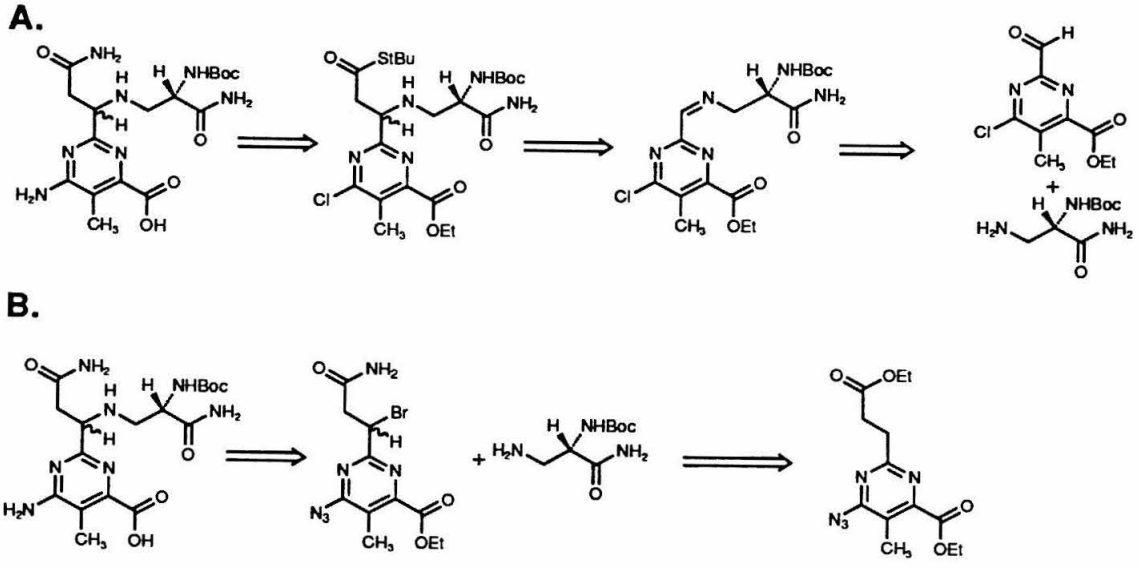


Figure 7. Retrosynthetic strategies for the synthesis of Boc-pyrimidoblamic acid.

Boc-pyrimidoblamic acid. Early chemical studies demonstrated that acylation of pyrimidoblamic acid occurs readily only at the primary amine of the β-aminoalanine moiety. Retrosynthetic analyses for two approaches to Boc-pyrimidoblamic acid are shown in Figure 7. In one approach (Fig. 7A), developed by Ohno and co-workers,[59] an imine was formed from a pyrimidine aldehyde and from a Boc-protected-amino-alanine-amide derivative. The acetamido side chain was then introduced via the addition of the dibutyl boronyl enolate derived from *tert*-butyl thioacetate, leading to a 1:1 mixture of diastereomers. Hecht and co-workers, on the other hand, chose to construct a bromopyrimidine moiety with the acetamido side chain intact (Fig. 7B).[56] The bromopyrimidine moiety was then treated with the Boc-protected-amino-alanine-amide to provide the nucleus of Boc-pyrimidoblamic acid as a 1:1 mixture of diasteromers.

More recent work by Boger and co-workers[58] has applied the general approach of the Ohno group with two major modifications: a diastereoselective introduction of the acetamido side chain was achieved using a stannous enolate of a chiral oxazolidinone species[69] (Fig. 8A), and the pyrimidine moiety was generated from the inverse-electron demand Diels-Alder reaction of 2,4,6-tris(ethoxycarbonyl)-1,3,5-triazine with propionamidine hydrochloride (Fig. 8B).[70]

erythro–β-Hydroxy-L-Histidine. A diastereoselective approach to the synthesis of β-OH-His has also been reported.[63] In this synthesis, the key step is the addition of the dibutylboronyl enolate of a chiral oxazolidinone to a protected imidazole-4-carboxaldehyde derivative (Fig. 9). This strategy has been modified by Boger and co-workers[64] to afford a β-OH-His suitable for coupling with Boc-pyrimidoblamic acid and for solid phase peptide synthesis.

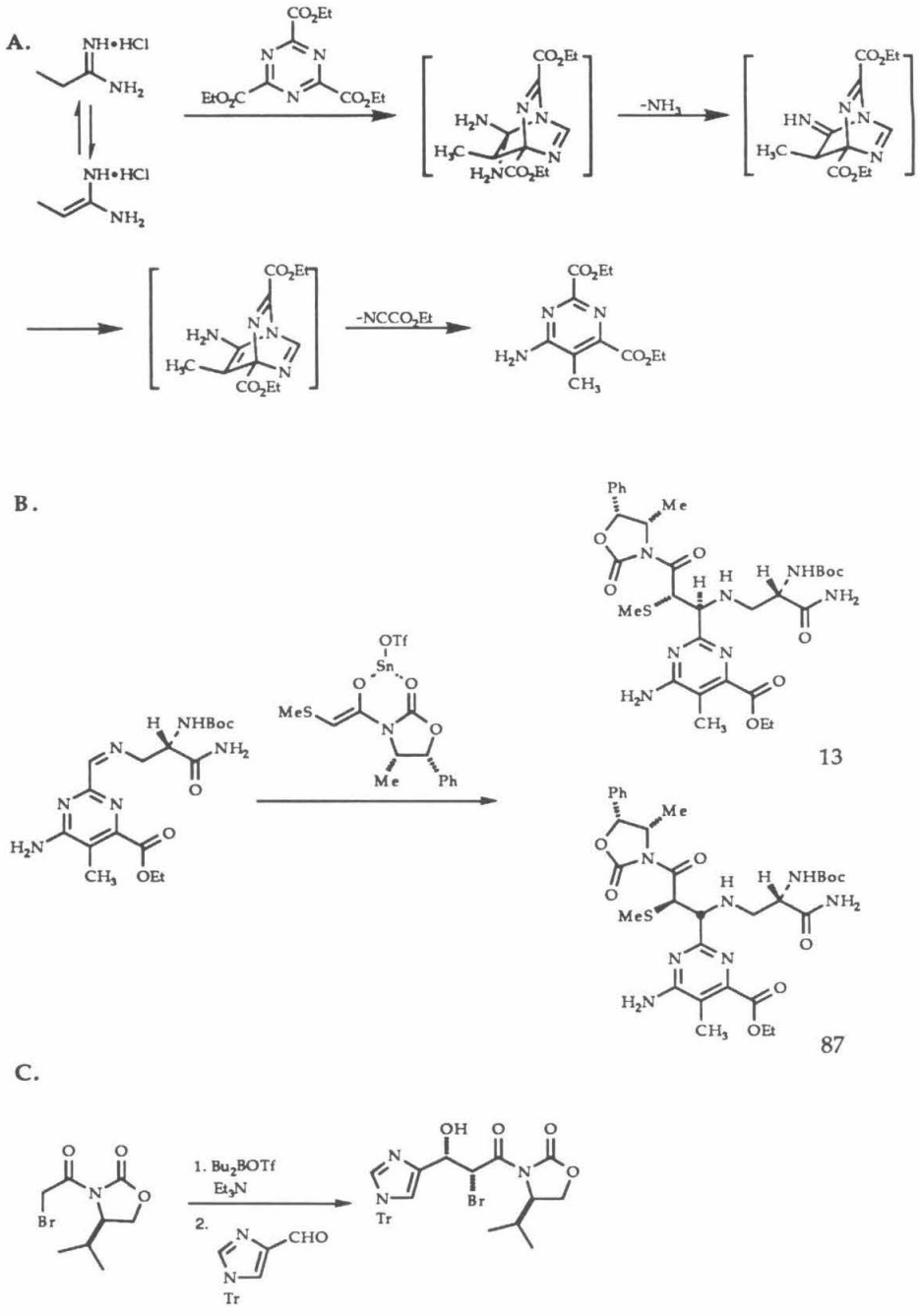


Figure 8. Key reactions in the synthesis of (**A, B**) Boc-pyrimidoblamic acid[52] and (**C**) β-OH-His.[57]

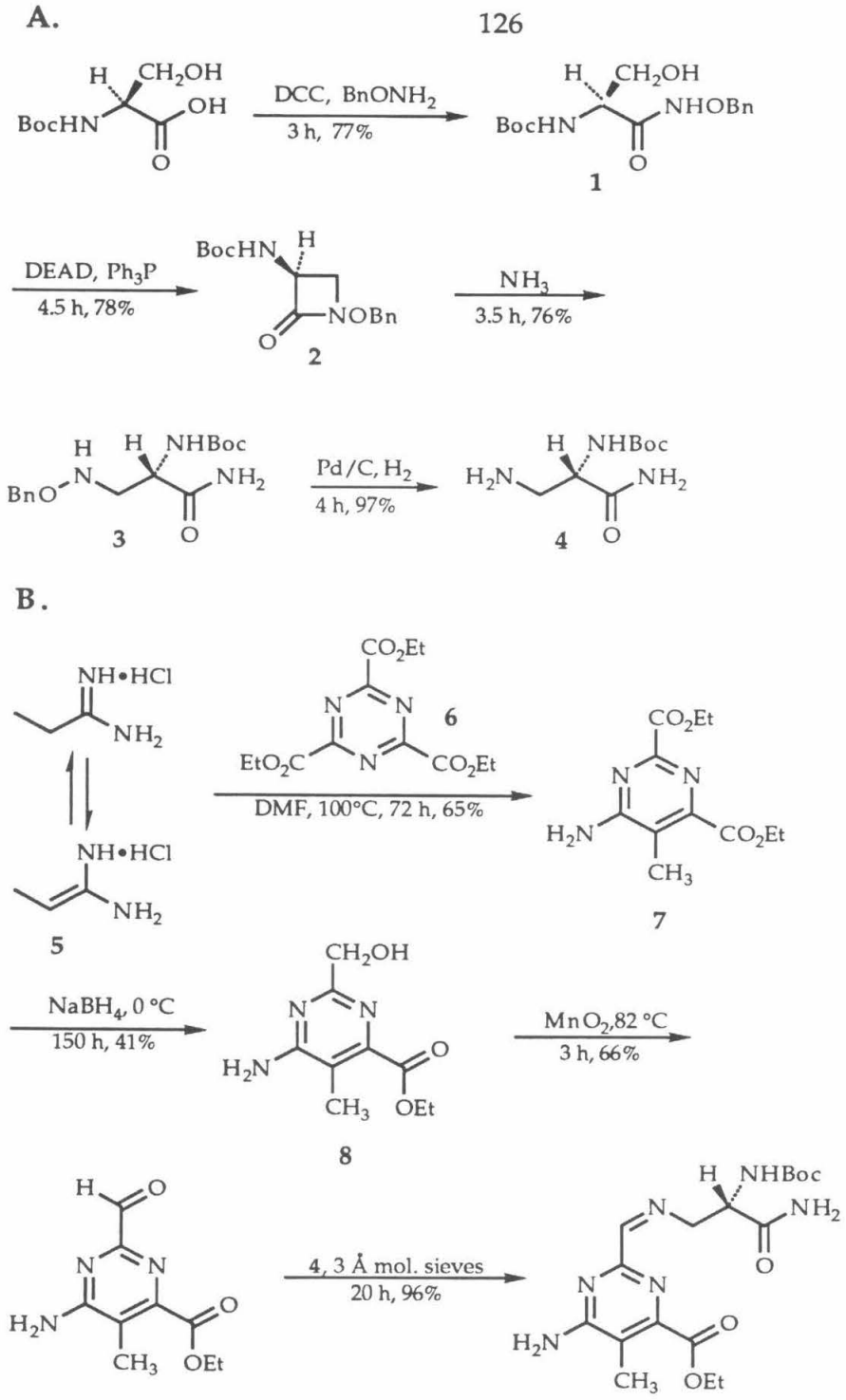


Figure 9. **(A)** Synthesis of the amine moiety **4**. **(B)** Synthesis of the pyrimidine moiety and the Schiff base **10**.

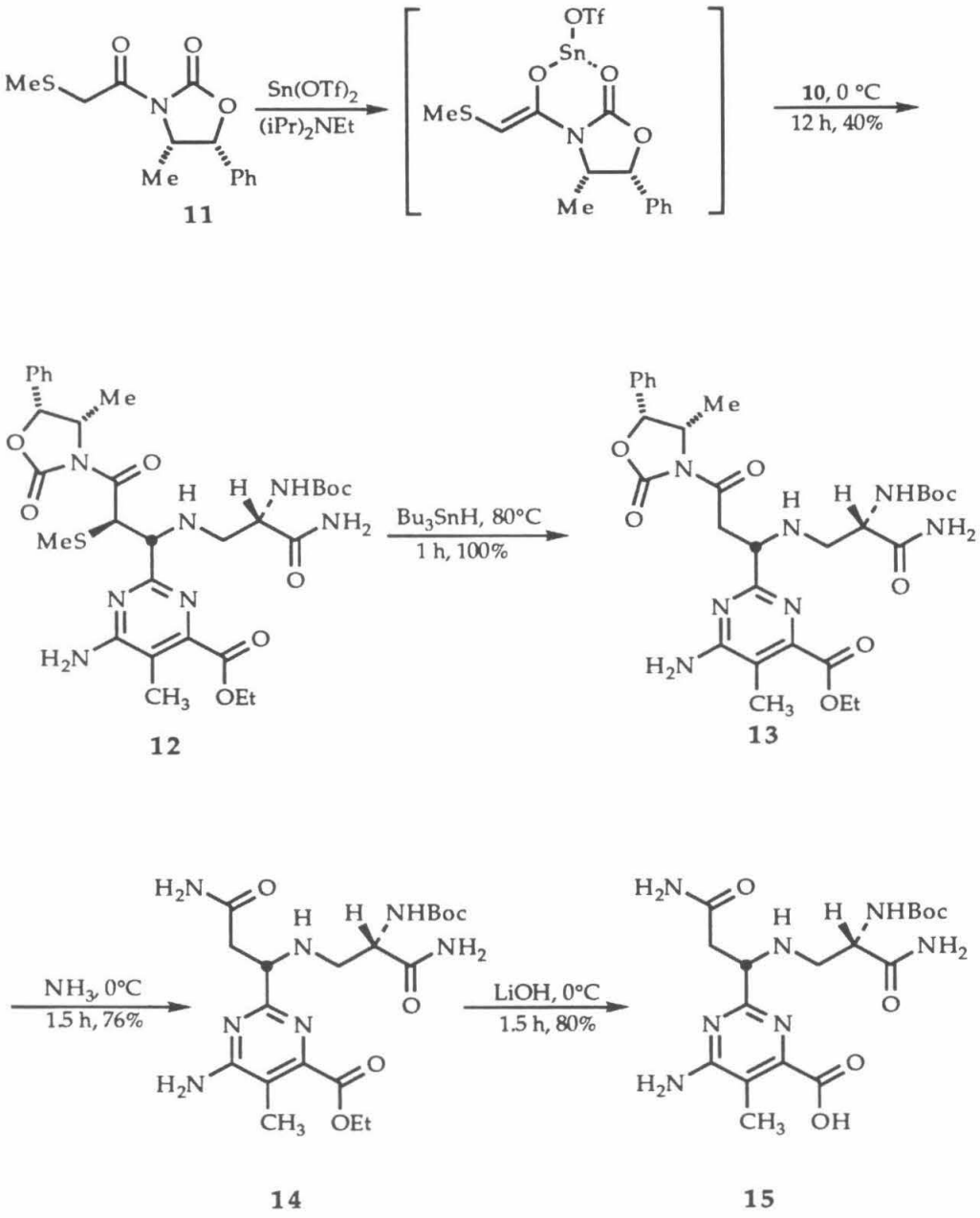


Figure 10. Diastereoselective enolate addition and subsequent functional group manipulation to yield Boc-pyrimidoblamic acid.

