Design, Synthesis and Characterization of Sequence Specific DNA Cleaving Agents

Thesis by

James P. Sluka

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosphy

California Institute of Technology

Pasadena, California

1988

(Submitted January 14, 1988)

Acknowledgments

This thesis is dedicated to my wife Anna, and my family, Jacob, Dorothy, David, Dan, Jeff, Jean, Mark and Damian, for their support during my studies.

I would like to thank Professor Peter Dervan for his guidance and support of this research. Additionally, I would like to thank the members of the Dervan group for helping to keep science, and life, interesting. Special thanks go to Bob Youngquist for keeping me sane, motivated, and in one piece. The assistance of Professor Mel Simon, Dr. Suzanna Horvath, Dr. Mike Bruist and Anna Glasglow was greatly appreciated. Finally, to Professor Gerald Selter go my thanks for getting it all started.

Financial support from the National Institute of Health is gratefully acknowledged.

Abstracts

Chapter I: Synthetic, Sequence Specific DNA Cleaving Peptides

Synthetic peptides based on the sequence specific DNA binding domain of Hin recombinase have been prepared which are equiped with ethylenediaminetetraacetic acid (EDTA) at the amino terminus. Covalent attachment of EDTA converts the sequence specific DNA binding peptides into peptides which are capable of DNA strand scission in the presence of iron II, a reducing agent and molecular oxygen. The EDTA–equiped peptides cleave DNA at Hin binding sites and provide information on the nucleotide postion and groove identity of the modified amino acid residue when bound to DNA. The EDTA–peptides were also competent at DNA double strand cleavage, which is a useful characteristic for identifying binding sites on large DNA fragments (>5000 base pairs). A general procedure for the incorporation of an EDTA equiped lysine residue at any position in a synthetic peptide is also described.

Results with 22 synthetic peptides indicate that Hin specifically recognizes a minor groove 5′-AAA sequence in its binding site with the peptide sequence Arg–Pro–Arg. Furthermore, the minor groove recognition event contributes substantially to the overall binding affinity. The possibility that minor groove recognition of tracks of A's by the peptide sequence Arg–Pro–Arg (and similar sequences), may be a common recognition motif in sequence specific DNA binding proteins is discussed.

Chapter II: Distamycin–Actinomycin Hybrid Affinity Cleaving Molecules

One approach to the design of sequence specific DNA binding molecules that read large sequences of double helical DNA is to couple DNA binding units of similar or diverse base pair specificities. Covalent attachment of bis- and tris-N-methylpyrrolecarboxamides (based on the sequence specific DNA binding antibiotics Netropsin and Distamycin) and an aromatic phenoxazone intercalator (based on the sequence specific DNA intercalator Actinomycin) creates hybrid DNA binding minor groove and intercalation compounds. Using the affinity cleaving method we find that a first generation hybrid, bis-(EDTA–distamycin–glycyl)phenoxazone, binds the sequence 5′-TTATGGTTAA-3′ which is consistent with simultaneous minor groove binding of the two tripyrrolecarboxamide units and intercalation of the phenoxazone moeity. Along with the targeted trimeric binding, the first generation hybrid also gave substantial monomeric/dimeric binding, and exihibited only modest sequence specificity on large DNA fragments.

In an attempt to reduce monomeric and dimeric binding, a second generation hybrid was prepared. Replacement of one pyrrolecarboxamide with a γ aminobutyric acid moeity, bis(EDTA–netropsin–glycyl)phenoxazone, generated a hybrid molecule which showed a reduced tendency for monomeric and dimeric DNA binding modes. The second generation hybrid binds to sequences of the form 5′-$(A/T)_4GT(A/T)_4$-3′ and 5′-$(A/T)_4CT(A/T)_4$-3′ and showed substantially improved sequence specificity on large DNA fragments relative to the first generation hybrid.

Chapter III: Bleomycin–Distamycin Hybrid Affinity Cleaving Molecules

Hybrids of the sequence specific DNA binding bithiazole domain of Bleomycin and tris-N-methylpyrrolecarboxamide of Distamycin are described. The hybrids address the issue of whether or not the bithiazole unit of the GC specific Bleomycin is a minor groove binding function. By the affinity cleaving method we find that hybrids in which the bithiazole is attached to either the amino or carboxy terminus of the tris-N-methylpyrrolecarboxamide, do not significantly alter the sequence specificity of the pyrrolecarboxamide domain. The results indicate that either the bithiazole unit is not a minor groove binding function in the same sense as the tripyrrolecarboxamide, or, alternatively, the bithiazole does not include all of the required sequence specific DNA recognition elements of Bleomycin.

Chapter IV: Studies in Metalloporphyrin Mediated Affinity Cleaving

Metalloporphyrins are known to cause DNA strand scission using a variety of central metals, and various reducing and oxidizing cofactors. To investigate if a metalloporphyrin can replace an EDTA•Fe chelate in an affinity cleaving molecule, a metalloporphyrin–Distamycin (**PD**) was prepared. **PD**•Fe showed very inefficient DNA cleavage compared to EDTA•Fe equiped affinity cleaving compounds. Additionally, **PD**•Fe showed no detectable sequence specific DNA cleaving ability. **PD**•Fe was also less efficient at DNA strand scission than ferriprotoporphyrin itself, which indicates that the DNA binding domain (tripyrrolecarboxamide) and cleaving domain (metalloporphyrin) are interfering with each other's function in the affinity cleaving compounds.

Table of Contents

CHAPTER I:

Synthetic, Sequence Specific DNA Cleaving Peptides

The interaction between DNA and DNA binding proteins is critical in the transfer of genetic information between the DNA "data base" and the "real word" of cellular function. Protein–DNA interactions are critical in DNA manipulations including packaging in nucleosomes, genomic rearrangements, transcription and replication. The sequence specificity of DNA binding proteins, i.e. the ability to differentiate between different DNA sequences, ranges from little or none in the histone like proteins responsible for chromatin formation, to the exquisite sequence specificity of the type I restriction endonucleases. The work described herein investigates the nature of DNA–protein interactions in general and sequence specific interaction in particular. Questions which are addressed include, the ability to sequence specifically deliver a chemicaly active group via a sequence specific DNA binding protein, the minimum peptide domain which is competent at sequence specific DNA recognition and the nature of particular DNA–protein interactions.

My primary focus will be the conversion of a sequence specific DNA binding peptide (based on a naturally occuring protein) into a sequence specific DNA cleaving peptide. This conversion was accomplished through incorporation of the metal chelator ethylenediaminetetraacetic acid (EDTA) in the peptide. I will also examine the effect of peptide size on sequence specific DNA recognition and the

feasibility of positioning the EDTA moiety at various locations in the peptide. Recently a *synthetic* 52 amino acid peptide (139→190; **HIN52**) based on the carboxy terminus of Hin recombinase (190 amino acids) has been shown to both bind to Hin recombination sites and to inhibit Hin activity.[1] I was interested in conversion of the **HIN52**, and related peptides, into sequence specific DNA cleaving peptides.

The general features of double stranded DNA and protein structures allow one to predict the types of interactions which may occur between the two.[2] (1) Ionic interactions between positively charged amino acid side chains and the DNA phosphate backbone would be strong, though generally nonspecific interactions, as would salt bridges with metal ions between negatively charged amino acid side chains and the phosphate backbone.[3] (2) Hydrogen bonding between peptide backbone and/or side chains and the DNA phosphate backbone or the exposed edges of the bases in the minor and major grooves of the DNA, (3) hydrophobic interactions between protein side chains and the bases or sugars, and (4) intercalation (stacking) of aromatic side chains between the nucleotide bases would all be potentially specific interactions. It appears that the greatest potential for sequence specific recognition occurs in the major and minor grooves of the DNA where the exposed edges of the bases present unique hydrogen bond donor/acceptor and hydrophobic contact patterns. Lists of potential hydrogen bond interactions between amino acid side chains and the exposed edges of the DNA bases have been compiled.[4–6]

Predictions can also be made as to the types of protein secondary structure which may be important in DNA binding. It has been observed that a protein β sheet should fit in/above the minor groove of B-form DNA.[7] The relative dimensions and twist, and the anti-parallel nature of both a two-stranded antiparallel β sheet and the phosphate backbone are similar. Side chains along each strand of

the β sheet would alternately be directed into and out of the DNA minor groove. Model building studies indicate that only proline and tryptophan could not be accommodated in the space between the minor groove and the β sheet.[7]

It has also been noted that a protein α helix has the proper width to fit in the B-form DNA major groove.[8] Since an α helix has a pronounced dipole moment which points toward the carboxy terminus, i.e. a net positive charge at the amino terminus,[9] and the side chains also point towards the amino terminus, it would seem that an α helix would prefer to point its amino terminus into the DNA groove more so than the carboxy terminus.

Another type of potential interaction is recognition of a particular DNA conformation by a protein. Structural recognition could range from detection of gross differences such as A versus B-form DNA to detection of more subtle differences such as sequence dependent variations in the groove widths, depth and helical rise of the DNA. Recently it has been proposed that sequence dependent DNA stiffness may also affect protein binding.[10] If tight protein binding induces a DNA conformational change then the plasticity of the DNA, both in regions directly contacted, and in flanking sequences, could be an important factor.

Several non-sequence specific DNA binding protein structures have been determined by x-ray crystallography. These include the 2.3Å resolution single stranded DNA binding protein product of Gene 5 from bacteriophage fd (87 amino acids),[11,12] the eukaryote nucleosome particle cocrystal with DNA which contains two copies each of four proteins (H2A, H2B, H3 and H4) at 7Å resolution,[13–15] the putative procaryote histone-like protein HU (90 amino acids) from *Bacillus Stearothermophilus*,[16] bovine pancreatic deoxyribonuclease I (DNase I, 257 amino

acids)[17] and the Klenow fragment of DNA ploymerase I from *E. Coli* (~600 amino acids, 3.2Å resolution).[18,19]

Sequence Specific DNA Binding Protein Structures

In recent years a number of x-ray crystal structures of sequence specific DNA binding proteins (or fragments thereof) have become available. The first three structures determined were DNA regulatory proteins: λ Cro (66 amino acids),[20–22] residues 1 through 92 (1→92) of λ repressor (236 amino acids, also called λ cI)[23,24] from λ phage and the 210 amino acid catabolite gene activator protein (CAP).[25–27]

It was soon recognized that all three crystal structures had several features in common (see figure 1). (1) All were dimeric and C_2 symmetric structures. (2) Each monomer contained an α helix–turn–α helix structure which was superimposable in the three structures. (3) The second helix of the helix–turn–helix domain had one face exposed to solvent and the other buried within the protein. (4) Both exposed helices within the dimeric unit were on the same face of the protein. (5) The seperation between the symmetrically related helices in the dimers was ~34Å which closely matches the major groove repeat distance of B-form DNA. (6) In both λ Cro and λ repressor the exposed helix is tilted approximately 30° with respect to the long dimension of the dimer, which closely resembles the slope of the B-form DNA backbone. The CAP helices, however, are tilted in the opposite direction.

These similarities, along with chemical protection data from DNA complexes and other solution studies, led to the hypothesis that the symmetrically related exposed helices of the dimers (second helix of the helix–turn–helix) bind in adjacent major grooves of B-form DNA and are the dominant sequence specific DNA

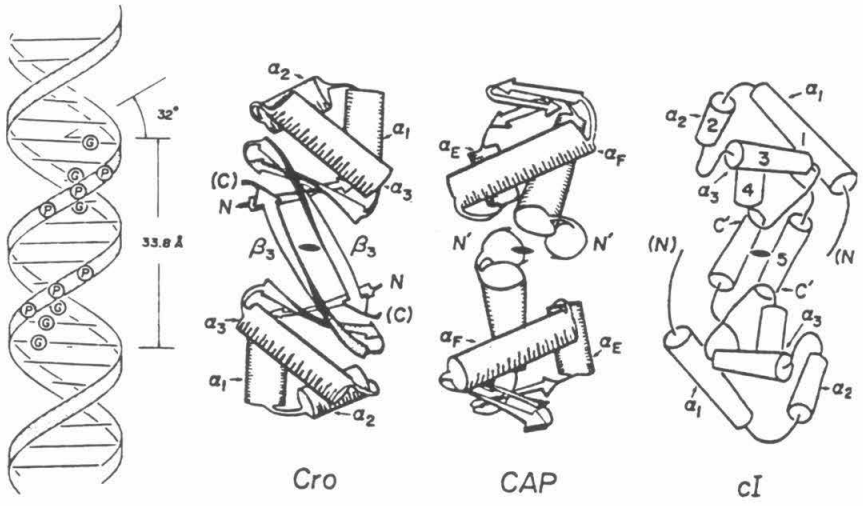


Figure 1: Schematic drawings of B-form DNA and the structures of λ Cro, CAP and λ repressor (cI). The helix–turn–helix domains are α_2–α_3 for λ Cro and λ repressor, and α_E–α_F for CAP.[28] The view of the proteins is of the face which interacts with the DNA. In the stylized DNA drawing, phosphates which when ethylated hinder Cro binding are marked, as well as G's which are protected from dimethyl sulfate by Cro (major groove).

recognition elements.[28–30] The first helix of the helix–turn-helix motif partially covers the binding helix and would stradle the DNA major groove.

The conservation of the helix–turn–helix motif between these three regulatory proteins prompted searches through other proteins for similar amino acid sequences. It was found that homologous sequences were very common in DNA regulatory proteins but rare in other proteins.[30] Table 1 shows an alignment of the putative helix–turn–helix domains of 16 sequence specific DNA binding proteins. The residues can be divided into three major classes: (1) residues involved (required) for the turn between helices and which may or may not contact the DNA, (2) residues which are internal to the protein, i.e. part of the hydrophobic core and do not contact the DNA and (3) residues which are positioned properly for use as DNA contacts.

			α_2												α_3												
				••	•									•	•••	•••			•	•••			•••			••	
PROTEIN	N	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	C
λ (cI) Repr	32	**LEU**	Ser	Gln	Glu⊖	Ser	**VAL**	**ALA**	Asp⊖	Lys⊕	**MeT**	Gly	**MeT**	Gly	Gln	Ser	Gly	**VAL**	Gly	**ALA**	**LEU**	**PHE**	Asn	Gly	**ILE**	Asn	56
λ Cro	14	**PHE**	Gly	Gln	Thr	Lys⊕	Thr	**ALA**	Lys⊕	Asp⊖	**LEU**	Gly	**VAL**	**TYR**	Gln	Ser	**ALA**	**ILE**	Asn	Lys⊕	**ALA**	**ILE**	His⊕	**ALA**	Gly	Arg⊕	38
λ & 434 cII	24	**LEU**	Gly	Thr	Glu⊖	Lys⊕	Thr	**ALA**	Glu⊖	**ALA**	**VAL**	Gly	**VAL**	Asp⊖	Lys⊕	Ser	Gln	**ILE**	Ser	Arg⊕	**Trp**	Lys⊕	Arg⊕	Asp⊖	**Trp**	**ILE**	48
P22(c2) Repr	19	**ILE**	Arg⊕	Gln	**ALA**	**ALA**	**LEU**	Gly	Lys⊕	**MeT**	**VAL**	Gly	**VAL**	Ser	Asn	**VAL**	**ALA**	**ILE**	Ser	Gln	**Trp**	Glu⊖	Arg⊕	Ser	Glu⊖	Thr	43
P22 Cro	11	Gly	Thr	Gln	Arg⊕	**ALA**	**VAL**	**ALA**	Lys⊕	**ALA**	**LEU**	Gly	**ILE**	Ser	Asp⊖	**ALA**	**ALA**	**VAL**	Ser	Gln	**Trp**	Lys⊕	Glu⊖				32
P22 cI	23	Arg⊕	Gly	Gln	Arg⊕	Lys⊕	**VAL**	**ALA**	Asp⊖	**ALA**	**LEU**	Gly	**ILE**	Asn	Glu⊖	Ser	Gln	**ILE**	Ser	Arg⊕	**Trp**	Lys⊕	Gly				44
434 (cI) Repr	16	**LEU**	Asn	Gln	**ALA**	Glu⊖	**LEU**	**ALA**	Gln	Lys⊕	**VAL**	Gly	Thr	Thr	Gln	Gln	Ser	**ILE**	Glu⊖	Gln	**LEU**	Glu⊖	Asn	Gly	Lys⊕	Thr	40
434 Cro	17	**MeT**	Thr	Gln	Thr	Glu⊖	**LEU**	**ALA**	Thr	Lys⊕	**ALA**	Gly	**VAL**	Lys⊕	Gln	Gln	Ser	**ILE**	Gln	**LEU**	**ILE**	Glu⊖	**ALA**	Gly	**VAL**	Thr	41
GAL Repr	2	**ALA**	Thr	**ILE**	Lys⊕	Asp⊖	**VAL**	**ALA**	Arg⊕	**LEU**	**ALA**	Gly	**VAL**	Ser	**VAL**	**ALA**	Thr	**VAL**	Ser	Arg⊕	**VAL**	**ILE**	Asn	Asn	Ser	Pro	26
LAC Repr	4	**VAL**	Thr	**LEU**	**TYR**	Asp⊖	**VAL**	**ALA**	Glu⊖	**TYR**	**ALA**	Gly	**VAL**	Ser	**TYR**	Gln	Thr	**VAL**	Ser	Arg⊕	**VAL**	**VAL**	Asn	Gln	**ALA**	Ser	28
TRP Repr	66	**MeT**	Ser	Gln	Arg⊕	Glu⊖	**LEU**	Lys⊕	Asn	Glu⊖	**LEU**	Gly	**ALA**	Gly	**ILE**	**ALA**	Thr	**ILE**	Thr	Arg⊕	Gly	Ser	Asn	Ser	**LEU**	Lys⊕	90
CAP	168	**ILE**	Thr	Arg⊕	Gln	Glu⊖	**ILE**	Gly	Gln	**ILE**	**VAL**	Gly	Cys	Ser	Arg⊕	Glu⊖	Thr	**VAL**	Gly	Arg⊕	**ILE**	**LEU**	Lys⊕	**MeT**	**LEU**	Glu⊖	192
HIN	158	His⊕	Pro	Arg⊕	Gln	Gln	**LEU**	**ALA**	**ILE**	**ILE**	**PHE**	Gly	**ILE**	Gly	**VAL**	Ser	Thr	**LEU**	**TYR**	Arg⊕	**TYR**	**PHE**	Pro	**ALA**	Ser	Ser	184
γδ Resolv	159	**LEU**	Gly	**ALA**	Ser	His⊕	**ILE**	Ser	Lys⊕	Thr	**MeT**	Asn	**ILE**	**ALA**	Arg⊕	Ser	Thr	**VAL**	**TYR**	Lys⊕	**VAL**	**ILE**	Asn	Glu⊖	Ser	Asn	183
Tnp Resolv	159	Thr	Gly	**ALA**	Thr	Glu⊖	**ILE**	**ALA**	His⊕	Gln	**LEU**	Ser	**ILE**	**ALA**	Arg⊕	Ser	Thr	**VAL**	**TYR**	Lys⊕	**ILE**	**LEU**	Gln				180
⊕		2	1	2	4	4	-	1	6	3	-	-	-	1	4	-	-	-	-	**10**	-	3	4	-	1	2	
⊖		-	-	-	2	7	-	-	3	2	-	-	-	1	2	1	-	-	1	-	-	3	1	2	1	1	
hydrophilic		4	**15**	**11**	**12**	**13**	2	4	**14**	7	-	**15**	2	**12**	**11**	**11**	**12**	-	**12**	**13**	1	7	**14**	9	6	**11**	
hydrophobic		**11**	-	4	3	2	**13**	**11**	1	8	**15**	-	**13**	3	4	4	3	**15**	3	2	**14**	8	1	3	6	1	
consensus		**phob**	phil	phil	phil	phil	**phob**	**phob**	phil	-	**phob**	phil	**phob**	phil	phil	phil	phil	**phob**	phil	⊕	**phob**	-	phil	-	-	phil	

Table 1: Helix–turn–helix alignment of some sequence specific DNA binding proteins. These are all regulatory proteins except for HIN, $\gamma\delta$ Resolvase and Tnp Resolvase which are recombinases. The sequences are shown N→C with hydrophobic residues in bold face capitals and charged residues marked appropriately. The columns marked N and C give the residue numbers for the residue that would normally occur in that position. The dots in the top line indicate positions that have been proposed as DNA contacts for λ Cro, λ repressor and CAP (one dot for each occurance). The brackets at the top of the table mark the α_2 and α_3 helices of λ Cro, α_3 is the binding helix. The bottom five rows tabulate the occurance of positive, negative, hydrophobic and hydrophilic residues and the consensus for each column, bold face numbers mark strong preferences for a given type of residue.

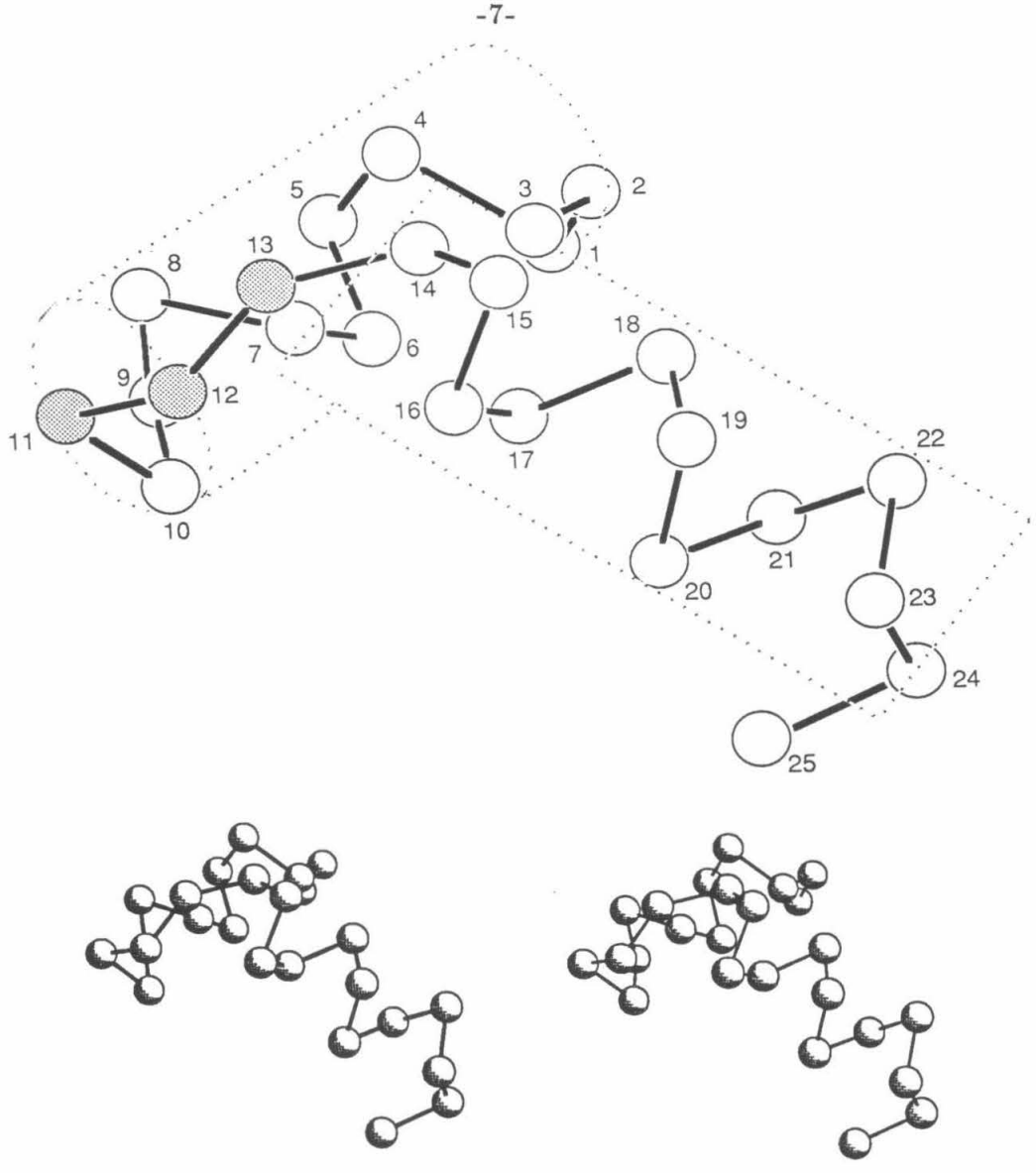


Figure 2: α Carbon trace of the helix–turn–helix domain of λ Cro (monomer B).[20,21] This domain is nearly superimposable on the equivalent regions of CAP and λ repressor. (Top) Two dimensional projection of the helix–turn–helix with the residues numbered N to C as in Table 1. The three residues in the turn are shaded and the approximate extent of the two α helices marked with dashed lines. (Bottom) Stereo view of the helix–turn–helix. In both drawings the view is the same as that which the DNA major groove sees; the binding helix is the nearer of the two and parallel to the plane of the page, the first helix is behind the binding helix.

The first class, residues in the turn, was originally thought to require a glycine at position 11 (see Table 1 and figure 2), the first position of the three residue turn, due to the unusual backbone conformation (ϕ=60°, ψ=44°) common for glycine but unusual for other residues.[30] However it has been shown that a double mutant of λ repressor containing Asp38→Asn (position 8, Table 1 and figure 2) and Gly41→Glu (position 11) is functional.[31] This example points out the difficulty of predicting structure from protein sequence similarities since multiple amino acid changes may be strongly interdependent.

The second class of residues, those contributing to the hydrophobic contacts between the helices, are positions 6, 7 and 10 in the first helix and 17 and 20 in the second. Small hydrophobic residues in positions 7 and 17 in van der Waals contact, as well as the nature of the turn, maintain the correct interhelical angle. In position 20 is usually a bulky aromatic residue which also contributes to the hydrophobic core. Interestingly tryptophan repressor (TRP repressor) contains an atypical glycine at position 20 whereas most other binding proteins have large hydrophobic residues with a particular preference for tryptophan. TRP repressor controls tryptophan biosynthesis by binding to the L-tryptophan biosynthetic operon in the presence of L-tryptophan.[32] Recently the x-ray crystal structure of the TRP repressor with bound Trp has been solved and it is clear that Trp binds in the cavity in the hydrophobic core caused by substitution of Gly for Trp at position 20.[33]

The third group of residues are those implicated as DNA contacts and are generally in positions 3, 4 and 8 in helix one, 13 (third residue of the turn), and 14, 15, 16, 18, 19 and 22 in the second helix. These are frequently hydrophilic residues and 19 is usually positively charged. For λ Cro various studies have implicated Gln

16 (position 3), Thr 17 (4), Lys 21 (7), Tyr 26 (13), Gln 27 (14) Ser 28 (15), Asn 31 (18), Lys 32 (19), His 35 (22) and Arg 38 (25) as DNA contacts. For λ repressor residues Gln 33 (13), Gln 44 (14), Ser 45 (15), Ala 49 (19), Asn 52 (22) and Asn 55 (25) and for CAP residues, Arg 180 (14), Glu 181 (15), Arg 185 (19) and Lys 188 (22) have also been proposed as DNA contacts.

DNA–Protein Cocrystal Structures

More recently two DNA–protein cocrystal x-ray structures have been solved. The structure of the 1→69 fragment of coliphage 434 repressor (a close relative to λ repressor) bound to a 14 base pair (bp) DNA oligomer supports the helix–turn–helix binding model described above.[34] In the cocrystal the 14bp oligomers are stacked end to end to form a psuedocontinous double helix and are B-DNA like.[35] Two proteins bind as a dimer to each 14bp DNA unit. Helices α_2 and α_3 are the helix–turn–helix domain and α_3 from each monomer lies in adjacent major grooves (see figure 3). The 14bp DNA fragment is overwound and slightly bent at its center causing the DNA to curl slightly around the peptide dimer and to narrow the minor groove. The DNA is slightly overwound at the ends.

The protein is folded in such a way in the complex that the amino terminus of helices 2, 3 and 4 all point towards the DNA. Important sequence specific major groove contacts are made by three glutamine side chains from α_3: Gln 28, Gln 29 and Gln 33 (positions 14, 15 and 19 in Table 1 and figure 2).

The other DNA–protein cocrystal determined recently is the complex between the 276 amino acid Eco RI endonuclease apoprotein (the protein is missing a magnesium ion cofactor required for DNA strand scission) and a 13bp cognate oligonucleotide.[36,37] The DNA recognition site for Eco RI (5′–GAATTC) is contained

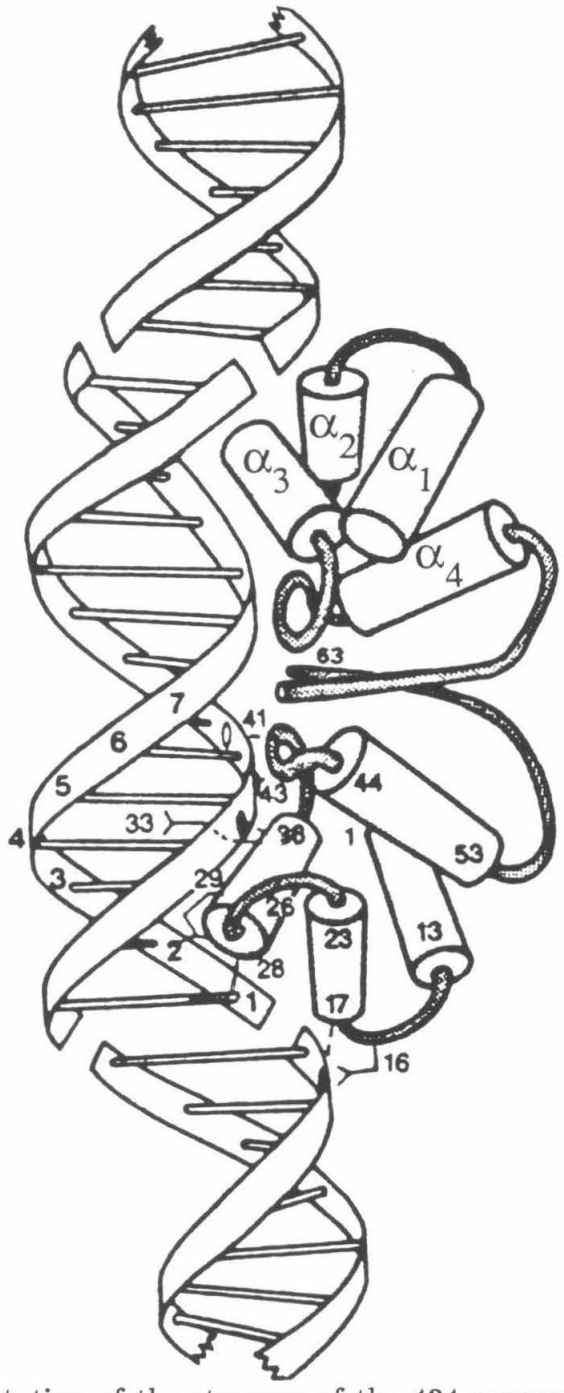


Figure 3: Representation of the strucure of the 434 repressor–DNA cocrystal.[34] The 14 bp oligomer sequence is; 5′–ACAATATATATTGT. α helices are shown as cylinders and all other protein structures as heavy lines. The lower protein monomer has some of the DNA contacting amino acids marked. The upper monomer shows the helix numbering. The DNA recognition helix is α_3.

within the self-complimentary oligonucleotide 5′–TCGCGAATTCGCG which has a one base pair 5′ overhang. As with the 434 repressor–DNA complex the DNA in the Eco RI complex forms a psuedocontinous double helix. The Eco RI interactions with DNA seen in the crystal structure are radically different from the helix–turn–helix motif seen in the 434 repressor complex which is believed to be common in the regulatory proteins. Figure 4 shows a schematic drawing of one subunit of Eco RI bound to the oligonucleotide.

The most striking features of the Eco RI complex are the absence of a helix–turn–helix domain, the severe distortions of the DNA backbone and the very tight packing of protein into a small DNA region. The Eco RI monomer consits of a five–stranded β sheet with six α helices distributed about it. The amino terminus of two of the helices ("inner" and "outer" in figure 4) from each subunit are directed into the DNA major groove and appear to be the main sequence specific recognition elements. In the dimer, then the amino terminus of a four–helix parrallel bundle is wedged into the major groove and interacts primarily with the six base pair recognition sequence.

The massive intrusion by the protein into the major groove causes the DNA backbone to be severely distorted. The backbone distortions are primarily evident as three major backbone kinks. The first kink, type I neokink (the "neo" is to differentiate these protein induced DNA kinks from DNA sequence induced kinks), is at the center of the recognition site and represents an unwinding of approximately 25°. The other two kinks (type II neokinks) are symmetrically positioned at positions -3 and +3 relative to the center of the recognition hexanucleotide. The type II neokinks span the scissile phosphate and are characterized by an increased phosphate to phosphate separation along the strand and a bend of 20–40°. The

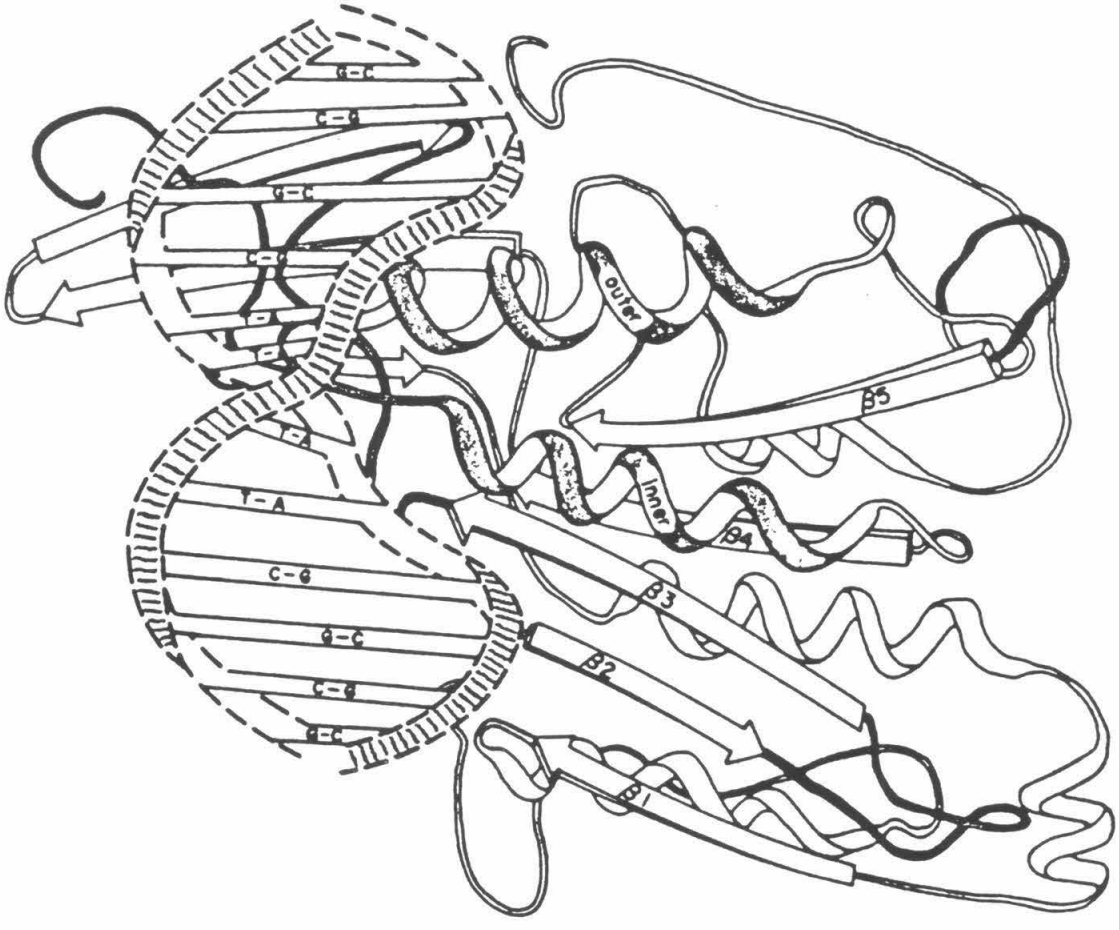


Figure 4: Schematic drawing of one subunit from the dimeric Eco RI–oligonucleotide complex.[36] The other subunit in the complex would be rotated 180 degrees about a horizontal line through the center of the first subunit in the plane of the page. The DNA oligomer sequence is 5′-TCTCGAATTCGCG-3′.

type II neokink, since it spans the scissile phosphate, appears to be intimately involved in the DNA cleavage reaction.

The radical differences between the interaction of the regulatory proteins and Eco RI with DNA may be partly due to the different requirements of the two systems. The regulatory proteins are dimeric and bind to 15–20 base pair, symmetric binding sites by inserting an α helix from each subunit into two adjacent major grooves. The intervening minor groove is covered by the protein–protein dimerization domains of the two protein monomers. Eco RI is also dimeric but its recognition sequence is only six base pairs. Helices from both protein monomers interact with the same groove and recognize the same six base pairs. The much higher degree of protein interaction per base pair observed with Eco RI may be required for two reason. First the distortions of the DNA caused by Eco RI binding may be a requirement for strand scission. Both 434 repressor and Eco RI distort the DNA, but the degree of distortion caused by Eco RI is much greater. In order to induce this degree of distortion a more substantial and tighter packed protein structure may be required. The second possible reason is the required fidelity of the recognition event. For the regulatory proteins an occasional "miss–reading" of an incorrect binding site would not be overly detrimental to the host organism. For Eco RI on the other hand, a miss–read sequence could lead to DNA strand scission, an event that is considered "fatal" to the DNA. For this reason the fidelity of an endonuclease recognition of DNA must be much higher than for the regulatory proteins, requiring a much higher degree of protein–DNA interaction packed tightly into a small DNA region.

DNA Binding Protein Fragments

Many sequence specific DNA binding proteins have been found to consists of two separable domains. One domain, frequently the smaller, contains all or most of the sequence specific DNA contacts. The other domain contains the remaining functions of the protein such as oligomerization or DNA modification. For example, limited proteolysis of 434 repressor (236 amino acids) gives a 69 amino acid amino terminal fragment which binds to repressor sites on DNA.[38] Chymotrypsin cleavage of the site-specific recombinase $\gamma\delta$resolvase (183 amino acids) gives a 43 residue carboxy terminus fragment which binds to $\gamma\delta$ recombination sites.[39] Similarly both 1→51[40] and 1→56[41] of Lac repressor (360 amino acids) bind to the 14bp LAC operator. More recently a *synthetic* 52 amino acid peptide (139→190; **HIN52**) based on the carboxy terminus of Hin recombinase (190 amino acids) has been shown to both bind to Hin recombination sites and to inhibit Hin activity.[1] Since we are interested in conversion of the **HIN52** into a sequence specific DNA cleaving molecule we shall first examine Hin recombinase.

Hin Recombinase

Hin is a 190 amino acid site specific recombinase isolated from *Salmonella Typhimurium*.[42] Its function is to switch expression of flagellin type by inverting a 995bp region of DNA which controls flagellin expression.[43,44] Figure 5 shows a proposed mechanism for Hin mediated switching.[45,46] Efficient recombination requires two protein cofactors, FIS which binds to a cis acting recombinational enhancer DNA sequence, and the histone like protein HU. The recombination event occurs between two crossover sites designated *hix*L and *hix*R, in inverted repeat configuration, when they are on a supercoiled substrate.[44] Each *hix* site is 26bp

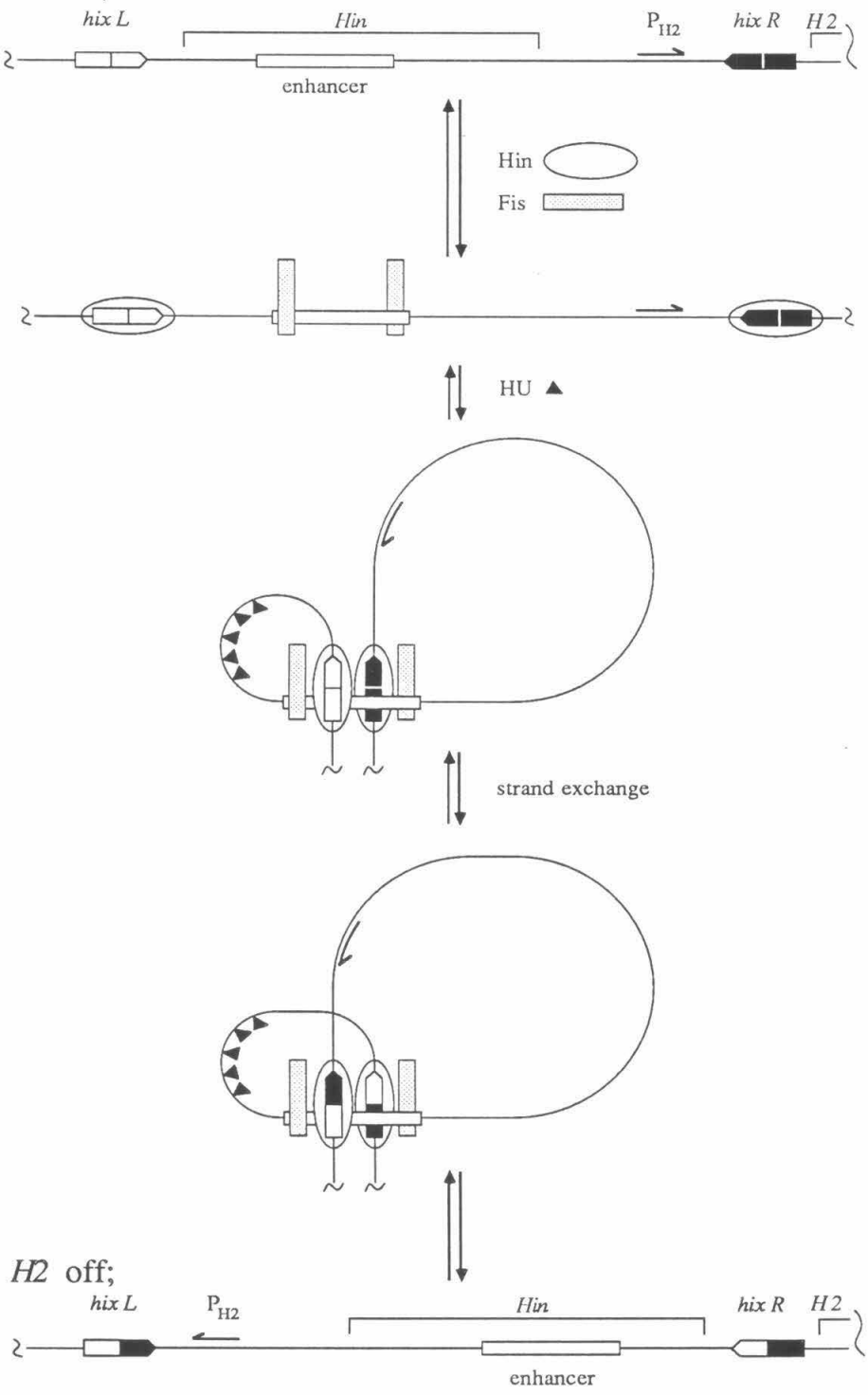
long, has near twofold symmetry, and shares a 14bp sequence with the other (see figure 6). Hin belongs to a family of recombinases that compliment one another and that includes Gin from phage Mu, Cin from phage P1 and Pin from the e14 element of *Escherichia coli*. The crossover sites for Hin, Gin, Pin and Cin are shown in figure 6 and the amino acid sequences of their carboxy–terminal domains corresponding to **HIN52** in figure 7.

hix L	ATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAA
hix R	TAAAATTTTCCTTTTGGAAGGTTTTTGATAACCAAT
gix R	... ACCGTTTCCTGTAAACCGAGGTTTTGGATAATGGTC
gix L	... ACCGTTTCCTGTAAACCGAGGTTTTGGATAAACAGC
pix R	GAGCCTTCTCCCAAACCAACGTTTATGAAAATGAAG
pix L	GCTCCTTCTCCCAAACCAAGGTTTTCGAGAGCCGTT
cix R	... CCGAGTTCTCTTAAACCAAGGTATTGGATAACAGCA
cix L	... CCGAGTTCTCTTAAACCAAGGTTTAGGATTGAAATG
consensus	-----TTcTC--AAACCAAGGTTT--GAtAA-----
consensus half-site	$TT^{a}_{c}TC--AAACC^{a}_{t}$
$\gamma\delta$ resolvase	$-a^{a}_{t}gTGTc^{c}_{t}g^{a}_{t}TA^{aa}_{tt}T^{c}_{t}AT^{a}_{g}A^{tt}_{aa}TA^{t}_{a}c^{a}_{g}gACAc^{t}_{a}t-$

Figure 6: Comparison of the cross-over sites of the Hin family of recombinases. The bases in the inverted repeats are underlined; note that both the *gix* and *cix* inverted repeat sequences extend further to the left. Also shown is the 26bp consensus site and the 13bp consensus half-site. The vertical line marks the psuedo-C_2 axis. The consensus for $\gamma\delta$ resolvase binding to the *res* subsites has also been included.

Hin binds cooperatively to a *hix* site as a dimer and the binding event has been investigated using chemical protection and interference assays. Figure 8 shows dimethyl sulfate protection and interference results for Hin binding to *hix*L.

Figure 5: Scheme for Hin mediated DNA inversion.



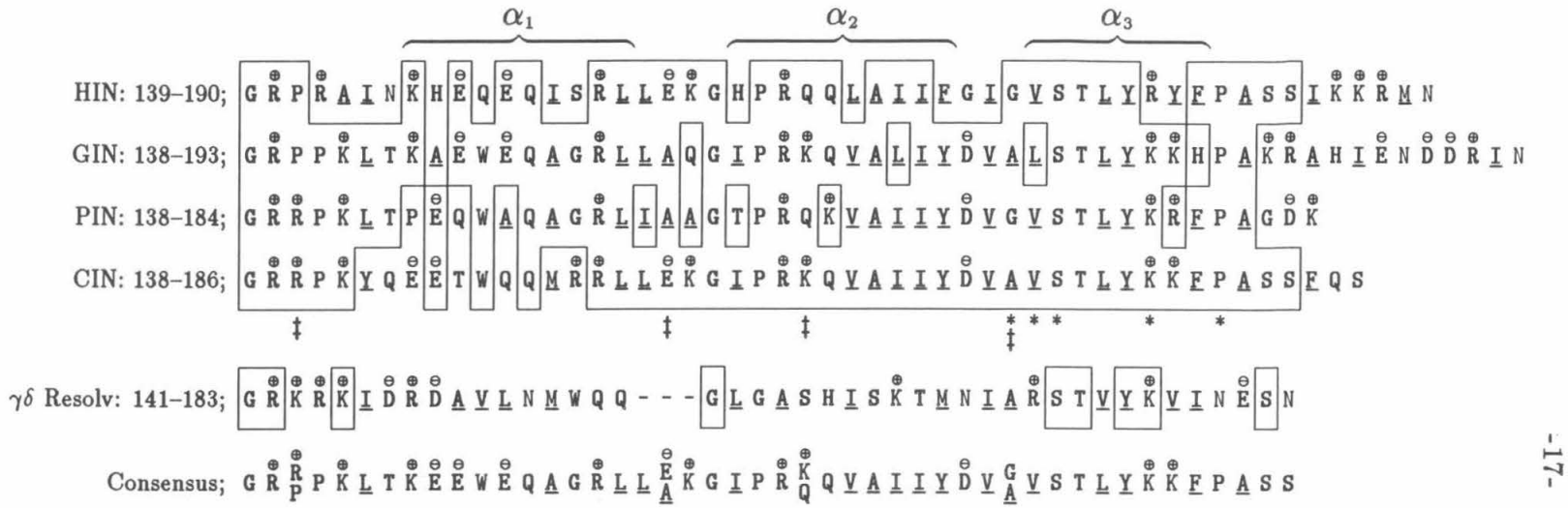
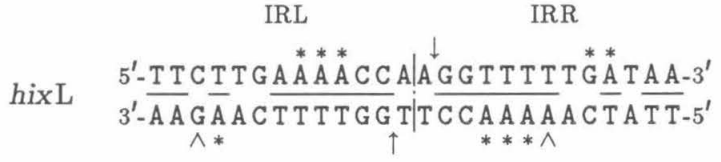


Figure 7: Protein sequence homology for the C-terminal domains of the HIN family of recombinases. Hydrophobic residues are underlined and positions that have been proposed to make DNA contacts for CAP, CRO and λ repressor are marked with an asterisk (*). There are four positions which contain two possible residues, these are marked with ‡ and considered homologous. For comparison the C-terminal 43mer chymotrypsin product of $\gamma\delta$ resolvase is included. Homologies with the consensus 46mer; HIN 31/46 (67%), GIN 39/46 (85%), PIN 38/46 (83%), CIN 39/46 (85%), $\gamma\delta$ resolvase 9/46 (20%). Net charges, excluding amino and carboxy terminal charges, for these domains are; HIN +7, GIN +5, PIN +6, CIN +6 and $\gamma\delta$ resolvase +5.

The dissociation constant for Hin binding to *hix*L has been estimated by quantitative DNase I footprinting to be 40nM at 100mM sodium chloride ($\Delta G_{assoc} \approx -10$kcal/mol). Hin protects base pairs -17 through +16 of *hix*L from Dnase I, dimethyl sulfate and methidiumpropyl–EDTA•Fe.

Dimethyl Sulfate Protection:



Dimethyl Sulfate Interference:

Figure 8: Dimethyl sulfate protection (top) and interference (bottom) results for Hin binding to *hix*L. Asterisks mark nucleotides protected from alkylation by Hin binding (top) or nucleotides which when methylated prevent Hin binding (bottom). Carets (∧) mark nucleotides which show enhanced reactivity towards dimethyl sulfate in the presence of Hin. The vertical line at the center of the sequence marks the psuedo–C_2 axis that seperates inverted repeat left (IRL) from inverted repeat right (IRR). Underlined nucleotides are conserved across the two half sites. The arrows mark the putative Hin cleavage sites. Dimethyl sulfate alkylates N3 of adenine (minor groove) and N7 of guanine (major groove).

To determine if Hin consists of two separable domains, one of which is responsible for sequence specific DNA recognition, two peptides from the carboxy terminus of Hin were chosen for total synthesis. The first was chosen by comparison of Hin to the related $\gamma\delta$ resolvase of transposon $\gamma\delta$. Since Hin and resolvase share 35% homology of amino acid sequence throughout their length,[47] the 52

amino acids from the carboxy terminus of Hin (residues 139→190, **HIN52**) are homologous to the 43 amino acid chymotrypsin cleavage product of resolvase (see figure 7). The second portion choosen for synthesis was the 31 amino acid sequence from the carboxy terminus which includes the putative helix–turn–helix domain of Hin (residues 160→190; **HIN31**) identified by protein sequence homology with the regulatory proteins such as those shown in Table 1.

HIN52 and **HIN31** were prepared by standard solid phase peptide synthesis procedures and the crude peptides' interactions with Hin recombination sites were investigated by footprinting experiments (RNase I, dimethyl sulfate and methidiumpropyl–EDTA•Fe) and by their ability to inhibit Hin mediated recombination.[1] **HIN52** inhibits Hin mediated recombination and protects the same portions of *hix* binding sites as Hin with the exception of the bases from -2 to +1 at the center of the site. The chemical protection studies indicate both minor and major groove protection for both Hin and **HIN52** (cf. figure 8). The dissociation constant for crude **HIN52** at *hix*L IRR and 100mM sodium chloride was estimated at 2μM ($\Delta G_{assoc} \approx$–7.9kcal/mol) by quantitative footprinting analysis.[1] More recently **HIN52** has been purified to homogeneity by reverse phase HPLC and its binding to *hix*L has been investigated. DNA gel mobility retardation assays indicate a dissociation constant of 140nM (20mM NaCl, $\Delta G_{assoc} \approx$–9.5kcal/mol) at *hix*L.[48] The results with **HIN52** suggest that Hin, like many other DNA binding proteins, consists of seperable domains and that the carboxy–terminal domain is responsible for sequence specific DNA binding.

HIN31 on the other hand gave no indication of sequence specific DNA binding and did not inhibit Hin mediated recombination. If Hin uses a helix–turn–helix

domain then the lack of specific binding by **HIN31** implies that though the helix–turn–helix may be required it is not sufficient for sequence specific binding.

Metal Chelators Coupled to Proteins

A number of workers have been interested in the preparation of metal chelates covalently linked to proteins, including monoclonal antibodies, for use as radiopharmaceuticals *in vivo* or as fluorescent tags for *in vitro* studies. In general this has been done by equipping a metal chelator with an electrophic function such as diazo,[49,50] isothiocyano,[51] α–bromoacetyl[51,52] or activated carboxyl[52–56] followed by random coupling to available nucleophilic sites (eg., lysyl ϵ–amino) of proteins isolated from natural sources.

More recently a modified 1,10–phenanthroline has been attached to the available lysyl ϵ–amines of TRP repressor.[57] As described previously, TRP repressor, in the presence of L-tryptophan, specifically binds to various DNA operon sites in order to control tryptophan biosynthesis. In the presence of L-tryptophan, cupric ion and 2-mercaptopropionic acid (as a reducing agent), the modified TRP repressor cleaves DNA at TRP operon sites. Analysis of the extent of 1,10–phenanthroline incorporation indicated that all four lysine residues of the TRP repressor had been modified. Despite the availability of a 2.6Å x–ray structure of the wild protein,[33] the 1,10–phenanthroline–modified lysyl residue(s) responsible for the nucleolytic activity could not be unambiguously identified.[57]

These "shotgun" approaches to chelator attachment to a protein can lead to considerable uncertainty as to the number of chelators and their positions in the modified protein.[51] Therefore, little can be said about what portions of the protein are in the vicinity of the chelator. Since we too are interested in the incorporation

of a metal chelator (EDTA) into a protein but would like to position it at a specific (known) location, we have developed a different approach.

Design of a Sequence Specific DNA Cleaving Peptide

As with the 1,10–phenanthroline•Cu described above, attachment of EDTA•Fe to a DNA binding molecule creates a DNA–cleaving molecule which functions at physiologically relevant pH, temperature and salt conditions.[58,59] If the DNA binding molecule is sequence specific then the EDTA•Fe is delivered to specific sites on the DNA.[60–63] EDTA•Fe equiped DNA binding molecules cleave DNA by oxidation of the ribose backbone via a diffusable oxidant, presumably hydroxyl radical. The cleavage reaction is dependent on a reducing agent and molecular oxygen, and the method has been called affinity cleaving. Affinity cleaving has been a successful approach to the study of sequence specific DNA recognition and has been applied to naturally occurring DNA binding antibiotics,[60,61] designed DNA binding molecules,[64–67] DNA double strand formation,[68] and DNA triple strand formation.[69]

My initial goal was to extend affinity cleaving to peptides and proteins by introduction of EDTA•Fe at a defined location. Incorporation of EDTA into a helix–turn–helix domain of a sequence specific DNA binding peptide should produce an affinity cleaving molecule which not only reports DNA binding sites of the peptide but also identifies the *groove identity* (major vs. minor for right handed B-form DNA) and *nucleotide position* of the amino acid covalently linked to the EDTA (see figure 9).

Since EDTA•Fe equiped DNA binding molecules cleave DNA via a diffusable species, cleavage is observed on both DNA strands and covers 3 to 6 base pairs.

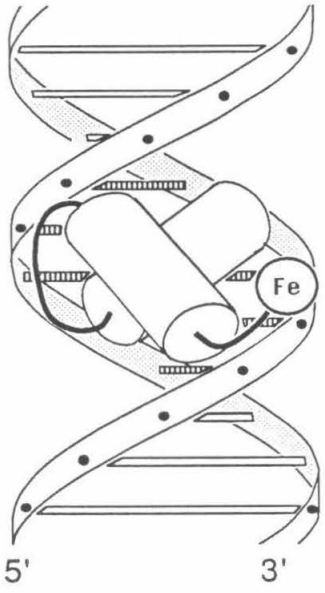


Figure 9: Stylized drawing of an EDTA•Fe equiped helix–turn–helix peptide bound in the major groove of B-form DNA.

Due to the double helical nature of right handed DNA, the groove in which the EDTA•Fe is bound can be identified by analysis of the cleavage patterns on both strands. An asymmetric cleavage pattern with maximal cleavage shifted to the 5′ or 3′ side on opposite strands corresponds to the EDTA•Fe located in/above the major and minor grooves, as shown in figure 10.[70]

Design and Synthesis of Tribenzyl–EDTA–GABA

In order to covalently link EDTA to **HIN52** at a *defined* location in the peptide, a protected EDTA derivative was prepared which is compatible with Merrifield solid phase peptide synthesis. Based on the carboxyl side chain protections of glutamic and aspartic acids used in solid phase peptide synthesis, I chose to protect three of the four carboxyl arms of EDTA as benzyl esters. The fourth carboxylate of the EDTA was coupled to a γ-amino isobutyric acid (GABA) linker to minimize interference between the EDTA chelate and the secondary or tertiary structure of

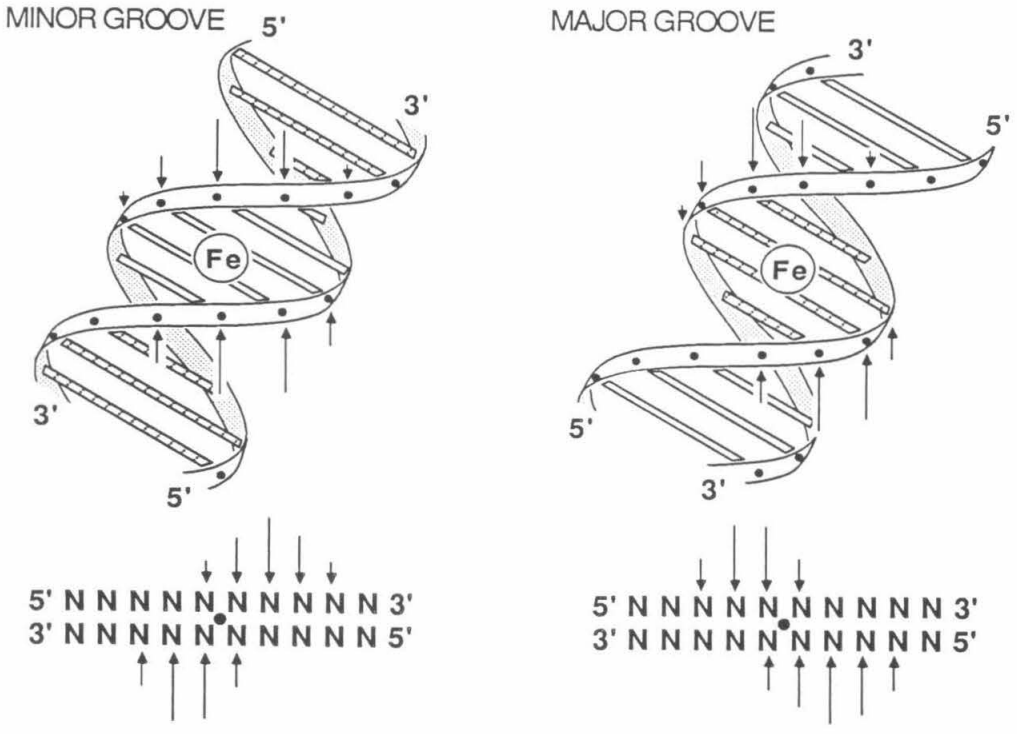
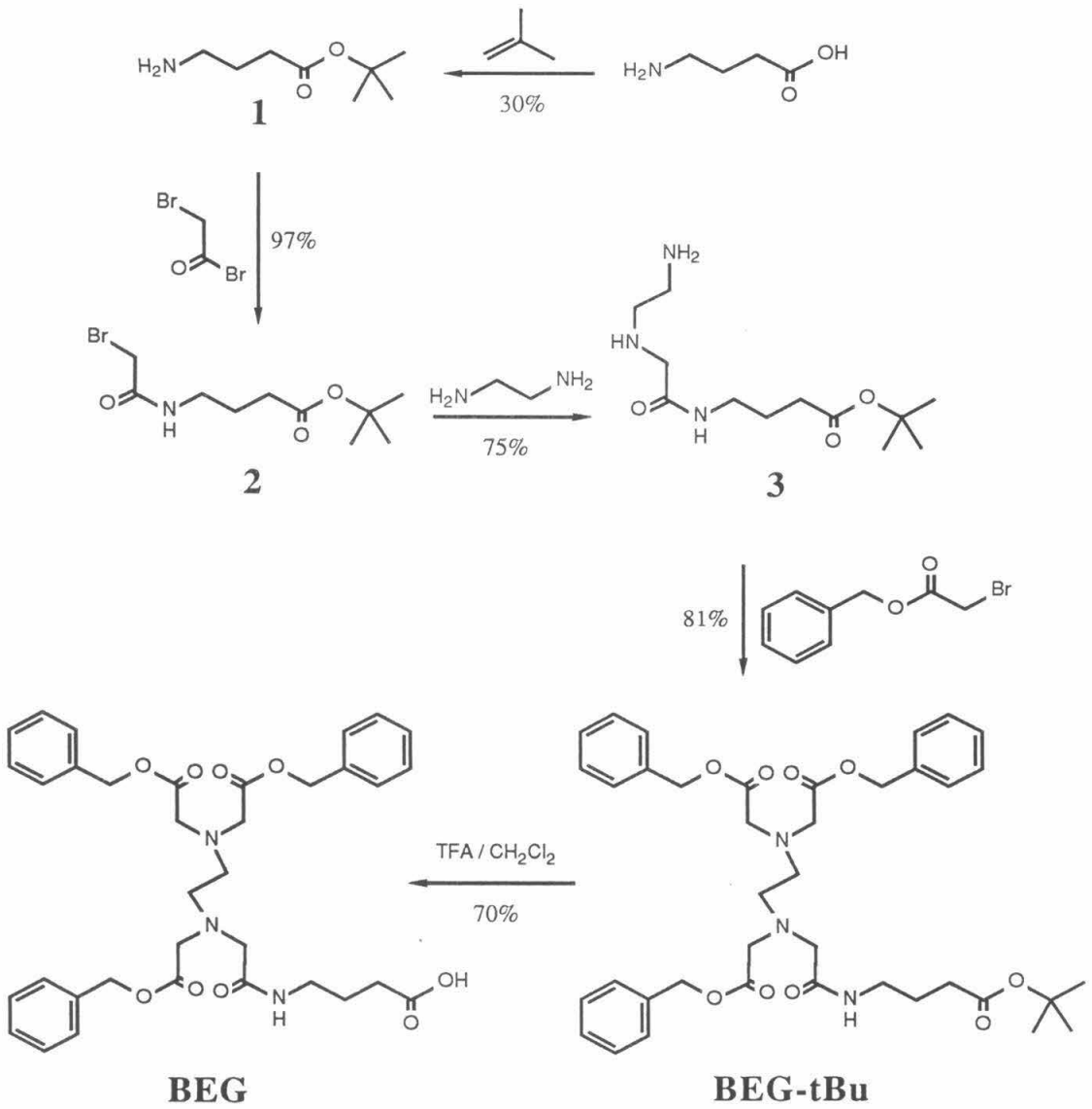


Figure 10: Cleavage patterns produced by a diffusable oxidant generated by EDTA•Fe localized in the minor and major grooves or right handed DNA. The edges of the bases are shown as open bars for the minor and crosshatched bars for the major grooves. The height of the arrows indicate the extent of cleavage expected at the indicated nucleotide.

HIN52. The resulting tribenzyl–EDTA–GABA (**BEG**) was suitable for coupling to the amino terminus of the protected, resin-bound, synthetic peptide. Though **BEG** is formally an amino acid dipeptide, it is only suitable for capping an amino terminus since it does not have a primary or secondary amino group for further chain extension.

The synthetic route to **BEG** is shown in Scheme 1. Esterification of GABA with isobutylene using the general procedure of Roeske[71] produces the carboxy protected GABA ester, **1**. Amidation of **1** with bromoacetyl bromide (**2**), followed by amination with excess ethylenediamine produces the ethylenediamine core of the

Scheme 1: Synthetic route to tribenzyl-EDTA-GABA.



$C_{35}H_{41}N_3O_9$
expected MW = 647.2842
expected + H = 648.2920
observed MW = 648.2914 ($+H^+$)
difference = 0.9 ppm

EDTA with one carboxyl arm in place, **3**. Trialkylation of **3** with three equivalences of benzyl bromoacetate in the presence of diisopropylethylamine as base generates the t-butyl ester of **BEG**. **BEG** is formed by deprotection with trifluoroacetic acid in dichloromethane followed by silica gel chromatography. All compounds were characterized by high field ^{1}H–NMR. **BEG** was also characterized by high resolution fast atom bombardment mass spectroscopy.

The synthetic target is the 52 amino acid carboxy-terminal domain of Hin (139→190) with EDTA–GABA covalently linked to the amino terminus as shown in figure 11. The coupling strategy used to attach **BEG** to the resin bound, protected **HIN52** is shown in Scheme 2. The amino–terminal t-BOC protecting group was removed, and the peptide neutralized, using standard procedures (see experimental section). **BEG** was attached to the resin bound peptide by dicyclohexylcarbodiimide (DCC) plus N-hyroxybenzotriazole (HOBt) mediated coupling in DMF. Ninhydrin analysis for primary amine indicated >99% reaction after 40 hours. The slow coupling of **BEG** to the peptide–resin may be due to the age of the peptide. This intitial coupling was carried out on peptide–resin that had been stored dry at –20°C for approximately one year. In later studies **BEG** coupling to fresh peptide–resins was complete in ≤3 hours which is comparable to standard DCC/HOBt/DMF coupling reactions used in the peptide synthesis. After washing and drying, the peptide was deprotected and cleaved from the resin using anhydrous hydrogen flouride in the presence of p-cresol and p-thiocresol scavengers. After removal of the HF under reduced pressure, the crude peptide mixture was precipitated and extensively washed with diethyl ether, redisolved in water and lyophilized affording **EG52**. Initial studies of **EG52** were carried out using this crude peptide.

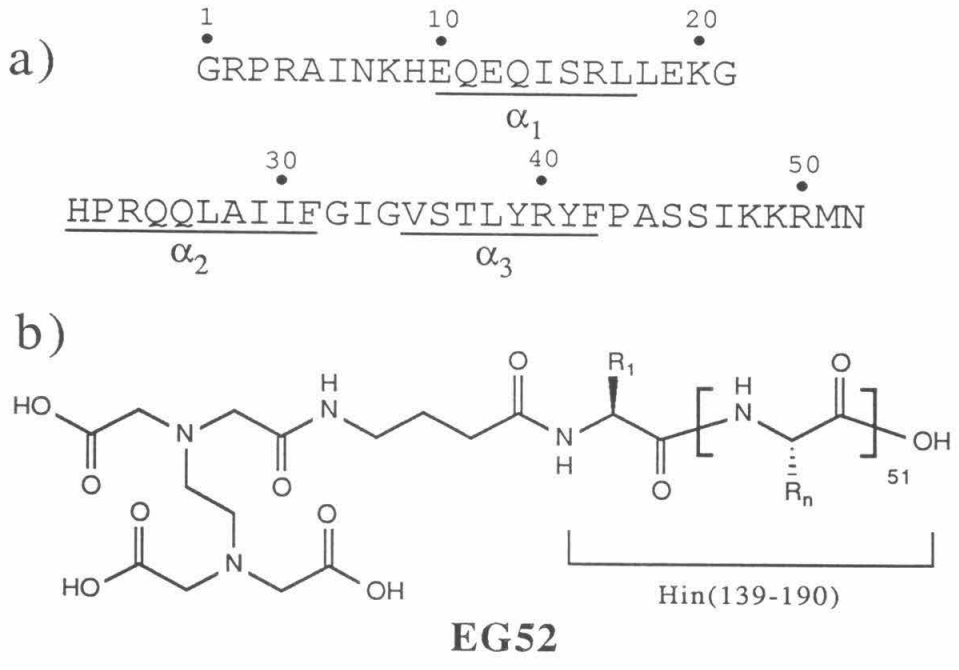


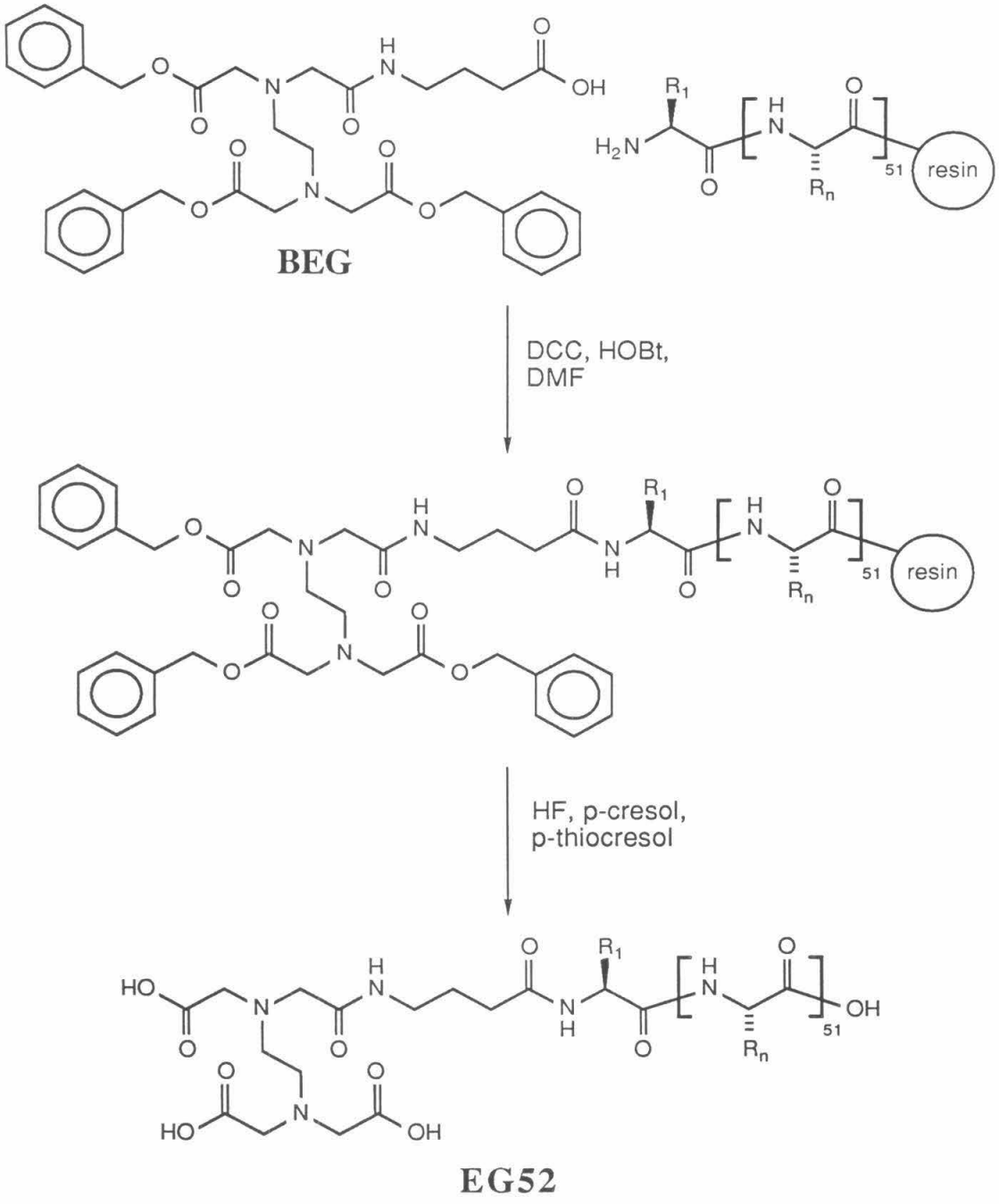
Figure 11: (a) The 52 residue peptide is the carboxy-terminal domain of Hin, amino acids 139 to 190. Tentitive α helix assignments are underlined; helices α_2 and α_3 are from protein sequence alignment with the DNA regulatory proteins in Table 1, α_3 was assigned using the secondary structure predicting algorithm of Garnier.[72] (b) The synthetic peptide **EG52**.

DNA Cleavage with EG52·Fe

The sequence specific cleavage of DNA by **EG52**•Fe was investigated by analysis of the cleavage patterns produced on a ^{32}P–endlabelled (both 5′ and 3′) DNA restriction fragment containing a shortened Hin inverting region of ~400 base pairs (the wild region is ~1kbp). The assay is outlined in figure 12.

Iron III was loaded into **EG52** by mixing one equivalence of 500μM ferric ammonium sulfate dodecahydrate with 500μM **EG52**, followed by dilution to the appropriate concentration. **EG52**•Fe was added to a solution of 5′ or 3′ endlabelled DNA buffered to pH 7.8 containing 400μM tRNA as carrier. The cleavage reac-

Scheme 2: Synthetic route to EG52.



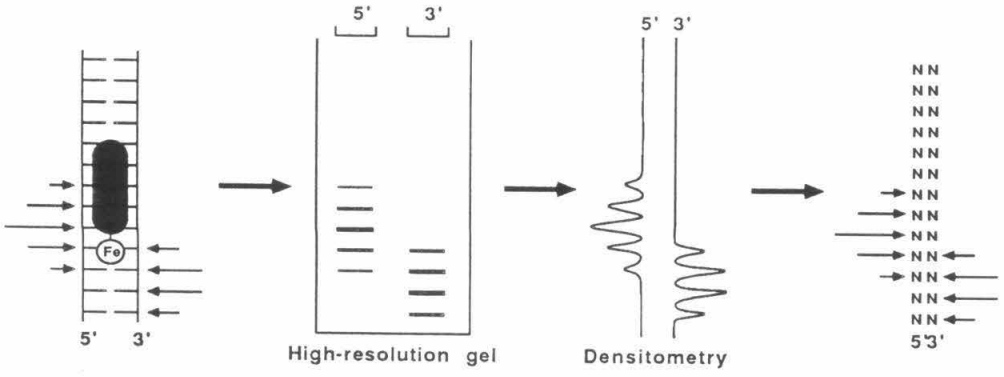


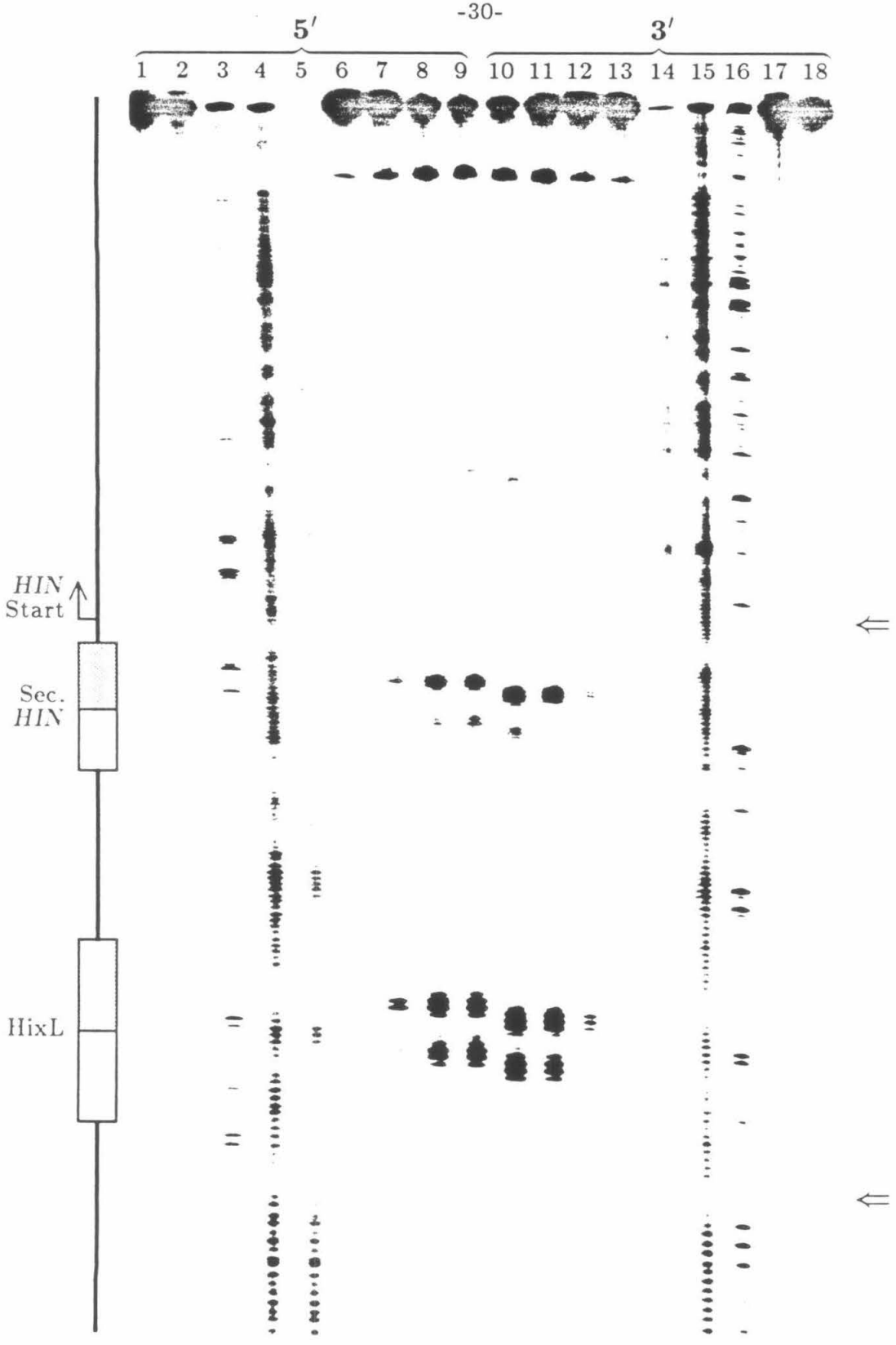
Figure 12: Outline of the high resolution DNA cleavage assay.

tion was initiated by addition of dithiothreitol (DTT) to a final concentration of 5mM. The reactions were allowed to proceed for 40–60 minutes at 22°C and terminated by ethanol precipitation. The ^{32}P–labelled DNA products were analyzed by denaturing polyacrylamide gel electrophoresis followed by autoradiography. Densitometric analysis of the gel autoradiogram and comparison of individual lanes with sequence marker lanes allowed assignment of DNA cleavage to nucleotide resolution as outlined in figure 12.

The gel autoradiogram of **EG52**•Fe induced cleavage on the 400bp fragment from pMFB36 is shown in figure 13. Two cleavage loci are observed at *hix*L corresponding to one **EG52**•Fe per half-site. Controls show that neither **HIN52** (see lanes 2 and 17 in figure 13) nor Hin give detectable DNA cleavage under these conditions (iron II/III plus DTT). For comparison with the affinity cleaving data, footprinting of **HIN52** with methidiumpropyl–EDTA•Fe (MPE•Fe)[58,59] is included in figure 13.