RESULTS

Synthesis of the Protected Bleomycin Ligand. Boc-pyrimidoblamic acid (**15**) was synthesized following the procedures reported by Boger and co-workers.[58] The synthetic scheme and the yields obtained for each reaction are shown in Figures 9 and 10. The protected β-OH-His derivative **16** was prepared, coupled to Boc-pyrimidoblamic acid, and deprotected by Dr. Ken Turnbull according to the procedure of Boger and co-workers[58,64] to afford the PBA-β-OH-His derivative **18** (Fig. 11).

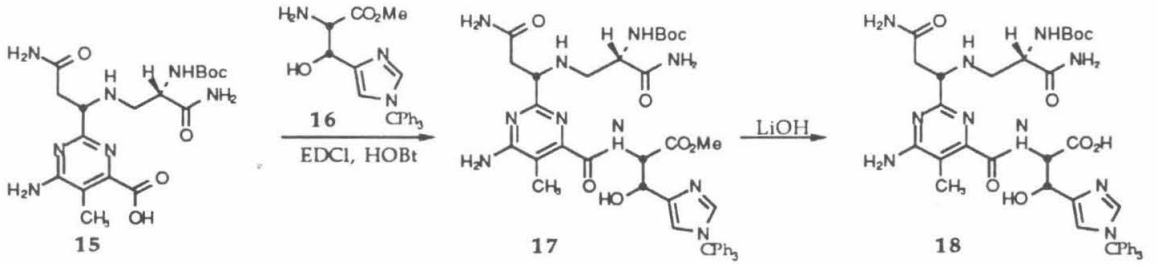


Figure 11. Synthesis of a suitably protected PBA-β-OH-His derivative for peptide synthesis.

Peptide Synthesis. **18** was coupled to the NH_2-terminus of resin-bound and protected Hin(139-190)[47-49,55] under standard coupling conditions to afford **19** in 88% yield (Fig. 12). The protein was deprotected with anhydrous HF and purified by preparative reverse-phase HPLC to afford PBA-β-OH-His-Hin(139-190) (Fig. 5C, 6). The synthesis of PBA-β-OH-His-Hin(139-190) was confirmed by laser desorption time-of-flight mass spectrometry. As a control to ensure that the PBA-β-OH-His dipeptide withstands HF treatment, **17**[58] was subjected to anhydrous HF under the conditions used for peptide deprotection to afford **20** in quantitative yield (Fig. 13).

DNA Binding and Cleaving. The DNA-binding and -cleaving properties of PBA-β-OH-His-Hin(139-190) were investigated by footprinting and affinity cleaving assays on a ^{32}P-end-labeled restriction fragment containing five 13

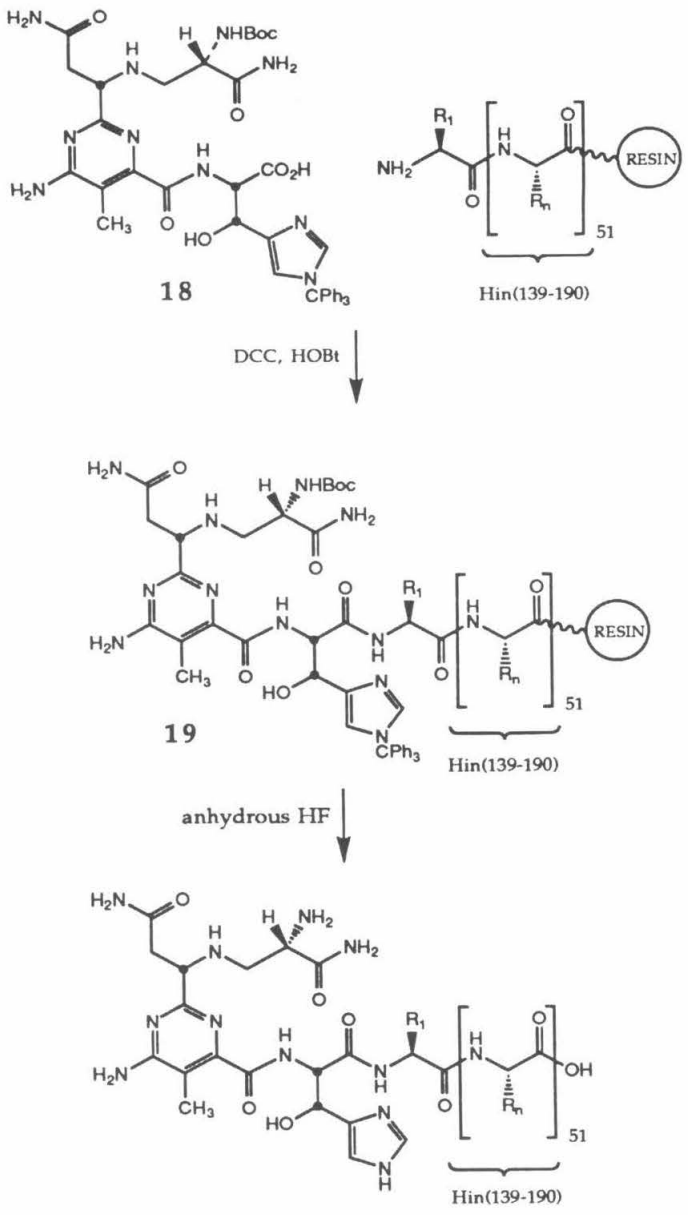


Figure 12. Synthetic scheme for the attachment of the protected PBA-β-OH-His derivative **18** to the NH2 terminus of resin-bound Hin(139-190) by solid phase methods and deprotection of the resulting protein.

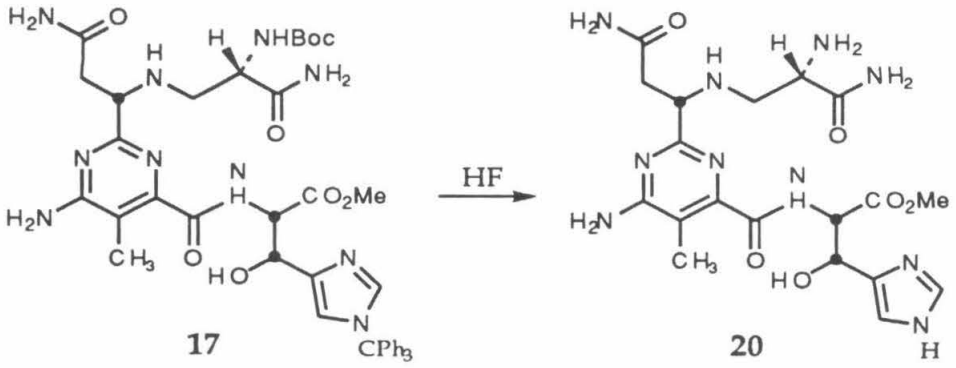


Figure 13. Scheme for the model HF deprotection of **17**.

bp binding sites for Hin recombinase: two full dimeric binding sites termed *hix*L and secondary, and a lower affinity site, tertiary (Fig. 14). The individual half sites are termed IRL and IRR, with the IRL sites on the left and IRR on the right side (Fig. 16; bottom and top, respectively, in Fig. 15) of the center of the full site. Deoxyribonuclease (DNase) I footprinting and affinity cleaving results are shown in Fig. 15, and histogram representations of the data are shown in Figure 16.

A.

major minor

5'-TTNTCNNAAACCA-3'

B.

Binding Site	Sequence
Consensus	5'-T**T**N**T**CNN**AA**ACCA-3'
*hix*L IRL	5'-T**T**C**T**TGA**AA**ACCA-3'
*hix*L IRR	5'-T**T**A**T**CAA**AA**ACCT-3'
secondary IRL	5'-T**C**T**T**CCT**TA**CTCG-3'
secondary IRR	5'-T**T**C**T**CCT**TT**ACAT-3'
tertiary	5'-T**TT**CTC**AT**GGAG-3'

Figure 14. A. Consensus Hin half-site showing positions contacted in the major and minor grooves.[42,50] **B.** Sequences of the Hin half-sites assayed. Bases implicated in specific contacts are printed in bold type.[50]

Figure 15. Autoradiogram of a high-resolution denaturing polyacrylamide gel showing DNase I footprinting reactions for Hin(139-190) and PBA-β-OH-His-Hin(139-190) and affinity cleaving reactions for Fe•PBA-β-OH-His-Hin(139-190), Fe•EDTA-Hin(139-190), Ni•GGH-Hin(139-190), and Fe•Bleomycin. Lanes 1, 3, 5-7, 11, 13, 15, and 17 contain 5'-^{32}P-end-labeled DNA; lanes 2, 4, 8-10, 12, 14, 16, and 18 contain 3'-^{32}P-end-labeled DNA. Lanes 1 and 2, intact DNA; lanes 3 and 4, Maxam-Gilbert G-specific sequencing reactions; lanes 5 and 8, DNase control lanes; lanes 6 and 9, DNase I cleavage protection by Hin(139-190) (5μM); lanes 7 and 10, DNase I cleavage protection by PBA-β-OH-His-Hin(139-190) (5μM); lanes 11 and 12, Fe•PBA-β-OH-His-Hin(139-190) at 10 μM; lanes 13 and 14, Fe•EDTA-Hin(139-190) at 5μM; lanes 15 and 16, Ni•GGH-Hin(139-190) at 5μM; lanes 17 and 18, Fe•Bleomycin at 2.5 μM.

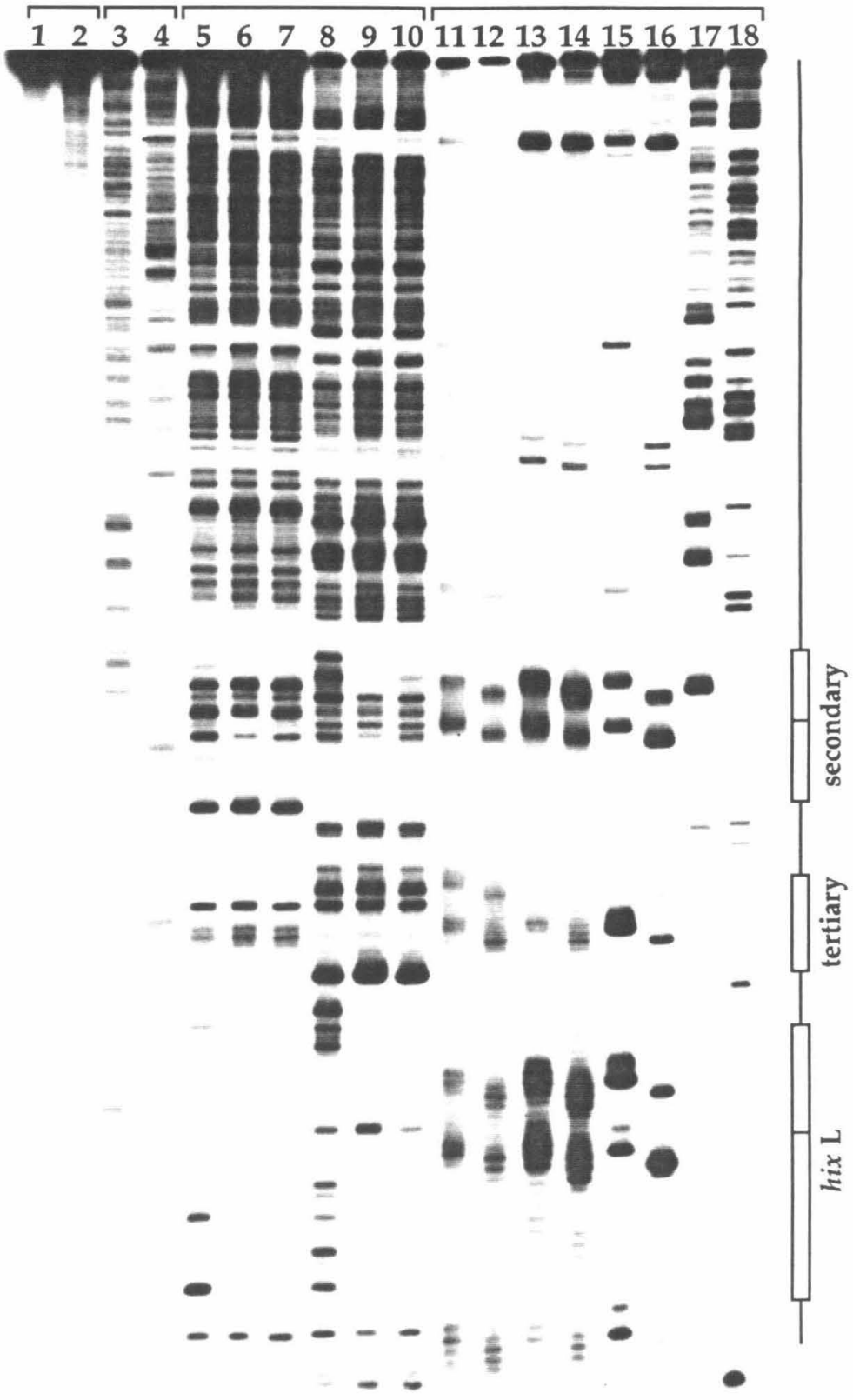
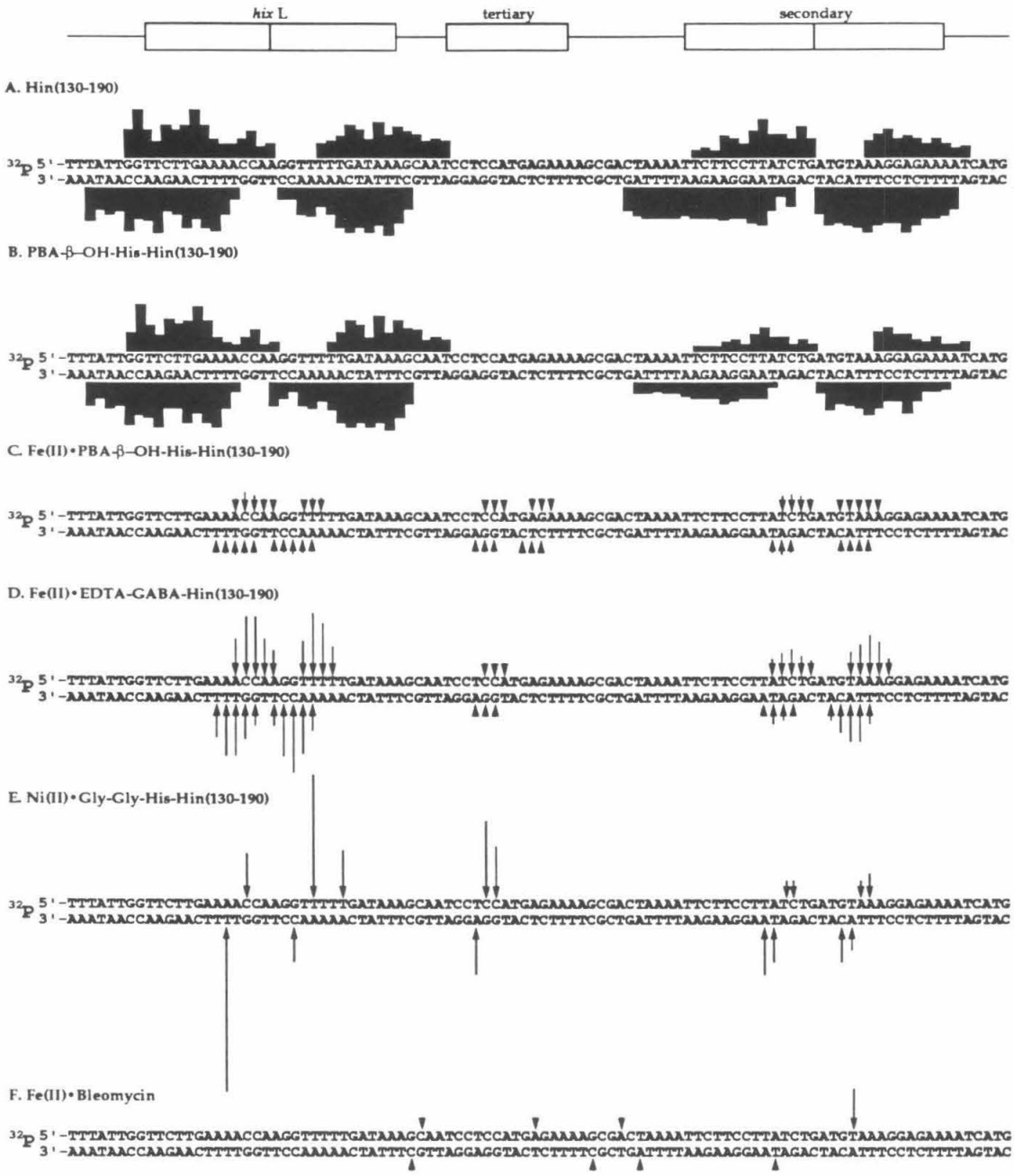