Histograms of MPE•Fe cleavage protection by **HIN52** and cleavage sites for **EG52**•Fe are shown in figure 14. Two **HIN52**s bound the two half-sites of *hix*L

Figure 13: Autoradiogram of high resolution denaturing polyacrylamide gel of **EG52**•Fe cleavage of a ^{32}P–endlabeled fragment (Xba I – Eco RI) from pMFB36.[1] Reaction conditions were 20mM NaCl, 20mM TRIS•HCl, pH 7.8, 400μM phosphate tRNA, 5mM DTT, and ≈15,000cpm (≈1ng) endlabelled DNA in a total volume of 10μl. Reactions were run for 40 minutes at 22°C for **EG52**•Fe and controls, and 8 minutes for MPE•Fe footprinting lanes. Cleavage products were analyzed on an 8%, 1:20 crosslinked, 50% urea denaturing polyacrylamide wedge gel (0.2mm thick at the top and 0.6mm thick at the bottom). Lanes 1–9 and 10–18 contain 5′ and 3′ endlabelled DNA respectively; lanes 1 and 18 are DNA controls; lanes 2 and 17 minus EDTA controls (10μM **HIN52**, 10μM Fe(II)); lanes 3 and 16 are Maxam-Gilbert chemical sequencing G lanes; lanes 4 and 15 are MPE•Fe controls (10μM); lanes 5 and 14 are MPE•Fe (10μM) footprints in the presence of **HIN52** (10μM); lanes 6 and 13 are **EG52**•Fe at 0.5μM, lanes 7 and 12 at 1.5μM, lanes 8 and 11 at 5μM and lanes 9 and 10 at 10μM. Concentrations of **EG52** and **HIN52** are based on the weight of crude product. (Left) Marked are *hix*L, the secondary Hin binding site (Sec HIN) and the start of the *hin* gene. (Right) Cleavage data between the arrows is shown in figure 14.

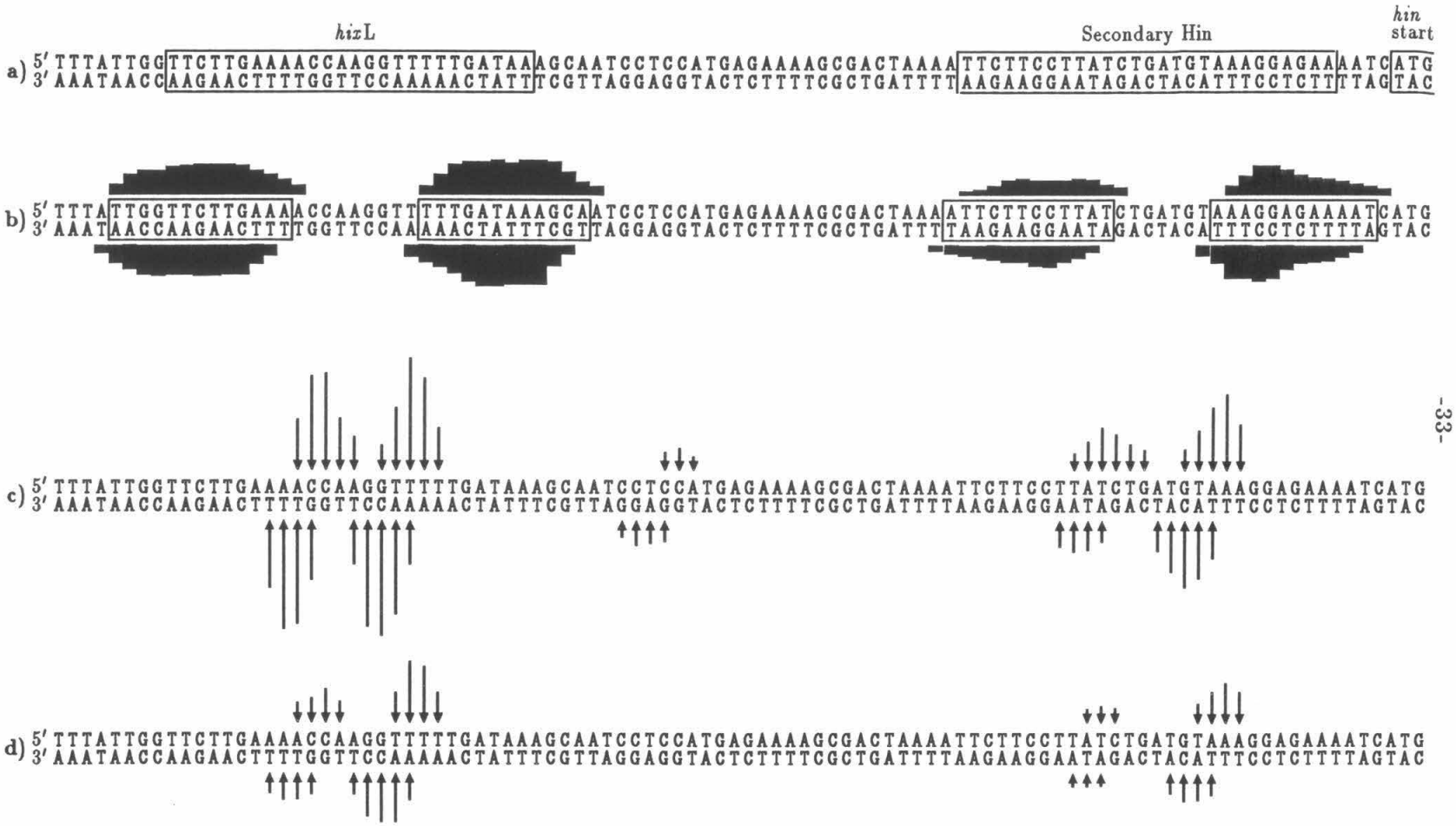

5′
3′
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
HIN
Start
Sec.
HIN
HixL

occupied by dimeric Hin.[1] **EG52**•Fe produced two cleavage patterns on the symmetry axis side of the indiviual half-sites. **EG52**•Fe appears to be similar to **HIN52**[1] in that the two half-sites were not equally bound. The other major cleavage site (secondary Hin) which is also bound by Hin, is located just upstream from the initiation codon of the *Hin* gene and though the site is not required for Hin mediated recombination its location suggests it is involved in autoregulation of Hin expression.[1]

The cleavage patterns on opposite strands of the DNA for **EG52**•Fe were shifted to the 3′ side at all sites (figure 14). This asymmetry, together with the model in figure 10, implies that the EDTA•Fe, and hence the amino terminus of the peptide, is located in or above the minor groove of the DNA. Positioning part of **EG52** in the minor groove is consistent with the protection from methylation by both **HIN52** and Hin at A(-6), A(-5), A(-4) and A(4), A(5), A(6) of *hix*L, shown in figure 8.

The gel in figure 13 was run using optimized conditions for **EG52**•Fe cleavage. These conditions were identified by studing the effects of various reaction conditions on the extent of DNA cleavage. The effect of salt concentration on the cleavage reaction was investigated at 2, 10, 20, 50, 200 amd 400mM added sodium chloride. The extent of cleavage is constant at $\leq$20mM and falls off sharply above 20mM giving ~13 times more cleavage in the 2–20mM range than at 400mM. This strong dependence on salt concentration is probably a result of the basic nature of **EG52** which is a +6 cation at neutral pH based on the peptide sequence. We have also found that inclusion of a polyanion carrier is critical for reducing the level of non-specific cleavage which again is probably related to the high net positive charge of the peptide. Calf thymus DNA, tRNA and poly-rC are all effective as carriers

Figure 14: The sequence left to right represents the cleavage data between the arrows in figure 13. (a) Boxes are the 26bp *hix*L binding site, 27bp secondary Hin binding site and the start of the *hin* gene. (b) Bars represent extent of protection from MPE•Fe cleavage in the presence of **HIN52** (figure 13, lanes 5 and 14). (c) Arrows represent the extent of cleavage for **EG52**•Fe at 10μM (figure 11 lanes 9 and 10). (d) As in (c) except 1.5μM **EG52**•Fe (figure 13, lanes 7 and 12). Extent of cleavage was determined by densitometric analysis of the gel autoradiogram.


hizL
Secondary Hin
hin start
a) 5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
b) 5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
c) 5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
d) 5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC

with 400μM phosphate tRNA giving about three times greater specific cleavage with **EG52**•Fe than 100μM bp calf thymus DNA.

The reaction time course was examined in the range of 5 to 80 minutes at 22°C and 5mM DTT. The initial rate of cleavage indicates an initial rate half-life of ~10 minutes. The cleavage reaction stops after 60–100 minutes. This is apparently due to inactivation (damage) of the EDTA and/or peptide domains and has also been observed with small molecule–EDTA compounds. The effect of equilibration time of **EG52**•Fe with DNA prior to addition of DTT was examined at 0, 10, 20 and 40 minutes (22°C). All equilibration times gave identical cleavage indicating that the equilibration rates are fast relative to the ~10 minute initial cleavage rate. As with the salt and carrier effects this is probably a result of the basic nature of the peptide which facilitates a "one dimensional" search of the DNA for binding sites, similar to the fast equilibration of DNA binding proteins.

Cleavage specificity as a function of temperature was examined at 0°C (120 minute reaction), 22° (30min) and 37° (8min) and over this temperature range no differences in specificity were detected. The effect of different reducing and oxidizing agents was also examined, compared to 5mM DTT and 1–4mM O_2 (water solubility), 1mM sodium ascorbate was 1.5 times and 1mM ascorbate plus 1mM hydrogen peroxide was three times more effective. Ascorbate plus H_2O_2, and to a lesser extent ascorbate alone, tend to give a much higher degree of background cleavage and increases the total salt concentration significantly, so that in subsequent studies only DTT was used.

After the initial report of the success of the EDTA coupling strategy[70] we acquired the capability of purifying the synthetic peptides to homogeneity by reverse

phase HPLC. The peptides were purified on a preparative C_4 reverse phase column using 0.1% trifluoroacetic acid (TFA) in aqueous acetonitrile gradients. Individual fractions from the preparative runs were checked for purity by re-injection on a C_4 analytical column and homogenous fractions were combined and lyophilized. The studies which follow were all carried out using the HPLC purified peptides. The purified peptides have all shown the same specificities as the crude peptides but have been approximately ten times more active. We have also changed our method of determining peptide solution concentrations. In initial studies concentrations were based on weight of crude peptide and the the molecular weight of the peptide amino acid sequence. In the studies which follow concentrations of purified peptides were determined by UV adsorption at 275nm using an extinction coefficient of $2800(M \cdot cm)^{-1}$ which is the expected extinction of a peptide containing two tyrosines. Typically the UV concentration obtained in this manner is 10–30% lower than that based on weight of peptide. We attribute the difference to two factors. The HPLC purified and lyophilized peptides are obtained as the TFA salts and probably contain 10–15 TFA's per peptide (15–20% by weight) based on the number of basic side chains in the peptide sequence. Secondly, the lyophilized peptides are extremely hygroscopic and readily acquire a static charge making accurate weighing difficult.

The robustness of the synthetic peptides has been somewhat surprising. No degradation of activity has been observed with lyophilized peptides stored at –20°C for 18 months, or for unbuffered solutions stored frozen, and after repeated freeze–thaw cycles, for six months. Even EDTA•Fe(III) containing peptides in unbuffered solutions keep several weeks at –20°C with no apparent change in activity.

To determine if the attachment of EDTA–GABA to **HIN52** was significantly affecting the peptides' binding ability, we examined the **HIN52** inhibition of **EG52**•Fe cleavage. Lanes 4 and 5 in figure 15 show DNA cleavage by 0.5μM **EG52**•Fe at the standard assay conditions in the absence and presence of 0.5μM **HIN52**, respectively. Densitometric analysis of cleavage at *hix*L and secondary binding sites indicates a 58($\pm$10)% reduction in **EG52**•Fe cleavage in the presence of equimolar **HIN52**. This result indicates that the EDTA–GABA is not significantly affecting the binding of the peptide at these binding sites. Even the loss of the positively charged amino terminus due to coupling of EDTA–GABA does not appear to affect DNA affinity. Interestingly, the weak cleavage by **EG52**•Fe at the tertiary site is unaffected by equimolar **HIN52**.

As mentioned earlier, the 3′ shift of the cleavage loci of **EG52**•Fe, as well as the dimethyl sulfate protection studies with **HIN52** and Hin, indicate that the amino terminus of the peptide is interacting with the minor groove of *hix*L at A_{-6}, A_{-5}, A_{-4} and A_4, A_5, A_6. To further investigate this apparent minor–groove recognition event, the ability of a minor–groove binding small molecule to inhibit **EG52**•Fe induced DNA cleavage was examined. Distamycin A (DST) is a naturally occurring tripyrrole antibiotic which binds to A+T regions of double stranded DNA.[60,61,73] Figure 15 shows the autoradiogram of **EG52**•Fe induced cleavage in the presence of DST, as well as MPE•Fe footprinting of DST and affinity cleaving by distamycin–EDTA•Fe (DE).[61] Histograms of the data in figure 15 are shown in figure 16.

Distamycin binds tightly to both A+T tracks near the center of *hix*L; $A_{-7}A_{-6}A_{-5}A_{-4}$ and $T_4T_5T_6T_7T_8$. Distamycin also binds to tracks of A+T just to the left of *hix*L, on either side of secondary Hin, and the end of the weak tertiary site. At high DST concentration (4μM), **EG52**•Fe (0.5μM) is strongly inhibited

Figure 15: Autoradiogram of high resolution denaturing polyacrylamide gel of the ^{32}P endlabelled ~400bp fragment from pMFB36 (Ava I labelled, Eco RI second cut). Lanes 1, 3→11, 16 and 18 are 5′ labelled and lanes 2, 12→15 and 17 are 3′ labelled DNA. Reaction conditions were 20mM TRIS•Cl, pH 8.0, 20mM NaCl, 400μM phosphate tRNA, 18kcpm/lane radiolabelled DNA and 5mM DTT. Reactions were run 1hr at 22°C for affinity cleaving lanes and 7.5min at 22°C for MPE•Fe footprinting lanes. All reactions were terminated by ethanol precipitation and analyzed on an 8% polyacrylamide, 1:20 crosslinked, 50% urea denaturing gel. Lanes 1 and 2 are 5′ and 3′ DNA controls, respectively. Lanes 3 and 18 are 5′ and lane 15 3′ Maxam-Gillbert chemical sequencing G marker lanes. Lanes 4→8 are 0.5μM **EG52**•Fe cleavage reactions, lanes 9→14 MPE•Fe (15μM) footprinting reactions. Lane 5 is **EG52**•Fe cleavage in the presence of 0.5μM **HIN52**. Lanes 6→8 are **EG52**•Fe cleavage in the presence of 4, 2 and 0.5μM DST, respectively. Lanes 9 and 14 are MPE•Fe control lanes, 10 and 13 MPE•Fe footprints of 2μM DST and 11 and 12 0.5μM DST. Lanes 16 and 17 are affinity cleaving with 0.5μM distamycin–EDTA•Fe (DE). Concentrations of the peptides were determined by UV adsorption at 275nm.

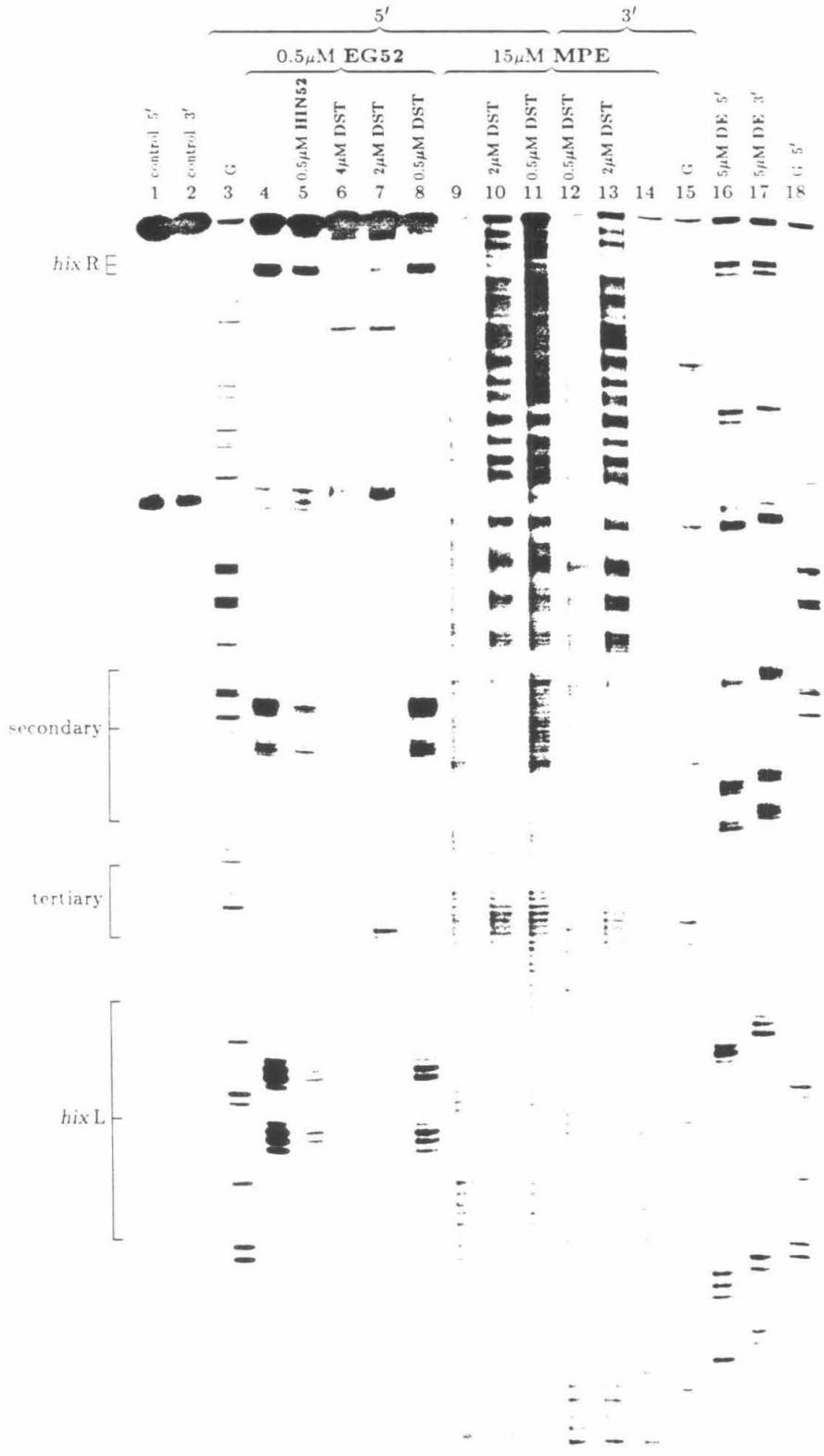
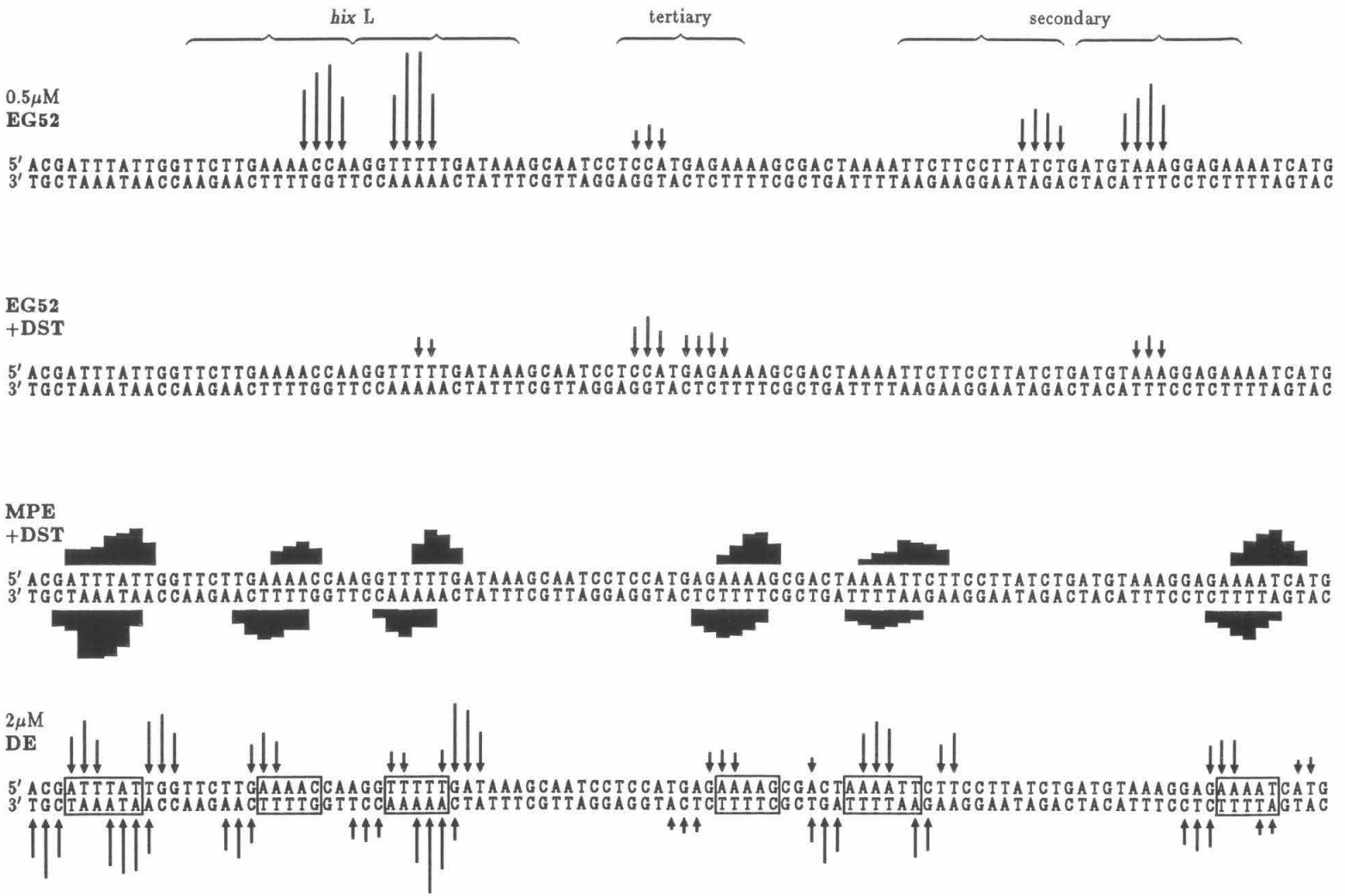

5′
3′
0.5μM EG52
15μM MPE
control 5′
control 3′
G
0.5μM HIN52
4μM DST
2μM DST
0.5μM DST
2μM DST
0.5μM DST
0.5μM DST
2μM DST
G
5μM DE 5′
5μM DE 3′
G 5′
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
hix R
secondary
tertiary
hix L

Figure 16: Histograms of the gel data in figure 15. The sequence left to right corresponds to the bottom of the gel to just above the secondary site. (Top) Brackets mark the 26bp *hix*L site, 10bp tertiary site and 27bp secondary site. Arrows represent the extent of cleavage resulting in removal of the indicated base. Bars represent the extent of protection from MPE•Fe cleavage. The top histogram is 5′ cleavage by 0.5μM **EG52**•Fe (figure 15, lane 4). The second histogram is 5′ cleavage by 0.5μM **EG52**•Fe in the presence of 2μM DST (figure 15, lanes 7). The third histogram is 5′ and 3′ MPE•Fe footprints of 0.5μM DST. The bottom histogram is 5′ and 3′ cleavage by distamycin–EDTA•Fe (DE) at 2μM, boxes represent binding sites assigned from the MPE–DST footprints and DE•Fe cleavage patterns.


hix L
tertiary
secondary
0.5μM
EG52
5′ ACGATTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ TGCTAAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
EG52
+DST
5′ ACGATTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ TGCTAAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
MPE
+DST
5′ ACGATTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ TGCTAAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
2μM
DE
5′ ACGATTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ TGCTAAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC

at all sites except tertiary, at 4:1 DST–**EG52**•Fe cleavage is inhibited at *hix*L and secondary, but enhanced at the tertiary binding site, and at 1:1 DST–**EG52**•Fe cleavage is only slightly inhibited at *hix*L and not affected at the secondary and tertiary sites.

DST binding at/near **EG52** sites can be broken down into two classes; binding at the A tracks of *hix* half–sites (see figure 16), i.e. the region where the peptide appears to interact with the DNA minor groove, and binding at sites on the ends of *hix*–like halfsites such as at *hix*L IRL, both sides of secondary and at the tertiary site. At high concentrations DST binding to all of these sites inhibits **EG52**•Fe cleavage. As the DST concentration is lowered, cleavage appears to be less affected at sites where it binds on the end, away from the center of the sites, such as at the secondary and tertiary sites. In fact, at 4:1 DST–**EG52**, binding of DST appears to enhance **EG52**•Fe cleavage at the tertiary site. On the other hand, DST binding to the central A tracks of *hix* half–sites seems to have a greater inhibitory affect on **EG52**•Fe cleavage. At *hix*L IRR the DST site is only these A's. It appears then that the strongest inhibition of **EG52**•Fe cleavage occurs when the central A track of a *hix* half–site binding site is blocked with DST.

To further investigate the effect of DST binding and to determine if **EG52**•Fe cleavage was being inhibited rather than peptide binding, the simultaneous binding of DST and **HIN52** were examined by competitive footprinting with MPE•Fe and EDTA•Fe.[74] Figure 17 shows densitometer scans of 5′ footprinting lanes at *hix*L in the presence of **HIN52**, DST and [**HIN52** plus DST]. It is clear from figure 17 that the footprint of DST in the presence of **HIN52** is identical to that of DST alone and unlike **HIN52** alone. The competitive footprinting and inhibition results indicate that DST competes for the same track of A's as is protected by

the peptide in the minor groove, and that inhibition of **EG52**•Fe cleavage is due to inhibition of peptide binding and not simply inhibition of EDTA•Fe cleavage.

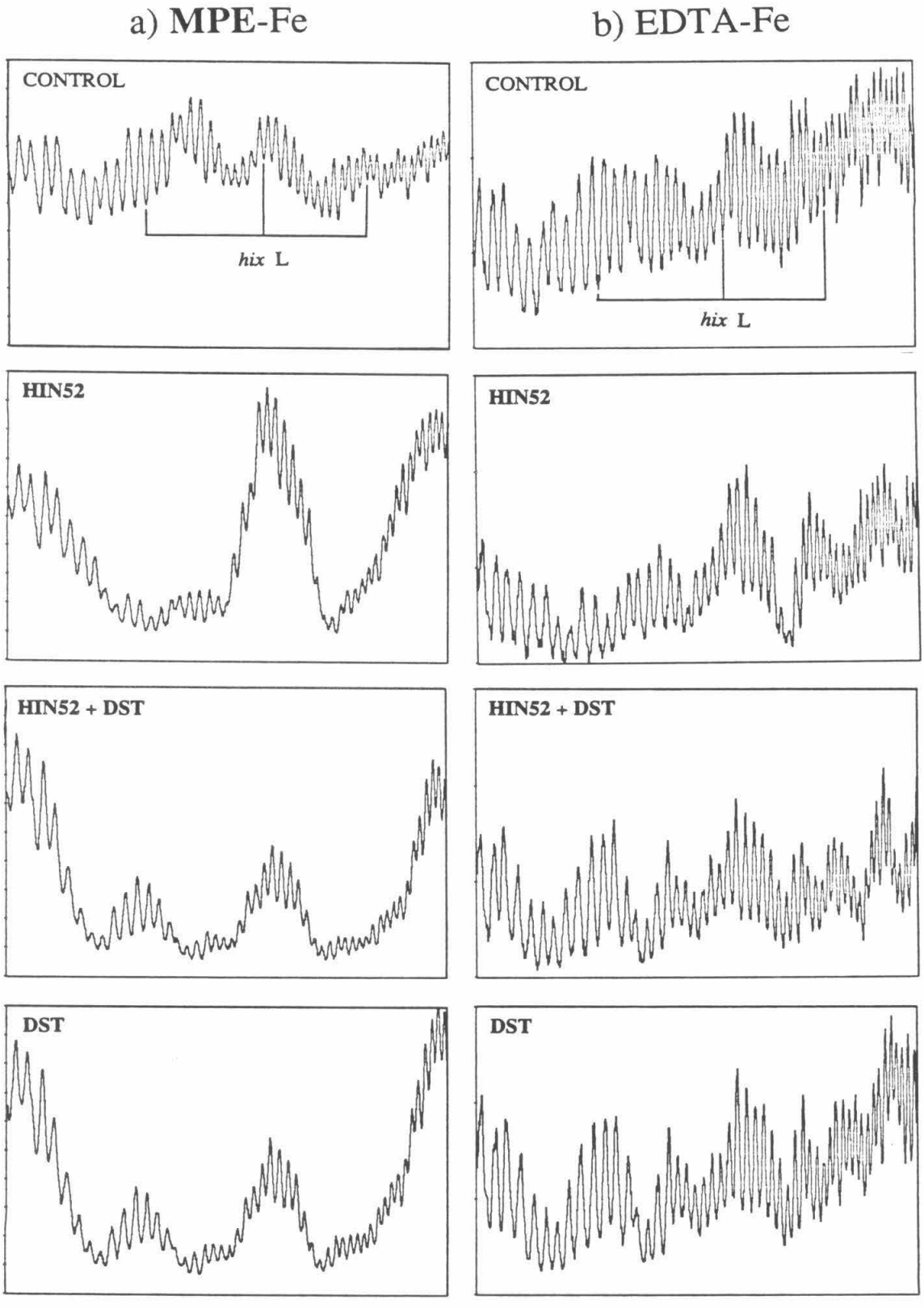
The EDTA•Fe footprinting of **HIN52** shown in figure 17 contains several unusually strongly protected bases. Analysis of the data in figure 17 (5′ strand) along with 3′ results (data not shown) indicates that the six strong protections of deoxyribose units in *hix*L by **HIN52** are symmetrically related between the two half–sites as shown in figure 18. Strong protections from EDTA•Fe cleavage would imply either very close interaction, perhaps contact, between the peptide and the deoxyribose at that base position or, alternatively, an altered DNA sugar structure which makes the sugar less susceptible to attack.

Figure 18: Asterisks indicate positions of *hix*L where the deoxyribose is strongly protected by **HIN52** from cleavage by EDTA•Fe.

DNA Double Strand Cleavage by EG52•Fe

The ability of **EG52**•Fe to cause double strand cleavage of DNA containing the *hix*L site was investigated. Double strand cleavage of linearized plasmid DNA containing the *hix*L site cloned into pBR322 is shown in figure 19. **EG52**•Fe at 5 and 10μM revealed two strong and three weak double strand cleavage sites. One strong site maps to the *hix*L site. The other strong sites maps to a location in pBR322 that is a perfect match to the outer 10bp of *hix*L IRR (5′–TTATCAAAAA–3′). The three weaker sites map to regions of pBR322 near sequences with lower degrees of homology to a *hix* halfsite (compare figure 6 and the sequences in figure

Figure 17: Densitometer scans of 5′ MPE•Fe and EDTA•Fe competitive footprinting studies. Reactions conditions were 20mM TRIS•HCl, pH 8.0, 20mM NaCl, 400μM phosphate tRNA and 20kcpm 5′ labelled DNA in 15μl total volume. The left hand column (a) is 15μM MPE•Fe footprints (5mM DTT, 7.5min at 22°C, no MPE•Fe equilibration) and the right hand column (b) 300μM EDTA•Fe footprints (2mM ascorbate, 15min at 22°C). The top two panels are control reactions in the absence of peptide and DST. The *hix*L site is marked by a bracket. The second pair of panels are footprints of 3μM **HIN52**. The third pair of panels are competitive footprints of 3μM **HIN52** and 2μM DST. The bottom pair of panels are 2μM DST footprints. All panels in a column are printed at identical scaling on both axis. Note the two unusualy strong protections by **HIN52** from EDTA•Fe cleavage in the right-hand halfsite (IRR) of *hix*L.


a) MPE-Fe
b) EDTA-Fe
CONTROL
hix L
CONTROL
hix L
HIN52
HIN52
HIN52 + DST
HIN52 + DST
DST
DST

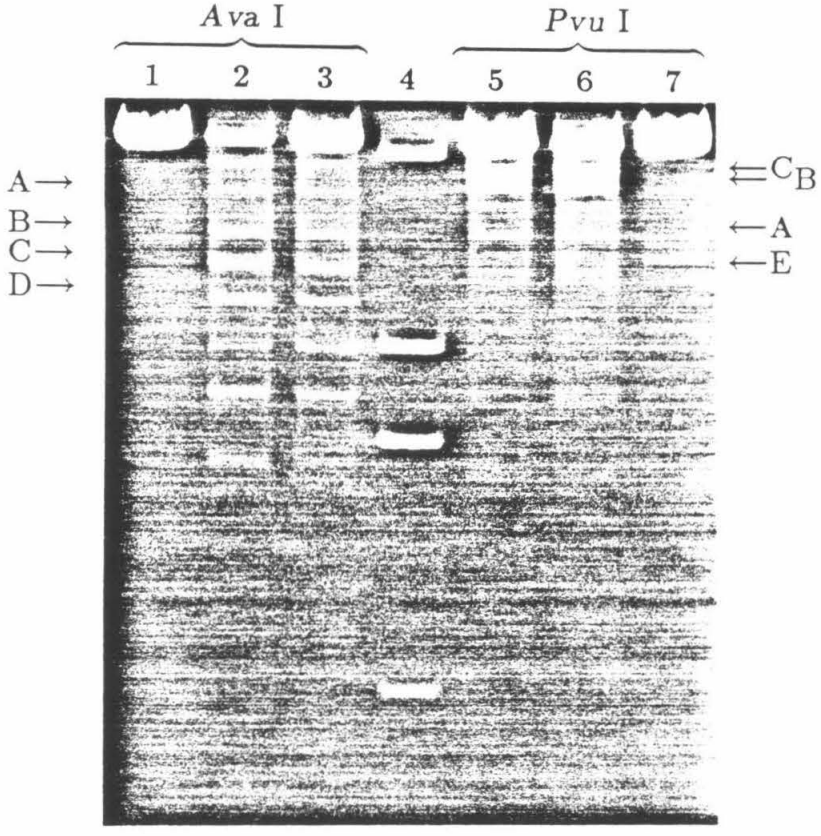
19). Identification of alternative DNA binding sites on large DNA fragments by gel electrophoresis is a useful feature of the affinity cleaving method for analyzing peptide–DNA recognition (see also Chen and Sigman[57]).

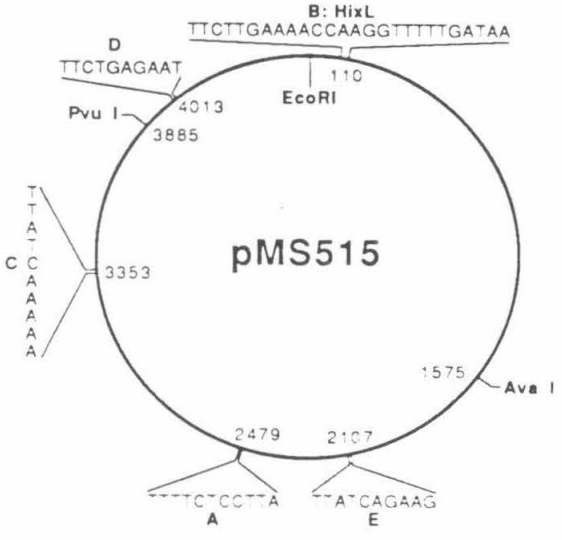
Initial Binding Model

Based on the amino acid sequence homology of Hin 160→190 (**HIN52** 22→43) with the regulatory proteins shown in Table 1, it seems likely that there is a helix–turn–helix domain within the 52 amino acid synthetic peptide. The footprinting results and binding site homology suggest that the helix–turn–helix binds at the outer five base pairs on each end of a *hix* site. In addition, the cleavage data for **EG52**•Fe has revealed the nucleotide postition and groove identity of the NH_2–terminal residue of the peptide. Using this information it is possible to propose DNA binding models for the synthetic peptides and, by extension, the DNA recognition domain of Hin binding to a *hix* site.

If Hin recognizes the binding site via specific contacts in the major groove, possibly with a helix–turn–helix structure, then the combination of the cleavage pattern with **EG52**•Fe, and the footprinting results with **HIN52**, indicates that the amino-terminal end of the peptide (Gly–$\overset{\oplus}{\text{Arg}}$–Pro–$\overset{\oplus}{\text{Arg}}$–) must have at least the Gly–$\overset{\oplus}{\text{Arg}}$– portion in/above the minor groove. Therefore, there must be a peptide structure which extends over the DNA phosphodiester backbone to follow the minor groove toward the center of the dimeric binding site. The region $C_{-2}A_{-1}A_1G_2$ at the center of a *hix* site is believed to be the location of the recombination event. The central six base pairs not bound by the 52-residue peptide would then be available for the recombination activity of the other ~140 residues of Hin. Therefore, the positions that are contacted by Hin in the major grooves are separated by one

Figure 19: (Top) Double strand cleavage from the reaction of **EG52**•Fe and linearized plasmid pMS515[75] analyzed on a nondenaturing agarose gel. A to E are fragments that resulted from one double strand break at five sites on pMS515. (Bottom) Locations of double strand breaks with Eco RI defined as the origin and the most probable recognition sequences are shown around the plasmid map along with the Eco RI, Ava I and Pvu I cleavage sites. Reaction conditions were 40mM TRIS–acetate, pH 7.9, 5mM NaOAc, 5mM DTT, 50μM bp DNA in a total volume of 10μl, 1 hour at 22°C. The products were analyzed on a 1% nondenaturing agarose gel with subsequent ethidium bromide staining and photographed under UV transillumination. Lanes 1 to 3 contain Ava I linearized plasmid, lanes 5 to 7 Pvu I linearized plasmid. Lane 4 consists of molecular weight markers of 4363, 2118, 1565 and 680bp. Lanes 3 and 5 are **EG52**•Fe at 10μM, lanes 2 and 6 at 2μM. Lanes 1 and 7 are DNA control lanes.