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
secondary
tertiary
hix L

Figure 16. The sequence from left to right represents the data from the bottom to the middle of the gel shown in Fig. 15. **(A)** Bars represent the extent of protection, derived from the ratio of cleavage at each band compared to the control lane, from DNase I cleavage in the presence of 5 μM Hin(139-190) (Fig. 6, lanes 6 and 9). **(B)** Bars represent the extent of protection from DNase I cleavage in the presence of 5 μM PBA-β-OH-His-Hin(139-190) (Fig. 6, lanes 7 and 10). **(C)** Arrows represent the extent of cleavage by 10 μM Fe•PBA-β-OH-His-Hin(139-190) (Fig. 6, lanes 11 and 12). **(D)** Arrows represent the extent of cleavage by 5 μM Fe•EDTA-Hin(139-190) (Fig. 6, lanes 13 and 14). **(E)** Arrows represent the extent of cleavage by 5 μM Ni•GGH-Hin(139-190) (Fig. 6, lanes 15 and 16). **(F)** Arrows represent the extent of cleavage by 2.5 μM Fe•Bleomycin (Fig. 6, lanes 17 and 18).


hix L
tertiary
secondary
A. Hin(130-190)
B. PBA-β-OH-His-Hin(130-190)
C. Fe(II)•PBA-β-OH-His-Hin(130-190)
D. Fe(II)•EDTA-GABA-Hin(130-190)
E. Ni(II)•Gly-Gly-His-Hin(130-190)
F. Fe(II)•Bleomycin
32P 5'-TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3'-AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC

In the presence of the reductant DTT, Fe•PBA-β-OH-His-Hin(139-190) at 10 μM cleaves the restriction fragment at every site cleaved by Fe•EDTA-Hin(139-190) at 5 μM, although with lower intensity (Fig. 15, lanes 11-14; Fig. 16C,D). Additional cleavage is not observed at Fe•Bleomycin sites, indicating that the DNA-binding and -cleaving specificity of Fe•PBA-β-OH-His-Hin(139-190) is largely determined by the Hin DNA-binding domain (Fig. 15, lanes 17, 18; Fig. 16F). Only one new set of cleavage bands is observed for Fe•PBA-β-OH-His-Hin(139-190), adjacent to the tertiary site. As is the case with Fe•EDTA-Hin(139-190), Fe•PBA-β-OH-His-Hin(139-190) cleavage occurs over several bases on each strand of DNA at each site. Densitometry traces comparing the cleavage patterns generated by Fe•EDTA-Hin(139-190) and Fe•PBA-β-OH-His-Hin(139-190) at *hixL* are shown in Fig. 17. Although cleavage by Fe•PBA-β-OH-His-Hin(139-190) is less efficient than that by Fe•EDTA-Hin(139-190), the shapes of the patterns are almost identical. This broad cleavage is in contrast to the cleavage observed with NiGGH-Hin(139-190), which cleaves each half site of *hix*L predominantly at one nucleotide on one strand (though cleavage at two nucleotide positions on each strand is seen at secondary, which is cleaved much less efficiently than *hix*L) (Fig. 15, lanes 15, 16; Fig. 16E). Cleavage by Fe•PBA-β-OH-His-Hin(139-190) is also much more diffuse than that by Fe•Bleomycin (2.5 μM), which occurs only at a single base on one strand of each site.

Qualitative DNase I footprinting reactions demonstrate that the reduced intensity of cleavage by Fe•PBA-β-OH-His-Hin(139-190) is not due to reduced binding affinity (Fig. 15, lanes 5-10). PBA-β-OH-His-Hin(139-190) protects both half sites of the *hix*L from cleavage by DNase I at approximately the same positions as Hin (139-190) and with roughly equal affinity at a concentration

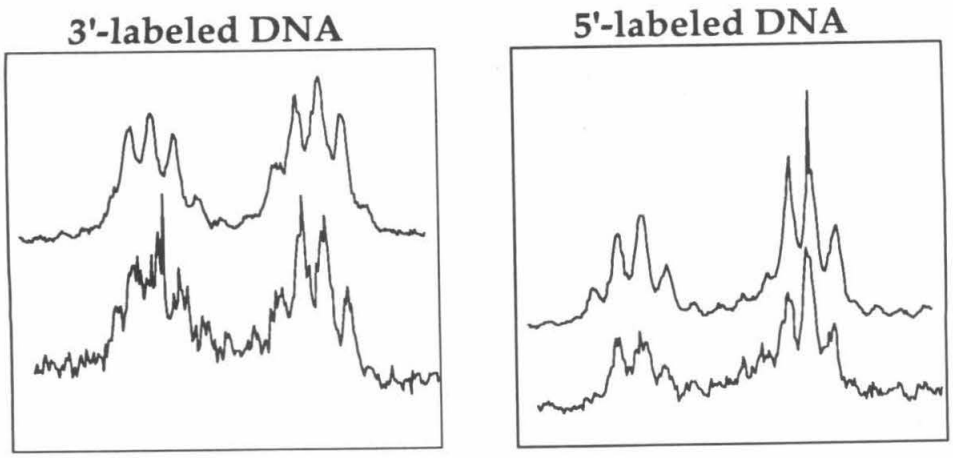


Figure 17. Densitometry traces of the cleavage by Fe•EDTA-Hin(139-190) (upper traces) and Fe•PBA-β-OH-His-Hin(139-190) (lower traces, expanded by a factor of 7) at the binding site *hix*L.

of 5 μM (Fig. 15, lanes 5-10; Fig. 16A,B). Protection at the secondary half sites appears to be reduced slightly relative to Hin(139-190); however, this reduction is insufficient to account for the greatly diminished cleavage intensity.

Time Course for the Reaction. The time course for the affinity cleaving reaction of Fe•PBA-β-OH-His-Hin(139-190) at 5 μM in the presence of DTT is shown in Figure 18. Cleavage intensity increased for several hours before leveling off after about six hours, in direct contrast to cleavage by bleomycin, which is complete within 30 minutes, the first time at which a measurement was taken. The Fe•PBA-β-OH-His-Hin(139-190) cleavage reaction is also much slower than that by Fe•EDTA-Hin(139-190), which has a half life of about ten minutes and levels off after an hour.[71]

Quantitative DNase I Footprint Titrations. To examine more fully the affinity of PBA-β-OH-His-Hin(139-190) for the various Hin binding sites and to determine whether metal complexation affects its binding to DNA, relative association constants for Fe•PBA-β-OH-His-Hin(139-190), PBA-β-OH-His-

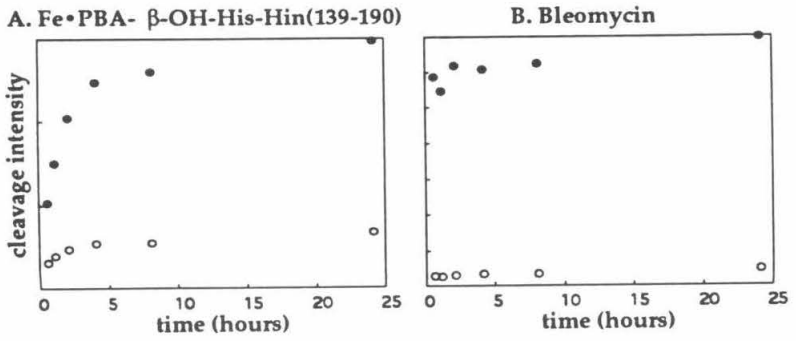


Figure 18. Time course for the Fe•PBA-β-OH-His-Hin(139-190) and Fe•Bleomycin cleavage reactions. Open circles represent background cleavage; filled circles represent specific cleavage.

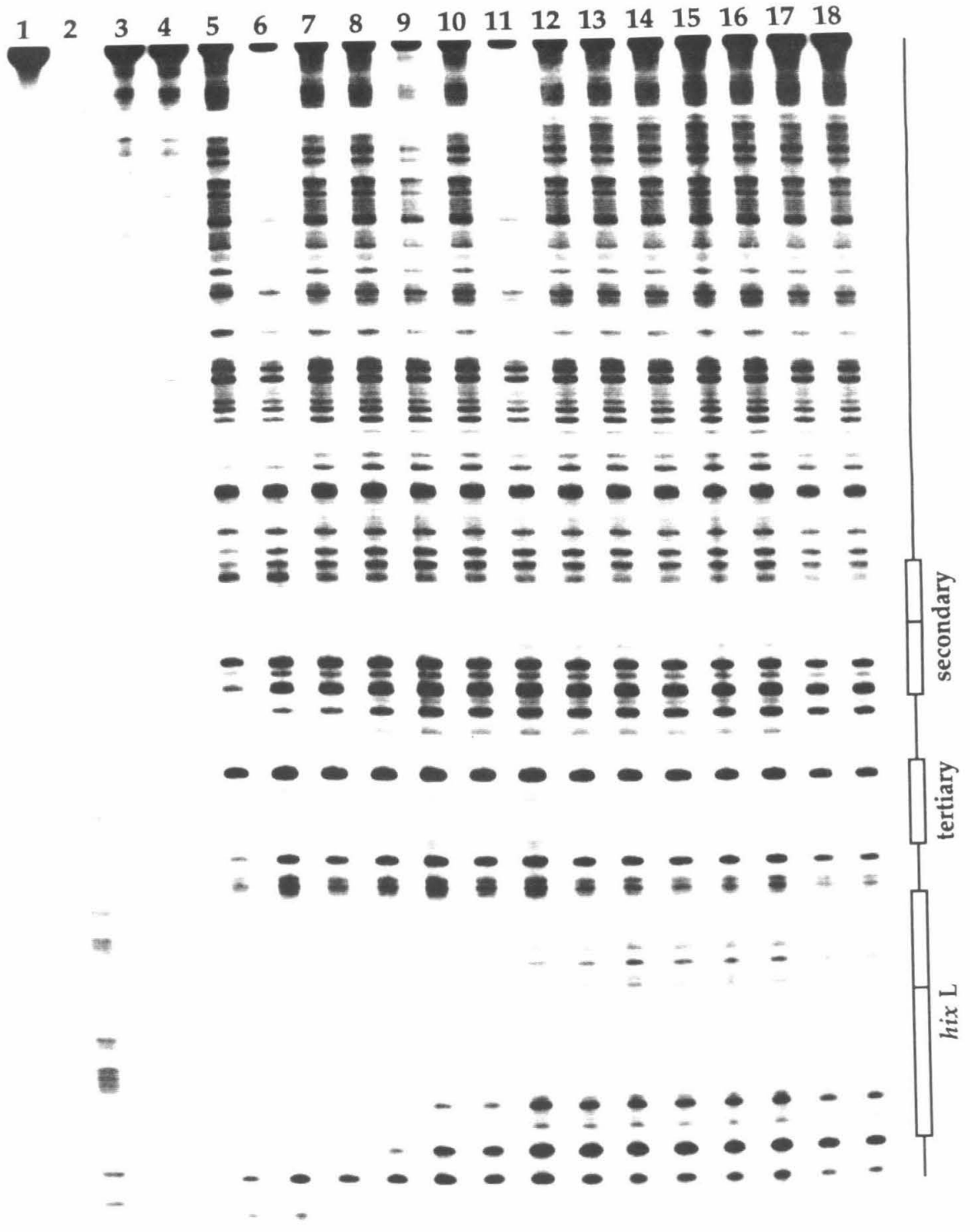
Hin(139-190), and Hin(139-190) were determined through quantitative DNase I footprint titration analyses (see experimental section).[72] A typical gel is shown in Figure 19. The relative association constants for each of the three proteins at four different Hin binding sites are shown in Table I; the corresponding binding isotherms for the *hix*L IRL site are shown in Figure 20. The

Table I. Binding constants for Fe•PBA-β-OH-His-Hin(139-190), PBA-β-OH-His-Hin(139-190), and Hin(139-190) at four DNA binding sites measured by quantitative DNase I footprint titrations.

Site	Protein	Relative K_a^{app} [a,b]
*hix*L IRL	Hin(139-190)	1.0 (0.3)
	PBA-β-OH-His-Hin(139-190)	1.1 (0.2)
	Fe•PBA-β-OH-His-Hin(139-190)	1.3 (0.4)
*hix*L IRR	Hin(139-190)	1.6 (0.5)
	PBA-β-OH-His-Hin(139-190)	1.9 (0.4)
	Fe•PBA-β-OH-His-Hin(139-190)	1.8 (0.4)
secondary IRL	Hin(139-190)	0.26 (0.04)
	PBA-β-OH-His-Hin(139-190)	0.11 (0.03)
	Fe•PBA-β-OH-His-Hin(139-190)	0.13 (0.03)
tertiary	Hin(139-190)	< 0.07
	PBA-β-OH-His-Hin(139-190)	<0.07
	Fe•PBA-β-OH-His-Hin(139-190)	<0.07

[a] Values reported are calculated from the average data from three or four footprint titration experiments. Relative values are based on the K_a^{app} for Hin(139-190) at the *hix*L IRL site. Numbers in parentheses indicate the standard deviation for the data set. [b] The assays were performed at room temperature, pH 7.6, in the presence of 20 mM bis-Tris-HCl, 20 mM sodium chloride, 10 mM magnesium chloride, and 100μM calf thymus DNA.

Figure 19. Autoradiogram of a high-resolution denaturing 6% polyacrylamide gel showing quantitative DNase I footprinting titration reactions for PBA-β-OH-His-Hin(139-190). All lanes contain 5'-^{32}P-end-labeled DNA. Lane 1, DNA; lane 2, A-specific sequencing reaction; lane 18, DNase control lane; lanes 3-17 show DNase I cleavage protection by PBA-β-OH-His-Hin(139-190) at concentrations of 30, 20, 10, 6, 3, 2, 1, 0.6, 0.3, 0.2, 0.1, 0.06, 0.03, 0.02, and 0.10 μM, respectively.


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
secondary
tertiary
hix L

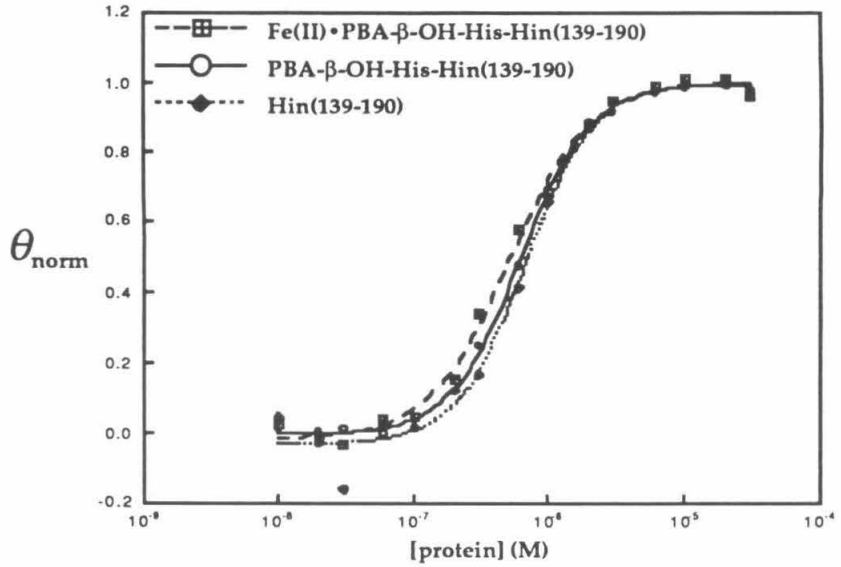
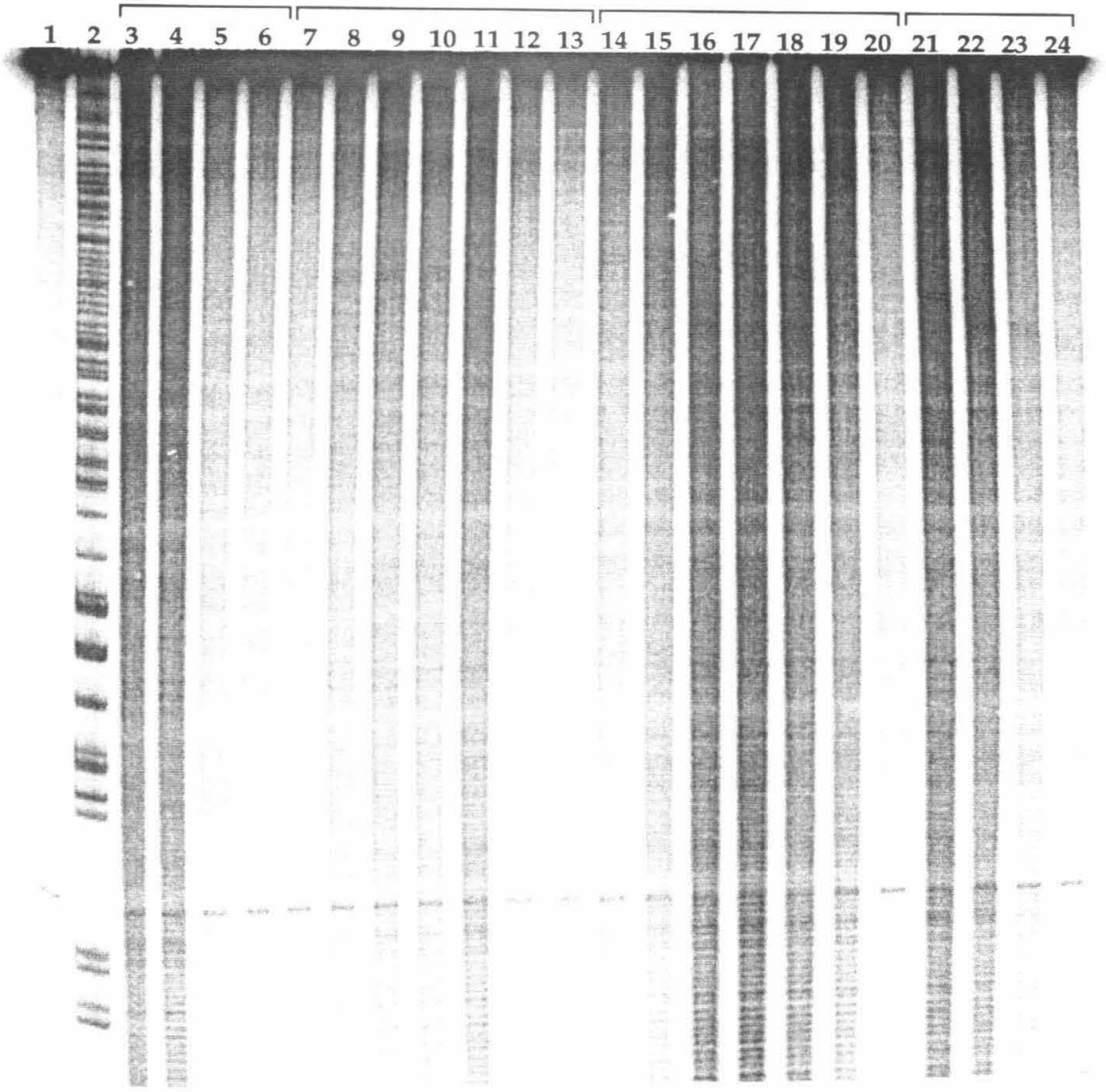


Figure 20. Langmuir binding isotherms derived from quantitative DNase I footprint titration experiments for Hin(139-190), PBA-β-OH-His-Hin(139-190), and Fe•PBA-β-OH-His-Hin(139-190) at *hix*L IRL. The θ_{norm} values were obtained using photostimulable storage phosphor autoradiography as described in the Experimental Section.

binding constants for all three proteins are within experimental error of one another in all cases except for secondary IRL, at which Hin(139-190) appears to display slightly stronger affinity than the other two proteins.

*Cleavage of DNA by Fe•**20*** A time course and concentration profile for the cleavage of DNA by Fe•**20** is shown in Figure 21 (lanes 7-13). Clearly, this complex cleaves non-specifically and only at high concentrations (> 100 μM). In fact, Fe•**20** cleaves no better than Fe(II) at the same concentrations (lanes 21-24) and slightly less efficiently than Fe•EDTA (lanes 14-20). As is the case with Fe•PBA-β-OH-His-Hin(139-190), Fe•**20** cleaves more slowly than its EDTA counterpart. Cleavage by the free acid counterpart of Fe•**20** (prepared

Figure 21. Gray-scale representation of a storage phosphor autoradiogram of a high-resolution denaturing polyacrylamide gel showing DNA cleaving reactions for Fe•**20** (Fe•PBA-β-OH-His-OMe), its corresponding acid complex (Fe•PBA-β-OH-His-OH), Fe•EDTA and Fe. Unless otherwise noted, reactions were allowed to proceed for 6 h. All lanes contain 5'-^{32}P-end-labeled DNA. Lane 1, intact DNA; lane 2, Maxam-Gilbert G-specific sequencing reaction. Lanes 3–6, Fe•PBA-β-OH-His-OH at concentrations of 180, 90, 18, and 9 μM, respectively. Lanes 7-13, Fe•**20**; lanes 7-10 contain 180 μM complex reacting for 0.5, 1, 2, and 6 h, respectively; lanes 11-13 contain 90, 18, and 9 μM complex, respectively. Lanes 14-20, Fe•EDTA; lanes 14-17 contain 180 μM complex reacting for 0.5, 1, 2, and 6 h, respectively; lanes 18-20 contain 90, 18, and 9 μM complex, respectively. Lanes 21–24, Fe(II) alone at concentrations of 180, 90, 18, and 9 μM, respectively.


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

by Dr. Ken Turnbull) is also shown (lanes 1-4); it is indistinguishable from that by Fe•**20**.

End Product Analysis by Gel Electrophoresis. The identity of the DNA end products of cleavage by Fe•PBA-β-OH-His-Hin(139-190) and Fe•**20** was investigated by 20% polyacrylamide gel electrophoresis on a restriction fragment containing the Hin binding site secondary.[54] The 5'-end products generated by Fe•Bleomycin, MPE•Fe, Fe•PBA-β-OH-His-Hin(139-190) and Fe•**20** are shown in Figure 22. In each case, the 5'-end products, known in the case of MPE and bleomycin to be phosphate,[46] have the same electrophoretic mobility. Treatment of the end products with alkaline phosphatase results in a slowing of the electrophoretic mobility as would be expected from the formation of hydroxyl ends. Although the Fe•PBA-β-OH-His-Hin(139-190) reaction has not proceeded to completion, other gels have shown complete reaction, indicating that the major 5'-end product is phosphate. The 3'-end products generated by the same affinity cleaving agents are shown in Figure 23. MPE•Fe is known to produce a roughly equal mixture of 3'-phosphate and -phosphoglycolate.[46] The two products are not clearly resolved in this gel. However, treatment of the end products with polynucleotide kinase to remove the 3'-phosphate groups (but leaving 3'phosphoglycolate intact)[73] results in an upward electrophoretic shift of approximately half the material, consistent with an initial equal mixture of phosphate and phosphoglycolate end products. In contrast, no electrophoretic shift is seen when the bleomycin-derived end products are treated with kinase in the absence of base treatment, consistent with the exclusive production of 3'-phosphoglycolate ends (Figs. 3, 4). Treatment with butylamine followed by kinase leads to an electrophoretic shift, indicating that phosphate is produced by the base treatment, as expected (Fig. 4). For both Fe•PBA-β-OH-His-Hin(139-190) and

Figure 22. Gray-scale representation of a storage phosphor autoradiogram of a denaturing 20% polyacrylamide gel showing the 5'-end products of DNA cleavage by Fe(II)•PBA-β-OH-His-Hin(139-190) and Fe(II)•**20**. Even numbered lanes 4-10 show reactions treated with calf alkaline phosphatase to remove 3'-phosphate groups. Lane 1, DNA standard lane. Lane 2, A- specific reaction. Lanes 3 and 4, MPE•Fe reactions, ~13, 000 cpm DNA. Lanes 5 and 6, Fe•Bleomycin reactions, ~10,000 cpm DNA. Lanes 7 and 8, Fe(II)•PBA-β-OH-His-Hin(139-190) reactions, ~17,000 cpm DNA. Lanes 9 and 10, Fe(II)•**20** reactions, ~24,000 cpm DNA.