Ava I
Pvu I
1
2
3
4
5
6
7
A
B
C
D
C
B
A
E


B: HixL
TTCTTGAAAACCAAGGTTTTTGATAA
D
TTCTGAGAAT
110
EcoRI
Pvu I
4013
3885
pMS515
3353
1575
Ava I
2479
2107
A
E

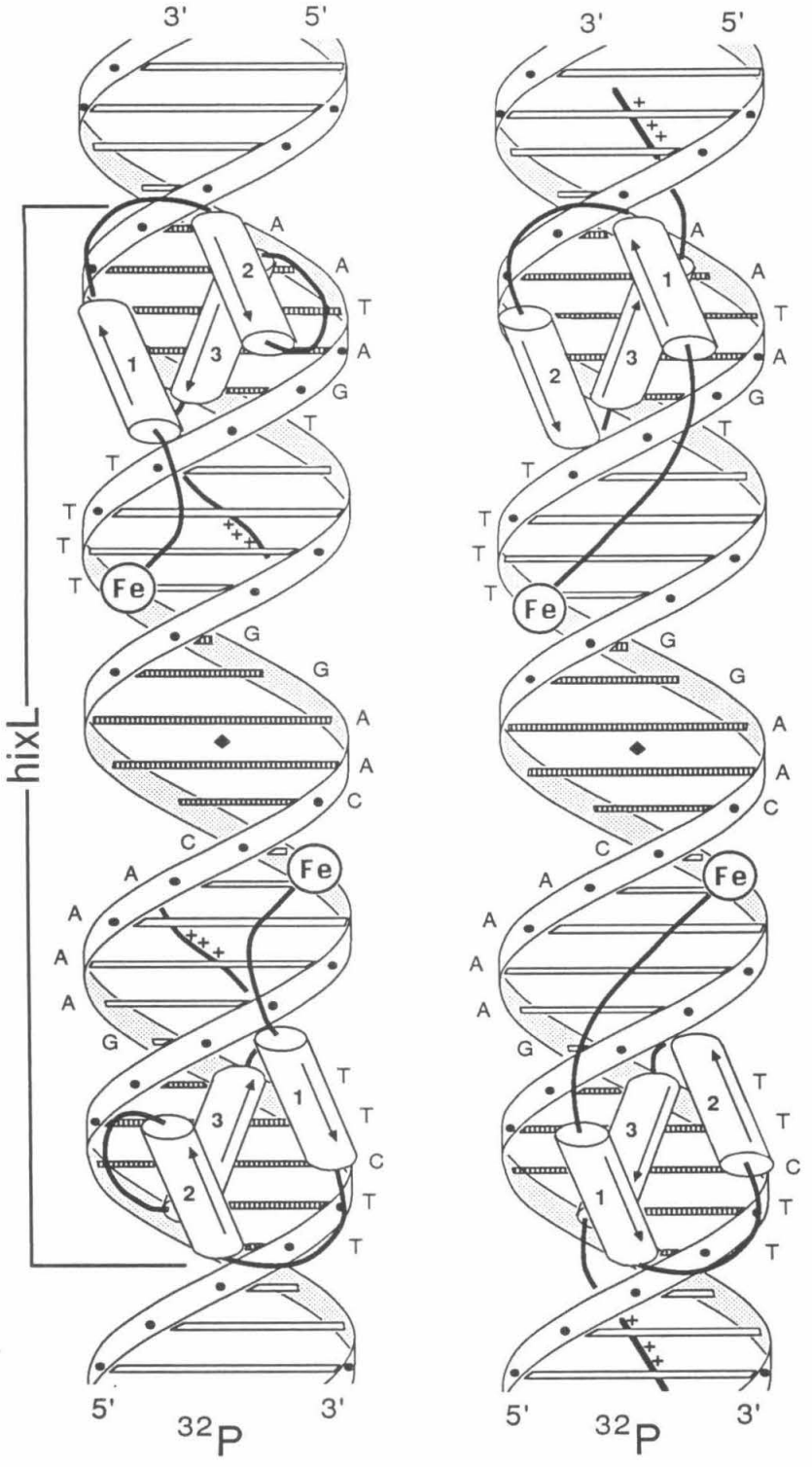
turn of the DNA helix. A similar model for $\gamma\delta$ resolvase has been proposed.[39] Furthermore, the COOH-terminal end of Hin contains several positively charged residues (–$\overset{\oplus}{\text{Lys}}$–$\overset{\oplus}{\text{Lys}}$–$\overset{\oplus}{\text{Arg}}$–Met–Asn-$\text{CO}_2^{\ominus}$) similar to the amino-terminus of λ repressor ($\text{NH}_3^{\oplus}$–Ser–Thr–$\overset{\oplus}{\text{Lys}}$–$\overset{\oplus}{\text{Lys}}$–$\overset{\oplus}{\text{Lys}}$–), which are believed to form an arm that wraps around the DNA following the major groove.[76] The carboxy-terminal region of Hin may be positioned to fulfill a similar role.

A binding model for the DNA recognition domain of Hin may involve three sets of interactions; (1) a helix–turn–helix domain binding in the major groove at the outer five base pairs on both ends of a *hix* site (5′-TTCTT-3′ and 5′-TTATC-3′ for *hix*L), (2) a region that connects the helix–turn–helix domain to the other ~140 residues of Hin by following the adjacent minor groove sequence 5′-AAA-3′ toward the center of the *hix* site, and (3) a carboxy-terminal arm which follows the major groove. Two models which incorporate these features are shown in figure 20.

The three α helices are based on the α_1-α_2-α_3 helices of Cro, and λ and 434 repressors. The α_3 helix is placed in the major groove. The two models differ in their orientation of the α_3 recognition helix relative to the C_2 axis of the *hix* binding site. For one model, the helix orientation is based on the 434 repressor cocrystal, and in the other, on the complex of LAC repressor headpiece with a 14bp LAC operator studied by two-dimensional nuclear magnetic resonance.[41]

Since **HIN31** contains the putative helix–turn–helix domain but does not bind to Hin recombination sites,[1] it would seem that the helix–turn–helix may be necessary, but not sufficient, for binding in this system. Binding may require a third α helix and/or structures that bind DNA in the minor groove to hold the DNA-binding helix–turn–helix in the proper configuration.

Figure 20: Schematic representation of two models for **EG52**•Fe and, by extension, the DNA recognition domain of Hin binding to *hix*L. The sequence of *hix*L is shown along the 5′ strand in the same direction as in figure 14. The psuedo–C_2 axis is indicated by a solid diamond. Putative α helices (see figure 11) are shown as cylinders with an arrow pointing from the NH_2- to the COOH-terminus. The model on the left is based on the 434 repressor fragment–14bp operator cocrystal[35] with the following changes; (i) 434 repressors α_4 helix is replaced with the ten residue carboxy teminus tail of Hin, (ii) the nine residue amino tail of **EG52**, which includes the EDTA, is added as an extension to the 434 repressor α_1 helix amino terminus. The model on the right represents the recognition helix orientation of the Lac repressor headpiece[41], which is rotated 180° from the 434 operator complex and Cro-OR3 model. This requires reorganization of the remaining helices in order to maintain the proper positioning of the amino terminus of the peptide. The α_3 helix of Lac repressor is replaced by the carboxy–terminal tail of Hin. α_2 of **EG52** is reversed so that the amino–terminus of the α_1 helix is towards the C_2 axis of *hix*L as required by the affinity cleaving data.


3'
5'
hixL
Fe
Fe
Fe
5'
32P
3'
3'
5'
Fe
Fe
Fe
5'
32P
3'

Extent of DNA Binding Domain

The success of the EDTA–GABA attachment to the synthetic peptide allowed us to address several other issues regarding the sequence specific DNA binding domain of Hin. In particular, is the entire 52 amino acid sequence required for sequence specific binding or do shorter peptides contain all of the required elements? Conversely does the 52 amino acid sequence contain *all* of the important carboxy terminus DNA recognition residues?

To investigate these two issues we, in collaboration with Dr. Suzanna Horvath and the Caltech microchemical facility, prepared a series of peptides which include shorter and longer versions of **HIN52**. Since it had already been shown that the 31 residue carboxy terminus of Hin (**HIN31**) does not bind sequence specifically to DNA we chose peptides ranging in size from 33 to 60 amino acids, all from the carboxy terminus of Hin. In particular we prepared the 33, 37, 45, 49, 50, 51, 56 and 60 amino acid carboxy–terminal sequences of Hin. We also prepared the 10 amino acid amino terminus of **HIN52** (Hin 139→148). The complete amino acid sequences of these synthetic peptides, along with the 31mer and 52mer already described, are shown in figure 21. Two versions of each peptide were prepared, one with EDTA–GABA at the amino terminus and one without.

All peptides were prepared as described for **EG52** and HF cleaved. The 49, 50, 51, 56 and 60mers, both with and without EDTA–GABA, were purified to homogeneity by reverse phase HPLC.

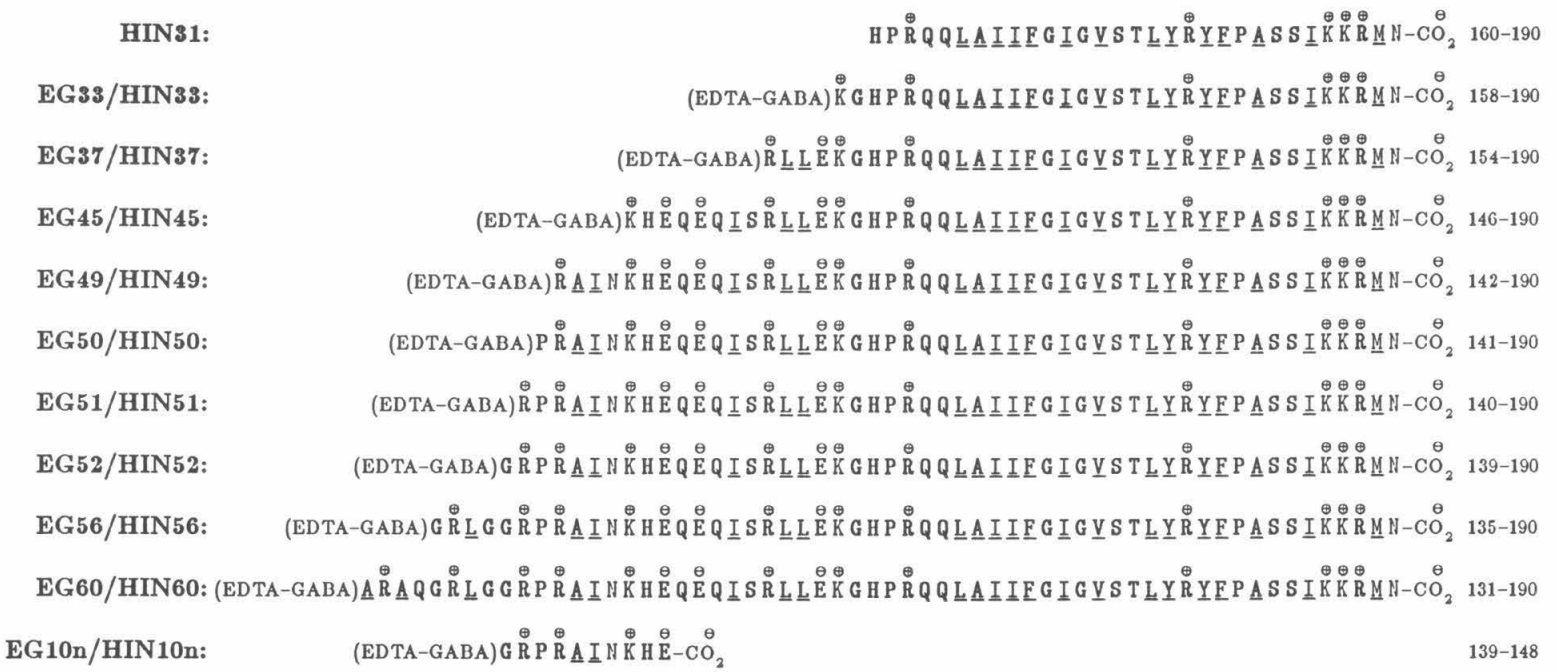


Figure 21: Amino acid sequences of the synthetic peptides. Hydrophobic residues are underlined and charged residues are marked with the appropriate sign. With the exception of **HIN31** each peptide was prepared with and without EDTA–GABA at the amino terminus. Note that the peptides without EDTA–GABA have an additional positive charge at the amino terminus. The right hand column lists the corresponding residue numbers for HIN.

DNA Binding

DNA Binding by 10n, 33, 37 and 45mers: These peptides showed no ability to sequence specifically cleave DNA. Even at 50μM EDTA–peptide concentrations, ~500 fold higher than the sequence specific cleavage detection limit for **EG52** (0.1μM), no specific cleavage was detected. Additionally **HIN10n**, **HIN33**, **HIN37** and **HIN45** at 50μM did not inhibit cleavage by 0.5μM **EG52**•Fe, or give detectable MPE•Fe footprints.

The EDTA–GABA equipped 49, 50, 51, 56 and 60 residue peptides all showed sequence specific DNA cleavage ability and **HIN49** through **HIN60** gave MPE•Fe footprints. Figure 22 shows DNA cleavage by **EG49**•Fe → **EG60**•Fe and figure 23 shows histograms of the sequence specific DNA cleavage. The sequence specific DNA binding abilities of **HIN49** through **HIN60** were examined by MPE•Fe footprinting. The high resolution gel autoradiogram for the MPE•Fe footprinting is shown in figure 24 and footprinting histograms in figure 25.

DNA Binding by 56 and 60mers: Unlike **EG52**, both **EG56** and **EG60** give very broad cleavage patterns at *hix*L which cover ~10 base pairs. However **HIN56** and **HIN60** give strong footprints at the same concentration as **HIN52**. The similar footprinting concentrations indicate that these three peptides have similar binding affinities. The 56 and 60mers also appear to have the same relative affinities for the secondary and tertiary sites as does the 52mer. It appears then that no additional DNA recognition elements are present in either the 56 or 60mers. Furthermore, the very broad cleavage patterns obtained with **EG56**•Fe and **EG60**•Fe suggest that the peptide backbone (hence the EDTA•Fe) is not held as close to the DNA as it is for **EG52**•Fe. In Hin then, these residues (131→138)

Figure 22: Autoradiogram of a high resolution denaturing polyacrylamide gel of affinity cleaving reactions on a ^{32}P–endlabeled fragment (Xba I – Eco RI) from pMFB36.[1] Reaction conditions were 20mM NaCl, 20mM TRIS•HCl, pH 8.0, 400μM tRNA, 5mM DTT, and ≈15,000cpm (≈1ng) endlabelled DNA in a total volume of 15μl. Reactions were run for 60 minutes at 22°C and terminated by ethanol precipitation. Cleavage products were analyzed on an 8%, 1:20 crosslinked, 50% urea denaturing polyacrylamide gel. Lanes 1 through 8 contain 5′ labelled and 9 through 16 3′ labelled DNA. Lanes 1 and 16 are DNA controls and 2 and 15 Maxam–Gilbert chemical sequencing marker lanes. Lanes 3 and 14 contain 10μM **EG49**•Fe, 4 and 13 contain 10μM **EG50**•Fe, 5 and 12 contain 0.5μM **EG51**•Fe, 6 and 11 contain 0.5μM **EG52**•Fe, 7 and 10 contain 2μM **EG56**•Fe and 8 and 9 contain 2μM **EG60**•Fe. (Left) Brackets mark *hix*R, secondary, tertiary and *hix*L binding sites.

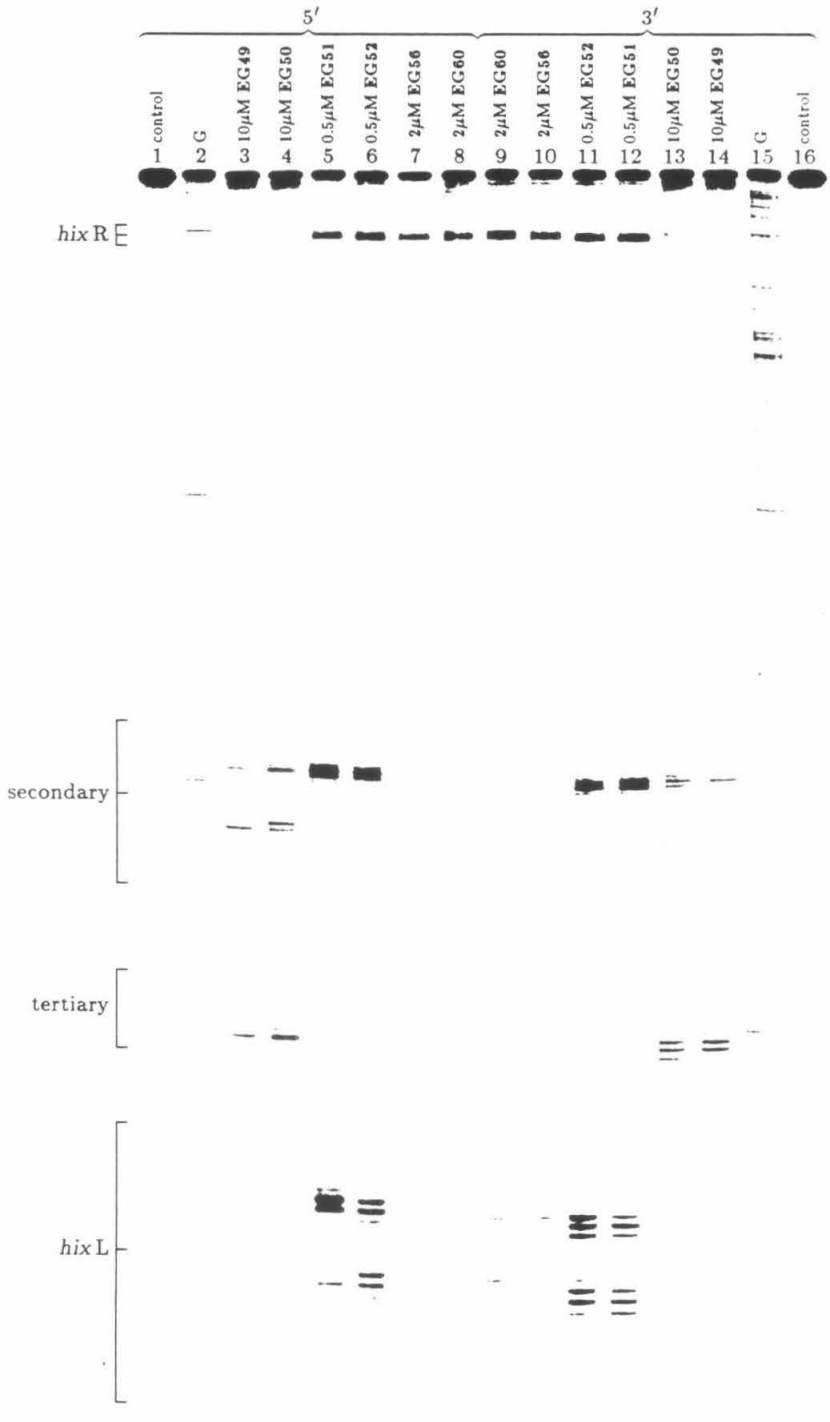

5′
3′
control
G
10μM EG49
10μM EG50
0.5μM EG51
0.5μM EG52
2μM EG56
2μM EG60
2μM EG60
2μM EG56
0.5μM EG52
0.5μM EG51
10μM EG50
10μM EG49
G
control
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
hix R
secondary
tertiary
hix L

Figure 23: Histograms of the gel data in figure 22. The sequence left to right is the bottom of the gel to just above the secondary site in figure 22. (Top) Brackets mark *hix*L, tertiary and secondary sites. Arrow indicate extent of cleavage resulting in removal of the indicated base. Dots mark the EDTA•Fe location assigned using the model in figure 10.

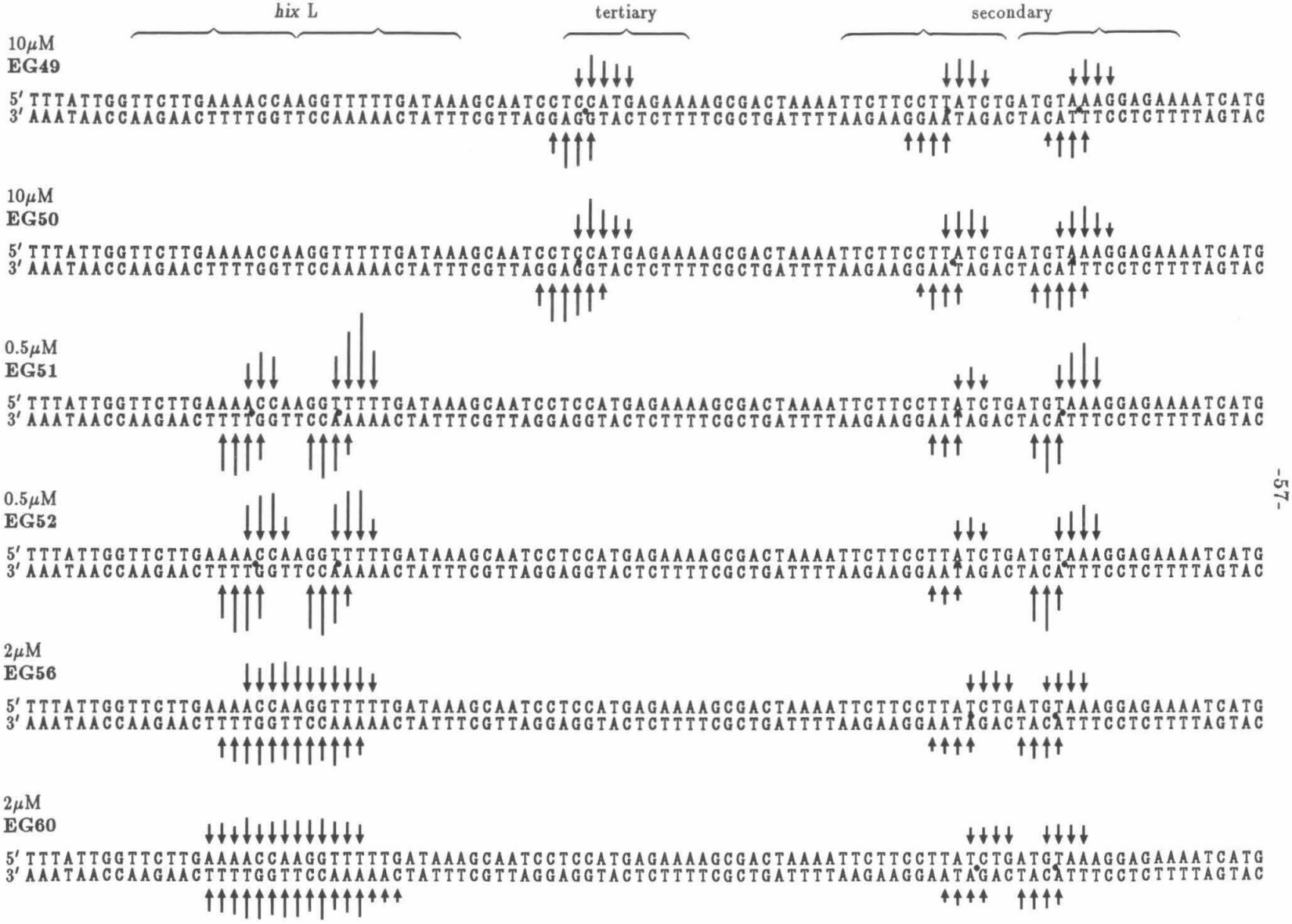

hix L
tertiary
secondary
10μM
EG49
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
10μM
EG50
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
0.5μM
EG51
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
0.5μM
EG52
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
2μM
EG56
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
2μM
EG60
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC

Figure 24: Autoradiogram of high resolution denaturing polyacrylamide gel of MPE•Fe footprinting of **HIN49**, **HIN50**, **HIN51**, **HIN52**, **HIN56** and **HIN60** on a ^{32}P–endlabelled fragment (Xba I – Eco RI) from pMFB36.[1] Reaction conditions were 20mM NaCl, 20mM TRIS•HCl, pH 8.0, 400μM tRNA, 5mM DTT, and $\approx$20kcpm ($\approx$1ng) endlabelled DNA in a total volume of 15μl. Reactions were run (without MPE equilibration) for 7.5 minutes at 22°C and terminated by ethanol precipitation. Cleavage products were analyzed on an 8%, 1:20 crosslinked, 50% urea denaturing polyacrylamide gel. Lanes 1 through 9 contain 5′ labelled and 10 through 18 3′ labelled DNA. Lanes 1 and 18 are DNA controls and 2 and 17 Maxam–Gilbert chemical sequencing marker lanes. Lanes 3 through 16 are footprinting reactions with 15μM MPE•Fe; Lanes 4 and 15 contain 20μM **HIN49**•Fe, 5 and 14 contain 20μM **HIN50**•Fe, 6 and 13 contain 4μM **HIN51**•Fe, 7 and 12 contain 2μM **HIN52**•Fe, 8 and 11 contain 2μM **HIN56**•Fe and 9 and 10 contain 2μM **HIN60**•Fe. (Left) Brackets mark *hix*R, secondary, tertiary and *hix*L binding sites.

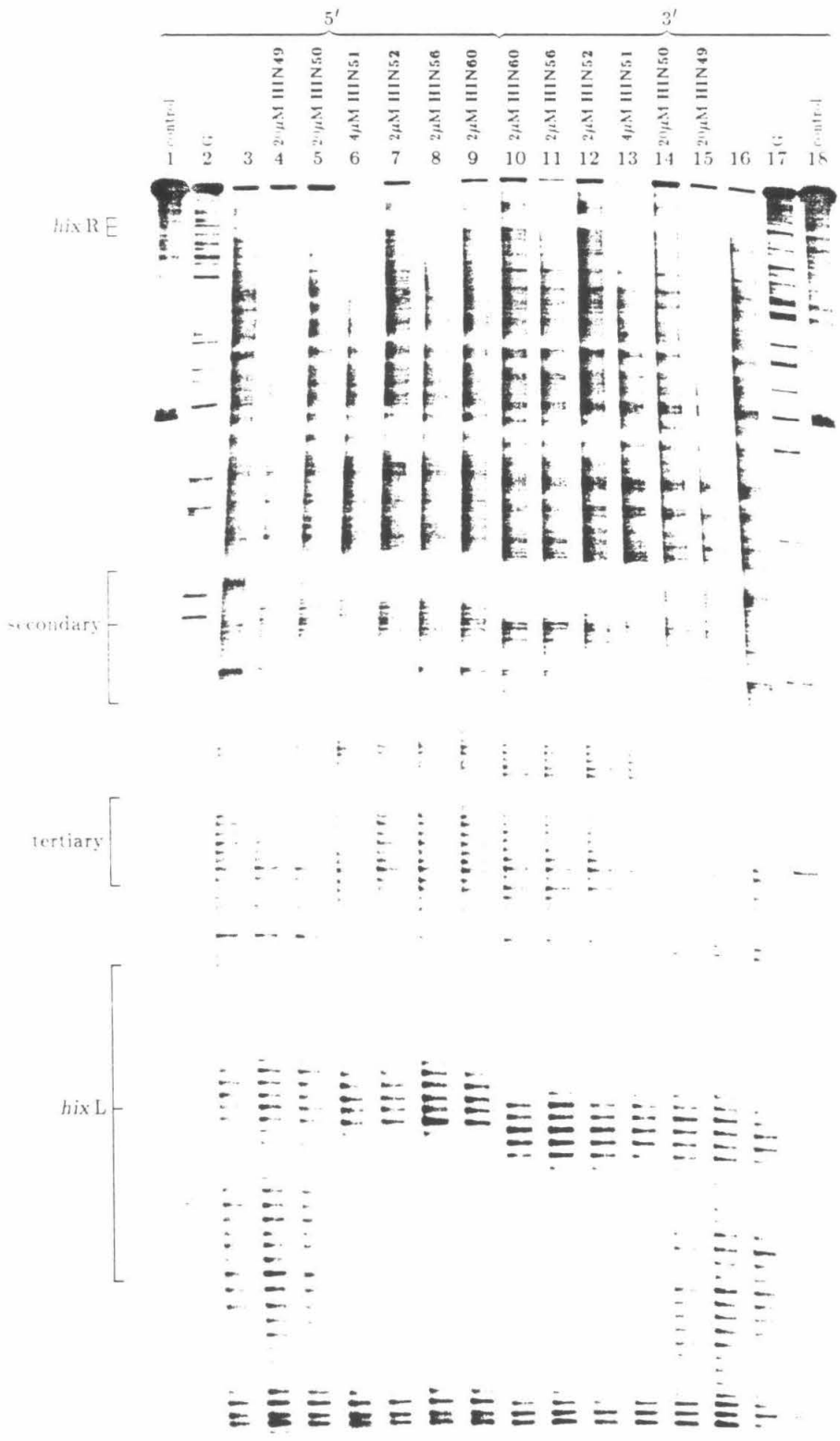
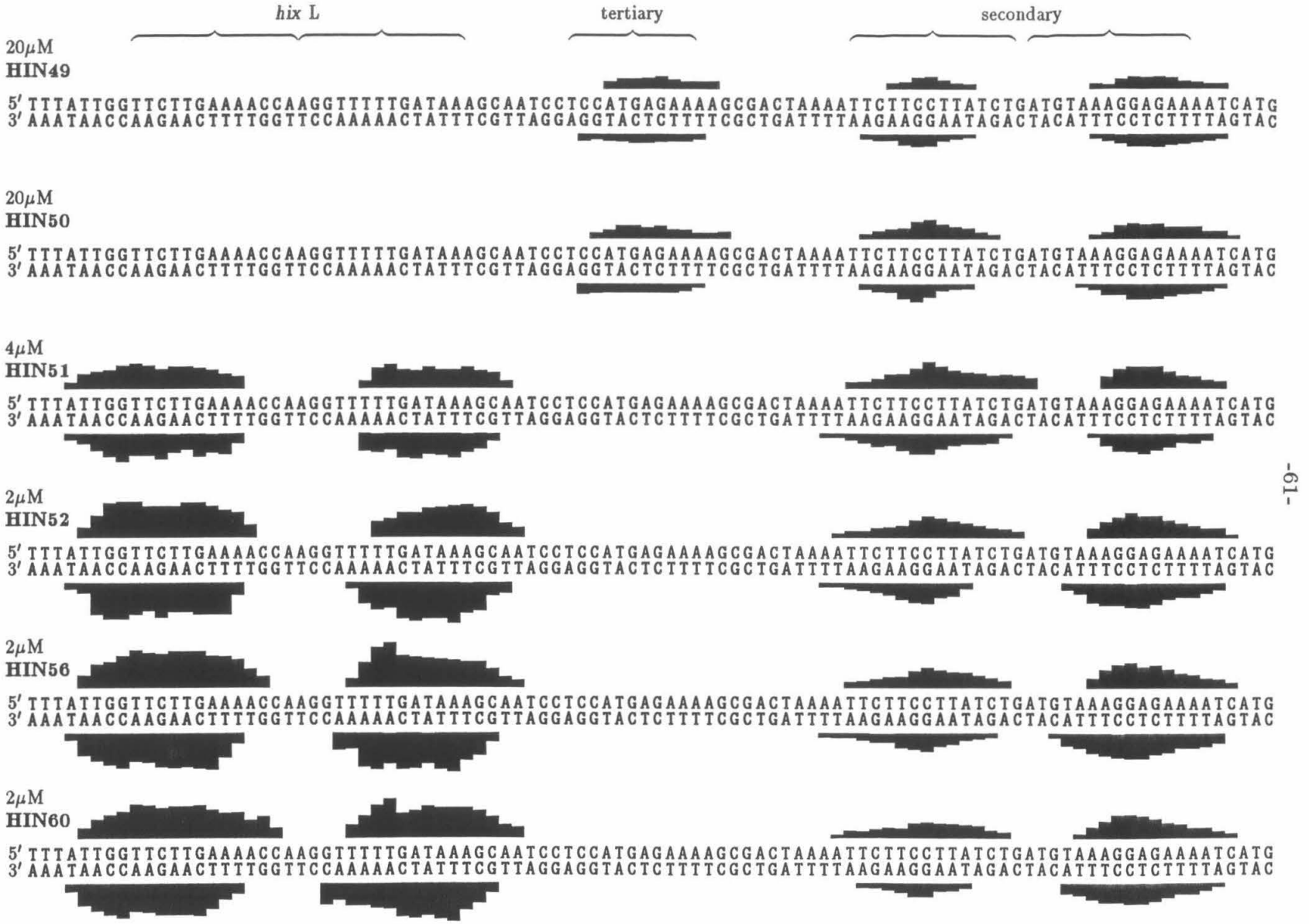

5′
3′
1 control
2 G
3
4 2μM HIN49
5 2μM HIN50
6 4μM HIN51
7 2μM HIN52
8 2μM HIN56
9 2μM HIN60
10 2μM HIN60
11 2μM HIN56
12 2μM HIN52
13 4μM HIN51
14 20μM HIN50
15 20μM HIN49
16
17 G
18 control
hix R
secondary
tertiary
hix L

Figure 25: Footprinting histograms of the gel data in figure 24. The sequence left to right is the bottom of the gel to just above the secondary site in figure 22. (Top) Brackets mark *hix*L, tertiary and secondary sites. Bars indicate extent of cleavage protection from MPE•Fe induced cleavage.


hix L
tertiary
secondary
20μM
HIN49
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
20μM
HIN50
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
4μM
HIN51
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
2μM
HIN52
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
2μM
HIN56
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC
2μM
HIN60
5′ TTTATTGGTTCTTGAAAACCAAGGTTTTTGATAAAGCAATCCTCCATGAGAAAAGCGACTAAAATTCTTCCTTATCTGATGTAAAGGAGAAAATCATG
3′ AAATAACCAAGAACTTTTGGTTCCAAAAACTATTTCGTTAGGAGGTACTCTTTTCGCTGATTTTAAGAAGGAATAGACTACATTTCCTCTTTTAGTAC

may be part of the linking region between the sequence specific DNA binding carboxy-terminal domain and the recombination and oligomerization domains of the protein.

DNA Binding by the 51mer: The 51mer differs from the 52mer only in the removal of one glycyl residue, 1% of the total mass and 2% of the peptide backbone. Both affinity cleaving with **EG51** and MPE•Fe footprinting of **HIN51** indicate essentially identical affinity and specificity. There is only a very small difference in apparent binding affinity, which changes from site to site, and is probably within experimental error (compare lanes 5 and 6 and lanes 11 and 12 in figure 22).

DNA Binding by the 49 and 50mers: The 50mer differs from the 51mer only in the loss of an arginyl residue, 3% of the total mass and 2% of the peptide backbone. The 49mer is missing an additional prolyl residue. The sequence specific DNA binding ability of these two peptides is radically different than that of the 51 and 52mers. For example, compare lanes 4 and 5 in figure 22. At *hix*L, even at 20 fold higher peptide concentration, specific cleavage has been completly lost. At the secondary site, cleavage with **EG49** and **EG50** is reduced by a factor of ~20. The only binding site at which the 49 and 50mers have affinities comparable to the 51 and 52mers is the tertiary site. DNA cleavage with **EG52**•Fe at 10μM gives comparable cleavage at the tertiary site, as does **EG49**•Fe and **EG50**•Fe at that concentration (data not shown).

Footprinting of the **HIN49** and **HIN50** indicates that these peptides only bind to the tertiary and secondary binding sites and not to *hix*L. The footprints indicate that the tertiary binding site extends to the right of the cleavage pattern. Reading right to left along the lower strand, 5′-TTCTCATGGA-3′. The first five base pairs

match the first five base pairs of the consensus *hix* half site, 5′-TT$^{A}_{C}$TC-3′ (see figure 6). The next two base pairs of the consensus half site are not conserved but the following 5′-AAA is strongly conserved and these are the three A's protected from dimethyl sulfate by Hin and the 52mer (see figure 8). The corresponding bases in the tertiary site are 5′-GGA. The affinity cleaving patterns in figure 23 indicate the EDTA•Fe is located between the two G's for **EG49**•Fe and on the central G for **EG50**•Fe. Therefore neither the 49mer nor the 50mer reaches the only conserved A in the tertiary site.

The cleavage patterns for **EG49**•Fe through **EG60**•Fe are 3′ shifted at all binding sites indicating that the EDTA•Fe of these six peptides are all located above the DNA minor groove. Furthermore the cleavage data at the two tertiary half–sites allows determination of the average EDTA•Fe position for each peptide. A plot of the average iron position versus peptide size is shown in figure 26. Note that the numbering of the base pairs has been reversed from the normal numbering of a *hix* site. The bases are numbered towards the C_2 axis instead of away from it, starting with the first conserved base on the outside of a *hix* site.

The plot in figure 26 indicates that each amino acid from 49 to 51 shifts the EDTA•Fe position by half a base pair. The 51 and 52mers place the EDTA•Fe at the same base position. The four additional amino acids of the 56mer shift the EDTA•Fe an additional base pair (0.25bp/amino acid). The effect of 4 more amino acids from 56 to 60mers is more difficult to define due to the diffuseness of the cleavage patterns. It appears however, that the four additional amino acids shift the EDTA•Fe location by half a base pair or less (~0.1bp/amino acid).

Since the nitrogen–nitrogen seperation in an extended chain (i.e. β sheet) protein conformation is ~3.4Å, which is the same as the base pair rise of B-form DNA,

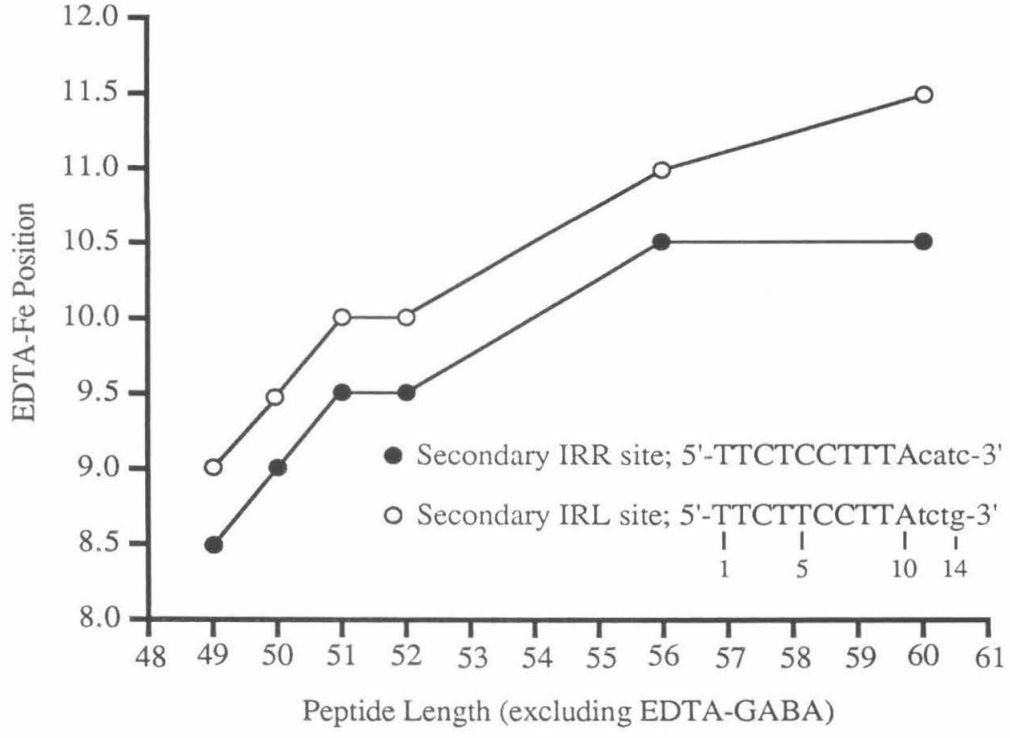


Figure 26: Plot of average iron location for **EG49**•Fe, **EG50**•Fe, **EG51**•Fe, **EG52**•Fe, **EG56**•Fe and **EG60**•Fe. Note that the numbering is from the outside of the site.

the 49, 50, 51 and 52mer data indicates the backbone conformation of this region of the peptide is not that of an extended chain. If the peptide backbone was in an extended conformation the EDTA•Fe postion would have been shifted one base pair per additional amino acid residue instead of the observed ~0.5bp/amino acid. Furthermore, a true extended conformation is excluded by the prolyl residue as the amino terminus of the 50mer. It appears then that the protein structure of this sub-domain is not that of a β sheet that follows the minor groove as described in the introductory discussion on potential types of protein–DNA interactions. Note that a double stranded, antiparallel β sheet is the only general protein structure

which has been proposed to potentially interact with the minor groove sequence specificaly. (An example of a β sheet interaction with the minor groove can be postulated for λ Cro, see figure 1).

DNA Binding Affinities of the Synthetic Peptides

The thermodynamics of HPLC purified **HIN49**, **HIN50**, **HIN51**, **HIN52**, **HIN56** and **HIN60** binding to *hix*L has been investigated by A. Glasglow *et al.*[48] DNA gel–mobility retardation assays[77,78] indicate the binding constants shown in Table 2. Gel–mobility retardation assays involve electrophoresis of a radiolabelled DNA fragment, in the presence of vary amounts of protein, on a nondenaturing polyacrylamide gel. The DNA•protein complex migrates through the gel more slowy than uncomplexed DNA. Table 2 also includes association constants for Hin, as well as $\gamma\delta$ resolvase and its 43 amino acid chymotrypsin product.

Table 2: Binding constants for **HIN52**, **HIN56**, **HIN60**, Hin, $\gamma\delta$ Resolvase and $\gamma\delta$–43mer at 22°C.

Site	Protein	[NaCl] (mM)	$K_{assoc.}$	$\Delta G_{assoc.}$ (kcal/mol)	Technique
*hix*L	**HIN52**[48]	20	7.1×10^6	-8.8	gel–mobility retardation
*hix*L	**HIN56**[48]	20	1.0×10^6	-7.7	gel–mobility retardation
*hix*L	**HIN60**[48]	20	1.4×10^6	-7.9	gel–mobility retardation
*hix*L	Hin[79]	100	2.5×10^7	-10.2	DNase I footprinting
II–L	$\gamma\delta$ Resolv[39]	50†	$>5.0 \times 10^9$	<-13.4	DNase I footprinting
II–L	$\gamma\delta$ 43mer[39]	50†	2.0×10^6	-8.5	DNase I footprinting

† 40mM NaCl + 5mM $MgCl_2$, 20°C.

The gel–mobility retardation assays did not detect descrete binding of **HIN49**, **HIN50** and **HIN51** to *hix*L. However, both **HIN51** and **EG51** bind to *hix*L as

revealed by footprinting and affinity cleaving assays. Furthermore the gel–mobility retardation assays indicate that **HIN56** and **HIN60** bind more weakly to *hix*L than does **HIN52** (a factor of 7 between **HIN52** and **HIN60**, $\Delta\Delta G \sim 1.1$kcal/mol, see Table 2) contrary to the apparently equivalent binding observed in the footprinting studies. The difference in the apparent bidning affinities, determined by gel–mobility retardation and footprinting or affinity cleaving, is perhaps due to changes in the on/off rates of the peptides. In particular the gel–mobility assay would be unreliable in systems where the binding rates were fast, since this would increase the likelihood of dissassociation of the DNA•protein complex during migration through the gel.

It is also possible to extract estimates of the *relative* binding affinities of the 49 through 60mers by comparison of the DNA cleavage with the EDTA•Fe equiped peptides. Table 3 lists relative cleavage efficiencies (RCE) and estimated $\Delta\Delta G$'s for **EG49**, **EG50** and **EG51** relative to **EG52**. Table 3 is divided into two sections. The upper section shows the RCE's between the indicated peptide at a particular site relative to **EG52**•Fe cleavage at the same site. The lower portion shows RCE's between the indicated peptide at a particular site relative to **EG52**•Fe at the best binding site (*hix*L IRR). The RCE's are valid if the efficiency of EDTA•Fe cleavage is the same for these peptides and is independent of the sequence being cleaved, i.e. observed cleavage efficence is only dependent on the peptide binding constant. This analysis was not extended to **EG56**•Fe and **EG60**•Fe since the EDTA•Fe moeities do not seem as tightly held to the DNA, and as a result, the EDTA•Fe cleavage efficience is certainly reduced.

A number of interesting comparisons can be made based on the data in Table 3. First, the relative affinities for **EG52** at the five sites is *hix*L IRR$\approx$*hix*L IRL>Sec

Table 3: Relative cleavage efficiencies (RCE) for **EG51**, **EG50** and **EG49** versus **EG52** at *hix* type binding sites.

		—**EG52**—		—**EG51**—		—**EG50**—		—**EG49**—	
Site	Sequence[a]	RCE[b]	~ΔΔG[c]	RCE[b]	~ΔΔG[c]	RCE[b]	~ΔΔG[c]	RCE[b]	~ΔΔG[c]
Relative to **EG52** at the same site:									
*hix*L IRL	ggTTCTTGAAAAcca	–	–	0.69	-0.22	0.0027	-3.5	0.0011	-4.0
*hix*L IRR	ctTTATCAAAAAcct	–	–	1.1	0.06	0.0068	-2.9	0.0065	-3.0
Sec IRL	aaTTCTTCCTTAtct	–	–	1.1	0.05	0.11	-1.3	0.087	-1.4
Sec IRR	ttTTCTCCTTTAcat	–	–	1.2	0.10	0.052	-1.7	0.032	-2.0
Tertiary	ttTTCTCATGGAgga	–	–	0.80	-0.13	0.59	-0.3	0.48	-0.4
Relative to **EG52** at *hix*L IRR:									
*hix*L IRL	ggTTCTTGAAAAcca	0.92	-0.05	0.64	-0.27	0.0025	-3.5	0.0010	-4.0
*hix*L IRR	ctTTATCAAAAAcct	–	–	1.1	0.06	0.0068	-2.9	0.0065	-3.0
Sec IRL	aaTTCTTCCTTAtct	0.27	-0.77	0.29	-0.72	0.029	-2.1	0.023	-2.2
Sec IRR	ttTTCTCCTTTAcat	0.68	-0.22	0.82	-0.12	0.036	-2.0	0.022	-2.2
Tertiary	ttTTCTCATGGAgga	0.035	-2.0	0.058	-1.7	0.043	-1.8	0.034	-2.0

a. Sequence ($5'\rightarrow3'$) towards the psuedo–C_2 axis.

b. Relative cleavage efficency. Cleavage for a given compound at a particular site ($A_{site}^{compound}$) is the sum of the integrated peak areas over the cleavage loci on both strands. RCE is the ratio of cleavage for two compounds corrected for differences in concentration. For example RCE for **EG51** at *hix*L IRR versus **EG52** at the same site is;

$$\mathrm{RCE} = \frac{(A_{hixL\ IRR}^{\mathbf{EG51}})}{[\mathbf{EG51}]} \bullet \frac{[\mathbf{EG52}]}{(A_{hixL\ IRR}^{\mathbf{EG52}})}$$

c. Calculated using $\Delta\Delta G = RT\, ln(K_1/K_2)$ (T=22°C) and assuming RCE $\equiv K_1/K_2$, in kcal/mol.

IRR>Sec IRL>>Tertiary, with the hix sites 1.5 to 3.7 times better than the two secondary sites and 30 times better than the tertiary site. Secondly, **EG51** is essentially identical to **EG52** both in differentiating between the five sites in and overall binding affinities. Thirdly, **EG50** and **EG49** are similar to each other but show inverted relative affinities compared to **EG52**, Tertiary≈Sec IRR≈Sec IRL>*hix*L IRR≈*hix*L IRL. Binding by **EG49** and **EG50** is reduced 150 to 1000

fold ($\Delta\Delta$G=3–4kcal/mol) at *hix*L relative to **EG52**. However at the tertiary site, **EG49** and **EG50** binding is reduced only by a factor of two relative to **EG52**.

Comparisons such as those listed above lead to two conclusions; (1) Since the first five base pairs of all five binding sites match the consensus half site (figure 6 and Table 3), the following two base pairs are not conserved in the consensus and the last 3bp of the consensus half–site are apparently not contacted by these peptides (see EDTA•Fe locations in figures 23 and 26), the remaining three consenus base pairs (5′-AAA, numbers 4, 5 and 6 from the C_2 axis) constitute the significant differences between the sites. Furthermore, the binding of **EG52** is directly correlated with the extent of deviation from the 5′-AAA consensus; 5′-AAA>5′-TTA>>5′-GGA. These are also the three A's protected from dimethyl sulfate by Hin (figure 8). (2) As the peptide is shortened from the amino terminus by removal of glycine, arginine and proline (**EG51**, **EG50** and **EG49** respectively) the effect of the 5′-AAA sequence on peptide binding changes from an enhancing to a *destabilizing* contribution to binding. For instance, **EG49** and **EG50** *prefer* the tertiary and secondary sites (5′-GGA and 5′-TTA) over *hix*L (5′-AAA) by a factor of 4–30 depending on the site. The EDTA•Fe cleavage patterns indicate that **EG49** and **EG50** do not even *reach* the A•T base pair nearest the C_2 axis, and perhaps the one before it as well.

It appears then that the final Arginine in the peptide, perhaps in conjunction with the following residues, not only specifically recognizes the 5′-AAA sequence via the minor groove, but also changes the DNA conformation in such a way that the peptide interaction at the outer five base pairs of a hix half–site is enhanced.

The amino terminus sequence of the 52mer; $NH_3^{\oplus}$–Gly–$Arg^{\oplus}$–Pro–$Arg^{\oplus}$–, appears to contain the element which recognizes the 5′-AAA track of a *hix* half–site. A

search of the GenBank protein sequence library (~2000 protein entries, ~500,000 amino acids) found 70 occurrences of $\overset{\oplus}{\text{Arg}}$–Pro–$\overset{\oplus}{\text{Arg}}$, 5 occurences of $\overset{\oplus}{\text{Arg}}$–Pro–$\overset{\oplus}{\text{Arg}}$–Gly and 4 occurences of Gly–$\overset{\oplus}{\text{Arg}}$–Pro–$\overset{\oplus}{\text{Arg}}$. Of the 70 occurrences of $\overset{\oplus}{\text{Arg}}$–Pro–$\overset{\oplus}{\text{Arg}}$, 11 (16%) were in protamines. Protamines are small (~30 amino acids) arginine rich peptides associated with DNA packaging and there was a very high correlation with the $\overset{\oplus}{\text{Arg}}$–Pro–$\overset{\oplus}{\text{Arg}}$ sequence located at the exact center of the peptide sequence. Nine (13%) of the $\overset{\oplus}{\text{Arg}}$–Pro–$\overset{\oplus}{\text{Arg}}$ were in Actins, proteins associated with muscular and other mechanical movements of the cell. These appear to be unrelated to DNA binding. Of the remaining 50 proteins, at least 14 ($\geq$20%) are DNA binding proteins and include the regulatory proteins; Arabinose Operon regulatory protein (both *E. Coli* and *Salmonella Typhimurium*), Galactose repressor (*E. Coli*), 434 Cro and 434 repressor. Other DNA binding proteins were, GAG's (three proteins), genome polyproteins (two proteins), kinase–related transforming proteins (three proteins), large T antigen (mouse polyomavirus) and the human chromasomal protein HMG–1. All four of the occurrences of Gly–$\overset{\oplus}{\text{Arg}}$–Pro–$\overset{\oplus}{\text{Arg}}$ are in DNA binding proteins; three in kinase–related transforming proteins (fgr–feline, mht and raf–murine) and one in HMG–1. It can be hypothesized that all of these proteins use the (Gly)–$\overset{\oplus}{\text{Arg}}$–Pro–$\overset{\oplus}{\text{Arg}}$ sequence to recognize A•T base pairs in the minor groove.

Two of the $\overset{\oplus}{\text{Arg}}$–Pro–$\overset{\oplus}{\text{Arg}}$ containing proteins deserve special mention. 434 repressor contains $\text{Lys}_{40}^{\oplus}$–$\text{Arg}_{41}^{\oplus}$–Pro_{42}–$\text{Arg}_{43}^{\oplus}$ near the carboxy terminus of the DNA binding helix of its helix–turn–helix domain (the small loop between α_3–α_4 in figure 3). This sequence interacts with the 3′ strand (ATA) of the DNA binding site. Ethylation of these three phosphates also interferes with repressor binding.[80] It is difficult to draw detailed conclusions from the 3.2–4.5Å resolution cocrystal, but

it appears that the peptide backbone crosses over the DNA phosphate backbone with the carboxy terminus in the major groove, $Arg^{\oplus}_{41}$ and Pro_{42} interacting with the DNA backbone at the crossover point and, perhaps, $Arg^{\oplus}_{43}$ interacting in the minor groove.[34] Anderson *et al.* point out that an $Arg^{\oplus}_{43}$→Ala mutant still preferentially binds to the wild type DNA sequence arguing against a specific minor groove contact by $Arg^{\oplus}_{43}$. This implies that the Arg$^{\oplus}$–Pro–Arg$^{\oplus}$ sequence is recognizing the DNA *conformation* rather than a specific hydrogen bonding pattern in the minor groove. They further point out that the minor groove is 2.5Å narrower than idealized B–form DNA at base pairs 6, 7, 8 and 9 (the center of the 14bp operator), and therefore $Arg^{\oplus}_{43}$ could have significant electrostatic attractions with the phosphates on the other side of the minor groove (phosphates 5 and 6 in figure 3). The Anderson model then, involves specific recognition by Arg$^{\oplus}$–Pro–Arg$^{\oplus}$ of a sequence specific DNA conformation which is characterized by a narrower minor groove. The narrow minor groove at the center of the operator also appears important in correctly positioning the the two protein monomers for proper protein–protein contact.[34,81]

Interestingly, if the interaction between the Arg$^{\oplus}$–Pro–Arg$^{\oplus}$ sequence and DNA observed in the repressor–DNA cocrystal were applied directly to **EG52**, the EDTA•Fe would be positioned in the DNA *major groove*.

The other DNA binding protein containing Arg$^{\oplus}$–Pro–Arg$^{\oplus}$ deserving special mention is the mammalian High Mobility Group–1 protein (HMG–1, also called α–protein). HMG–1 contains two palindromic sequence of Pro–Arg$^{\oplus}$–Gly–Arg$^{\oplus}$–Pro as well as a Gly–Arg$^{\oplus}$–Pro–Arg$^{\oplus}$ sequence.[82] Ethylation and methylation interference studies indicate that HMG–1 binds to stretches of five or more A•T base pairs in

the DNA minor groove.[83] Solomon *et al.* speculate that the minor groove recognition is primarily a result of the unique DNA conformation of stretches of five or more A•T base pairs, i.e. a narrower minor groove.[83]

The DNA cleaving ability of **EG49**•Fe, **EG50**•Fe, **EG51**•Fe and **EG52**•Fe, binding affinities of **HIN49**, **HIN50**, **HIN51** and **HIN52** and dimethyl sulfate studies with Hin, along with the proceeding discussion, indicate that the $\overset{\oplus}{\text{Arg}}$–Pro–$\overset{\oplus}{\text{Arg}}$ amino terminus sequence of the 52mer is responsible for recognizing, via the minor groove, the conserved 5′-AAA in a *hix* half–site. It appears from the protein homologies for Hin, Gin, Pin and Cin (see figure 7) that there are perhaps three protein sequences which are capable of recognizing the same DNA sequence from the minor groove. These amino acids sequences are shown in figure 27.

Figure 27: Protein sequences which apparently recognize three or more A•T base pairs from the minor groove. Note that this alignment differs slightly from the one in figure 7 in that Hin and Gin have been shifted right one residue.

Hin: 139→143		Gly	$\overset{\oplus}{\text{Arg}}$	Pro	$\overset{\oplus}{\text{Arg}}$	(Ala)				
Gin: 139→143		Gly	$\overset{\oplus}{\text{Arg}}$	Pro	Pro	$\overset{\oplus}{\text{Arg}}$				
Pin: 139→143	Gly	$\overset{\oplus}{\text{Arg}}$	$\overset{\oplus}{\text{Arg}}$	Pro	$\overset{\oplus}{\text{Lys}}$					
Cin: 139→143	Gly	$\overset{\oplus}{\text{Arg}}$	$\overset{\oplus}{\text{Arg}}$	Pro	$\overset{\oplus}{\text{Lys}}$					
HMG–1:[82] a)	Pro	$\overset{\oplus}{\text{Lys}}$	$\overset{\oplus}{\text{Arg}}$	Pro	$\overset{\oplus}{\text{Arg}}$	Gly	$\overset{\oplus}{\text{Arg}}$	Pro	$\overset{\oplus}{\text{Lys}}$	Gly
b)	Gly	$\overset{\oplus}{\text{Arg}}$	$\overset{\oplus}{\text{Lys}}$	Pro	$\overset{\oplus}{\text{Arg}}$	Gly	$\overset{\oplus}{\text{Arg}}$	Pro	$\overset{\oplus}{\text{Lys}}$	$\overset{\oplus}{\text{Lys}}$
c)	?	Gly	$\overset{\oplus}{\text{Arg}}$	Pro	$\overset{\oplus}{\text{Arg}}$	$\overset{\oplus}{\text{Lys}}$				
434 Repressor: 40→44	(Thr)	$\overset{\oplus}{\text{Lys}}$	$\overset{\oplus}{\text{Arg}}$	Pro	$\overset{\oplus}{\text{Arg}}$	(Phe)				

It appears from the amino acid sequences shown in figure 27 that the conserved element is an $\overset{\oplus}{\text{Arg}}$–Pro–($\overset{\oplus}{\text{Arg}}$/$\overset{\oplus}{\text{Lys}}$) sequence which occurs in all but Gin. The variation in amino acid sequence suggest that there is either more than one mode

of recognition or that the common $\overset{\oplus}{\text{Arg}}$–Pro–($\overset{\oplus}{\text{Arg}}$/$\overset{\oplus}{\text{Lys}}$) sequence is competent at recognition of 5′-AAA in the minor groove. The *mode* of recognition may be based on the minor groove width at runs of four or more A•T base pairs and not on specific hydrogen bonds in the minor groove. The $\overset{\oplus}{\text{Arg}}$–Pro–($\overset{\oplus}{\text{Arg}}$/$\overset{\oplus}{\text{Lys}}$) sequence may act as a "molecular caliper" which "measures" the minor groove width by ionic interactions accross the two phosphate backbones. The central proline would be required for this type of interaction due to the unique (and limited) backbone torsional angles of the pyrrolidine moeity which would uniquely fix the positively charged first and third residues.

Location of Carboxy Terminus

The results with the 49 through 60 series of peptides identified the location of the amino termini of these peptides when bound to DNA. The two models shown in figure 20, which are based on the assumption of an helix–turn–helix domain in the major groove, differ significantly in their positioning of the 10 amino acids at the carboxy terminus. The amino acid sequence homology between the carboxy terminus of Hin and the amino terminus of λ repressor suggest that these 10 amino acids may form an arm which wraps around the DNA following the major groove. In order to test the "arm" hypothesis, as well as to differentiate between the two models, a strategy for incorporating **BEG** at any position in a synthetic peptide was developed.

It was decided to replace Methionine–189 (Hin numbering), the next to last residue from the carboxy teminus, with a lysine residue which had been modified by covalent attachment of EDTA–GABA to the lysyl–N^{ϵ} amino group. The target compounds were the 52mer with N^{ϵ}-(EDTA–GABA)–Lysyl at position 51 (**EG*52**), a double EDTA–GABA peptide ((**EG**)$_2$**52**), as well as a peptide equivalent to **EG50** (which did not cleave at *hix* half sites); **EG*50**. Figure 28 shows the complete amino acid sequence of these three peptides as well as the points of EDTA–GABA attachment.

The synthetic route to the **EG*** compounds is shown in Scheme 3. Starting with PAM–Asn resin, N^{ϵ}-FMOC-N^{α}-tBOC-L-Lysine was coupled to the Asn resin in place of the normal Hin methionine–189 residue. Incorporation of N^{ϵ}-FMOC-N^{α}-tBOC-L-Lysine allowed the N^{ϵ} amino group to be deprotected (piperidine in DMF) without affecting the N^{α}-tBOC or the attachment of the peptide to the

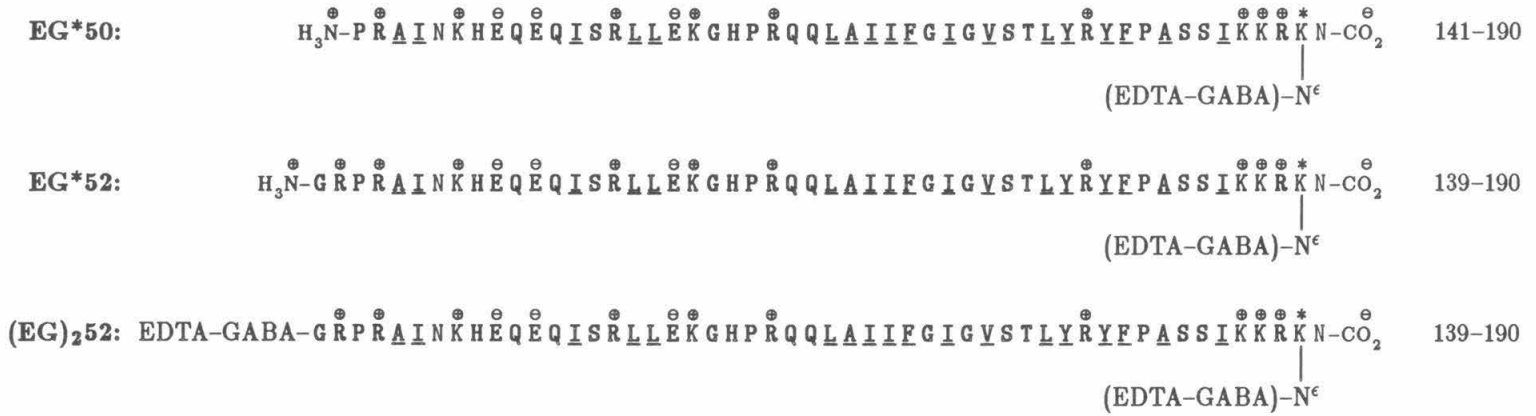


Figure 28: Amino acid sequences of the synthetic peptides with N^{ϵ}–(EDTA–GABA)lysine near the carboxy terminus. Hydrophobic residues are underlined and charged residues are marked with the appropriate sign. The right hand column lists the corresponding residue numbers for Hin.

resin. **BEG** was then coupled to the free lysyl N^{ϵ} amino via DCC/HOBt mediated coupling in DMF. Ninhydrin analysis indicated 99.8% reaction after 100 minutes. The lysyl N^{α}-tBOC was then removed and the next amino acid (Arg-TOS) was coupled to the peptide–resin using the standard procedures. The remainder of the synthesis was carried out as described for the previous peptides. At the 52mer stage a portion of the peptide–resin was again coupled to **BEG** to give the double EDTA–GABA peptide; $(\mathbf{EG})_2\mathbf{52}$. During the synthesis, samples of 3mer, 30mer, 50mer and 52mer were removed, subjected to HF cleavage, and purified by HPLC. The 3mer, Arg–N^{ϵ}-(EDTA–GABA)-Lys–Asn, was characterized by high field NMR and high resolution FAB mass spectroscopy, both of which were consistent with the target structure. The 30mer was also characterized by NMR which indicated that the EDTA–GABA was still present in the peptide at that stage of the synthesis.

DNA Cleavage with EG*50•Fe, EG*52•Fe and $(EG)_2$52•2Fe

The ability of the carboxy-terminal EDTA–GABA peptides to cleave DNA was investigated by high resolution gel electrophoresis of ^{32}P–endlabelled DNA fragments as described earlier for **EG52**•Fe. The autoradiogram of DNA cleavage by **EG*50**•Fe, **EG*52**•Fe and $(\mathbf{EG})_2\mathbf{52}$•2Fe is shown in figure 29 and cleavage histograms in figure 30.

DNA cleavage with **EG*52**•Fe gives three very broad cleavage patterns at *hix*L. $(\mathbf{EG})_2\mathbf{52}$•2Fe gives a cleavage pattern which is identical to the sum of the **EG52**•Fe and **EG*52**•Fe patterns. The fact that $(\mathbf{EG})_2\mathbf{52}$•2Fe gives the **EG52**•Fe pattern superimosed on the **EG*52**•Fe pattern leads to two conclusions; (1) The synthetic method used to incoporate EDTA–GABA at the carboxy terminus was succesful and the primary structure of both **EG*52** and $(\mathbf{EG})_2\mathbf{52}$ is correct. (2) Since

Scheme 3: Synthetic route to EG*52.

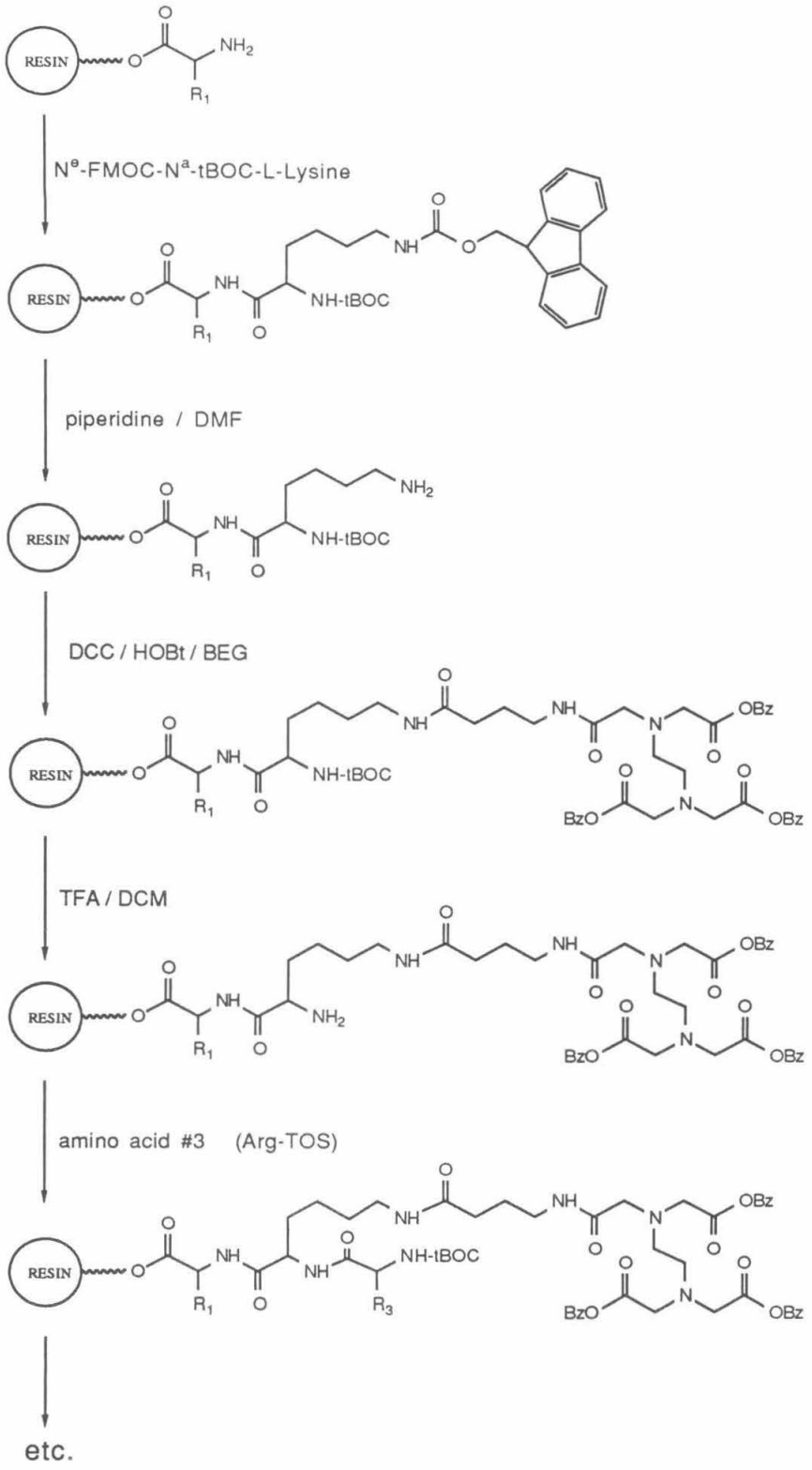
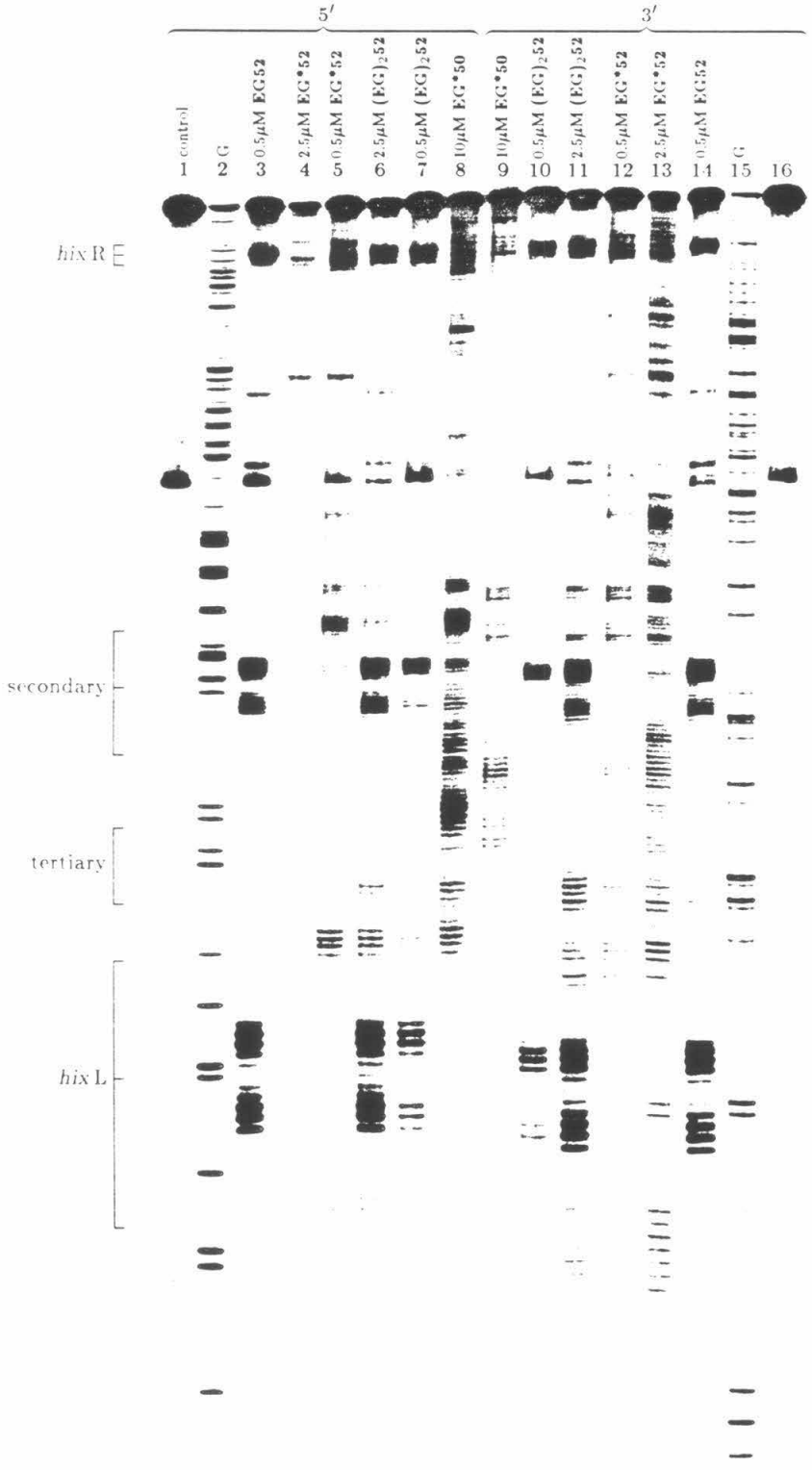


Figure 29: Autoradiogram of a high resolution denaturing polyacrylamide gel of affinity cleaving reactions on a ^{32}P–endlabeled fragment (Xba I – Eco RI) from pMFB36.[1] Reaction conditions were 20mM NaCl, 20mM TRIS•HCl, pH 8.0, 400μM tRNA, 5mM DTT, and ≈15,000cpm (≈1ng) endlabelled DNA in a total volume of 15μl. Reactions were run for 60 minutes at 22°C and terminated by ethanol precipitation. Cleavage products were analyzed on an 8%, 1:20 crosslinked, 50% urea denaturing polyacrylamide gel. Lanes 1 through 8 contain 5′ labelled and 9 through 16 3′ labelled DNA. Lanes 1 and 16 are DNA controls and 2 and 15 Maxam–Gilbert chemical sequencing marker lanes. Lanes 3 and 14 contain 0.5μM **EG52**•Fe, 4 and 13 contain 2.5μM **EG*52**•Fe, 5 and 12 contain 0.5μM **EG*52**•Fe, 6 and 11 contain 2.5μM $(\mathbf{EG})_2$**52**•2Fe, 7 and 10 contain 0.5μM $(\mathbf{EG})_2$**52**•2Fe and 8 and 9 contain 10μM **EG*50**•Fe. (Left) Brackets mark *hix*R, secondary, tertiary and *hix*L binding sites.


5′
3′
control
G
0.5μM EG52
2.5μM EG*52
0.5μM EG*52
2.5μM (EG)252
0.5μM (EG)252
10μM EG*50
10μM EG*50
0.5μM (EG)252
2.5μM (EG)252
0.5μM EG*52
2.5μM EG*52
0.5μM EG52
G
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
hix R
secondary
tertiary
hix L

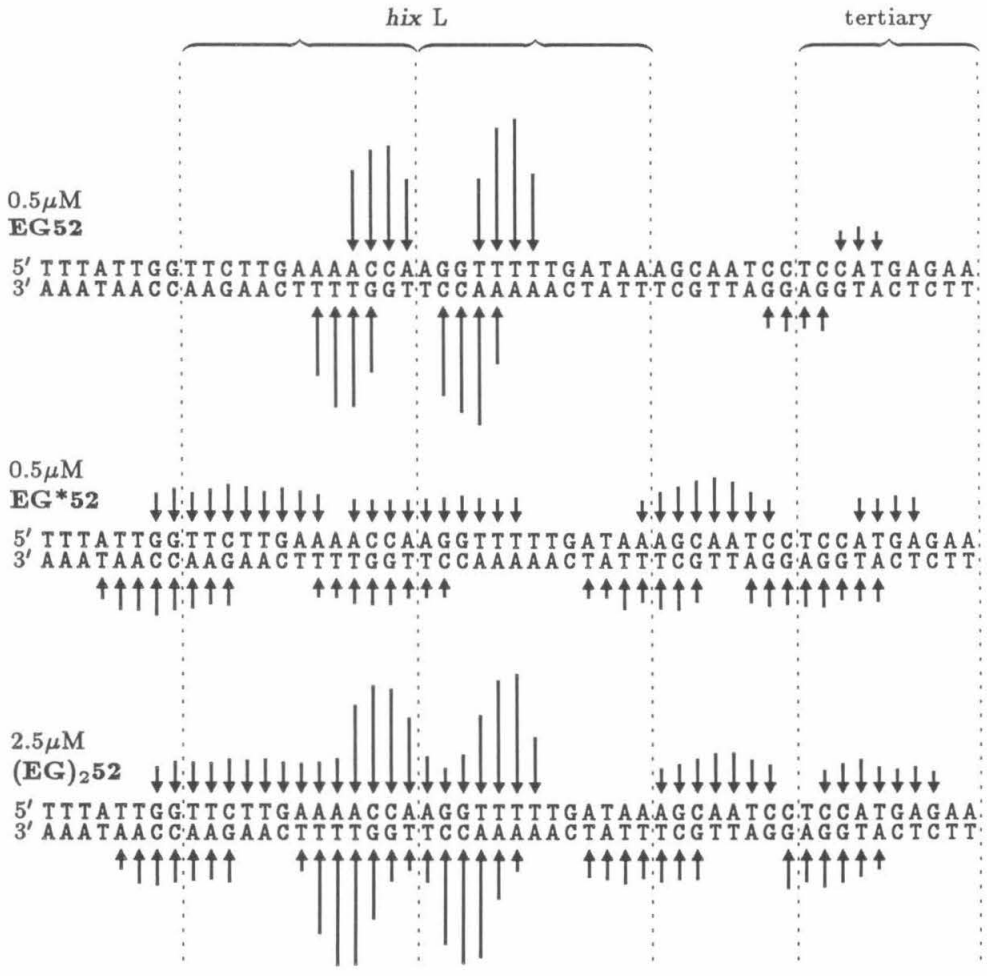


Figure 30: Histograms of the gel data in figure 29. The sequence left to right is the bottom of the gel to just above the tertiary site in figure 29. (Top) Brackets mark *hix*L and tertiary binding sites. Arrows indicate extent of cleavage resulting in removal of the indicated base.

$(\mathbf{EG})_2\mathbf{52}$•2Fe gives the **EG52**•Fe pattern as a subset of its total cleavage pattern, inorporation of EDTA–GABA at the carboxy terminus has not grossly altered the DNA binding ability of the 52mer. The carboxy–terminal EDTA–GABA has, however, lowered the binding affinity of the 52mer somewhat. Comparison of the extent of cleavage of **EG52**•Fe with the portion of the $(\mathbf{EG})_2\mathbf{52}$•2Fe pattern apparently due to the amino–terminal EDTA–GABA indicates that the cleavage efficiency has been reduced by a factor of five. Assuming the reduction of cleavage efficiency is due to a reduced binding constant, the $\Delta\Delta G_{assoc}$ for $(\mathbf{EG})_2\mathbf{52}$•2Fe relative to **EG52**•Fe would be ~1kcal/mol. The reduced binding affinity of $(\mathbf{EG})_2\mathbf{52}$•2Fe could result from three factors; (1) addition of one negative charge due to EDTA•Fe (either $EDTA^{-3}$•Fe^{+2} or $EDTA^{-3}$•Fe^{+3}•OH^-), (2) loss of a contact made by the methionine residue in the wild amino acid sequence to either the DNA or the peptide or (3) an incorrect steric, hydrogen bonding or hydrophobic interaction between the EDTA–GABA•Fe and the DNA or peptide.

Assuming that the observed **EG*52**•Fe cleavage occurs in the same DNA binding conformation as the **EG52**•Fe cleavage and not some other binding mode, two conclusions can be made about the location of the carboxy terminus of the 52mer relative to the DNA binding site; (1) As with **EG56**•Fe and **EG60**•Fe the broad cleavage pattern implies that the EDTA•Fe, and hence the carboxy terminus of the peptide, is not held close to the DNA. (2) Though the cleavage patterns are broad they are clearly 3′ shifted implying *cleavage* is occurring from the minor groove. This indicates that the carboxy–terminal domain of the peptide does not form an arm which wraps around the DNA following the major groove as proposed in the intial binding models.

In retrospect, though the sequence homology between then amino–terminal arm of λ repressor and the carboxy terminus of Hin is a strong one, the fact is that the carboxy–terminal domains of Hin, Gin, Pin and Cin are not conserved. (see figure 7). Since these four proteins almost certainly have the same DNA binding structure, the lack of sequence homology between the six carboxy–terminal residues of Hin argues against a significant contribution to binding by these residues.

In summary, the **EG*** results indicate that the synthetic strategy used to incorporate EDTA–GABA at the carboxy terminus of the 52mer peptide was successful. Furthermore, the observed cleavage patterns rule out the possibility of a carboxy–terminal arm that wraps around the DNA following the major groove, and instead indicate that this domain is not intimately associated with the DNA. The **EG*** results do not however, preclude a helix–turn–helix domain as shown in the initial models, nor do they specifically differentiate between the two binding helix orientations.

Revised DNA Binding Model

The **EG*** results indicate that the binding models shown in figure 20 require revision. The original models were based on a helix–turn–helix domain and a carboxy–terminal arm in the major grove and the minor groove location of the amino terminus. The **EG*** results indicate that the carboxy terminus is positioned in such a way that the diffusable oxidant generated by EDTA•Fe does not have direct access to the DNA backbone. Inspection of the structure of λ Cro (figure 1) indicates that Cro has a peptide backbone which turns after leaving the carboxy terminus of the recognition helix and runs back along the DNA backbone, turns again and forms a β sheet with itself, then crosses over the DNA phosphate

backbone and runs along the minor groove. In dimeric λ Cro, the peptide chain forms an antiparallel β sheet with the equivalent strand from the other monomer above the minor groove (see figures 1 and 31). If the 10 amino acid carboxy terminus of Hin follows a similar path, the carboxy terminus is located above, and somewhat between, helices 1 and 2 as shown in figure 31. The interaction between the carboxy terminus and the DNA or remainder of the protein is not required for DNA binding. This model is consistent with the broad and weak cleavage pattern observed with **EG*52**•Fe if the binding of the three α helices protects the major groove directly below the carboxy–terminal EDTA•Fe. With the major groove protected from cleavage by the peptide, the only cleavage that is observed is that which results from diffusion away from the major groove binding site and into the minor grooves above and below. It is interesting to note that the two regions of Hin 139→190 that deviate most from the Hin/Gin/Pin/Cin protein consensus are the six carboxy–terminal residues and residues near the amino terminus of the putative α_1 helix (see figure 7). The revised model suggests that these two regions may interact.