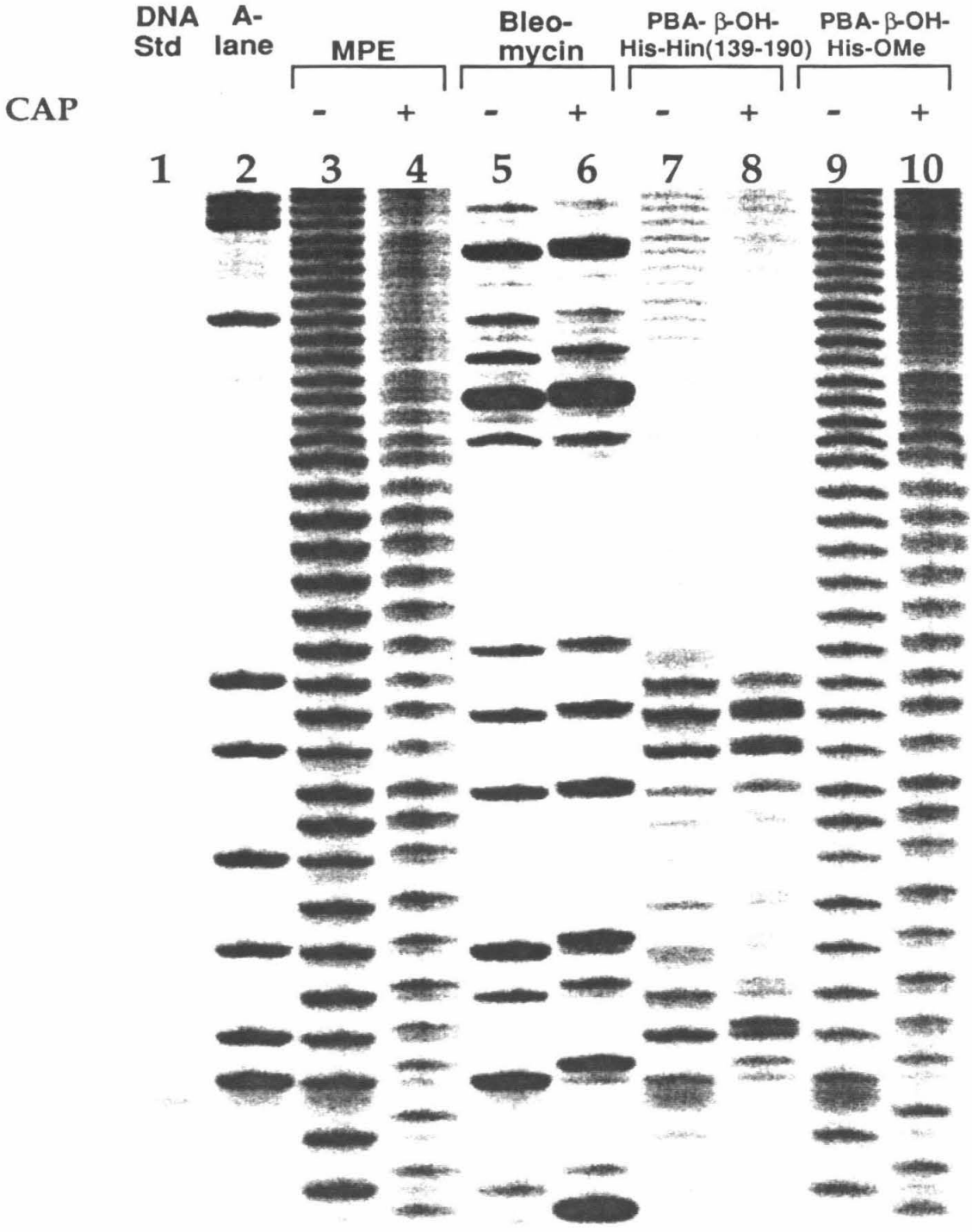
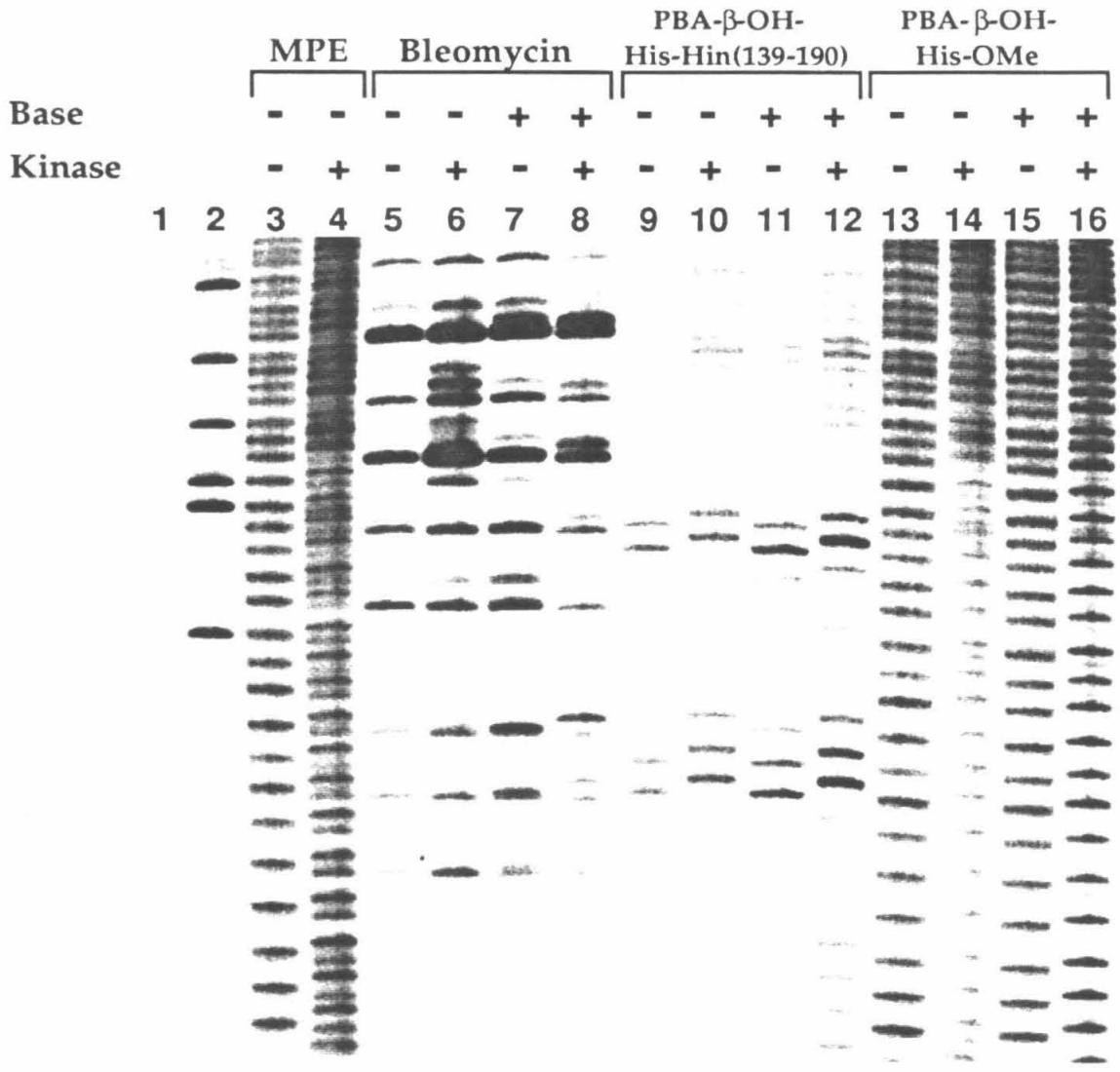

DNA Std
A-lane
MPE
Bleo-mycin
PBA-β-OH-His-Hin(139-190)
PBA-β-OH-His-OMe
CAP
- + - + - + - +
1 2 3 4 5 6 7 8 9 10

Figure 23. Gray-scale representation of a storage phosphor autoradiogram of a denaturing 20% polyacrylamide gel showing the 3'-end products of DNA cleavage by Fe(II)•PBA-β-OH-His-Hin(139-190) and Fe(II)•**20**. Even numbered lanes 4-16 show reactions treated with T4 polynucleotide kinase to remove 3' phosphate groups. Lane 1, DNA standard lane. Lane 2, Maxam-Gilbert G-reaction. Lanes 3 and 4, MPE•Fe reactions, ~9000 cpm DNA. Lanes 5-8, Fe•Bleomycin reactions, ~9000 cpm DNA. Lanes 7 and 8 are with base workup. Lanes 9-12, Fe(II)•PBA-β-OH-His-Hin(139-190) reactions, ~9000 cpm DNA. Lanes 11 and 12 are with base workup. Lanes 13-16, Fe(II)•**20** reactions, ~17,000 cpm DNA. Lanes 15 and 16 are with base workup.


MPE
Bleomycin
PBA-β-OH-His-Hin(139-190)
PBA-β-OH-His-OMe
Base
Kinase
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Fe•20, phosphate ends are the major product both in the presence and in the absence of base treatment.

DISCUSSION

The hybrid protein Fe•PBA-β-OH-His-Hin(139-190) containing the DNA-binding domain of Hin recombinase and the metal-binding and oxygen-activating domain of bleomycin binds to and cleaves DNA with the desired specificity. However, the cleavage pattern is much broader than that expected for cleavage by a nondiffusible oxidant such as that found in bleomycin. Formally, this data can be interpreted in two ways: cleavage could occur through a diffusible oxidizing species, such as hydroxyl radical, or a number of nucleotide positions could be accessible to a highly localized oxidant. The second explanation would require significant mobility of the NH_2-terminal segment of Fe•PBA-β-OH-His-Hin(139-190). However, the available evidence indicates that this segment is firmly anchored in the minor groove. The Ni•GGH-Hin(139-190) cleavage pattern, the result of cleavage by a localized oxidant, suggests that the NH_2-terminal segment is held in place in the minor groove.[52,53] Affinity cleaving studies have shown the NH_2-terminal Gly-Arg-Pro-Arg sequence to be necessary for DNA binding.[74,75] For example, deletion of the two NH_2-terminal residues, Gly-139 and Arg-140, leads to a dramatic reduction in DNA affinity and alteration of DNA-binding specificity.[49] No such change in binding is observed with Fe•PBA-β-OH-His-Hin(139-190), suggesting that the best explanation for the observed cleavage pattern is that the DNA is cleaved by a diffusible oxidant.

The DNA end products generated by Fe•PBA-β-OH-His-Hin(139-190) have been analyzed by gel electrophoresis. The major products are 5'- and 3'-phosphate, consistent with oxidative degradation of the deoxyribose

backbone. However, the appearance of 3'- phosphate termini in the absence of base treatment clearly demonstrates that DNA degradation by Fe•PBA-β-OH-His-Hin(139-190) does not proceed through the same pathway as bleomycin-induced degradation, for which no 3'-phosphate is observed in the absence of base treatment. It is well established that degradation of DNA by bleomycin leads to two sets of products: base propenals are released along with 5' phosphate and 3' phosphoglycolate termini; and free base release is accompanied by formation of an alkali-labile lesion[2-4,16,76,77]. Only upon treatment with n-butylamine does this lesion produce 3'- and 5'- phosphate termini.[22]

There are at least two possible explanations for cleavage by Fe•PBA-β-OH-His-Hin(139-190) through a diffusible oxidant rather than through the localized oxidant found in bleomycin. First, the PBA-β-OH-His ligand might be insufficient to form the iron-oxo complex found in "activated bleomycin." Although we cannot rule out this explanation, the data suggests that complexes even simpler than PBA-β-OH-His can activate oxygen. The spectral properties of such bleomycin analogs bound to Fe and oxygen analogs are very similar to those of bleomycin, suggesting a similar ternary complex is formed.[24] The activity of these analogs has also been confirmed by a spin trapping assay for the presence hydroxyl and superoxide radicals.[24] More compelling evidence for bleomycin-like oxygen activation is the demonstration that analogs of bleomycin can transfer oxygen to small organic substrates. For example, an iron complex consisting of the pyrimidoblamic acid, hydroxy-histidine and methylvalerate moieties has been shown to transfer oxygen to stilbene in the presence of iodosobenzene or oxygen and ascorbate.[25] Similarly, Hecht and co-workers have demonstrated that deglycobleomycin analogs in which one of the two thiaziole units has been

replaced with cysteine also give non-specific cleavage and produce 3'-phosphate termini in the absence of a base workup.[78] As the thiazole units are not likely to participate in iron binding and oxygen activation, this study demonstrates that bleomycin analogs with intact oxygen activation domains can display behavior similar to that seen for Fe•PBA-β-OH-His-Hin(139-190).

An alternative explanation for diffusible cleavage is that the appropriate activated iron-oxygen complex is formed but is not held in the appropriate position to permit reaction of the non-diffusible oxidant with the deoxyribose backbone. Breakdown of the complex could then produce a diffusible oxidant that could cleave DNA proximal to the position of the metal complex. In support of this, Fe•Bleomycin has been shown by spin-trapping experiments to generate hydroxyl radicals in the presence of oxygen.[79] Previous studies with nickel and copper complexes of GGH-Hin(139-190) illustrate the sensitivity of the cleavage reaction to small changes in the metal-ligand complex and the DNA structure. For example, although GGH-Hin(139-190) binds efficiently to both the *hix*L and secondary full sites, Cu•GGH-Hin(139-190) cleaves primarily at the secondary IRR site while Ni•GGH-Hin(139-190) cleaves much more efficiently at the *hix*L site.[51,52,79]

Because of the precise spatial requirements for DNA degradation by a nondiffusible oxidant, we chose a DNA-binding protein that has been shown to support oxidative DNA cleavage by a reactive metal complex through 4'-H abstraction. A comparison of the proposed structures of the metal complexes for bleomycin[2] and GGH[80-83] (Fig. 5B and 2C, respectively) shows that the putative bleomycin equatorial ligands occupy very similar positions in the primary amino acid sequence of Fe•PBA-β-OH-His-Hin(139-190) to those of the nickel ligands in Ni-GGH-Hin(139-190). However, the iron complex of the bleomycin ligand is considerably bulkier than the nickel complex of GGH,

perhaps resulting in its exclusion from the minor groove. Interestingly, the only Hin site at which cleavage by Fe•PBA-β-OH-His-Hin(139-190) is comparable to that of Fe•EDTA-Hin(139-190) is the tertiary site. At this site the typical A,T rich minor groove Hin-binding sequences are replaced with G,C base pairs,[84] leading to a wider minor groove and perhaps allowing the metal complex to position itself closer to the DNA backbone (Fig. 14).[42,50] This would explain the relatively efficient cleavage by Fe•PBA-β-OH-His-Hin(139-190) at this low affinity site. For Fe•EDTA-Hin(139-190), the smaller Fe•EDTA complex would allow the cleavage intensity to reflect its binding affinity more accurately.

Another possibility is that the metal binding ligand of bleomycin has sequence-specific requirments for DNA binding or cleaving that are not met by the five Hin binding sites assayed in this study. Several lines of evidence support the view that the iron-binding domain is a determinant of bleomycin sequence selectivity. A number of bleomycin analogs having identical CO_2H-termini but differing in the metal-binding domain exhibit modified sequence and strand selectivity.[12,29] Similarly, DNA unwinding by bleomycin is markedly reduced in the absence of a metal ion.[85] In addition, bleomycin analogs containing linkers of variable lengths (0 to 4 glycine residues) between the metal-binding ligand and the bithiazole moiety cleave DNA at the same nucleotides.[86] Finally, a recent report demonstrates that Fe(II) and Fe(II) complexes of a bleomycin model ligand cleave DNA with reasonable efficiency and with a cleavage specificity remarkably similar to that of bleomycin.[34]

There are very few good bleomycin cleaving sites in proximity to the Hin binding sites assayed in this work (Fig. 15, lanes 17-18; Fig 16A). The only exception occurs at secondary, where very efficient cleavage by Fe•bleomycin

is seen at a 5'-GT-3' sequence. However, cleavage by Fe•PBA-β-OH-His-Hin(139-190) is no more efficient at this site than at any of the others. This raises the possiblity that the NH_2-terminal segment of the protein is insufficiently flexible to allow the complex to position itself appropriately to react with the deoxyribose backbone at this site.

In light of these considerations, we suggest that future experiments on hybrid DNA-cleaving molecules containing the iron-binding domain of bleomycin might include a DNA binding site with bleomycin recognition sites proximal to the recognition sequence of the DNA-binding molecule. In addition, in the absence of detailed structural information about the bleomycin-DNA complex, it may be necessary to optimize a linker moiety by trial and error for a suitable orientation of the bleomycin-derived DNA-cleaving domain in the hybrid molecule.

EXPERIMENTAL SECTION

Materials.

Preparation of Boc-pyrimidoblamic acid. ^1H and ^{13}C NMR spectra were recorded on a General Electric QE-300 NMR. High resolution mass spectra for organic compounds were recorded using fast atom bombardment techniques at the Mass Spectrometry Facility of the University of California, Riverside. Infrared spectra were recorded on a Perkin-Elmer FTIR. Optical rotations were measured on a Jasco DIP-180 digital polarimeter. Flame- or oven-dried glassware was used for all reactions run under an argon atmosphere. Reagent grade chemicals were obtained from Aldrich and used as received unless otherwise noted. Tetrahydrofuran (THF) and diethyl ether (Et_2O) were distilled under nitrogen from sodium/benzophenone ketyl. Dichloromethane, diisopropylethyl amine, acetonitrile (ACN) and benzene were

distilled under nitrogen or argon from powdered calcium hydride. Dimethylformamide (DMF) was obtained as an anhydrous solvent from Aldrich. Tin(II) triflate was a generous gift from Rob Singer and Prof. Erick Carreria. Flash column chromatography was carried out using EM Science Kieselgel 60 (230-400 mesh) unless otherwise stated. Thin layer chromatography was performed on EM Reagents silica gel plates (0.5 mm thickness). Compounds were visualized with shortwave ultraviolet light or iodine. Preparative TLC was performed on E. Merck reverse-phase 18 plates (0.25 mm thick) with a 254 nm fluorescent indicator.

Protein Synthesis. Asn-(Phenylacetamido)methyl (PAM) resin, DMF, diisopropylethylamine (DIEA), dicyclohexylcarbodiimide (DCC) in dichloromethane, *N*-hydroxybenzotriazole (HOBt) in DMF, trifluoroacetic acid (TFA), and protected amino-acid derivatives were purchased from Applied Biosystems. Boc-L-His(DNP) was obtained from Fluka. Dichloromethane and methanol (HPLC grade) were obtained from Mallinckrodt; HOBt, *p* -cresol and *p* -thiocresol from Aldrich, and diethyl ether (low peroxide content) from Baker. Automated peptide synthesis was performed on an ABI 430A peptide synthesizer. Protein mass spectra were recorded at the Biotechnology Instrumentation Facility of the University of California, Riverside, on a laser desorption time-of-flight mass spectrometer. UV-vis spectra were recorded on a Hewlett-Packard Diode Array spectrophotometer. GGH-Hin(139-190) was prepared by David Mack.

DNA Reactions. Doubly distilled water was further purified through the Milli Q filtration system from Millipore. Sonicated, deproteinized calf thymus DNA was purchased from Pharmacia and dissolved in H_2O to a concentration of 2 mM in base pairs. Enzymes were obtained from Boehringer-Mannheim or New England Biolabs and used with the buffers

supplied unless otherwise noted. Glycogen was obtained from Boehringer-Mannheim. Deoxyadenosine- 5'- [α- ^{32}P]-triphosphate and adenosine- 5'-[γ -^{32}P]-triphosphate were obtained from Amersham. Bleomycin sulfate was purchased from Sigma and solutions were prepared based on an extinction coefficient derived from an $E_{1\%}$ of 105 at 292 nm and the molecular weight of Bleomycin A_2 sulfate. Monoperoxyphthalic acid (magnesium salt hexahydrate) was obtained from Aldrich. The plasmid pMFB36 was a gift of Dr. M. F. Bruist and Prof. M. I. Simon. The plasmid pDPM34 was prepared by David Mack. Cerenkov radioactivity was measured with a Beckman LS 3801 Scintillation Counter. Storage phosphor technology autoradiography was accomplished using a Molecular Dynamics 400S Phosphorimager and ImageQuant software. Storage phosphor screens (Kodak #S0230) were purchased from Molecular Dynamics.

Methods.

***O*-Benzyl-α-*N*-(*tert*-butoxycarbonyl)-L-serinehydroxamic Acid (1).**[87] *N*-Boc-L-serine (6.4 g, 31 mmol, US Biochemical Corporation) was dissolved in THF (~65 mL). A solution of *O*-benzylhydroxylamine (10 g, 63 mmol) in H_2O (~250 mL, pH 4.5) was added, and the apparent pH was readjusted to 4.5 by addition of 50% NaOH. A solution of DCC (12.8 g, 62 mmol) in THF (~200 mL) was added dropwise over 30-40 min while the apparent pH was maintained by the dropwise addition of 1.0 M HCl. A white precipitate began to form almost immediately. The reaction mixture was stirred until the pH stopped rising (~2.5 h). THF was removed under reduced pressure and the slurry was filtered to afford a white solid residue consisting of the desired product and dicyclohexyl urea (DCU). This residue was suspended in hot ethyl acetate (~ 200 mL) and filtered to remove the DCU. The filtrate was concentrated under reduced pressure and purified by flash column

chromatography (1:1 ethyl acetate-hexanes, followed by 2:1 ethyl acetate-hexanes) to afford **1** (7.4 g, 77%). The ^{1}H NMR data were in accord with those previously reported.[87] MGO VII-61.

3-[(*tert*-Butoxycarbonyl)amino]-1-(benzyloxy)-2-azetidinone (2).[87] A flame-dried flask was charged with argon, **1** (3.5 g, 11 mmol), Ph_3P (3 g, 11 mmol) and THF (60 mL). To this solution was added dropwise (under argon and with stirring) a solution of diethylazadicarboxylate (DEAD) (1.8 mL, 11 mmol) in THF (50 mL). After the addition was complete, the reaction mixture was heated to 60 °C for 4.5 h. THF was removed under reduced pressure and the residue was purified by flash column chromatography (2:1 hexanes-ethyl acetate) to afford **2** (2.6 g, 78%) as a white solid. The ^{1}H NMR data were in accord with those previously reported .[87] MGO VI-55.

(2S)-3-[(Benzyloxy)amino]-2-[(*tert*-butoxycarbonyl)-amino]proprionamide (3).[56] **2** (12.2 g, 41.8 mmol) was dissolved in methanol (~70 mL), and the solution was cooled to 0 °C and saturated with NH_3 for 10 min. The reaction mixture was stirred at 0 °C for 3.5 h and the solvent was removed under reduced pressure. The residue was recrystallized from ether-hexanes to afford **3** (9.83 g, 76%) as a white solid. The ^{1}H NMR data were in accord with those previously reported.[56] MGO IV-28.

(2S)-3-amino-2-[(*tert*-butoxycarbonyl)-amino]proprionamide (4).[56] A solution of **3** (1.1g, 3.7 mmol) in methanol (12 mL) was treated with 10% palladium-on-charcoal (218 mg) and maintained under a hydrogen atmosphere at room temperature for 4 h. The reaction mixture was filtered through Celite and the filtrate was concentrated under reduced pressure. The residue was purified by flash column chromatography (10%-MeOH- 2% Et_3N-$MeCl_2$) to afford **4** (730 mg, 97%) as a white foam. The ^{1}H NMR data were in accord with those previously reported.[56] MGO VII-121.

Propionamidine Hydrochloride (5).[70] A mixture of propionitrile (13 g, 0.24 mol) and methanol (9.6 mL, 0.24 mol) was cooled to -20 °C and HCl (10.7 g, 0.30 mol) was bubbled in while maintaining the temperature between -5 °C and -20 °C. The mixture was stirred at 0 °C for 45 min, during which time a white precipitate formed. Anhydrous Et_2O (30 mL) was added and the solution was left unstirred for 3 h. The solution was filtered to afford crude methyl propionimidate (22 g) as a pale yellow solid. The crude methyl propionimidate was suspended in absolute ethanol (100 mL), and a solution of NH_3 (3 g, 0.2 mol) in absolute ethanol (60 g) was added; the resulting suspension was stirred at room temperature for 12 h. Solvent was removed under reduced pressure to afford **5** as a white solid (18 g, 70%). ^1H NMR (DMSO-d_6) revealed an unidentified contaminant. The partially purified material (13.6 g) was dissolved in hot methanol and the contaminating material was allowed to recrystallize. The solution was filtered and the filtrate was concentrated under reduced pressure to afford **5** (10.1 g, 71%) with only trace contaminant. This removal of the contaminating material was not found to affect the yield of the subsequent reaction. ^1H NMR (DMSO-d_6) was consistent with the structure of the desired compound. MGO VII-46.

2,4,6-tris(Ethoxycarbonyl)-1,3,5-triazine (6).[70] Into a flask containing neat ethyl cyanoformate (16.4 g, 166 mmol) was bubbled HCl (0.71 g, 20 mmol). The flask was tightly stoppered and allowed to react for 36 h, forming a pale yellow solid. The crude solid was recrystallized from absolute ethanol to afford pure **6** (11.1 g, 67%) as a white solid. Analysis of the mother liquor revealed a substantial quantity of unreacted starting material, suggesting that the reaction had not gone to completion. ^1H NMR data were in accord with those previously reported.[70] MGO VII-117.

6-Amino-2,4-bis(ethoxycarbonyl)-5-methylpyrimidine (7).[58] To a solution of **6** (2.7 g, 9.1 mmol) in DMF (45 mL) under an Ar atmosphere was added **5** (2.0 g, 19 mmol). The resulting reaction mixture was warmed at 100 °C for 66 h. The solution was filtered, the solid residue was washed with methylene chloride and the filtrate was concentrated *in vacuo* to afford a dark brown solid. The crude residue was purified by flash column chromatography (2:1 EtOAc-hexanes) to afford **7** (1.5 g, 65%) as an off-white solid. ^{1}H NMR data were in accord with those previously reported.[58] MGO VII-85.

6-Amino-4-ethoxycarbonyl-2-hydroxymethyl-5-methylpyrimidine (8).[58] A solution of **7** (1.0 g, 4.0 mmol) in anhydrous ethanol was cooled to -5 °C and treated with sodium borohydride (159 mg, 4.0 mmol) under an Ar atmosphere. After stirring for 5 days at 0 °C, the reaction mixture was quenched by the addition of saturated $NaHCO_3$ (20 mL) and aqueous H_2O_2 (5%, 20 mL) at 0 °C. The mixture was stirred 5 h and extracted with 20% isopropanol-$CHCl_3$ (5 x 40 mL). The combined organic extracts were dried ($MgSO_4$) and concentrated *in vacuo* to afford a white solid. The crude residue was purified by flash column chromatography (1:1 EtOAc-hexanes until the desired product began to elute, then 2:1 EtOAc-hexanes) to afford **8** (340 mg, 41%) as a white solid. ^{1}H NMR data were in accord with those previously reported.[58] MGO VII-148.

6-Amino-4-ethoxycarbonyl-5-methylpyrimidine-2-carboxaldehyde (9).[58] A solution of **8** (50 mg, 0.24 mmol) in acetonitrile was treated under an Ar atmosphere at room temperature with activated MnO_2 and the resulting reaction mixture was warmed at 82 °C for 3 h. The mixture was allowed to cool to room temperature and filtered through a Celite pad (5 x 10 mL acetonitrile). The solvent was removed *in vacuo* and the residue was purified

by flash column chromatography (EtOAc) to afford **8** (33 mg, 66%) as a white foam. ^{1}H NMR data were in accord with those previously reported.[58] MGO VII-153.

N^{α}-(*tert*-Butoxycarbonyl)-N^{β}-[(6-ethoxycarbonyl-4-amino--5-methylpyrimidin-2-yl) methyl]-(S)-β-aminoalanine amide (10) .[58] To a flame-dried flask charged with Ar and crushed activated 3 Å molecular sieves (520 mg) were added solution of **9** (30.0 mg, 0.144 mmol) in acetonitrile (2 mL) and a solution of **4** (37.9 mg, 0.188 mmol) in acetonitrile (3 mL). The resulting mixture was stirred at room temperature for 19.5 h, then filtered through a Celite pad. After washing with methylene chloride, the fitrate was concentrated *in vacuo* to afford **10** (54.3 mg, 96%) as a white foam. ^{1}H NMR data were in accord with those previously reported.[58] MGO VII-154.

(4*S*, 5*R*)-3-(Methylthioacetyl)- 4-methyl- 5- phenyl -2- oxazolidinone (11).[88,89] To a cooled solution (-78 °C) of (4*S*, 5*R*)-4-methyl-5-phenyl-2-oxazolidinone (1.0g, 5.6 mmol) in THF (12 mL) under an Ar atmosphere was added dropwise *n*-butyl lithium (1.6 M in hexanes, 3.5 mL, 5.6 mmol). The resulting solution was stirred at -78 °C for 2 h. 2-Thiomethyl acetyl chloride[90] (550 μL, 5.6 mmol) was added to the solution, followed by stirring for an additional 90 minutes and allowed to warm to room temperature. The solution was treated with saturated aqueous ammonium chloride (6 mL), and the THF was removed under reduced pressure and the water layer was extracted with ether (3 x 10 mL). The combined organic extract was washed successively with saturated aqueous sodium bicarbonate (10 mL), brine (10 mL), and water (10 mL) and dried ($MgSO_4$). Removal of the solvent under reduced pressure followed by purification of the residue by flash column chromatography (4:1 hexanes-ethyl acetate) yielded **11** (850 mg, 57%) as a

crystalline product. ^{1}H NMR data were consistent with those previously reported.[88-90] MGO VI-146.

Ethyl (*S*)-2-[1-[[(*S*)-2- *tert*-Butyloxycarbonyl)amino]- 2-carbamoylethyl]-amino]-2-[((4*S*, 5*R*)-4-methyl- 5- phenyl- 2- oxazolidinyl) carbonyl]- 3-((*S*)-2-methylthio-1-ethyl]-6-amino-5-methylpyrimidine-4-carboxylate (12).[58] Tin(II) trifluoromethanesulfonate (500 mg, 1.2 mmol) was dissolved in THF (2 mL) under an Ar atmosphere and cooled to -78 °C. The cooled solution was treated sequentially with **11** (210 mg, 0.80 mmol) in THF (2 mL) and freshly distilled *i*Pr$_2$NEt (236 μL, 1.35 mmol). The mixture was stirred 1 h at -20 °C then recooled to -78 °C. A solution of **10** (95 mg, 0.24 mmol) in THF (2 mL) was added and the resulting mixture was allowed to warm to 0 °C where it was allowed to stir for 12 h under an Ar atmosphere. The reaction mixture was poured into a two-layer solution of CH_2Cl_2 (30 mL) and saturated aqueous sodium bicarbonate (15 mL) with vigorous shaking. The organic layer was washed with brine (2 x 10 mL), dried ($MgSO_4$), and concentrated under reduced pressure. Flash column chromatography (2% MeOH-CH_2Cl_2) of the resulting residue afforded partially purified **12** (250 mg, theoretical yield 160 mg). This residue was dissolved in CH_2Cl_2, filtered (Nylon, 0.2 μM) and subjected to preparative HPLC purification (SiO_2, 3% MeOH-$CHCl_3$). Fractions containing the compound eluting in the major peak (~11.4-12.4 min) were pooled and the solvent was removed in vacuo to afford **12** (63 mg, 40%) as a white solid. ^{1}H NMR data were in accord with those previously reported.[58] MGO VII-109.