The results with **EG49** through **EG60** allow refinement of the amino terminus of the model. The affinity cleaving studies indicate a minor groove recognition of the $A_4A_5A_6$ sequence with a protein structure which is neither a β sheet or an α helix. Additionally, amino acid locations can be specified for the amino terminal residues of the 49, 50, 51 and 52mers within the minor groove and for the amino–terminal residues of the 56 and 60mers located away from the DNA.

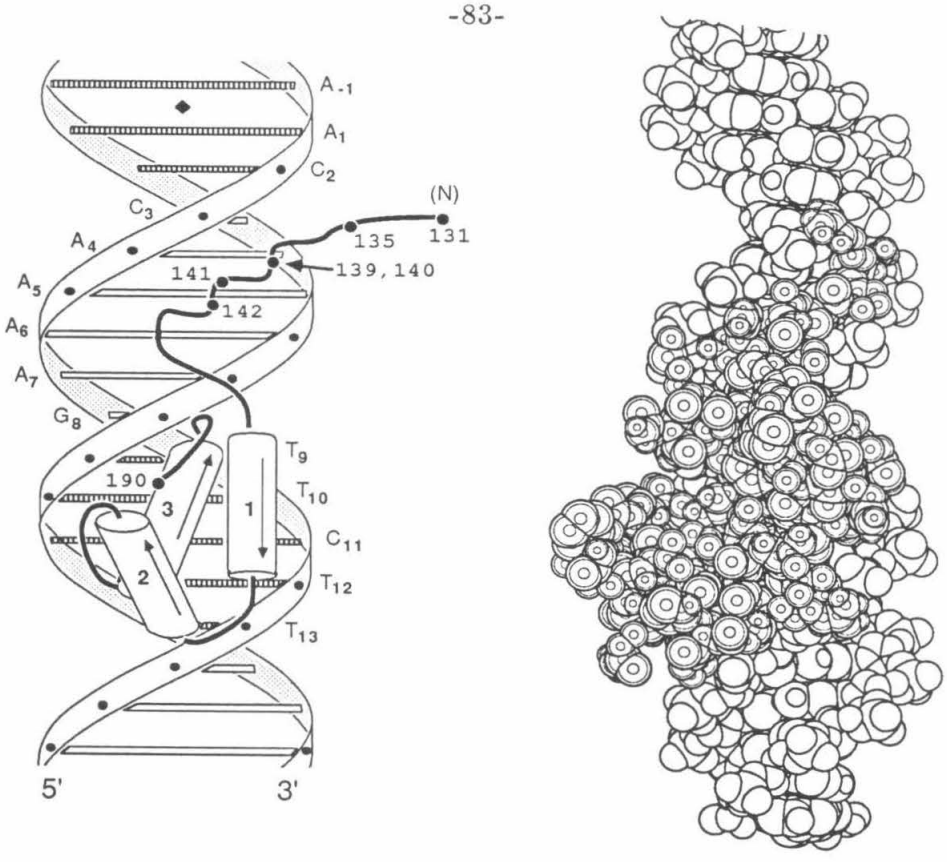
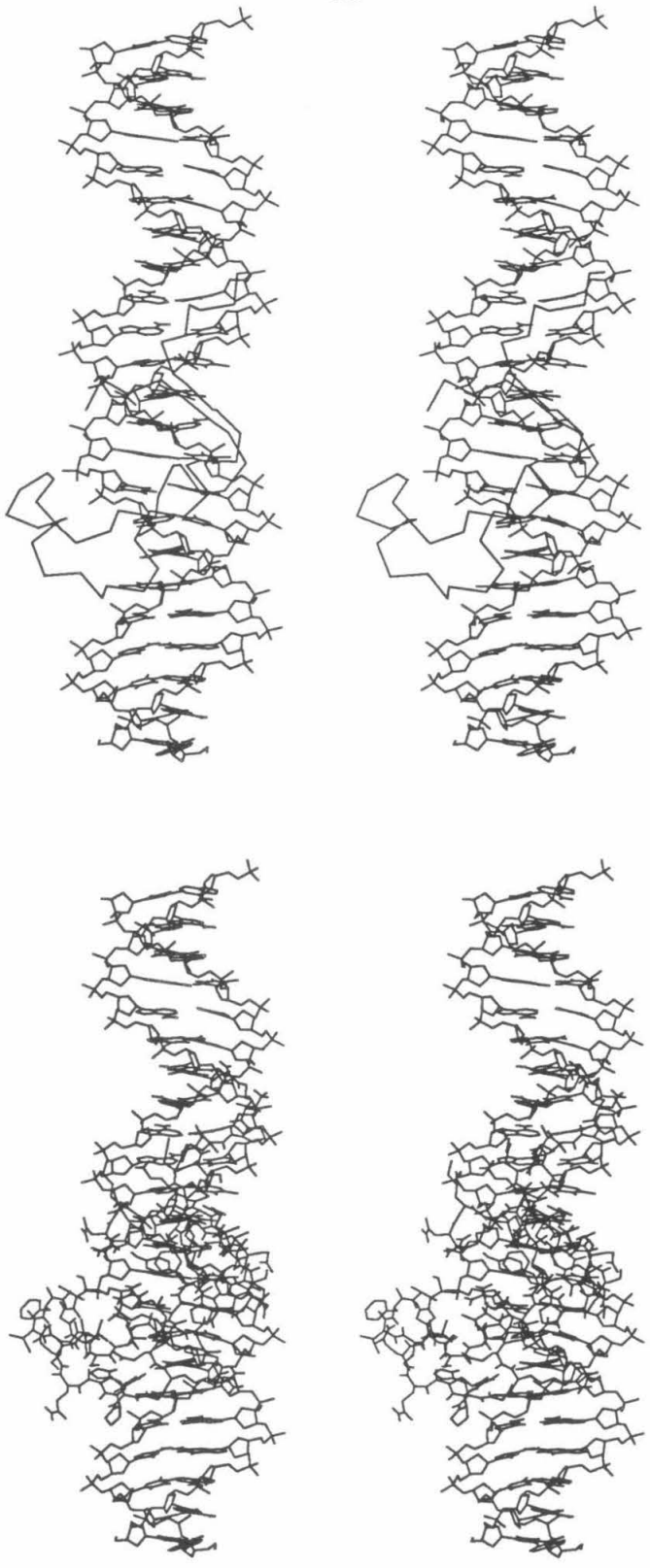


Figure 31: Revised binding model of the interaction between the synthetic peptides (and by extension, Hin) and a *hix* half-site. The left hand drawing shows the approximate course of the peptide backbone and the location of the three helices. The DNA sequence shown along the 5′ strand is that of *hix*L IRL and the solid diamond marks the psuedo-C_2 axis. The numbers along the peptide indicate the approximate locations of peptide alpha-carbons (Hin numbering) relative to the DNA. The wavy backbone near the amino terminus represents a specific interaction with the minor groove. Residue 142, the second Arg of the Arg–Pro–Arg sequence, interacts with A_5 and A_6 in the minor groove. Residue 141 is the proline. Residue 140 is the first Arg in the tripeptide sequence and interacts with the phosphate on the 5′ strand between C_3 and A_4. The right hand drawing is a spacefilling model of **HIN52**. The peptide atoms are highlighted. The Arg_{140} interaction with the phosphate is visible at the top of the peptide. In the spacefilling model only non-hydrogen and hetero-hydrogen atoms are shown. Shown on the following page are stereo drawings of this complex. The upper drawing shows *hix*L and the **HIN52** alpha-carbon trace. The lower drawing includes all peptide atoms except hydrogens bonded to carbons. In the lower drawing the interaction between the DNA and the Arg–Pro–Arg sequence is clearly seen.



Conclusion

It has been shown that attachment of EDTA at a specific amino acid in a synthetic DNA binding peptide creates a peptide which not only reports its binding sites on DNA but also the groove identity and nucleotide position of the modified amino acid. The methods required for placement of the EDTA at any position(s) in the peptide sequence have been described. The EDTA attachment strategy has been used to investigate the sequence specific DNA binding domain of Hin recombinase and has helped to identify a unique, sequence specific DNA recognition which occurs in the minor groove. Furthermore, the EDTA•Fe equipped peptides are competent at double strand cleavage. This is useful for identifying alternate DNA binding sites on large pieces of DNA. The method of conversion of either the amino terminus or a single lysyl side chain to a non-standard one is general and should allow the construction of other synthetic peptides capable of delivering a reporter function (not necessarily EDTA•Fe) to a target molecule. The target molecules could themselves be proteins or other types of macromolecules.

Materials and Methods

Nuclear magnetic resonance spectra were recorded on a JEOL JNM-GX 400 (^{1}H=399.65) and are reported in parts per million downfield from tetramethylsilane. Ultraviolet-visible spectra were recorded on a Beckman Model 25 spectrophotometer, a Cary Model 219 spectrophotometer or a Hewlett-Packard Model 8451A spectrophotometer. Mass spectra were recorded using fast atom bombardment (FAB) techniques at the Midwest Center for Mass Spectrometry at the University of Nebraska.

Agarose gels stained with ethidium bromide were photographed with Polaroid type 55 film and the negatives were scanned at 485nm on an LKB Ultroscan XL densitometer. Autoradiography of dried polyacrylamide gels with radiolabelled DNA was carried out at room temperature on Kodak X-Omat AR film. Dithiothreitol (DTT) was purchased from Boehringer Mannheim. All water used in biological reactions and dilutions was deionized and distilled using a Corning Mega-Pure system. Aqueous (α^{32}P) dATP and dCTP, triethylammonium salts, 3000 Ci/mmol, were from Amersham and aqueous (γ^{32}P)ATP, 5000-9000 Ci/mmol, was from ICN. Nucleotide triphosphates were from Pharmacia. All enzymes were from Boehringer Mannheim except the Klenow fragment of DNA polymerase I which was from New England Biolabs. Calf thymus DNA from Sigma was sonicated, deproteinized, and extensively dialyzed.[84] tRNA from Sigma was deproteinized by phenol extraction prior to use. Plasmids pMFB36 and pMS515 were gifts from Melvin Simon and were prepared using standard procedures.[85]

BEG Synthesis

All reagents were commercial grades and used without further purification except: dimethyl formamide (DMF), pumped *in vacuo* until there was no odor of amines and stored over 4A molecular sieves, and diaminoethane distilled from sodium hydroxide under argon. TLC's were performed on EM reagents silica gel plates (with UV indicator) and visualized by both UV and iodine staining; solvent systems were (A) 12% concentrated ammonium hydroxide in methanol, (B) 3% methanol in dichloromethane, (C) 3% concentrated ammonium hydroxide in methanol and (D) 6% methanol in dichloromethane. Chromatography was carried out with EM Reagents silica gel 60 (230–400 mesh). All non-aqueous reactions were run under argon with rigorous exclusion of water unless otherwise noted.

Initial attempts to prepare **BEG** via mono-saponification of the copper II chelate of EDTA tetrabenzyl ester similar to the preparation of the EDTA triethyl ester from tetra ethyl ester[86] were unsuccesful due to the low solubility of the Cu-EDTA tetrabenzyl ester in the aqueous ethanol saponification mixture. Attempted mono-hydrogenation (H_2, palladium on carbon) was also unsuccesful.

EDTA Tetrabenzyl Ester: To a solution of 2.9g (10mmol) EDTA and 5.5g (40mmol, 4eq.) freshly powdered potassium carbonate in 50ml hexamethyl phosphoramide (HMPA) was added 14.2g (80mmol, 8eq.) benzyl bromide. The mixture was held at 90°C for 22hrs, cooled to room temperature, diluted with 60ml water and extracted with diethyl ether (2x200ml). The combined ether layers were washed with 0.1N sodium hydroxide (2x60ml), water (60ml) and saturated NaCl (60ml), dried (Na_2SO_4) and the ether removed under reduced pressure. The crude product was recrystallized from 20ml 1:1 diethyl ether/petroleum ether (30–60°) at

-20°C and from 150ml diethyl ether at -20°C giving 4.9g (75%) **EDTA tetrabenzyl ester** as white needles. MP = 67.2–68.0°C, NMR ($CDCl_3$) δ7.33 (m, 20H), 5.10 (s, 8H), 3.63 (s, 8H), 2.90 (s, 4H).

***t*-Butyl-γ-aminobutyrate (1):** To a cold solution of 20ml concentrated sulfuric acid and 200ml dry dioxane (prepared cold) in a 500ml Parr bottle was added 12g (116mmol) γ-aminobutyric acid. The mixture was cooled to -70C in a dry ice/acetone bath and approximately 100ml of isobutylene was condensed in. The bottle was stoppered and shaken on a Parr–rocker (**no** connection to H_2 tank!) at room temperature. After 6hrs the bottle was removed (caution, the bottle is under pressure), cooled in ice, poured into cold ether (500ml) plus 250ml cold 5N sodium hydroxide and extracted. The aqueous layer was extracted with 2x200ml ether. The combined ether extracts were washed with 1N sodium hydroxide (2x200ml), dried (Na_2SO_4) and the ether and bulk of the dioxane removed by carefull rotovaping at 35C. The crude product was distilled at 35C and 40-50μm vacuum to a -70C trap giving 5.6g (30%) **1** as a colorless liquid which contains ca. 5% (wht) dioxane. NMR ($CDCl_3$) δ2.72 (t, 2H, j=7Hz), 2.27 (t, 2H, j=7Hz), 1.73 (p, 2H, j=7Hz), 1.45 (s, 9H), 1.14 (broad s, 2H). TLC(A) Rf 0.56.

***t*-Butyl-γ-aminobutyrate bromoacetamide(2):** To a solution of 9.5g (47mmol, 1.5eq) of bromoacetyl bromide in 175ml dichloromethane at -50C was slowly added a cold solution of 5g (47mmol, 1.5eq) triethylamine, 70mg dimethylaminopyridine and 4.7g (30mmol, 1eq) **1** in 50ml dichloromethane. The addition funnel was rinsed with 20ml dichloromethane. After 1hr at -50C the flask and cooling bath were allowed to warm to room temperature over 2 hours. The dichloromethane was removed *in vacuo* at 25C, and the residue taken up in 800ml ether and extracted with cold 0.1N sodium hydroxide (2x100ml), cold 0.1N hydrochloric acid

(100ml) and cold water (50ml). Drying ($MgSO_4$) and removal of solvents gave 8.1g (97%) **2** as a light yellow liquid. The product can be chromatographed with 3% methanol–dichloromethane. **2** does not keep very well and should be stored cold under nitrogen. NMR ($CDCl_3$) δ6.8 (broad s, 1H), 3.88 (s, 2H), 3.33 (q, 2H, j=6Hz), 2.23 (t, 2H, j=7Hz), 1.84 (p, 2H, j=7Hz), 1.46 (s, 9H). TLC(B) Rf 0.35.

3: To 27ml (≈40eq.) freshly distilled 1,2–diaminoethane at 10C was slowly added 1.6g (5.8mmol) **2** in 2ml dimethylformamide. The solution was stirred for 1hr at 10C and 2hr at 22C. Excess diaminoethane was removed by distillation under high vacuum at 25C and the crude product chromatographed with 5% concentrated ammonium hydroxide in methanol to give 1.1g (75%) **3** as a yellow oil. **3** does not keep very well and should be stored cold under nitrogen. NMR ($CDCl_3$) δ7.27 (broad s, 1H), 3.30 (q, 2H, j=7Hz), 3.28 (s, 2H), 2.82 (t, 2H, j=6Hz), 2.66 (t, 2H, j=6Hz), 2.27 (t, 2H, j=7Hz), 1.81 (p, 2H, j=7Hz), 1.68 (broad s, 3H). 1.46 (s, 9H). TLC(C) Rf 0.33.

BEG-tBu: To a solution of 400mg (1.54mmol) **3** and 632mg (4.9mmol, 3.2eq.) diisopropylethylamine in 20ml DMF was slowly added a solution of 1.18g (4.9mmol, 3.2eq) 97% benzyl bromoacetate in 10ml DMF. The mixture was stirred 1hr at room temperature and 15hrs at 35C. The DMF was removed *in vacuo* at 35C and the crude product taken up in 100ml diethyl ether, washed with water (3x50ml) and saturated sodium chloride (50ml), dried (Na_2SO_4) and the ether removed under reduced pressure. Chromatography with 3% methanol in dichloromethane gave 890mg (81%) **BEG-tBu** as a yellow oil. **BEG-tBu** can also be chromatographed with 25% ethyl acetate/dichloromethane plus 0.5% water. The column must be completely equilibrated with solvent. NMR ($CDCl_3$) δ8.07 (t, 1H, j=6.0Hz), 7.33 (m, 15H), 5.11 (s, 6H), 3.58 (s, 4H), 3.40 (s, 2H), 3.27 (s, 2H), 3.23 (q, 2H, j=6.7Hz),

2.80 (t, 2H, j=5.1Hz), 2.75 (t, 2H, j=5.1Hz), 2.22 (t, 2H, j=7.5Hz), 1.76 (p, 2H, j=7.5Hz), 1.42 (s, 9H). TLC(B) Rf 0.25.

BEG: To a solution of 310mg (440μmol) **BEG-tBu** in 3ml dichloromethane was added 2ml trifluoroacetic acid with vigorous stirring. The mixture was stirred for 30min at room temperature, the solvents removed under vacuum at room temperature and the residue taken up in 100ml ethyl acetate. The solution was extracted with cold saturated sodium bicarbonate (2x150ml, if the mixture emulsifies allow it to sit overnight at 4C), saturated sodium chloride (150ml) followed by drying (Na_2SO_4, do not use $MgSO_4$) and ether removal under reduced pressure giving 230mg (78%) **BEG** as a light yellow oil. If needed, **BEG** can be chromatographed on a well equilibrated silica column using 8% methanol in dichloromethane with $\approx$0.8% water (saturated). **BEG** does not keep very well and should be stored under vacuum. NMR (D_6–DMSO+TFA) δ8.50 (t, 1H, j=6.0Hz), 7.37 (m, 15H), 5.23 (s, 2H), 5.12 (s, 4H), 4.40 (s, 2H), 4.08 (s, 2H), 3.76 (s, 4H), 3.39 (t, 2H, j=7.0Hz), 3.16 (m, 4H), 2.26 (t, 2H, j=7.3Hz), 1.68 (p, 2H, j=7.1Hz). TLC(D) Rf 0.32. FAB–MS; expected M^*+H=648.2920, observed=648.2914 (error=1ppm).

Peptide Synthesis

Peptides in the **HIN31** to **HIN60** series were prepared by the California Institute of Technology Microchemical Facility under the direction of Dr. Suzanna J. Horvath.[1] The general procedure is described below. Reagents, used without further purification, were: Burdick and Jackson HPLC grade DMF, Mallincrodt "Ar analytical grade" DCM, BOC-L-amino acids (with the exception of N^{α}-BOC-N^{im}-DNP-L-histidine which was from Fluka), PAM-Asn resin and 2M DCC/DCM

from Peninsula Laboratories. Diisopropyl ethyl amine was from Aldrich (98%). A Peninisula Laboratories reactor and wrist–shaker were used for the synthesis.

In the general procedure described below "wash" refers to a flow wash, i.e. the wash solvent is added with the reactor connected to a source of vacuum so that the solvent passes imediately through the reactor. In general, the peptide/resin mixture was not drained completely. Instead, the solution was drained only until it was level with the top of the resin. This avoids drawing air through the resin which, particularly when the resin is wet with with DCM or TFA/DCM, can condense considerable moisture on the peptide.

Abbreviations:

DMF = dimethylformamide
DCC = N,N′–dicycloheylcarbodiimide
DCM = dichloromethane
DIPEA = diisopropyl ethyl amine
HOBt = N-hydroxybenzotriazole
TOS = tosylate
OBzl = benzyl ester
Bzl = benzyl ether
Cl-Z = p-chlorobenzyloxy carbonyl
Br-Z = p-bromobenzyloxy carbonyl
DNP = 2,4-dinitrophenyl
BOC = t-butyloxycarbonyl

Amino Acids:

BOC–Arg–Tos
BOC–Asp–OBzl
BOC–Cys–4-MeBzl
BOC–Glu–OBzl
BOC–His–DNP
BOC–Lys–Cl-Z
BOC–Ser–Bzl
BOC–Thr–Bzl
BOC–Tyr–Br-Z
rest BOC–AA's

NINYDRIN Assay for primary amines:[87]

The ninhydrin assay does not work with N-terminal secondary amino acids (ie., proline) and is unreliable with N-termiminal tryptophanes.

Reagents:

A. 40g 100% phenol in 10ml absolute ethanol.

B. 0.2mM potassium cyanide in pyridine (pyridine is distilled from ninhydrin).

C. 500mg ninhydrin in 10ml absolute ethanol.

D. 0.5M tetraethylammonium chloride in DCM.

Procedure:

1. Remove ~5mg peptide/resin from reactor.

2. Wash well with DCM and 1:1 methanol/DCM on a sintered glass funnel.
3. Allow to dry on funnel ~3min.
4. Accurately weigh the sample into a small test tube and add:
   a. 2 drops *A*.
   b. 4 drops *B*.
   c. 3 drops *C*.
5. Heat at 100°C, 5min.
6. Cool to room temperature.
7. If there is substantial color in the beads filter through a pasteur pipet plugged with glass wool and wash with 60% ethanol (2x200λ), 2x200λ *D*, and 2x200λ 60% ethanol.
8. Dilute with 60% ethanol to 3ml.
9. Mix thoroughly and measure O.D. at 570nm versus 60% ethanol blank;

$$\mu\text{mol amine/g resin} = \frac{(OD_{570} \times 200)}{\text{wht(mg)}}$$

COUPLING CYCLE:[88]

I. DCM Flow Wash (~5 vessel volumes).

II. 65% TFA/DCM tBOC Removal;
   A. 1min
   B. 15min.

III. DCM Flow Wash, ~10 vessel volumes.

IV. 10% DIPEA/DCM Neutralization (also rinse the reactor cap);
   A. 1min
   B. DCM flow wash, ~4 vessel volumes.
   C. 1min
   D. DCM flow wash, ~10 vessel volumes.

V. Coupling Cycles;
   A. **Symmetric Anhydride (SA), all residues except Asn, Gln and Arg.**
      1. Add a solution of 2mmol BOC–AA in 2ml DCM (may need a little DMF to get completely into solution).
      2. Add 1mmol DCC/2ml DCM, add DCM until inversion wets everything.
      3. After 10min do a ninhydrin test and add DMF to reactor until 2/3 – 3/4 full. If the coupling passes the ninhydrin test go to step I.
      4. After 30min take another ninhydrin. If ninhydrin still less than ~99.6% early in the synthesis or ~99% later, then recouple.

Preformed Symmetric Anhydride (PSA) Recouple:

a. DMF flow wash, 5 vessel volumes.
b. 10% DIPEA/DMF neutralization (prepare fresh), 1min.
c. DMF flow wash, 10 vessel volumes.
d. If removing DCM from **PSA** then DMF wash 5min.
e. **PSA** preparation:
   i. Dissolve 4eq. AA in minimum DCM.
   ii. Add 2eq. DCC/DCM, stir 5min.
   iii. Filter to remove DCU (strip DCM). Add to reactor.
   iv. Fill reactor with DMF.
   v. Ninhydrin test at 30min.

B. **DCC/HOBt Coupling for Arg–TOS, Gln and Asn:**

1. 2mmol HOBt in 2ml DMF.
2. Add 2mmol DCC in 4ml DCM.
3. Add 2mmol AA in 2ml DMF.
4. Add to resin, top reactor off with DMF.
5. Ninhydrin at 30–45 minutes. If needed, this reaction can proceed for up to 2 hours.
6. If needed, recouple Gln or Asn with the same procedure (recouple Arg using **PSA**):
   a. DMF wash, 5 vessel volumes.
   b. 10% DIPEA/DMF wash (prepared fresh), 1min.
   c. DMF wash, 10 vessel volumes.
   d. Recouple as above to a 1hr ninhydrin, then a 2hr ninhydrin.

DNP Removal:

Resin Bound Peptide:

1. 20% 2-mercaptoethanol, 10% DIPEA in DMF, 2x30min.
2. DMF wash, 5 vessel volumes.

Residual DNP Removal (after HF cleavage):

1. Peptide at 10~20mg/ml in 4M guanidine-HCl, 50mM TRIS, pH 8.5, 20% thioethanol.
2. 50°C, 1hr.

Optional Small Molecule Removal:

Mini G-10 Sephadex column, elute with 5% HOAc.

HF Cleavage:

1. Remove terminal t-BOC, wash, neutralize with DIPEA/DCM as above, wash well with DCM then 1:1 methanol/DCM.
2. Vacuum dry, 30min.
3. For a 100mg peptide/resin sample add:
    a. 0.5ml p-thiocresol (works best if this is added first).
    b. 0.5ml p-cresol.
4. Cool to -70°C and condense in 10ml HF.
5. Stir 1hr at 0°C.
6. Remove HF under reduced pressure.
7. Add, with stirring, 20ml diethyl ether to precipitate the peptide.
8. Filter, wash the precipitated peptide well with ether (20~50ml) on the funnel.
9. Wash crude peptide through funnel with DD water or 5% HOAc.
10. Freeze and lyophilize.

Peptide Purification

All peptides described in this report were purified by reverse phase HPLC on a Beckman Instuments HPLC using a 421A controller, 165 dual wavelength UV/VIS detector, two 114M analytical pumps and a Spectra Physics SP8780XR auto-injector controlled by the 421A controller.

HPLC:

Analytic:

column: Vydac C_4, 210x4mm analytic column.
flow: 0.7ml/min
temperature: 37°C
solvent **A**: 0.1% TFA/water
solvent **B**: 0.1% TFA/60% acetonitrile/water
gradient: 0% to 100% **B** in 60min
detection: UV at 215nm (peptide) 2AUFS
: UV at 254nm (aromatics) 0.2AUFS
sample: ~200μg in water, 5% HOAc or **A** for crude material,
: ~30μg for purified material.

Sample straight from 5% HOAc after HF or after residual DNP removal.

Prep:

column: Vydac C_4, 210x25mm semiprep column.
flow: 3.0ml/min
temperature: 37°C
solvent **A**: 0.1% TFA/water
solvent **B**: 0.1% TFA/60% acetonitrile/water
gradient: 0% to 100% **B** in 240min
detection: UV at 215nm (peptide) 2AUFS
: UV at 254nm (aromatics) 2AUFS
sample: Up to 200mg in water, 5% HOAc, **A** or DNP mixture.

After residual DNP removal load the entire solution in line as solvent **A**. Run solvent **A** until initial large peak is almost completely off then start gradient and fraction collection. Re-inject (analytical column) 20–100μl samples of each fraction, pool good fractions and lyophilize.

BEG Coupling to Resin Bound Peptides.

The general procedure for the preparation of the N-terminal EDTA–GABA peptides was as follows.[70] A ~100mg sample of peptide/resin (loading ~100μmol/g, total peptide ~10μmol) was placed in a 12x80mm reactor, and the resin was swollen in DCM for 15 minutes. The resin was washed with DCM (5x), the terminal t-BOC was removed and the resin neutralized as described above. **BEG** (125mg, 194μmol, 20eq.) and 50mg (330μmol, 34eq.) N-hydroxybenoztriazole monohydrate was dissolved in 2ml DMF and 42mg (203μmol, 21eq.) DCC was added. The solution was stirred 15 minutes at room temperature and added to the peptide/resin along with sufficient DMF to fill the reactor 2/3 full. Ninhydrin assays were used to follow the reaction and typically indicated ≥97% reaction after 3–4 hours. The peptide/resin was then washed with DMF (4x) followed by DNP removal (see above). The final washes of the peptide/resin were; DMF (4x), 1:1 DCM/DMF (2x), DCM (4x) and

1:1 Methanol/DCM (4x). Though there is no terminal t-BOC group remaining, a final wash with TFA/DCM helps to remove any remaining exogenous materials. The peptide/resin is then dried under vacuum and subjected to HF cleavage as described earlier. The crude peptide is then treated for residual DNP and purified by preparative HPLC (see above). Typically 3–6mg (5–10% theoretical) of purified peptide was obtain after HPLC.

EG*50, EG*52 and $(EG)_2$52 Synthesis

The peptides with N^{ϵ}-(EDTA–GABA)-L-lysine near the C terminus were prepared using the same peptide synthesis protocol described earlier except for the incorporation of $\overset{*}{\text{Lys}}$. Reagents were the same except; "Biograde" TFA from Halocarbon Inc. and N^{ϵ}-FMOC-N^{α}-t-BOC-L-lysine ($\overset{*}{\text{Lys}}$) which was from Chemical Dynamics Corp.

The synthesis was started with 500μmol PAM-Asn-t-BOC resin (670μmol amine/g). The peptide/resin, after DCM swelling and washing, gave a background ninhydrin of 1.9μmol/g (99.7% t-BOC protected), standard t-BOC removal gave a ninhydrin of 110μmol/g. It may be advantageous at this point to determine free amine concentration using a picric acid titration to avoid the uncertanties of variable ninhydrin yields on very small peptides.[87] The N^{ϵ}-FMOC-N^{α}-t-BOC-L-lysine was coupled to the peptide/resin using the standard SA procedure and gave a ninhydrin value of 1.6μmol/g (99.8%). The peptide/resin was washed with DCM (5x) and DMF (5x). The FMOC group was removed with 20% piperidine/DMF[89] (1min and 10min treatments) and gave a ninhydrin value of 370μmol/g.

BEG (1.3g, 2mmol, ~3eq.) was activated with 460mg (3mmol, ~4.5eq.) HOBt and 412mg (2mmol) DCC in 4ml DMF for 30min and added to the reactor with

DMF. **BEG** coupling, as monitored by ninhydrin analysis, was 7.7μmol/g (98.4%) after 45min and 1.1μmol/g (99.8%) after 100min. The peptide/resin was then washed with DMF (5x) and DCM (5x), the N^{α}-t-BOC removed and the resin neutralized using the standard procedures. The deprotected peptide/resin gave a ninhydrin value of 320μmol/g.

Arg-TOS was coupled to the peptide/resin using the standard HOBt procedure and gave a ninhydrin value of 1.9μmol/g (99.5%) after 30min. At this point a small sample of peptide/resin was removed, HF cleaved and purified by HPLC to afford **EG*3**. High resolution fast atom bombardment (FAB) mass spectra; calculated M^{*}-H^{+} = 776.390, observed = 776.391, difference = 1.3ppm. The 400MHz NMR of **EG*3** in 10% TFA/D_6–DMSO, with partial peak assignments, is shown in figure 32.

Though it did not appear to be necessary in this synthesis, it may be advantagous to "cap" the peptide after the **BEG** coupling to Lys^{*} and the subsequent amino acid coupling with acetic anhydride or acetyl imidazole[90] to minimize potential side chain elongation during the remainder of the synthesis.

The synthesis was continued to the 30mer stage at which time a sample of **EG*30** was removed, HF cleaved and purified by HPLC. The 400MHz NMR of **EG*30** in (10% TFA/D_6–DMSO) is shown in figure 33. Integration indicates that the expected amount of EDTA–GABA is still present in the peptide at this stage of the synthesis.

The synthesis was then continued to the 50mer stage, where another small sample was removed (**EG*50**), and to the 52mer stage (**EG*52**). Finally, another **BEG** coupling was performed on the amino terminus to afford $(\mathbf{EG})_2\mathbf{52}$. The

Figure 32: 400MHz NMR of **EG*3**.

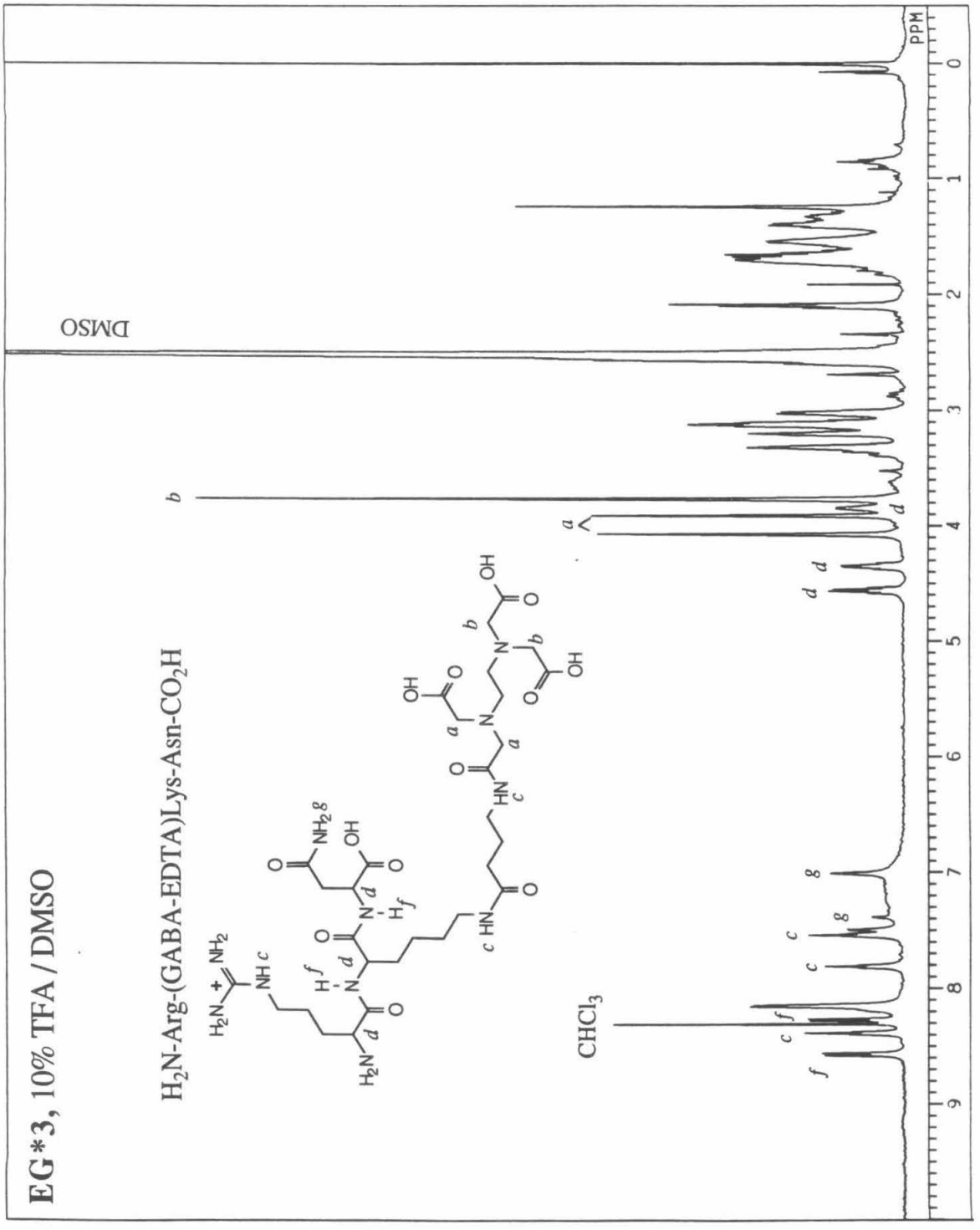


Figure 33: 400MHz NMR of **EG*30**.

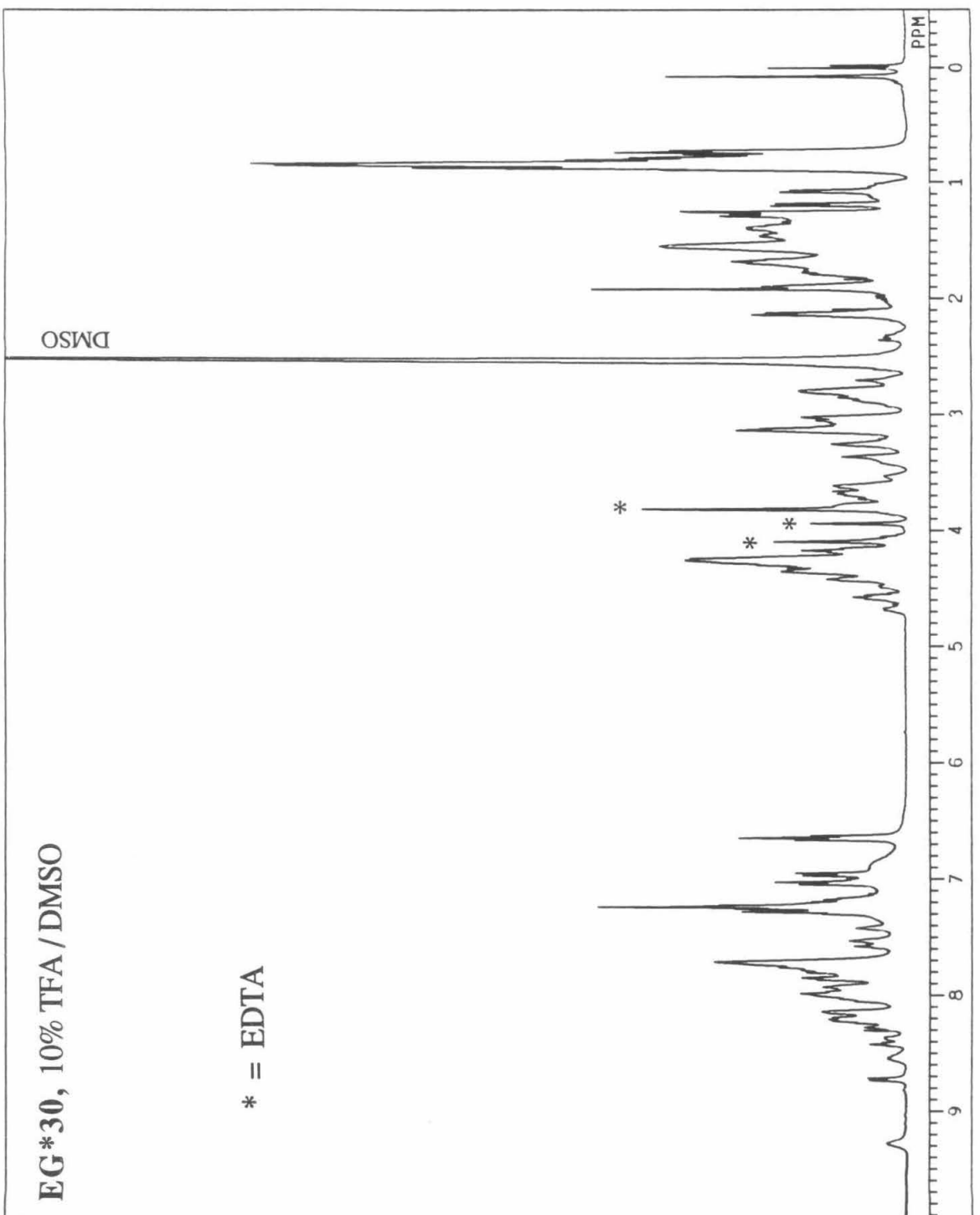


Table 4: Synthesis record for **EG*52**.

number	AA	cycle type	couples	ninydrin μmol/g	% yield	comments
1	Asn	–	1	"1.9"	"99.7"	background ninhydrin
2	Lys*	SA	1	1.6	99.8	
2.5	**BEG**	HOBt	1	1.1	99.8	
3	Arg	HOBt	1	1.9	99.5	60mg sample **EG*3**
4	Lys	SA	1	1.2	99.7	
5	Lys	SA	1	0.6	99.8	
6	Ile	SA	1	0.6	99.8	
7	Ser	SA	1	1.3	99.5	
8	Ser	SA	1	1.9	99.3	
9	Ala	SA	1	1.5	99.4	
10	Pro	SA/PSA	3	1.9	99.3	
11	Phe	SA/PSA	2	0.7	99.7	
12	Tyr	SA/PSA	2	0.6	99.8	
13	Arg	HOBt/PSA	2	0.7	99.7	
14	Tyr	SA/PSA	2	1.0	99.5	
15	Leu	SA	1	1.0	99.5	
16	Thr	SA/PSA	2	2.0	98.9	
17	Ser	SA/PSA	2	1.8	99.0	
18	Val	SA/PSA	2	1.5	99.2	
19	Gly	SA/PSA	2	1.6	99.1	
20	Ile	SA/PSA	2	1.6	99.1	
21	Gly	SA	1	1.6	99.1	
22	Phe	SA/PSA	2	1.6	99.0	
23	Ile	SA/PSA	2	2.5	98.5	
24	Ile	SA/PSA	3	2.5	98.5	
25	Ala	SA/PSA	2	1.8	98.8	
26	Leu	SA/PSA	2	1.8	98.8	
27	Gln	HOBt	2	2.3	98.5	
28	Gln	HOBt	2	2.2	98.6	
29	Arg	HOBt/PSA	3	1.5	99.0	
30	Pro	SA/PSA	2	2.5	98.2	97mg sample **EG*30**
31	His	SA/PSA	2	"1.5"	"98.9"	background ninhydrin
32	Gly	SA	1	1.0	99.3	
33	Lys	SA/PSA	2	1.8	98.6	
34	Glu	SA/PSA	2	1.3	99.0	
35	Leu	SA	1	1.5	98.8	
36	Leu	SA	1	1.4	98.8	
37	Arg	HOBt/PSA	2	1.4	98.8	
38	Ser	SA/PSA	2	2.3	98.0	
39	Ile	SA/PSA	2	1.8	98.4	
40	Gln	HOBt	2	1.2	98.9	
41	Glu	SA/PSA	2	1.8	98.4	
42	Gln	HOBt	2	1.4	98.7	
43	Glu	SA/PSA	2	1.6	98.5	
44	His	SA	1	1.6	98.4	
45	Lys	SA/PSA	2	1.8	98.2	
46	Asn	HOBt	2	1.1	98.8	
47	Ile	SA/PSA	2	1.4	98.6	
48	Ala	SA/PSA	2	2.4	97.5	
49	Arg	HOBt/PSA	2	2.0	97.8	
50	Pro	SA	1	2.6	97.2	30mg sample **EG*50**
51	Arg	HOBt/PSA	2	"1.0"	"98.8"	background ninhydrin
52	Gly	SA	1	1.4	98.4	**EG*52**
53	BEG	HOBt	1	3.5	96.1	**$(EG)_2$52**

average % yield = 98.9±0.6

0.989^{52} = 57%

peptides were all subjected to DNP removal, HF deprotection, residual DNP removal and purified by HPLC using the standard procedures. Table 4 contains the complete ninhydrin and coupling procedure record for the synthesis.

Plasmid Preparation

The following procedure was used to prepare both pMFB36[91] and pBR322 (strain HB101).

Cell Growth:

Prepare 2L VB minimal media (1L each in two 2L flasks)

per 1L:	
2 g	Casamino acids
20 ml	50X V-B salts
50 ml	autoclaved thymine (1mg/ml)

H_2O to give ~980 ml total volume

Autoclave on wet cycle, allow to cool to r.t., add:

20 ml	Sterile (autoclaved) 20% (w/v) D-glucose
2 ml	Ampicillin (25mg/ml, filtered sterile 0.45μM)
1 ml	Thiamine·HCl (1mg/ml, filtered sterile 0.45μM)

Inoculate 1L of VB minimal with 4ml of an overnight culture of bacteria in L-broth and 50μg/ml ampicillin (to keep selective pressure on, increases the yield of plasmid). Shake at 37°C at 200rpm.

Amplify when $O.D._{590nm}$ = 0.6–0.9 (best to amplify at higher O.D.) Amplify by adding 2ml Chloromycetin sodium succinate (100mg/ml) per 1L of broth (final concentration = 200μg/ml)

> Chloromycetin sodium succinate = Chloroamphenical (Cm) and the 100mg/ml solution is made by adding 10ml of sterile H_2O(*via* syringe) to a 1g vial of Cm and shaking until it has completely dissolved.

Amplify for 12–17 hours at 37°C with shaking.

Harvest Cells:

Chill 2L flasks on ice bath with frequent agitation

Load chilled cells into 500 ml bottles, balance with double pan balance.

Spin at 5000 rpm, 4°C for 20′, discard supernatants and wipe bottles dry.

Lysis:

Resuspend cell pellet in minimum volume (4–6ml) of:

25% Sucrose

50mM Tris pH 8

Transfer resuspended cells to a 70Ti (polycarbonate) tube and add:
1.6ml lysozyme (5mg/ml prepared just before use)
–Let sit 5′ on ice with occasional stirring
5ml 0.5M EDTA pH 8
–Let sit 5′ on ice with occasional stirring

One tube at a time, add:

10%	Triton X-100
50mM	Tris pH 8
60mM	EDTA pH 8

Solution to almost fill tube – (air bubble good for agitation)

Shake vigorously (!!) until the lysates "clear"
Looks like shampoo that's been shaken, usually 20–30sec is O.K.
Try to avoid clumping of lysates

Balance tubes to ±0.01g, load into Ti70 rotor
Spin 30Krpm, 4°C for 30′

Pour off supernatants into 50ml polypropylene centrifuge tube
** Avoid gluey junk near cell debris **

Add 350λ of 2mg/ml RNase A (DNase free as in Maniatis), incubate 20′ at 37°C

Add 1/3 volume (~6–7ml) of 30% PEG 8000/1.5M NaCl to each tube, shake to mix and incubate 0.5–8hr on ice (longer times give better yields).

Centrifuge PEG/NaCl mixture: 20′ at 6Krpm, 4°C, discard supernatants

Resuspend pellet in 2–5ml cold 10X TE using a glass rod
100mM Tris pH 8
10mM EDTA pH 8

CsCl spins:

Set up 1st CsCl isopycnic banding run:
(use 1x40ml quick seal tube per 500ml of original broth)
59.52g CsCl (ultrapure)
+ Pellets in 10X TE
+ 3.2ml ethidium bromide (10mg/ml)
+ 1X TE buffer to give a final weight of : 124.00g
This equals 80ml of 48% (w/w) CsCl and has a ρ ~1.55.

Stir to dissolve CsCl
Load into VTi50 quickseal tubes, balance to ±0.01g and seal
Run in VTi50 rotor at 42Krpm, 17°C for 20hr.

Vent tube at top with an 18G needle and pull lower (plasmid) band (using UV illumination) by puncturing quick seal tube ~1 cm below band with an 18G needle and 10ml syringe. Keep needle bevel up and protect DNA from excessive exposure to UV light.

Transfer bands into VTi65 tubes, fill tubes with CsCl filling solution (ρ = 1.55, 400μg/ml ethidium bromide).

Balance tubes to ±0.01 g, seal and load into VTi65.2 Rotor.

Spin at 55Krpm, 17°C for 12hr.

Extract plasmid (lower) band as before, but use a 3ml syringe.

Extract ethidium bromide with 1X TE saturated isoamyl alcohol:

about 12-15 extractions of 1 volume each are needed

(DNA is in bottom layer)

check for EtBr removal with UV light

Dialyze plasmid versus 1X Reaction Buffer (40mM Tris acetate, pH 7.9, 5mM sodium acetate) at 4°C.

High Resolution DNA Studies

Preparation of Specifically Labelled DNA Fragments. Endlabelling was carried out in a similar way for each fragment. The following is a description of the procedure for the labelling and isolation of the ~400bp fragment from pMFB36. Superhelical pMFB36 plasmid DNA was digested with the restriction endonuclease Xba I, then labeled at the 3′ end with (α^{32}P)dATP using the Klenow fragment of DNA polymerase I. A second digest with restriction endonuclease Eco RI yielded the 3′ labeled ~400bp fragment. Cleavage of pMFB36 with Xba I and successive treatment with calf alkaline phosphatase (CAP), (γ^{32}P)ATP and T4 polynucleotide kinase, and Eco RI yielded the 5′ labelled fragments. The restriction fragments were isolated by non-denaturing polyacrylamide gel electrophoresis. Guanine specific chemical cleavage lanes were carried out according to Maxam and Gilbert.[92] The specific reaction procedures and conditions are outlined below.

3′ and 5′ Labelling of ~400bp Fragment of pMFB36

General Procedures

Ethanol Precipitation:

$\frac{1}{10}$vol 20% NaOAc
vortex
2.5vol EtOH
invert 5x

For fragments larger than 1kb spin at 17krpm for 10min at 4°C.
For smaller fragments spin 20min at 4°C.
Discard supernatants

Ethanol Re-precipitation:

50-200λ 0.3M NaOAc
Vortex
2.5vol EtOH
invert 5x

Centrifuge as above.
Discard supernatants

Phenol Extraction:

1vol Phenol saturated with 1x TE
Vortex 1min
Centrifuge 20sec
Discard LOWER layer

Ether Extraction:

~3vol H_2O saturated ether
Vortex 1min
Centrifuge 20sec
Discard UPPER layer

Choroform Extraction:

1vol 24:1 $CHCl_3$/*i*-amyl alcohol
Vortex 1min
Centrifuge 20sec
Discard LOWER layer

Xba I Digest of 15μg pMFB36:

100λ	pMFB36 (150μg/ml)
20λ	10x Xba I buffer
78λ	DD H_2O
2λ	Xba I (20u/λ)
200λ	37°, ~3hr

Phenol extract 2x.

Ether extract 3x.

Etanol precipitate 2x.

70% Ethanol wash.

3′ Labelling:

15μg	Xba I'd pMFB36
6λ	DD H_2O
5λ	10x Hae III buffer
5λ	66mM DTT (10mg/ml)
25λ	dTTP+dGTP+dCTP (5λ, 10mM)
10λ	α-^{32}P dATP
4λ	polymerase (5u/λ)
50λ	25°, 20min
1λ	dATP (10mM)
1λ	dTTP (10mM)
1λ	dGTP (10mM)
1λ	dCTP (10mM)
54λ	25°, 10min

CAPing for 5′ labelling:

15μg	Xba I'd pMFB36
80λ	DD H_2O
10λ	10x CAP Buffer
10λ	CAP (1uλ)
100λ	37°, 30 min

Phenol Extract 2x.

Chloroform Extract.

Ether Extract 3x.

Ethanol Precipitate 2x.

70% Ethanol Wash.

5′ Labelling:

15μg	Xba I'd & CAP'd pMFB36
34λ	DD H_2O
5λ	10x Kinase buff
5λ	66mM DTT (10mg/ml)
~4λ	γ-^{32}P dATP
2λ	Kinase (NBL, 2u/λ)
50λ	37°, 45 min

Ethanol precipitate 2x.

70% Ethanol wash.

Eco RI Digest:

15μg	labelled DNA
41λ	DD H_2O
5λ	10x Eco RI buffer
4λ	Eco RI (20u/λ)
50λ	37°, 1-3hr

Ethanol precipitate 2x.

70% Ethanol wash.

Resuspend in 5λ ficol buffer.

Prep Polyacrylamide Gel (5%, 1:30):

Plug (10%, 1:30):

2.5g	acrylamide
85mg	bisacrylamide
30mg	ammonia persulfate
5ml	5x TBE
to 25ml	with DD H_2O

Stir ~30min, millipore filter, add 40λ TEMED and pour plug.

Gel (5%, 1:30):

4.0g	acrylamide
136mg	bisacrylamide
48mg	ammonia persulfate
16ml	5x TBE
to 80ml	with DD H_2O

Stir ~30min, millipore filter, add 40λ TEMED, pour gel, let stand $\geq$2 hours.

Load labelled nucleotides and electrophorese at 240V until BPB reaches the bottom of the gel.

Excise bands after 3min autoradiogram, smash ploy acrylamide fragments between parafilm and transfer to eppendorf tubes.

500λ elution buffer:
0.25% sodium dodecyl sulfate
1mM disodium EDTA
10mM magnesium chloride
500mM ammonium acetate

Seal with parafilm, 10–20hrs, 37° shaking incubator.
Centrifuge through a quick-sep column (0.45μm CA, 15min).
600λ Phenol extract, spin 5min over phenol.
Dry butanol extract until volume ~200λ.
Ethanol precipitate 2x.
70% Ethanol wash.
Vacuum dry.
Resuspend in DD H_2O at ~18kcpm/λ.

Maxam-Gilbert G-Reaction:

Start Buffer:

500λ 100mM sodium cacodylate (pH 8.0)
100λ 100mM magnesium chloride
10λ 100mM disodium EDTA
390λ DD H_2O

1000λ

Stop Buffer:

94λ 4N sodium acetate
18λ thioethanol
126λ DD H_2O
12λ tRNA (2mg/ml)

250λ

G Reaction:

200λ Start Buffer
5λ ^{32}P labelled DNA

205λ

1λ dimethyl sulfate
20°, 20sec

Stop:

50λ stop buffer, vortex
750λ ethanol, invert 5x
centrifuge 20min at 4°C
Discard supernatants

Ethanol re-precipitate.
70% Ethanol wash.

100λ 1M Piperidine (99λ piperidine + 900λ DD H_2O)
90°, 30 min
freeze, lyophilize

50λ DD H_2O
freeze, lyophilize

10λ DD H_2O
freeze, lyophilize

Resuspend in formamide loading buffer at ~9kcpm/λ.

Affinity Cleavage Reactions

The concentration of each peptide compound was determined spectroscopically assuming an extinction coefficient of $2800M^{-1}$ at 275nm (two tyrosines). Peptide-EDTA•Fe(III) solutions were prepared by mixing 0.5mM solutions of peptide-EDTA and freshly prepared $Fe(NH_4)(SO_4)_2 \cdot 12H_2O$ followed by dilution to the appropriate concentration. The cleavage reactions were run with ~18kcpm of ^{32}P–endlabelled restriction fragment made up to a total concentration of 400μM phosphate tRNA (final concentration). To 9μl of ^{32}P–DNA/tRNA/buffer solution, 3μl of peptide-EDTA•Fe(III) complex was added. Cleavage was initiated by the addition of 3μl of a 25mM dithiothreitol (DTT) solution to the reaction mixture to give a total reaction volume of 15μl and a final concentration of 5mM DTT. The reaction buffer was prepared to 20mM TRIS•HCl (pH 8.0) and 20mM sodium chloride. The reactions were typically run for one hour at room temperature and were terminated by ethanol precipitation.

Footprinting Reactions

The footprinting reactions were run in a buffer of pH 8.0 containing 20mM TRIS•HCl and 20mM sodium chloride and were made up to a final tRNA phosphate concentration of 400μM. The footprinting reagents, MPE•Fe(III) and EDTA•Fe-(III), were prepared by mixing at 1mM with one equivalent of $Fe(NH_4)(SO_4)_2$ and diluting immediately. To a 6μl solution of DNA and buffer, 3μl of the molecule to be footprinted at an appropriate concentration was added. Next, 3μl of footprinting reagent (10μM MPE•Fe(III) or 1.5mM EDTA•Fe(III)) was added. The cleavage reaction was initiated by the addition of 3μl of 15mM DTT or 10mM sodium ascorbate to give final concentrations of 3mM DTT and 2mM sodium ascorbate, respectively. MPE Footprinting reactions were allowed to continue for 7min at room temperature, and EDTA reactions for 20min at room temperature. Both were terminated by ethanol precipitation.

High Resolution Gels

The precipitated reactions were suspended in 4μl of 1X TBE, 80% (vol/vol) formamide at pH 8.3 containing bromophenol blue and xylene cylanol tracking dyes. Recrystalization of the formamide is important to remove any traces of formic acid which may cause depurination upon heat denaturation. Storing the TBE/formamide loading buffer for extended periods of time results in the formation of trace formic acid. However, the dye/formamide mixture alone appears to be stable when frozen. Gel buffer (TBE) consisted of 89mM TRIS base, 89mM boric acid and 2mM Na_2EDTA, pH 8.3. The gels were either 0.4mm thick, 40cm long or wedge-shaped 0.2-0.6mm thick, 40cm long and were 5-8% polyacrylamide (1:20 crosslinked), 50% urea. The gels were pre-electrophoresed for 30 minutes

and then the samples were heat denatured at 95° for 2min, cooled to 0°, loaded onto the gel, and electrophoresed at 35mA (1400–1700V). For the ~400bp fragment from pMFB36, gels were typically run until the xylene cylanol tracking dye was 110mm from the bottom of the gel. The gels were dried using a BIO-RAD Model 483 Slab Gel Dryer. They were then exposed to Kodak X-Omat photographic film at -70° with a duPont Cronex Lighting Plus intensification screen for 8-12 hours. The gels were subsequently re-exposed at room temperature without an intensification screen. The time for the second exposure was determined by assuming the intensification screen decreased the exposure time by a factor of ten. The gel was scanned and the data were recorded as a plot of absorbance versus distance scanned, quantitated and converted into histogram form.

Double Strand Cleavage

DNA Preparation. pMS515 is a pBR322 based plasmid which contains a 150bp insert containing *hix*L cloned in at the *Cla* I site.[91] The plasmid was linearized separately with Pvu I and Ava I, phenol extracted (2x), ether extracted (3x) and ethanol precipitated (2x). The concentration of digested plasmid was adjusted by normalizing the UV adsorptions at 270nm.

Agarose Gels The DNA cleaving reactions were carried out as described for the high resolution gels except the reaction buffer was 40mM TRIS-acetate, 5mM sodium acetate, pH 7.9. The reactions were initiated by the addition of a DTT solution and carried out at 22°C for 2 hours. The final reaction volumes were typically 15λ and to this reaction mixture was added 2λ 10X ficol loading buffer. 10X Ficol loading buffer was made up as follows:

0.25% BPB	40 mg
0.25% XC	40 mg
25% ficol	3.75 gm
	to 15 mL

The samples were loaded onto a vertical 4mm thick, 1% agarose gel and electrophoresed versus 40mM TRIS-acetate, pH 7.9, 100μM Na_2EDTA at 120V until the bromophenol blue tracking dye reached the bottom of the gel. The DNA was visualized by staining with ethidium bromide (0.25μg/mL) followed by a 30 minute running water de-stain. The gel was then photographed with Polaroid type 55 positive/negative film under UV trans-illumination.

The molecular weight marker lane for the radiolabeled DNA was produced by individual digestions of pBR322 with Rsa I and Eco RI. The general digestion procedure was as described in the labelling procedures. After digestion, the reactions were ethanol precipitated and combined to produce the molecular weight marker lane. The molecular weights of the fragments in the final marker lane mixture are; 4363, 2118, 1565 and 680bp.

Negatives of the gel photographs were scanned with an LKB densitometer and quantitated with the accompanying data analysis software. The molecular weights of the DNA fragments from the cleavage reactions were determined by comparison of the mobility of the bands observed with the mobility of the molecular weight standard. To obtain greater accuracy in molecular weight determination, some cleavage reactions were spiked with the molecular weight standard. This allowed direct comparison of the mobilities of the reaction fragments with the molecular weight standard and avoided error introduced by lane to lane variations in electrophoretic mobility.

Because each double strand cleavage event produces two fragments of DNA which sum to the fragments original size, the cleavage of linearized plasmid DNA

should produce a symmetrical banding pattern. Therefore the bands resulting from cleavage at each site can be paired so that they sum to 4663 (the size of pMS515).

References and Notes

1. F. B. Bruist, S. J. Horvath, L. E. Hood, T. A. Steitz, *Science*, **235**, 777(1987).
2. D. L. Ollis, S. W. White, *Chem. Rev.*, **87**, 981 (1987).
3. C. Helene, *Nucleic Acids Res.*, **2**, 961 (1975).
4. C. Helene, G. Lancelot, *Prog. Biophys. Molec Biol.*, **39**, 1 (1982).
5. N. C. Seemen, J. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci., USA*, **73**, 804 (1976).
6. P. H. von Hippel, O. G. Berg, *Proc. Natl. Acad. Sci., USA*, **83**, 1608 (1986).
7. G. Church, J. Sussman, S. Kim, *Proc. Natl. Acad. Sci., USA*, **74**, 1458 (1977).
8. G. Zubay, P. Doty, *J. Molec. Biol.*, **1**, 1 (1959).
9. W. G. Hol, P. T. van Duijnen, H. J. Berendsen, , **273**, 443 (1978).
10. M. E. Hogan, R. H. Austin, *Nature*, **329**, 263 (1987).
11. G. D. Brayer, A. McPherson, *J. Mol. Biol.*, **169**, 565 (1983).
12. G. D. Brayer, A. McPherson, *Biochem.*, **23**, 340 (1984).
13. J. T. Finch, R. S. Brown, D. Rhodes, T. Richmond, B. Rushton, L. C. Lutter, A. Klug, *J. Mol. Biol.*, **145**, 757 (1981).
14. T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes, A. Klug, *Nature*, **311**, 532 (1984).
15. R. Burlingame, W. Love, B. Wang, R. Hamlin, N. Xuong, E. Moudranakis, *Science*, **228**, 546 (1985).
16. I. Tanaka, K. Appelt, J. Dijk, S. W. White, K. S. Wilson, *Nature*, **310**, 376 (1984).
17. D. Suck, C. Oefiner, W. Kabsch, *EMBO J.*, **3**, 2423 (1984).
18. P. Brick, D. L. Ollis, T. Steitz, *J. Mol. Biol.*, **166**, 453 (1983).
19. D. L. Ollis, P. Brick, R. Hamlin, N. G. Xuong, T. Steitz, *Nature*, **313**, 762 (1985).

20. D. H. Ohlendorf, W. F. Anderson, Y. Takeda, B. W. Matthews, *Nature*, **290**, 754 (1981).

21. B. W. Matthews, D. H. Ohlendorf, W. F. Anderson, R. G. Fisher, Y. Takeda, *Cold Spring Harbor Symp, Quant. Biol.*, **47**, 427 (1982).

22. B. W. Matthews, W. F. Anderson, R. G. Fisher, Y. Takeda, D. H. Ohlendorf, *Nature*, **298**, 718 (1982).

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

24. M. Lewis, A. Jeffrey, J. Wang, R. Ladner, M. Ptashne, C. O. Pabo, *Cold Spring Harbor Symp. Quantitative Biol.*, **47**, 435 (1982).

25. T. A. Steitz, I. T. Weber, J. B. Matthews, *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 419 (1982).

26. D. B. McKay, I. T. Weber, T. A. Steitz, *J. Bio. Chem.*, **257**, 9518 (1982).

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

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

29. T. T. Sauer, R. R. Yocum, R. F. Doolittle, M. Lewis, C. O. Pabo, *Nature*, **298**, 447 (1982).

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

31. A. Horchschild, N. Irwin, M. Ptashne, *Cell*, **32**, 319 (1983).

32. A. Joachimiak, R. Kelly, R. Gunsalus, C. Yanofsky, P. Sigler, *Proc. Natl. Acad. Sci., USA*, **80**, 668 (1983).

33. R. W. Schevitz, Z. Otwinowski, A. Joachimiak, C. L. Lawson, P. B. Sigler, *Nature*, **317**, 782 (1985).

34. J. E. Anderson, M. Ptashne, S.C. Harrison, *Nature*, **326**, 846 (1987).

35. J. E. Anderson, M. Ptashne, S.C. Harrison, *Nature*, **316**, 596 (1985).

36. J. A. McClarin, C. A. Frederick, B. C. Wang, P. Greene, H. Boyer, J. Grable, J. Rosenberg, *Science*, **234**, 1526 (1986).

37. C. A. Frederick, *et al.*, *Nature*, **309**, 327 (1984).

38. J. E. Anderson, M. Ptashne, S.C. Harrison, *Proc. Natl. Acad. Sci., USA*, **81**, 1307 (1984).

39. S. A. Abdel-Meguid, N. D. F. Grindley, N. Smyth Templeton, T. A. Steitz, *Proc. Natl. Acad. Sci., USA*, **81**, 2001 (1984).

40. R. Scheek, E. Zuiderweg, K. Klappe, J. van Bloom, R. Kaptein, H. Rüterjans, K. Beyreuther, *Biochem*, **22**, 228 (1983).

41. R. Boelens, R. M. Scheek, J. H. van Bloom, R. Kaptein, *J. Mol. Biol.*, **193**, 213 (1987).

42. J. Zeig, M. Silverman, M. Hilmen, M. Simon, *Science*, **196**, 170 (1977).

43. M. Bruist, R. Johnson, M. Glaccum, M. Simon, *Genome Rearangments*, 63 (1985).

44. R. C. Johnson, M. F. Bruist, M. B. Glaccum, M. I. Simon, *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 751 (1984).

45. M. Bruist, A. Glasglow, R. Johnson, M. Simon, *Genes and Development*, **1**, 762 (1987).

46. R. Johnson, A. Glasglow, M. Simon, *Nature*, **329**, 462 (1987).

47. M. I. Simon, J. Zieg, M. Silverman, G. Mandel, R. Doolittle, *Science*, **209**, 1370 (1980).

48. A. C. Glasglow, M. I. Simon, J. P. Sluka, P. B. Dervan, unpublished results.

49. M. W. Sundberg, C. F. Meares, D. A. Goodwin, C. I. Diamanti, *Nature*, **250**, 587 (1974).

50. M. W. Sundberg, C. F. Meares, D. A. Goodwin, *J. Med. Chem.*, **17**, 1304 (1974).

51. C. F. Meares, M. J. McCall, D. T. Reardan, D. A. Goodwin, C. I. Diamanti, M. McTigue, *Anal. Biochem.*, **142**, 68 (1984).

52. S. M. Yeh, D. G. Sherman, C. F. Meares, *Anal. Biochem.*, **100**, 152 (1979).

53. B. A. Khaw, J. T. Fallon, H. W Strauss, E. Haber, *Science*, **209**, 295 (1980).

54. D. A. Scheinberg, M. Strand, O. A. Gansow, *Science*, **215**, 1511 (1982).

55. R. M. Rainsbury, J. H. Westwood, R. C. Coombes, A. H. Neville, R. J. Ott, T. S. Kalirai, V. R. McCready, J.-C. Gazet, *Lancet*, **1**, 934 (1983).

56. G. Krejcarek, K. Tucker, *Biophys. Res. Commun.*, **77** , 581 (1977).

57. C. B. Chen, D. S. Sigman, *Science*, **237**, 1197 (1987).

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

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

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

61. J. S. Taylor, P. G. Schultz, P. B. Dervan, *Tetrahedron*, **40**, 457 (1984).

62. P. B. Dervan, R. S. Youngquist, J. P. Sluka, in *Stereochemistry of Organic and Bioorganic Transformations,* W. Bartman and K. B. Sharpless, Eds, VCH Verlagsgesellschaft mbH, Weinheim, West Germany, 1987, pp. 221–234.

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

64. J. H. Griffin, P. B. Dervan, *J. Am. Chem. Soc.*, **108**, 5008 (1986).

65. R. S. Youngquist, P. B. Dervan, *J. Am. Chem. Soc.*, **24**, 7564 (1987).

66. J. P. Sluka, P. B. Dervan in *New Synthetic Methodology and Functionally Interesting Compounds, Proceedings of the 3rd International Kyoto Conference on New Aspects of Organic Chemistry*, Z. -I. Yoshida, Ed, Elsevier, NY, vol. 25, 1986, pp. 307–322.

67. See also chapters 2 and 3.

68. G. B. Dreyer, P. B. Dervan, *Proc. Natl. Acad. Sci., USA*, **82**, 968 (1985).

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

70. J. P. Sluka, S. J. Horvath, M. F. Bruist, M. I. Simon, P. B. Dervan, *Science*, **238**, 1129 (1987).

71. R. Roeske, *J. Org. Chem.*, **28**, 1251 (1963).

72. J. Garnier, J. Osguthorpe, B. Robson, *J. Mol. Biol*, **120**, 97 (1978).

73. Ch. Zimmer, U. Waehnert, *Prog. Biophys. Molec. Biol.*, **47**, 31 (1986).

74. T. D Tullius, B. A. Dombrowski, *Proc. Natl. Acad. Sci., USA*, **83**, 5469 (1986).

75. pMS515 is a pBR322 based clone with a 150bp insert containing *hix*L cloned in at the Cla I site.

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

77. M. M. Garner, A. Revzin, *Nucleic Acids Res.*, **9**, 3047 (1981).

78. M. Freid, D. M. Crothers, *Nucleic Acids Res.*, **9**, 6505 (1981).

79. R. C. Johnson, M. I. Simon, *Cell*, **41**, 781 (1985).

80. R. Bushman, J. E. Anderson, S. C. Harrison, M. Ptashne, *Nature*, **316**, 651 (1985).

81. R. P. Wharton, M. Ptashne, *Nature*, **326**, 888 (1987).

82. T. Lund, K. Dahl, E. Mork, J. Holtlund, S. Lalannd, *Biochem. Biophys. Res. Comm.*, **146**, 725 (1987).

83. M. Solomon, F. Strauss, A. Varshavsky, *Proc. Natl. Acad. Sci., USA*, **83**, 1276 (1986).

84. P.G. Shultz, Ph.D. Thesis, California Institute of Technology, Pasadena, California, 1984.

85. T. Tanka, B. Weisblum, *J. Bacteriol.*, **121**, 354 (1979).

86. R. Hay, K. Nolan, *J. Chem. Soc., Dalton Trans.*, 1348 (1975).

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

88. This procedure for peptide synthesis is that of Steve B. Kent.

89. N. Ono, I. Hamamoto, A. Kamimura, A. Kaji, *J. Org. Chem.*, **51**, 3732 (1986).

90. J. M. Stewart, J. B. Young, "Solid Phase Peptide Synthesis", 2nd. ed., Pierce Chemical Co., Rockford, Illinois, 1984.

91. M. I. Simon, unpublished results.

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

CHAPTER II:

Distamycin–Actinomycin Hybrid Affinity Cleaving Molecules

One approach to the design of sequence specific DNA binding molecules that read large sequences of DNA, is to couple DNA binding units of similar or diverse base pair specificities. The natural product distamycin is a *tris*-N-methylpyrrolecarboxamide tripeptide that binds in the minor groove of DNA at A•T rich sequences that are five base pairs in size. The natural product actinomycin D, which consists of an aromatic phenoxazone coupled to two identical cyclic pentapeptide lactones is an intercalator which binds four base pairs with a preference for 5′–NGCN–3′ sequences. The synthesis of hybrids of distamycin and actinomycin, *bis*(distamycin)phenoxazone is described, where the cyclic pentapeptides of actinomycin are replaced by the tripeptide of distamycin. The synthetic hybrid DNA binding molecule should have the following capabilities; (1) bind by intercalation to a 5′-GC sequence, (2) bind in the minor groove to the adjacent A+T region on both sides of the central GC, and (3) cleave the DNA to facilitate identification of binding sites.

Before discussing the design and analysis of the hybrid molecules, we will examine small molecule/DNA interactions in general, as well as the two natural products.

Small Molecule–DNA Interactions

Minor Groove Binding: The interaction of a small molecule with the minor groove of DNA can be broken down into several types of interactions including:

1. Electrostatic interaction of charged molecules with the polyanion phosphate backbone.[1]
2. Hydrophobic interactions with the sugar portion of the backbone.[2]
3. Hydrogen bonding with the nitrogens and oxygens of the bases.

Of these three types of interactions, the third would be expected to be the most sensitive to the base sequence at the binding site. Figure 1 shows A•T and G•C base pairs as viewed down the duplex axis. Potential minor groove hydrogen bond acceptor sites for an A•T base pair are the adenine N(3) and thymine O(2). The G•C base pair is significantly different when viewed from the minor groove, primarily due to the amino function at guanine C(2). The cytosine O(2) and guanine N(3) are still hydrogen bond acceptors, but the minor groove is partially blocked by the amino group at guanine C(2), which is potentially both a donor and an acceptor site. Interestingly, if the guanine amino function is removed, i.e. guanine is substituted by inosine, the I•C base pair is qualitatively the same as the A•T base pair when viewed from the minor groove. In synthetic DNA's, the substitution of inosine for guanine gives a structure which behaves more like an A•T than a G•C base pair with respect to resistance to forming A-form DNA.[3]

For a small molecule to hydrogen bond to the bases in the minor groove, it must first displace the spine of hydration, particularly in A+T regions. The difference in hydrogen bond enthalpies between the water•DNA complex and small

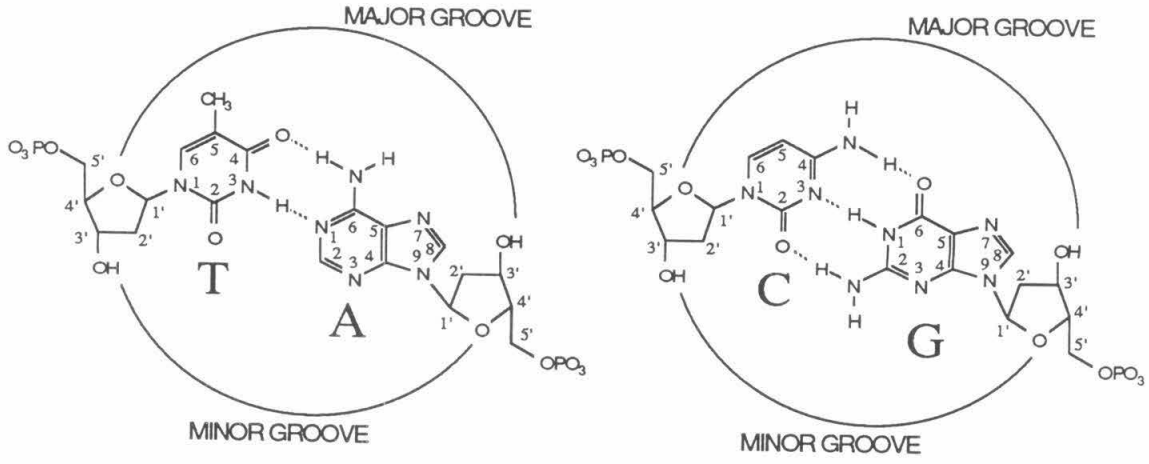


Figure 1: A•T and G•C base pairs.

molecule•DNA complex ($\Delta\Delta H$) may be quite small, even positive. However, if several waters are displaced by a single small molecule, then the entropic contribution to $\Delta\Delta G$, hence the binding constant, could be substantial.

Intercalative Binding: There are many small molecules containing planar aromatic ring systems which can be inserted between adjacent base pairs of duplex DNA in a process called intercalation.[4] The classical intercalation model, based on model building studies, involved extending the B-form double helix with simple bond rotations to create a space for insertion of a planar aromatic molecule of approximately 3.4Å thickness.[5] This intercalation model has the following features:

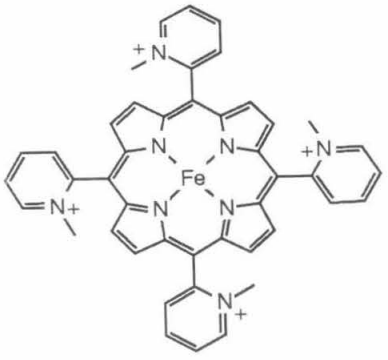
1. The rod-like structure, hydrogen bonding and covalent bonding of B-form DNA is retained.
2. Separation between two adjacent base pairs is approximately doubled, from 3.4Å to ~6.8Å, creating space for intercalation of an aromatic ring.
3. An unwinding of the double helix is required, (unwinding was estimated at 36°)[5] leaving adjacent base pairs positioned more directly over each other.

4. The intercalator is inserted into the opening parallel to the adjacent bases in such a way as to fill the space as fully as possible.

Extensive studies aimed at verifying the preceding model have been carried out in the last 20 years. In particular, ambiguities in the determination of unwinding angles based on viscosity and sedimentation coefficients have been largely eliminated with the development of methods employing superhelical closed circular (plasmid) DNA's.[6,7] Unwinding angles (per intercalation) determined on plasmid DNA's have ranged from a high of 26° for ethidium and related phenanthridines,[8] to a low of 10° for anthracyclines.[9]

Evidence has also been found for a nonclassical intercalation event.[4] Asymmetric distribution of bulky substituents around an intercalator, as well as a small or bent aromatic system, can cause asymmetric unwinding of the two strands. If the unwinding angle of the two strands is not equal, the resulting complex has a bend in the helix.[10,11]

An important part of the driving force for intercalation appears to be transfer of the hydrophobic intercalator from the aqueous medium to the hydrophobic intercalation cavity. Although the kinetics of intercalation are slowed down by bulky substituents on both sides of the aromatic nucleus, binding constants are frequently not affected.[12] For example, tetra-meso-pyridylporphyrins are known to intercalate readily, even though the pyridine rings seem far to bulky to pass through the intercalation cavity. Apparently, the DNA duplex denatures, i.e. the strands separate, and the bulky substituents are passed through the denatured region. The duplex is then reformed around the intercalated porphyrin nucleus.[13]



Iron(II) meso-tetra-2-(N-methyl)pyridylporphyrin

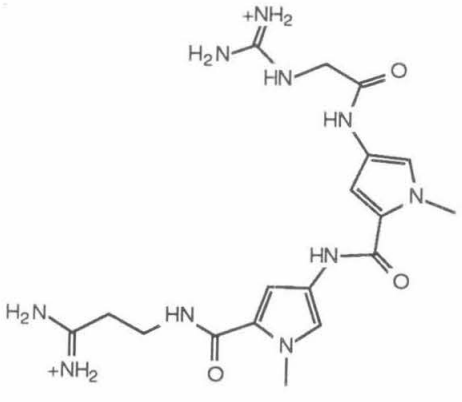
The thermodynamics of intercalation are generally dependent on the ionic strength of the solution. For charged intercalators, it appears that much of the observed free energy of intercalation at low ionic strength, calculated from observed binding constants, is due primarily to cation release from the phosphate backbone, and not to any especially favorable interaction between the intercalator and DNA.[4]

Most mono-intercalators studied to date have exhibited from slight to quite pronounced specificity for G•C base pairs.[4] Sequence specificity appears to be due to stacking interactions and/or interaction of pendant groups, i.e. groups not part of the aromatic system, in the major or minor grooves adjacent to the intercalation site.

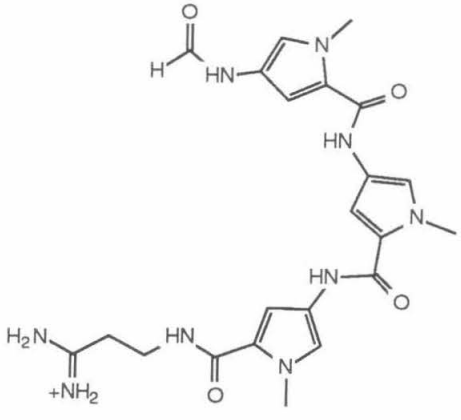
Distamycin A

Distamycin A (**DST**) is a naturally occurring basic oligopeptide antibiotic isolated from *Strepomyces Distallicus.*[14] It shows antifungal, antimitotic and anti(DNA)viral activities,[15] and is used clinically as a topical ointment in the treatment of chickenpox, herpes zoster and eruptions resulting from smallpox vaccinations.