Ethyl (*S*)-2-[1-[[(*S*)-2-*tert*-Butyloxycarbonyl)amino]-2-carbamoylethyl]-amino]-2- [((4*S*, 5*R*)-4- methyl- 5- phenyl- 2- oxazolidinyl)carbonyl]ethyl]-6-amino-5-methylpyrimidine-4-carboxylate (13).[58] A solution of **12** (68.7 mg, 0.10 mmol) in degassed, freshly distilled benzene (2 mL) under an Ar

atmosphere was subjected to a steady stream of Ar for ~10 min. The solution was treated with Bu_3SnH (81 μL, 0.30 mmol) and AIBN (~3 mg, 0.02 mmol) at room temperature and warmed at 80 °C for 1 h. More Bu_3SnH (50 μL, 0.19 mmol) was added and the reaction mixture was heated for 1 h. The reaction mixture was concentrated *in vacuo* and the residue was purified by flash column chromatography (2% MeOH-CH_2Cl_2) to afford **13** (64 mg, quantitative) as a white solid. ^1H NMR data were in accord with those previously reported.[58] MGO VII-146.

N^{α}-(*tert*-Butoxycarbonyl)-N^{β}-[1-amino-3-(*S*)-(4-amino-6-ethoxycarbonyl-5-methylpyrimidin-2-yl) propion-3-yl]-(S)-β-aminoalanine Amide (14).[58] Solid **13** (64 mg, 0.10 mmol) was treated with an ethanolic solution of NH_3 (16%, 16 mL) and the solution was stirred at 0 °C for 1.5 h. The solvent was removed *in vacuo* and the residue was purified by flash column chromatography (0% --> 20% MeOH-CH_2Cl_2 gradient) to afford **14** (34 mg, 76%) as a white solid. ^1H NMR data were in accord with those previously reported.[58] MGO VII-149.

N^{α}- (*tert*-Butoxycarbonyl)- N^{β}-[1-amino- 3-(*S*)- (4-amino-6-carboxy--5-methylpyrimidin-2-yl) propion-3-yl]-(S)-β-aminoalanine Amide (Boc-pyrmidoblamic acid) (15).[58] A solution of **14** (34 mg, 0.076 mmol) in THF-MeOH-H_2O (3:1:1, 0.83 mL) at 0 °C was treated with aqueous 1N LiOH (0.11 mL, 0.11 mmol) and the mixture was stirred for 1.5 h. The organic solvent was removed in vacuo and the aqueous phase was acidified by the addition of 1 N HCl (0.13 mL). The solvent was removed in vacuo and the residue was purified by flash column chromatography (Reverse Phase-18 gel, 0% --> 50% MeOH-H_2O) to afford **15** (26 mg, 80%) as a white solid. ^1H and ^{13}C NMR data were in accord with those previously reported.[58] The ^{13}C NMR spectrum was recorded by Dr. Ken Turnbull. MGO VII-150.

***N*-β-[1-amino-3(S)-(4-amino-(6-amido-erythro–β-hyxdroxy-L-histidine methyl ester)-5-methylpyrimidin-2-yl)-propion-3-yl]-(S)-β-aminoalanine Amide (20)** Compound **17** was prepared by Dr. Ken Turnbull as previously reported by Boger et al.[58] **17** (2.1 mg, 2.5 μmol) *p*-cresol (1 drop), and *p*-thiocresol (1 drop) were added to a Teflon tube equipped with a Teflon-coated magnetic stir bar. The reaction vessel was evacuated, cooled to -78 °C, and charged with HF (~2.0 mL). The mixture was allowed to stir at 0 °C (1h). HF was removed in vacuo and the resulting yellow residue was taken up in H_2O (0.5 mL). The aqueous portion was extracted with $CHCl_3$ (1 x 0.5 mL) and the organic portion was washed with H_2O (1 x 0.5 mL). The combined aqueous portions were concentrated in vacuo to afford an off-white foam. TLC (analytical RP-18, 5 x 10 cm, 0.25 mm thickness, 25% CH_3CN/H_2O, MeOH extraction) afforded a white powder which was dissolved in H_2O (1 mL) and filtered (0.2 μm) to remove particulate contaminants. Concentration *in vacuo* afforded **4** (1.2 mg, quantitative) as a white foam. The compound was characterized by Dr. Ken Turnbull. $[\alpha]^{25} = +12.2$ (c 0.03, CH_3OH); ^1H NMR (300 MHz, D_2O) δ 8.73 (s, 1H), 7.50 (s, 1H), 5.44 (d, J = 5.7 Hz, 1H), 5.12 (d, J = 5.7 Hz, 1H), 4.51 (m, 1H), 4.41 (m, 1H), 3.78 (s, 3H), 3.18 (m, 2H), 3.04 (m, 1H), 2.84 (m, 1H), 2.09 (s, 3H); IR (neat) 3384, 1736, 1664, 1460, 1380, 850, 698; FABHRMS (SGly), *m/e* 493.2272 (MH^+, $C_{19}H_{29}N_{10}O_6$ requires 493.2267).

Protein Syntheses. Hin(139-190) was synthesized by automated stepwise solid phase chemical synthesis on an ABI 430A peptide synthesizer as previously reported.[47-49,55] An approximately 50-mg sample of resin bound peptide (ca. 100 μmol/g, total peptide ca. 5 μmol) was placed in a 12 x 80 mm reaction vessel. The terminal Boc protecting group was removed and the resin neutralized by standard procedures. **18** was prepared by Dr. Ken

Turnbull as previously described.[58] **18** (16.0 mg, 20 μmol) and HOBt (5.5 mg, 42 μmol) were dissolved in DMF (1 mL) and DCC (4.6 mg, 22 μmol) was added. The solution was allowed to stir at room temperature for 30 min and added to the peptide/resin along with additional DMF (1 mL). Ninhydrin analysis[91] indicated 88% reaction with no further improvement in yield after 5 h. The resulting protected PBA-β-OH-His-Hin(139-190) was deprotected by anhydrous HF and purified by preparative reverse-phase HPLC as previously described.[47-49,55] Laser desorption time-of-flight mass spectrometry (α-cyano-4-hydroxy-cinnamic acid matrix, 70% acetonitrile, 2% TFA in H_2O) is consistent with the desired structure (calculated mass, 6495.5, observed, 6493.3). A second, much smaller peak, due to Hin(139-190) (calculated mass, 6036.1, observed, 6035.6) was also present. Hin(139-190) and EDTA-GABA-Hin(139-190) were also synthesized, deprotected and purified according to previously reported procedures.[47-49,55]

Fe•PBA-β-OH-His-Hin(139-190) and Fe•EDTA-Hin(139-190) were generated immediately prior to use by equilibration of 250 μM protein with an equal volume of 250 μM ferrous ammonium sulfate (FAS) for 15 min before dilution to the appropriate concentrations. Bleomycin was not incubated with FAS at high concentrations but diluted immediately after the addition of iron. Ni-GGH-Hin was generated in a similar fashion by incubation with $Ni(OAc)_2$. Fe•21, its corresponding free acid complex, and Fe•EDTA were all incubated with FAS at ≥1 mM concentrations before dilution.

DNA Manipulations. The 450 bp *Xba* I-*Eco* RI restriction fragment of the plasmid pMFB36[44] were isolated and labeled at the 5' or 3' ends by standard procedures.[92] A-specific[93] and Maxam-Gilbert G-specific[94] sequencing reactions were performed according to established protocols. The 450 bp *Xba* I-

Eco RI restriction fragment of the plasmid pMFB36 was used for all assays except end-product analysis. All footprinting and affinity cleaving reactions were performed at room temperature.

DNase I footprinting[95] reaction conditions were 20 mM phosphate, 20 mM NaCl, 10 mM $MgCl_2$, 100 μM DTT, 0.004 units DNase I, 100 μM calf thymus DNA, and ~20,000 cpm labeled DNA, pH 7.5. The DNA binding molecules were allowed to equilibrate with the DNA for 15 min. Reactions were then initiated by the addition of a stock solution of DNase I, $MgCl_2$, and DTT and allowed to proceed 10 min at room temperature. Reactions were quenched by the addition of a 3 M ammonium acetate solution containing 250 mM EDTA followed by ethanol precipitation.

Fe•PBA-β-OH-His-Hin(139-190) and Fe•EDTA-Hin(139-190) affinity cleaving reaction conditions were 20 mM phosphate, 20 mM NaCl, 5 mM DTT, 100 μM calf thymus DNA, ~10,000 cpm labeled DNA, and 5 μM Fe•EDTA-Hin(139-190) or 10 μM Fe•PBA-β-OH-His-Hin(139-190), pH 7.5. Ni•GGH-Hin(139-190) affinity cleavage reaction conditions were 20 mM phosphate, 20 mM NaCl, 5 μM monoperoxyphthalic acid, 100 μM calf thymus DNA, ~10,000 cpm labeled DNA, and 5 μM Ni•GGH-Hin, pH 7.5. Fe•Bleomycin affinity cleavage reaction conditions were 20 mM phosphate, 20 mM NaCl, 5 mM DTT, 100 μM calf thymus DNA, ~10,000 cpm labeled DNA, and 2.5 μM Fe•Bleomycin, pH 7.5. All affinity cleaving reagents were allowed to equilibrate with the DNA for 15 min before reactions were initiated with the addition of DTT or monoperoxyphthalic acid. Fe•EDTA-Hin(139-190) and Ni•GGH-Hin reactions were allowed to proceed for 1 h, Fe•PBA-β-OH-His-Hin(139-190) reactions for 6 h, and Fe•Bleomycin reactions for 15 min before they were terminated by ethanol precipitation. Residues were dried and resuspended in 100 mM tris-borate-EDTA/80% formamide

loading buffer. Reaction products were analyzed by electrophoresis on 6% polyacrylamide denaturing gels (5% crosslink, 7M urea). After electrophoresis, gels were dried and autoradiographed. Gels were analyzed using storage phosphor technology. Storage phosphor screens were pressed against dried gels and exposed. A Molecular Dynamics 400S PhosphorImager was used to obtain data from storage screens, and data were analyzed by performing volume integrations using the ImageQuant v.3.0 software running on an AST Premium 386/33 computer.

End Product Analysis. For 3'-end product analysis, reactions were performed on the 562 bp 5' end-labeled *Hind* III/ *Rsa* I restriction fragment from pDPM34[54] containing the Hin binding site secondary. Fe•PBA-β-OH-His-Hin(139-190) and Fe•**4** affinity cleaving reaction conditions were 20 mM phosphate, 20 mM NaCl, 5 mM DTT, 100 μM calf thymus DNA, and 100 μM Fe•**4** or 10 μM Fe•PBA-β-OH-His-Hin(139-190), pH 7.5. Fe•Bleomycin affinity cleavage reaction conditions were 20 mM phosphate, 20 mM NaCl, 5 mM DTT, 100 μM calf thymus DNA, and 2.5 μM Fe•Bleomycin, pH 7.5. MPE•Fe reaction conditions were 20 mM phosphate, 20 mM NaCl, 5 mM DTT, 5 μM MPE•Fe. Cleaving agents were allowed to equilibrate with the DNA for 15 min before reactions were initiated by the addition of DTT. MPE•Fe and Fe•Bleomycin reactions were allowed to proceed 15 min and Fe•PBA-β-OH-His-Hin(139-190) and Fe•**4** reactions for 4 h at room temperature before termination by lyophilization. Where base treatment was appropriate, residues were resuspended in 50 μL of 0.1 N butylamine and heated at 90 °C for 30 min. Reaction solutions were frozen and lyophilized.

The presence of 3'-phosphate groups was tested using T4 polynucleotide kinase to remove the 3'-phosphate groups.[73] After the DNA reactions and base treatment, DNA residues were resuspended in 25 μL water, heated to 90

°C for 5 min, and cooled on ice. 25 μL of a buffer containing 20 mM Tris-HCl, pH 6.6, 20 mM $MgCl_2$, and 10 mM β-mercaptoethanol were added followed by 2 μL T4 polynucleotide kinase (10 units/mL). Reactions were incubated 1 h at 37 °C and terminated by ethanol precipitation. Residues were prepared for gel electrophoresis (20% acrylamide, 5% crosslink, 7M urea) as described above. Storage phosphor screens were pressed against the gels and exposed. Gels were analyzed as described above.

For 5'-end product analysis, reactions were performed on the 562 bp 5' end-labeled *Eco* RI/ *Rsa* I restriction fragment from pDPM34. Reaction conditions were the same as above, except 5 μM Fe•Bleomycin was used. Incubation time before reaction initiation with DTT was 15 min in all cases. MPE•Fe and Fe•Bleomycin reactions were allowed to proceed 20 min and Fe•PBA-β-OH-His-Hin(139-190) and Fe•**4** reactions for 4 h at room temperature before quenching by phenol (2 x 2:1 phenol:chloroform, equal volume) extraction and ethanol precipitation. The presence of 5'-phosphate groups was tested using calf alkaline phosphatase to remove the 5'-phosphate groups. The residues were resuspended in H_2O (45 μL) and heat denatured at 90 °C for 5 min. 10x dephosphorylation buffer (Boehringer-Mannheim, 7 μL) was added, followed by 20 μL of enzyme (Boehringer Mannheim, 1 unit/μL). The reaction mixture was incubated at 37 °C for 1 h and ethanol precipitated.

Quantitative Footprint Titrations. The above reaction conditions were modified for quantitative footprinting as follows. Stock solutions of Hin(139-190), PBA-β-OH-His-Hin(139-190), or Fe•PBA-β-OH-His-Hin(139-190) were diluted serially to give 15 solutions of concentrations ranging from 50 nM to 150 μM, leading to final reaction concentrations of 10 nM to 30 μM. Final reaction conditions were 20 mM bis-Tris•HCl, 20 mM NaCl, 5 mM DTT, 100 μM bp calf thymus DNA , 0.004 units DNase I, and 15,000 cpm 5'-^{32}P end-

labeled DNA, pH 7.6. The proteins were allowed to equilibrate with the DNA for 1 h before the reactions were initiated by the addition of a stock solution of DNase I, $MgCl_2$, and DTT and allowed to proceed 10 min at room temperature. Reactions were quenched by the addition of a 3 M ammonium acetate solution containing 250 mM EDTA. Reaction mixtures were diluted 2.5 fold and extracted with an equal volume of phenol and then butanol before addition of 5 μL 20 mg/mL glycogen and ethanol precipitation. Residues were prepared for gel electrophoresis as described above.

Footprint Titration Fitting Procedure. The footprint titration gels were quantitated using storage phosphor technology, and data were analyzed by performing volume integrations of the target and reference sites using the ImageQuant v.3.0 software running on an AST Premium 386/33 computer. The target sites chosen were the five Hin half sites found in the 450 bp *Xba* I-*Eco* RI restriction fragment of the plasmid pMFB36. Specifically, bands at the following sequences were quantitated, moving from left to right in the sequence (Fig. 5): *hix*L IRL, 5'-TTCTT-3' (upper strand); *hix*L IRR, 5'-TATCAA-3' (lower strand); tertiary, 5'-TCTCATGGA-3' (lower strand); secondary IRL 5'-TTCCTT-3' (upper strand); and secondary IRR, 5'-TTCTCC-3' (lower strand). A reference site between the tertiary and secondary sites of 5'-ACTAA-3' was also quantitated.

The data from the DNase I footprint titrations were analyzed as previously described.[72] The apparent DNA target site saturation, θ_{app}, was calculated using the following equation:

$$\theta_{app} = 1 - \frac{I_{tot}/I_{ref}}{I_{tot}^{\circ}/I_{ref}^{\circ}} \qquad (1)$$

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values for a DNase I control lane to which no protein has been added. At higher concentrations of protein (> 10 μM), the reference sites become partially protected, resulting in low θ_{app} values. For these data points, I_{ref} values were corrected by multiplying the amount of radioactivity loaded per lane (CPM_{tot}) by the mean ratio of I_{ref} /CPM_{tot} for all data points from lanes with <10 μM protein. The ($[L]_{tot}$, θ_{app}) data points were fit to a Langmuir binding isotherm by minimizing the difference between θ_{app} and θ_{fit}:

$$\theta_{fit} = \theta_{min} + (\theta_{max} - \theta_{min}) \cdot \frac{K_a^{\,n}[L]_{tot}^{\,n}}{1 + K_a^{\,n}[L]_{tot}^{\,n}} \qquad (2)$$

where $[L]_{tot}$ corresponds to the total protein concentration, K_a corresponds to the apparent monomeric association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively.

Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 3.1, Abelbeck software) running on a Macintosh IIfx computer with K_a, θ_{max}, θ_{min} and n as the adjustable parameters. All lanes from each gel were used unless visual inspection of the computer image from a storage phosphor screen revealed a flaw at the target or reference site, or the θ_{app} value was greater than two standard errors away from values in both neighboring lanes, using a standard error in θ_{app} of 0.1. The goodness-of-fit of the binding curves was evaluated by the correlation coefficient, with R> 0.97 as the criterion for an acceptable fit. Three or four sets of acceptable data were obtained for all sites except for secondary

IRR, for which no data are reported. Data were normalized using the following equation:

$$\theta_{norm} = \frac{\theta_{app} - \theta_{min}}{\theta_{max} - \theta_{min}} \quad (3)$$

The association constants were calculated by fitting the average θ_{norm} values from three or four gels to eq. 2 as described above. The standard deviation of calculated association constants from each gel is also given (Table I). The Hill coefficient, n, varied between the sites *hix*L and secondary. Values ranged from 1.5 to 1.9 for the three proteins at the two *hix*L half sites. At secondary IRL, n was ~1.0 for Fe(II)•PBA-β-OH-His-Hin(139-190) and PBA-β-OH-His-Hin(139-190) and ~1.5 for Hin(139-190). The standard deviations for n were 0.1-0.3.

REFERENCES

1. H. Umezawa, in *Bleomycin: Current Status and New Developments* S. K. Carter, S. T. Crooke, H. Umezawa, Eds. (Academic Press, New York, (1978) p. 15.

2. J. Stubbe and J. W. Kozarich, *Chem. Rev.* **87**, 1107 (1987).

3. S. M. Hecht, *Acc. Chem. Res.* **19**, 383 (1986).

4. P. C. Dedon and I. H. Goldberg, *Chem. Res. Toxicol.* **5**, 311 (1992).

5. M. Takeshita, A. P. Grollman, E. Ohtsubo, and H. Otsubo, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5983 (1978).

6. A. D. D'Andrea and W. A. Hasetine, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3608 (1978).

7. J. W. Kozarich et al., *Science* **245**, 1396 (1989).

8. J. C. Wu, J. W. Kozarich, and J. Stubbe, *Biochemistry* **24**, 7562 (1985).

9. R. M. Burger, J. Peisach, and S. B. Horwitz, *J. Biol. Chem.* **256**, 11636 (1981).

10. G. Pratviel, J. Bernadou, and B. Meunier, *Biochem. Pharmacol.* **38**, 133 (1989).

11. H. Kuramochi, K. Takahashi, T. Takita, and H. Umezawa, *J. Antibiot.* **34**, 576 (1981).

12. H. Sugiyama, R. E. Kilkuskie, L.-H. Chang, L.-T. Ma, and S. M. Hecht, *J. Am. Chem. Soc.* **108**, 3852 (1986).

13. R. M. Burger, J. S. Blanchard, S. B. Horwitz, and J. Peisach, *J. Biol. Chem.* **260**, 15406 (1985).

14. L. J. Worth et al., *Biochemistry* **32**, 2601 (1993).

15. G. H. McGall, J. Stubbe, and J. W. Kozarich, *J. Org. Chem.* **56**, 48 (1991).

16. L. Giloni, M. Takeshita, F. Johnson, C.Iden, and A. P. Grollma, *J. Biol. Chem.* **256**, 8608 (1981).

17. G. H. McGall et al., *J. Am. Chem. Soc.* **114**, 4958 (1992).

18. H. Sugiyama, H. Kawabata, T. Fujiwara, Y. Dannoue, and I. Saito, *J. Am. Chem. Soc.* **112**, 5252 (1990).

19. L. E. Rabow, G. H. McGall, J. Stubbe, and J. W. Kozarich, *J. Am. Chem. Soc.* **112**, 3202 (1990).

20. J. W. Sam and J. Peisach, *Biochemistry* **32**, 1488 (1993).

21. L. E. Rabow, J. Stubbe, and J. W. Kozarich, *J. Am. Chem. Soc.* **112**, 3196 (1990).

22. H. Sugiyama, C. Xu, N. Murugesan, and S. M. Hecht, *Biochemistry* **27**, 58 (1988).

23. Y. Iitaka et al., *J. Antibiot.* **31**, 1070 (1978).

24. Y. Sugiura et al., *J. Biol. Chem.* **258**, 1328 (1983).

25. R. E. Killkuskie, H. Suguna, B. Yellin, N. Murugesan, and S. M. Hecht, *J. Am. Chem. Soc.* **107**, 260 (1985).

26. H. Umezawa et al., *Tetrahedron* **44**, 501 (1984).

27. H. Sugiyama, R. E. Kilkuskie, and S. M. Hecht, *J. Am. Chem. Soc.* **107**, 7765 (1985).

28. H. Sugiyama et al., *J. Nat. Prod.* **48**, 869 (1985).

29. J. B. Shipley and S. M. Hecht, *Chem. Res. Toxicol.* **1**, 25 (1988).

30. J. Kross et al., *Biochemistry* **21**, 3711 (1982).

31. M. Chien, A. P. Grollman, and S. B. Horwitz, *Biochemistry* **16**, 3641 (1977).

32. L. M. Fisher, R. Kuroda, and T. Sakai, *Biochemistry* **24**, 3199 (1985).

33. M. Ohno and M. Otsuka, in *Recent Progress in the Chemical Synthesis of Antibiotics* G. Lukacs, M. Ohno, Eds. (Springer-Verlag, Berlin, 1990) p. 387.

34. R. J. Guajardo, S. E. Hudson, S. J. Brown, and P. K. Mascharak, *J. Am. Chem. Soc.* **115**, 7971 (1993).

35. D. S. Sergeyev, T. S. Godovikova, and V. F. Zarytova, *FEBS Lett.* **280**, 271 (1991).

36. V. F. Zarytova, D. S. Sergeyev, and T. S. Godovikova, *Bioconjugate Chem.* **4**, 189 (1993).

37. M. Otsuka et al., *J. Am. Chem. Soc.* **112**, 838 (1990).

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

39. P. G. Schultz, J. S. Taylor, and P. B. Dervan, *J. Am. Chem. Soc.* **104**, 6861 (1982).

40. J. Taylor, P. Schultz, and P. Dervan, *Tetrahedron* **40**, 457 (1984).

41. R. C. Johnson, M. B. Bruist, M. B. Glaccum, and M. I. Simon, *Cold Spring Harbor Sym. Quant. Biol.* **49**, 751 (1984).

42. A. C. Glasgow, M. F. Bruist, and M. I. Simon, *J. Biol. Chem.* **264**, 10072 (1989).

43. R. G. Brennan, *Curr. Opin. Struct. Biol.* **2**, 100 (1992).

44. M. F. Bruist, S. J. Horvath, L. E. Hood, T. A. Steitz, and M. I. Simon, *Science* **235**, 777 (1987).

45. R. P. Hertzberg and P. B. Dervan, *J. Am. Chem. Soc.* **104**, 313 (1982).

46. R. P. Hertzberg and P. B. Dervan, *Biochemistry* **23**, 3934 (1984).

47. J. Sluka, S. Horvath, M. Bruist, M. Simon, and P. Dervan, *Science* **238**, 1129 (1987).

48. D. P. Mack et al., *Biochemistry* **29**, 6561 (1990).

49. J. P. Sluka, S. J. Horvath, A. C. Glasgow, M. I. Simon, and P. B. Dervan, *Biochemistry* **29**, 6561 (1990).

50. K. T. Hughes, P. C. W. Gaines, J. E. Karlinsey, R. Vinayak, and M. I. Simon, *EMBO J.* **11**, 2695 (1992).

51. D. P. Mack, B. L. Iverson, and P. B. Dervan, *J. Am. Chem. Soc.* **110**, 7572 (1988).

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

53. D. P. Mack and P. B. Dervan, *Biochemistry* **31**, 9399 (1992).

54. D. P. Mack, Ph. D. Thesis, California Institute of Technology (1991).

55. S. Kent, *Annu. Rev. Biochem.* **57**, 957 (1988).

56. Y. Aoyagi, M. S. Chorghade, A. A. Padmapriya, H. Suguna, and S. M. Hecht, *J. Org. Chem.* **55**, 6291 (1990).

57. H. Arai, K. W. Hagmann, H. Suguna, and S. M. Hecht, *J. Am. Chem. Soc.* **102**, 6631 (1980).

58. D. L. Boger, R. F. Menezes, and T. Honda, *Angew. Chem. Int. Ed. Engl.* **32**, 273 (1993).

59. M. Otsuka et al., *Chem. Pharm. Bull.* **33**, 520 (1985).

60. Y. Umezawa et al., *J. Am. Chem. Soc.* **102**, 6630 (1980).

61. S. M. Hecht, K. Rupprecht, and P. M. Jacobs, *J. Am. Chem. Soc.* **101**, 3982 (1979).

62. T. Owa, M. Otsuka, and M. Ohno, *Chem Lett.* **1988**, 1873 (1988).

63. T. Owa, M. Otsuka, and M. Ohno, *Chem Lett.* **1988**, 83 (1988).

64. D. L. Boger and R. F. Menezes, *J. Org. Chem.* **57**, 4331 (1992).

65. T. Takita et al., *Tetrahedron Lett.* **23**, 521 (1982).

66. Y. Aoyagi et al., *J. Am. Chem. Soc.* **104**, 5537 (1982).

67. N. J. Oppenheimer, L. O. Rodriguez, and S. M. Hecht, *Biochemistry* **18**, 3439 (1979).

68. C.-H. Huang, L. Galvan, and S. T. Crookes, *Biochemistry* **18**, 2880 (1979).

69. D. L. Boger and T. Honda, *Tetrahedron Lett.* **34**, 1567 (1993).

70. D. L. Boger and Q. Dang, *Tetrahedron* **44**, 3379 (1988).

71. J. P. Sluka, Ph. D. Thesis, California Institute of Technology (1988).

72. M. Brenowitz, D. F. Senear, M. A. Shea, and G. K. Ackers, *Methods Enzymol.* **130**, 132 (1986).

73. V. Cameron and O. C. Uhlenbeck, *Biochemistry* **16**, 5120 (1977).

74. J. P. Sluka, J. H. Griffin, D. P. Mack, and P. B. Dervan, *J. Am. Chem. Soc.* **112**, 6369 (1990).

75. J. A. Shin, Ph. D. Thesis, California Institute of Technology (1992).

76. R. M. Burger, A. R. Berkowitz, J. Peisach, and S. B. Horwitz, *J. Biol. Chem.* **255**, 11832 (1980).

77. E. A. Sausville, R. W. Stein, J. Peisach, and S. B. Horwitz, *Biochemistry* **17**, 2746 (1978).

78. N. Hamamichi, A. Natrajan, and S. M. Hecht, *J. Am. Chem. Soc.* **114**, 6278 (1992).

79. Y. Sugiura, *J. Am. Chem. Soc.* **102**, 5208 (1980).

80. N. Cameraman, A. Cameraman, and B. Sarkar, *Can. J. Chem.* **54**, 1309 (1976).

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

82. C. E. Bannister, J. M. T. Raycheba, and D. W. Margerum, *Inorg. Chem.* **21**, 1106 (1982).

83. T. Sakurai and N. Nakahara, *Inorg. Chim. Acta.* **34**, L243 (1979).

84. The Hin binding site at tertiary is 5'-TTTTCTC**A**TG-3, compared to 5'-TTCTTG**AAA**A-3' at hixL IRL, where bold type indicates essential bases according to extensive mutational analysis.[50] The new cleavage bands at tertiary correspond to a Hin binding site of 5'-TCCTCCATGA-3'.

85. M. J. Levy and S. M. Hecht, *Biochemistry* **27**, 2647 (1988).

86. B. J. Carter, V. S. Murty, K. S. Reddy, S.-N. Wang, and S. M. Hecht, *J. Biol. Chem.* **265**, 4193 (1990).

87. P. G. Mattingly and M. J. Miller, *J. Org. Chem.* **45**, 410 (1980).

88. D. A. Evans, J. Bartroli, and D. J. Mathre, *J. Am. Chem. Soc.* **103**, 2127 (1981).

89. D. A. Evans, M. D. Ennis, and D. J. Mathre, *J. Am. Chem. Soc.* **104**, 1737 (1982).

90. T. L. Shih, Ph. D. Thesis, California Institute of Technology (1984).

91. V. K. Sarin, S. B. H. Kent, J. P. Tam, and R. B. Merrifield, *Anal. Biochem.* **117**, 147 (1981).

92. J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratory Manual, 2nd Edition* (Cold Spring Harbor Press, New York, (1989).

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

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

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