Extensive studies have been carried out on **DST**, including total synthesis[16] and numerous homologs.[17] An x-ray crystal structure of the related dipyrrole netropsin (**NET**) (as the sulfate salt) indicates that the majority of the molecule is essentially flat, with a pronounced curvature in the pyrrole/peptide backbone. The amide nitrogens are on the concave side and carbonyl and methyl groups on the convex side.[18]



NET



DST

Numerous studies of the **DST**•DNA interaction have appeared. The following is a brief synopsis of some of the relevant data. Equilibrium binding constants, obtained from sedimentation data, indicate a binding constant for preferred sites of $2.4 \times 10^9 M^{-1}$ (1 **DST**:8 base pairs) as well as $\sim 10^6 M^{-1}$ (1:3) for a weaker binding process.[19] **DST** binding to double stranded DNA increases the thermal denaturation temperature of the DNA[19] and causes a bathocromatic shift in the **DST** adsorption spectra,[15] as well as cotton effects in the circular dichroism spectra.[20] Binding site size, determined by a number of techniques, shows **DST** to bind to a six base pair length of DNA for preferred sites.[21]

Binding is non-intercalative[22,23] and methylation (minor and major groove) and glycosylation (major groove) studies indicate binding in the minor groove.[24,25]

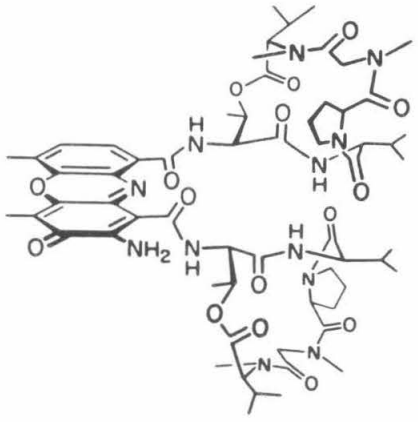
Sequence specificity appears to be for A+T rich regions of DNA,[24] with a minimal number of A/T switches, 5′-AAAAAA > 5′-ATATAT).[26] Structural studies indicate that the basic N-methylpyrrole oligopeptide antibiotics bind in the minor groove of A+T rich regions of DNA, with both specific and non-specific interactions. The amide hydrogens are involved in hydrogen bonding with the O(2) of thymine and/or N(3) of adenine, and contribute to the sequence specific binding component.[17] Electrostatic interaction of the positively charged end group(s) of the antibiotic and the polyanion backbone of the DNA contribute to the non-specific interaction. There also appears to be a significant entropic (from release of bound water) and hydrophobic contribution to the free energy of binding. Conformational changes in the DNA caused by G•C base pairs, and/or steric interactions with N(2) of guanine, substantialy reduce the affinity for G+C rich regions of DNA, though I+C rich synthetic DNA's have high affinity for **DST**.[27]

NMR studies of an oligonucleotide–netropsin complex confirm the groove and sequence specificity, hydrogen bonding contacts and increased resistance to thermal denaturation required by the preceding model for N-methylpyrrole oligopeptide binding to DNA.[28,29] Additionally, the x-ray crystal structure of a complex between netropsin and a 14bp oligonucleotide further confirms the groove, hydrogen bonding and hydrophobic interactions.[30,31]

Footprinting, with both DNase I and MPE-Fe,[32,33] and affinity cleaving[34–36] methodologies have been applied to **DST** binding to heterogenous DNA. Both studies confirm the A+T sequence specificity, as well as giving a more direct determination of the binding site size, which is nominally five base pairs.[32,33]

Actinomycin D

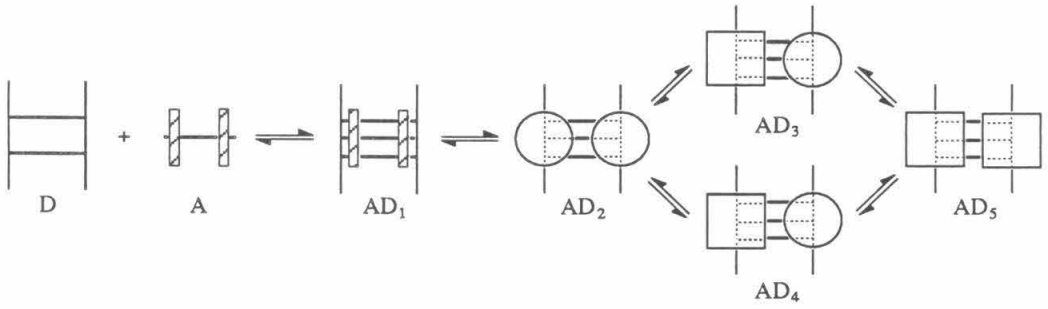
Actinomycin D (**ACT**)($D=C_1=X_1=I_1$) is a red, crystalline antibiotic isolated from the culture of *Streptomyces*. It consists of a planar aromatic system (phenoxazone) coupled to two identical cyclic pentapeptide lactones. Numerous total syntheses of **ACT** have been reported[37-40] as well as many derivatives.[41,42]



Actinomycin D

ACT exhibits some antibacterial activity, especially against Gram (+) bacteria such as *B. Subtilis*, and has also been shown to inhibit replication of double stranded DNA viruses, although RNA viruses are unaffected.[43] Studies on the interaction of **ACT** with DNA have been extensive.[41,42] Spectrophotometric[2] and equilibrium dialysis[44] studies of **ACT** binding to both natural and synthetic DNA's have indicated an apparent binding constant of $0.6–3.3 \times 10^6 M^{-1}$ and a binding site size of 4.5–6 base pairs.[2] In a classic study, Muller and Crothers determined the kinetics of **ACT** binding to heterogenous DNA.[2] The data indicated at least five different DNA•**ACT** complexes present during the course of the reaction and at equilibrium. Based on the observed kinetics and the effect of chemical modification of actinomycin C_3 (C_3 differs from D in that both D-valine residues in the

cyclic peptides are replaced by D-alloisoleucine)[45], the binding mechanism shown in scheme 1 was proposed. The first two steps are very fast ($k_1 \approx 1.5 \times 10^4 Msites^{-1}$, $k_2 \approx 5 sec^{-1}$), the third and forth slower ($k_3 \approx k_4 \approx 0.2 sec^{-1}$) and the final two very slow ($k_5 \approx k_6 \approx 0.002 sec^{-1}$). The slowness of the last two steps was attributed to the rupture of an internal hydrogen bond in each of the peptide rings. The kinetic data allows the determination of an intrinsic binding constant for actinomycin C_3 of $3.3 \times 10^5 M^{-1}$.



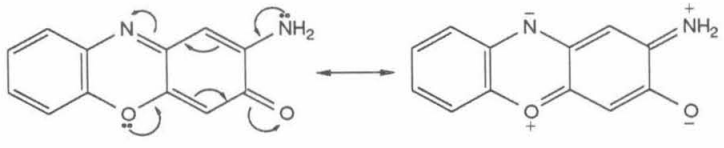
Scheme 1: Schematic drawing for the proposed mechanism for the reaction of **ACT** with DNA.[2] The reaction begins at the left with DNA (D) and actimomycin (A); the DNA base pairs are represented by horizontal lines, as is the actinomycin chromophore. The actinomycin peptide rings are viewed from the edge; it is supposed that their faces interact with each other in solution. In complex AD_1, the chromophore has been inserted between the base pairs, and in complex AD_2 the peptide rings reorient themselves to interact more favorably but non-specificly with the DNA backbones; the rings are viewed from their faces. In complex AD_3 and AD_4, one or the other of the peptide rings undergoes a conformational change which adapts it to interact specifically with the DNA, and in form AD_5 both rings have been so rearranged. At equilibrium, only AD_3, AD_4 and AD_5 are present in significant amounts, with about half the complex present as AD_5.

X-ray crystal structures for 1:2 complexes of **ACT** with d(GpC)[46] and dG[47,48] show an intercalative like structure and specific hydrogen bonds from the guanine 2-amino group to the carbonyl oxygen of the L-threonine residue, as well as two

hydrogen bonds between the D-valine residues of the peptide rings. Further evidence for an intercalative mode of binding comes from changes in superhelicity of ϕX174 RF DNA on binding **ACT**, giving an unwinding angle of 26° (analytical ultracentrifugation).[22] As an intercalator, **ACT** is unusual in several respects:[4]

1. It is one of the few known intercalators which is uncharged.[2]
2. Its binding constant is nearly ionic strength independent.
3. It exhibits unusually slow kinetics for an intercalator.
4. ΔH of binding to DNA is positive, +2kcal/mol.[49,50]
5. It has a large negative temperature coefficient of solubility in water (128 mg/ml at 1°C, 0.8mg/ml at 20°C)[33] which may be related to the large entropic contribution to binding.[51]
6. Based on light scattering experiments[52] and a 2:1 co-crystal structure with d(GpC),[46] **ACT** may also crosslink DNA (non-covalently).

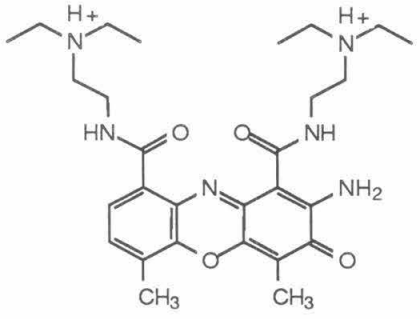
Sequence preference is for nucleotides with 2-amino functions such as guanine or diaminpurine though there are exceptions.[48] The alternation of purine and pyrimidine also appears important.[44] **ACT** affinity for repeating homopolymers has been determined; 5′-GC>GGG>GT,GTT>GA,GAA,GTA>GAT-3′.[44] Stacking interactions also appear to be important in determining **ACT**'s sequence specificity. The dipole moments of a G•C base pair are expected to interact favorably with the charge distribution in the phenoxazone ring system:[2,53]



NMR has provided additional data on preferred stacking interactions. Studies with dinucleotides suggest that the quinoid portion of the phenoxazone ring system prefers to stack with guanine, while the benzoid portion binds adenine and

guanine equally well.[54,55] It has been suggested that stacking interactions are more responsible for sequence specificity than is hydrogen bonding between the peptide and guanine 2-amino function.[54,55]

As noted earlier, **ACT** is unique in that ΔH of binding is +2kcal/mol. This implies a large contribution from entropy effects resulting from release of partially bound water molecules associated with the peptide rings and/or DNA.[55] The ΔH of binding for actinomine at low binding densities is ~-7 kcal/mol ($K_{app} \approx 10^6 M^{-1}$).[2]



Actinomine

It is unclear however, how much of the difference between **ACT** and actinomine is due to the removal of the peptides from **ACT** and how much is due to the addition of two positive charges in actinomine. Unlike **ACT**, actinomine binding is highly dependent on the ionic strength of the solution.[4]

ACT–DNA interactions have also been studied on heterogenous DNA using MPE-Fe and DNase I footprinting methods.[32,33] The footprinting results agree well with the homopolymer binding studies, and further indicate a preference for 5′-GC-3′ flanked by A•T base pairs as an optimal site.[56] Figure 2 shows footprinting results for both **ACT** and **DST**.[32,57–59]

Distamycin A / MPE•Fe

Actinomycin D / MPE•Fe

Figure 2: (Top) Binding sites of **DST** and (bottom) **ACT** on the 517bp restriction fragment of pBR322 from footprinting data.

Simultaneous DST and ACT Binding

Given the radically different binding modes of **ACT** and **DST** (intercalation versus minor groove binding) the question still remains whether the DNA helix can accommodate both of these ligands within close proximity, despite their unique distortions to the DNA structure. Fortunately, there is sound literature evidence indicating that proximal binding of the two ligands is possible.

Spectral titration, circular dichroism and methylation studies on the effect of **ACT** binding to **DST**–DNA complexes indicate that binding of **ACT** does not give rise to appreciable dissociation of **DST** at normal binding densities.[25] A reasonable, but not theoretically rigorous, treatment of spectral titration data for the simultaneous binding of **ACT** and **NET** indicates that adjacent binding is possible and that each ligand's total neighbor exclusion is consistent with its van der Waals radius.[24] This implies that longer range effects are small enough to not significantly affect the adjacent binding of another ligand, even if the ligand has a different binding mode.

An NMR study of simultaneous **NET** and **ACT** binding to a synthetic dodecamer, 5′-CGCGAATTCGCG-3′, provides further evidence for adjacent binding on a defined sequence.[60] Analyses of changes in chemical shift and half-life of the guanine H(1) and thymine H(3) imino resonances, as well as of backbone phosphorous shifts, indicate that two 40and a **NET** can simultaneously bind to the dodecamer as shown in figure 3.

Of further interest is the observation that the binding of two 40destablizes the A+T core of the dodecamer (lowers melting temperature and decreases imino proton half lives from 56 to ~28msec). **NET** binding, however, stabilizes the same

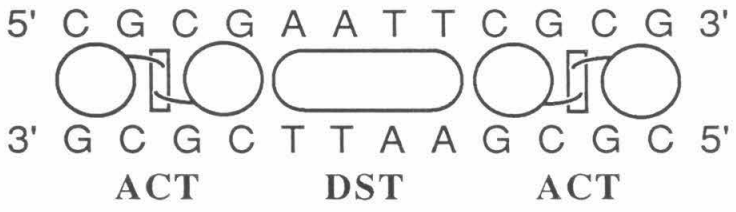


Figure 3: Simultaneous binding of two **ACT**'s and one **DST** to a 14bp oligonucleotide.

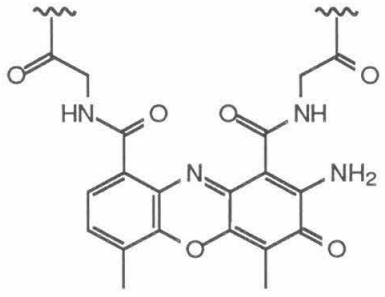
region (imino half live increases from 56 to 165msec) and causes small chemical shift differences in the adjacent G+C regions. Simultaneous binding of both antibiotics (two 40and one **NET**) gives results which are very nearly equal to the sum of the independent effects. For example dA•dT core imino proton half lives increase from 56 to 130msec ($\Delta t_{1/2}$=+74) as compared to $\Delta t_{1/2}$'s of +109 msec for just **NET** or -28msec for two **ACT**'s.

The preceding discussion indicates that intercalation of **ACT** and minor groove binding of **DST**in close proximity are indeed possible even though the two binding modes exert different constraints on the duplex DNA structure. The final remaining question then is, whether an intercalator and a minor groove binder can be covalently linked in such a way that the sequence specificity and the ability to simultaneously bind both moeities is retained. During the course of the work described here, Krivtsova and coworkers have reported the synthesis of a series of hybrid molecules, called distactins, which have one, two and three N-methylpyrrolecarboxamides *directly* attached to the 1,9 positions of the phenoxazone chromophore.[61] From spectrophotometry, viscometry and flow birefringence studies they concluded that the distactin containing three N-methylpyrrole rings is bound to DNA only on the outside of the double helix through the peptide fragments, and that intercalation of the phenoxazone does not take place. Distactins

with one or two N-methylpyrroles appear to have both intercalation and groove binding modes.

BEDP: Design and Synthesis[62]

Based on the preceding discussion of **ACT** binding and the apparent sources of G•C specificity two important points emerge: stacking interaction of adjacent G•C base pairs with the phenoxazone ring system contributes to specificity, as does the hydrogen bond between the guanine 2-amino and the carbonyl oxygen of the peptide L-threonine residue. The smallest portion of **ACT** which may still retain these capabilities appears to be:



The phenoxazone ring system is completely retained. The threonine residue is replaced with glycine to simplify the structure while retaining the amide carbonyl function with a similar spatial arrangement. The rest of the cyclic peptide lactones are removed. The remaining question then, is how to couple the intercalator with the affinity cleaving molecule EDTA–distamycin (**ED**) without affecting both their abilities to bind to DNA.

A combination of model building and judicous use of Occam's Razor[60] suggest that a reasonable first generation target structure involves direct coupling

of the intercalator (shown above) and groove binding moieties giving *bis*(EDTA–distamycin–glycyl)phenoxazone (**BEDP**). For comparison, the iron chelates of EDTA–distamycin[35] and *bis*(EDTA–distamycin)fumaramide[63] (**BEDF**) will also be included (see figure 4).

It is hoped that the hybrid of **ACT** and **DST** will bind to a DNA sequence such as 5′–$(A•T)_4(G•C)_2(A•T)_4$ as shown in the binding model in figure 5.

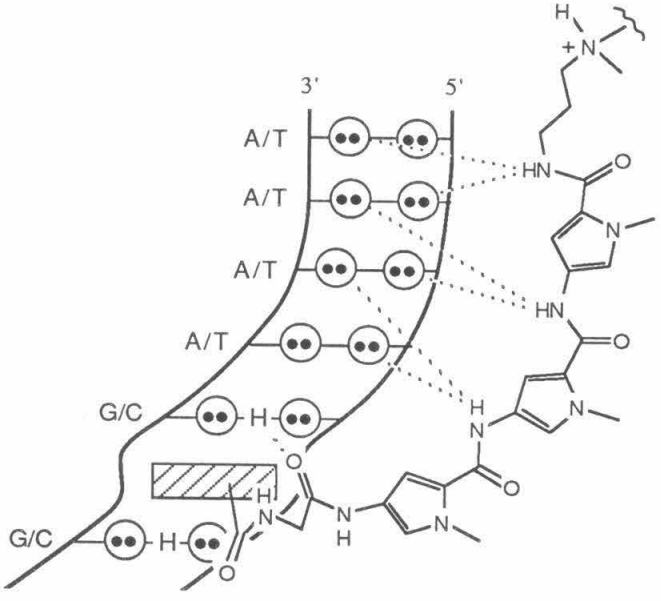


Figure 5: Stylized drawing of the DNA binding model of a hybrid **DST**–intercalator molecule showing the targeted hydrogen bonding pattern.

Synthesis

The synthetic route to *bis*(EDTA–distamycin–glycyl)phenoxazone (**BEDP**) is shown in Scheme 2. Compound **1** was prepared in eight steps (27% overall) from N-methylpyrrole-2-carboxylic acid using a modification of a published procedure.[16] N,N′-Carbonyldiimidazole mediated coupling of **1** with mono-*t*BOC

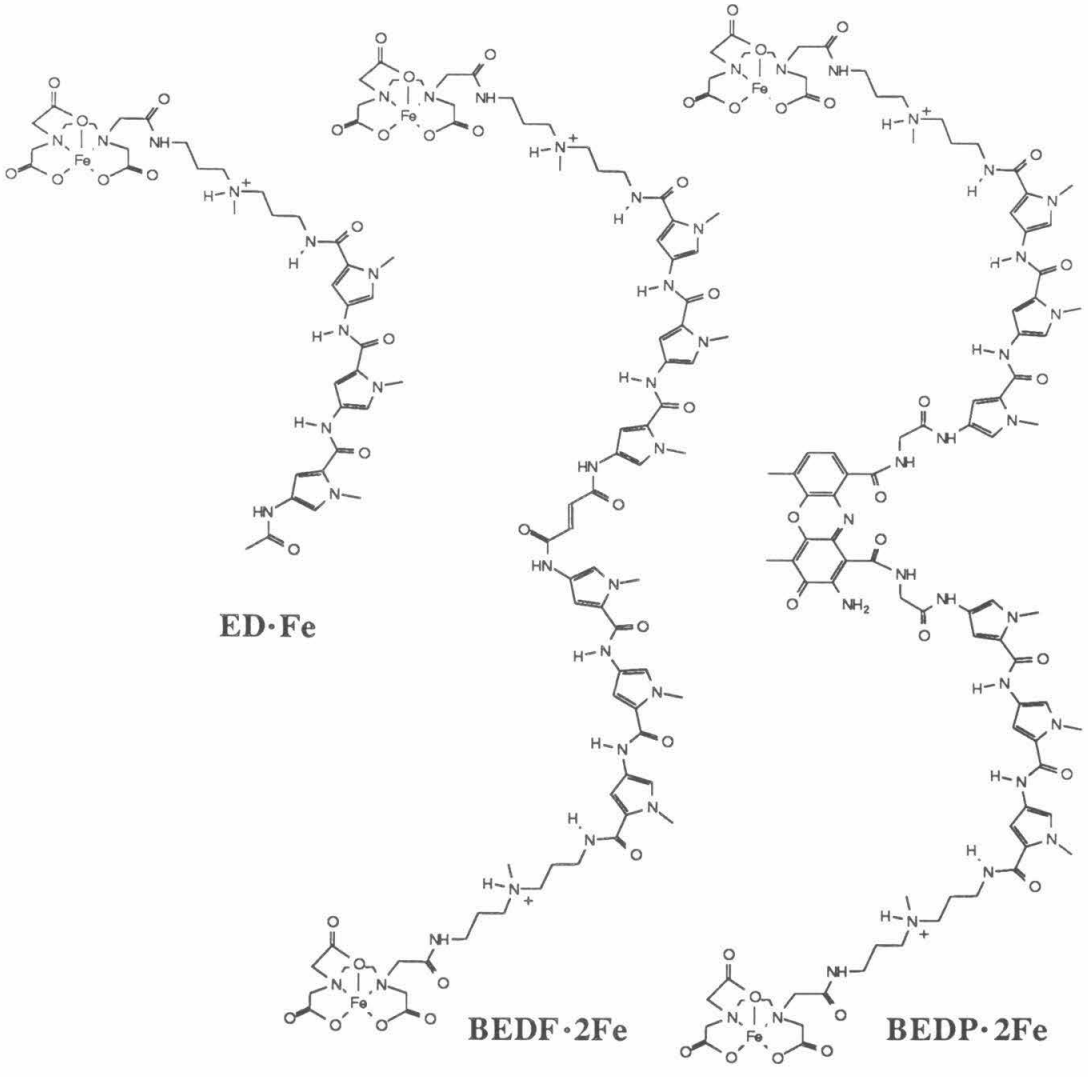


Figure 4: The iron chelates of EDTA-distamycin (ED), bis(EDTA-distamycin)-fumaramide (BEDF) and bis(EDTA-distamycin-glycyl)phenoxazone (BEDP).

protected 3,3′-diamino-N-methyldipropylamine (**7**) afforded **2** in 79% yield. The protected aromatic precursor to the phenoxazone ring system (**4b**) was prepared in three steps from the commercially available 2-nitro-3-hydroxy-4-methylbenzoic acid in 83% overall yield. Catalytic hydrogenation of **2** and subsequent condensation with the carbonyldiimidazole adduct of **4b** gave **5a** in 48% yield. Deprotection of **5a** to afford **5b** (93%) and condensation with excess carbonyldiimidazole activated EDTA-triethyl ester afforded **6a** in 95% yield. Saponification of **6a**, catalytic hydrogenation of **6b** and subsequent oxidative coupling with potassium ferricyanide in 1:1 methanol/phosphate buffer (pH 7.1) followed by chromatography on Amberlite XAD–2 gave the red-orange hexazwitterion of **BEDP** in 60% yield.

Affinity Cleaving Assays

The DNA cleavage efficiency of **BEDP**•Fe(II) was investigated by monitoring the conversion of supercoiled pBR322[64] plasmid DNA (form I) to open circular (form II) and to linear form (form III) (Table 1). One single strand scission converts form I to form II. In the presence of O_2 and dithiothreitol (DTT), 0.1μM concentration of **BEDP**•2Fe(II) converts form I to 82% form II and 18% form III.

The sequence and site size of the DNA binding sites were examined by analysis of DNA cleavage on ^{32}P end-labelled DNA restriction fragments by denaturing polyacrylamide gel electrophoresis as shown in figure 6. A 517 base pair Rsa I–Eco RI restriction fragment (base pairs 3848 to 4362) from plasmid pBR322 DNA was labelled separately with ^{32}P on the 5′ and 3′ ends. The DNA fragments were allowed to react with **BEDP**•2Fe(II) (5μM concentration) at **BEDP**/DNA base pair ratio of 0.05 in the presence of DTT (5mM) for two hours (37°C, pH 7.9). The DNA cleavage sites were visualized by high resolution gel electrophoresis (figure 7).

Scheme 2: Synthetic route to **BEDP**.

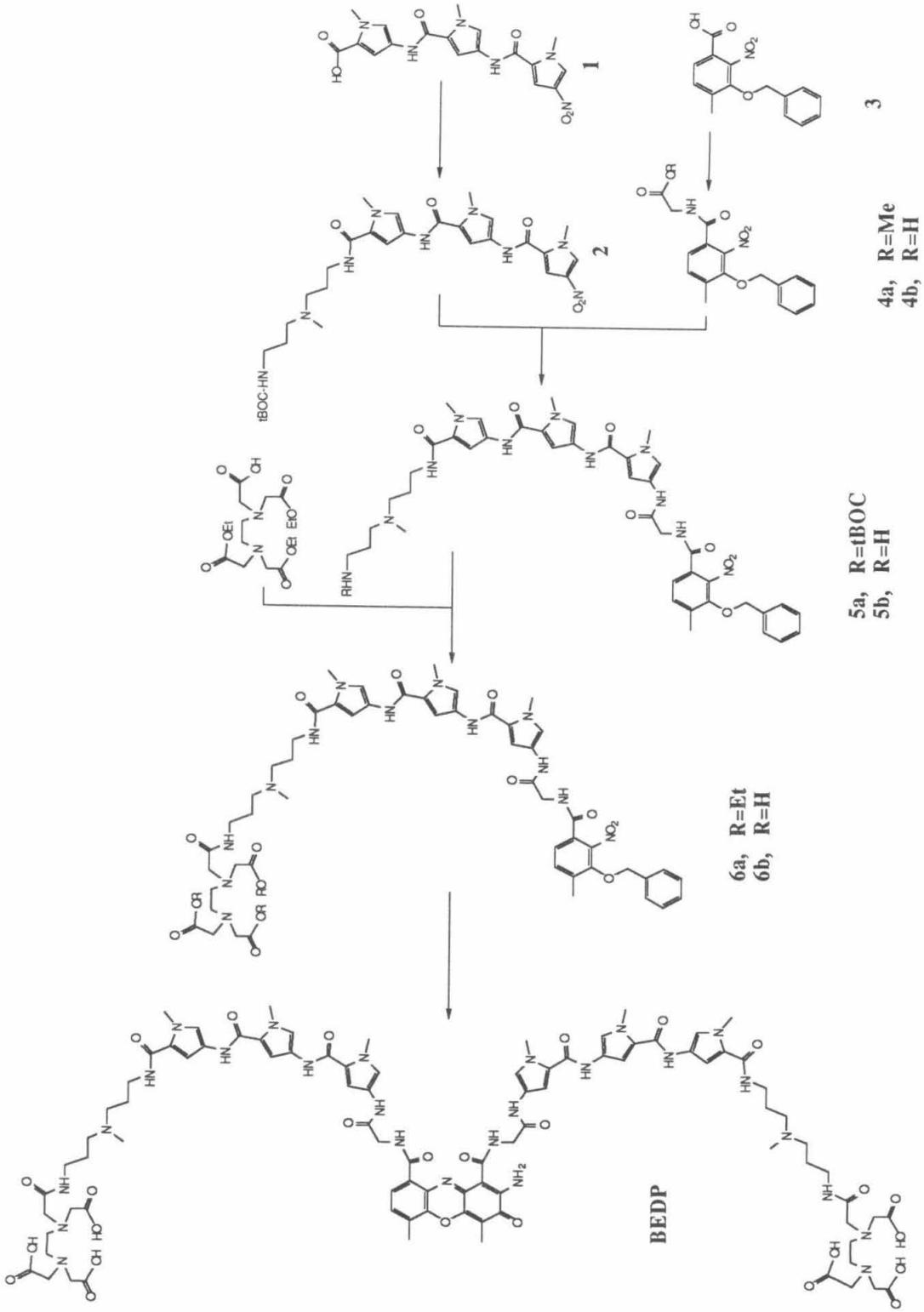


Table 1: Cleavage[a] of pBR322 by **BEDP**•2Fe(II).

BEDP•2Fe(II) (μM)	% I[b]	% II	% III	% low mw linear DNA
1.0	0	15	28	58
0.1	0	82	18	0
0.01	58	39	3	0
0.001	86	14	0	0

a. 10μM bp, buffer (40mM TRIS, 5mM NaOAc, pH 7.9), 5mM DTT. Reactions were run 1hr at 25°C, **BEDP**•2Fe(II) annealed with pBR322 5min at 65°C prior to addition of DTT.

b. Corrected for decreased stainability of form I DNA.

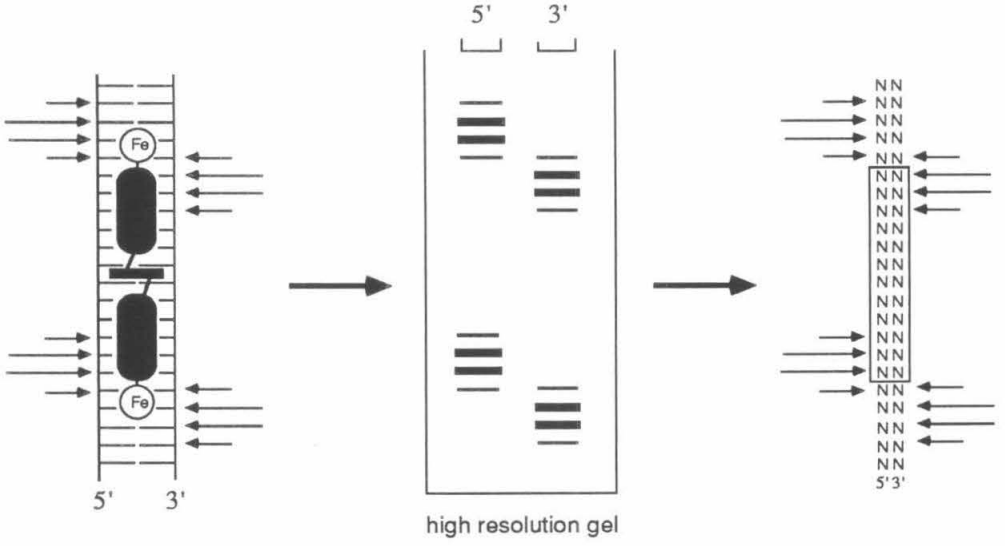


Figure 6: Scheme for the affinity cleaving method. (Left) A molecule equipped with two EDTA functions gives two cleavage loci which are asymmetric on opposite strands of a ^{32}P end-labelled restriction fragment. Lengths of arrows indicate frequencey of cleavage. (Center) Cleavage patterns visualized on the autoradiogram of a high resolution denaturing polyacrylamide gel. (Right) Assignment of binding site boundaries based on the model of Schultz *et al.*[35,36]

From densitometric analysis of the autoradiogram (lower half of figure 7), the DNA cleavage patterns reveal a major cleavage site flanking the ten base pair sequence, 5′-TATAGGTTAA-3′ (figure 8). One interpretation of the data suggests that the

tripyrroles are binding simultaneously at A•T rich sequences, four base pairs in size, flanking a central 5′-GG-3′ phenoxazone binding site. Although not proven, this is consistent with the groove binder–intercalator–groove binder mode.

However, at two other sites (middle of the autoradiogram) single cleavage loci are observed. One interpretation of this data is that only one tripyrrole or possibly a tripyrrole–phenoxazone (groove binding–intercalation) is binding at these sites. The absence of two cleavage loci prohibits the assignment of binding location and site size.

From the histograms and binding locations illustrated in figure 8 we see that the tripyrrole EDTA-distamycin (**ED**) binds the five base pair site, 5′-TTTTT-3′. The octaamide, *bis*(EDTA-distamycin)fumaramide (**BEDF**), binds the *same* location but of larger site size, 5′-ATTTTTATA-3′. In contrast, *bis*(EDTA-distamycin–glycyl)phenoxazone (**BEDP**) *shifts* the binding location to a sequence that is ten base pairs in size, 5′-TATAGGTTAA-3′. Although we present no data here on whether the phenoxazone is intercalating, the apparent binding site size of ten base pairs and the binding sequence of four A•T flanking *both* sides of a GG sequence suggest that this may be a site where the hybrid molecule is binding in a simultaneous (though not necessarily equal) groove binding–intercalation–groove binding mode. This is in contrast to the corresponding distactin studied by Krivtsova *et al.*[61] The molecules have in common a *tris*-N-methylpyrrole covalently tethered to the 1,9 positions of the phenoxazone. However, the distactins[61] are *directly* coupled whereas **BEDP** described here has a *glycine tether* which acts as a hinge and preserves the postulated critical carbonyl recognition functional group.

Because there are 524,800 unique sequences of double helical DNA that are ten base pairs in size, it is likely that we have not yet identified the optimal 10 base pair

Figure 7: Autoradiogram of 3′ (lanes 1–5) and 5′ (lanes 6–10) ^{32}P end-labelled 517bp DNA restriction fragment on a high resolution denaturing wedge gel. Lanes 1 and 6, intact DNA; lanes 2 and 7, Maxam-Gilbert G reaction; lanes 3 and 8, **ED**•Fe(II), 5μM; lanes 4 and 9, **BEDF**•2Fe(II), 1μM; lanes 5 and 10, **BEDP**•2Fe(II), 5μM. Bottom to the middle of the gel is the sequence left to right in figure 8.


3′
5′
control
G
5μM ED
1μM BEDF
5μM BEDP
control
G
5μM ED
1μM BEDF
5μM BEDP
1
2
3
4
5
6
7
8
9
10

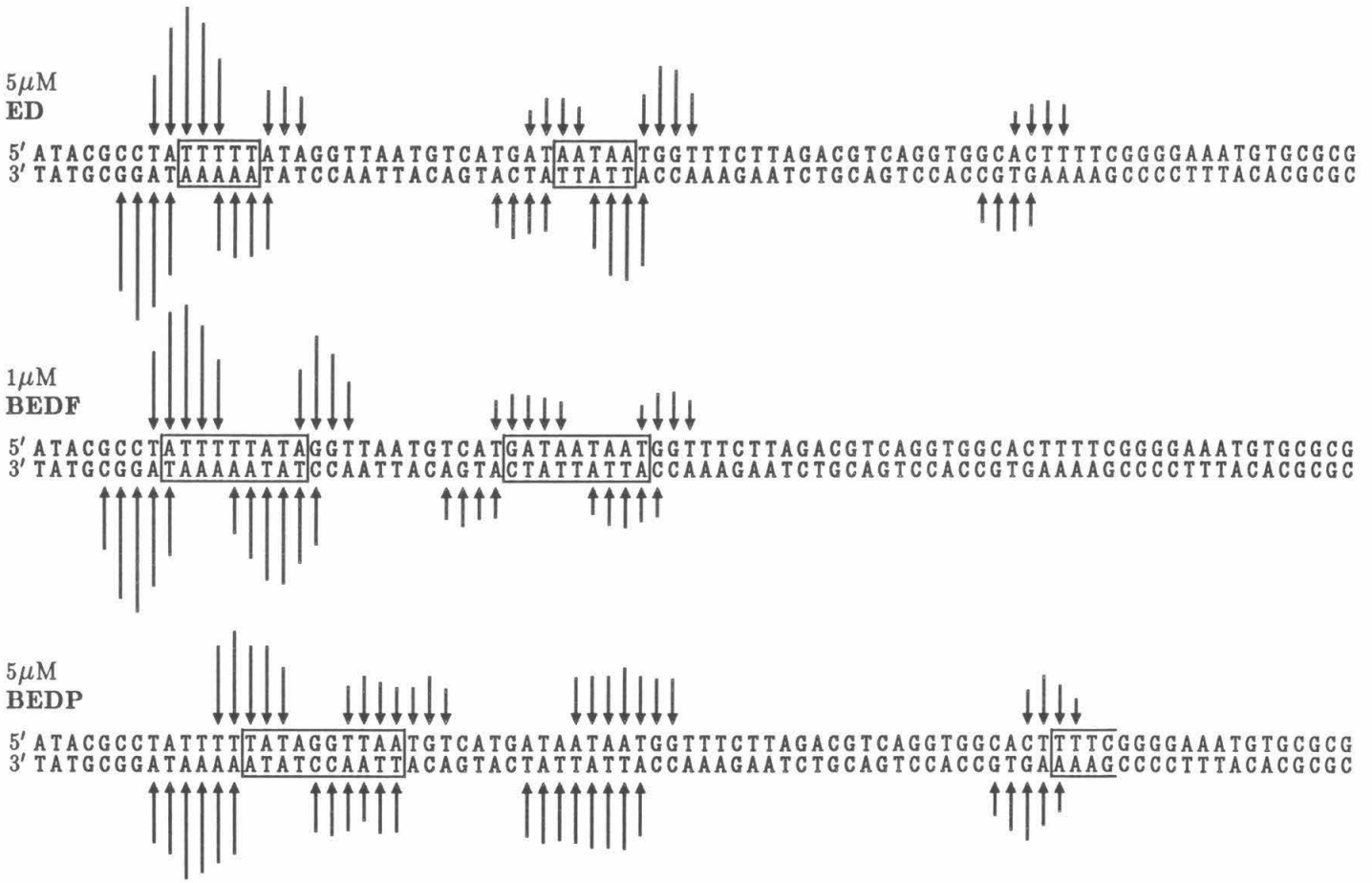


Figure 8: Histograms of the DNA cleavage on the 517bp (base pairs 3848 to 4362) restriction fragment from pBR322 from figure 7. The height of the arrows indicate extent of cleavage resulting in removal of the indicated base. Boxes define binding site location and size as described in figure 6.

recogntion site for *bis*(distamycin–glycyl)phenoxazone. The data presented here only suggests that 10 base pair recognition may be possible with this molecule. Indeed, this is not an obvious result when one considers that actinomycin, which most likely unwinds flanking sequences (by intercalation), and distamycin, which most likely binds B form DNA, could present a potentially incompatible situation.

We also see in figure 8 that the single cleavage loci at the other sites suggest that simultaneous groove binding of the tripyrroles is *not* occurring at these sites. Perhaps, at these sites, there is intercalation of the phenoxazone with sequence dependent distortions on both sides of the intercalation site. This leaves only one site compatible with tripyrrole binding.

Second Generation Hybrid

The results with **BEDP** suggest that though simultaneous groove binding and intercalation occur at some sites, at other sites this is not the case. Furthermore, the observation by Kristova *et al.*, that the directly coupled tripyrrole–phenoxazone only binds the tripyrrole units and not the phenoxazone suggests that the nature of the *linker* between the phenoxazone and tripryrrole unit is of critical importance. Re-analysis of the binding model in figure 5 indicates that in essence, **BEDP** uses a *pyrrole* unit as a linker function to connect the glycyl–phenoxazone to a dipyrrole unit. The glycyl unit is not a linker since its carbonyl oxygen was designed to interact, via a specific hydrogen bond, with the guanine 2-amino group. If the glycyl carbonyl is directed *into* the DNA minor groove, the amide hydrogen of the amino terminus tripyrrole unit must be directed *out* of the minor groove. This orientation of the amide unit is reversed relative to the normal **DST** binding conformation and is problematic in two ways. First, the carbonyl recognition element

is shifted by one backbone bond relative to the normal pyrrole carboxamide NH recognition element. Secondly, the carboxamide orientation is rotated 180° relative to the orientation observed in both the **NET** and **NET**•DNA crystal structures. This rotated carboxamide is presumably a higher energy conformation. These two problems indicate that **BEDP** must assume a higher energy conformation in order to adopt the targeted binding conformation. Furthermore, if the glycyl carbonyl does assume the proper configuration, then the amino–terminal pyrrole unit must be forced out of register compared to the normal **DST**-like binding. Since a pyrrole is a rigid unit, it would be difficult for the two remaining pyrroles to get back into proper register with the DNA minor groove.

The preceding discussion suggests that the amino terminus pyrrole is acting as an excessively rigid linker function, and, additionally, is biasing the glycyl carbonyl towards an incorrect conformation. Therefore, we have redesigned the hybrid molecule in an attempt to remove these difficulties. Replacement of the amino terminus pyrrole unit with a γ-aminobutyric acid (GABA) linker should minimize these problems. Once the pyrrole unit is removed, the incorrect biasing of the glycyl carbonyl should no longer be a problem. Additionally, the increased flexibility of GABA should facilitate the proper registering of the two remaining pyrroles within the minor groove. The second generation hybrid is formally related to the dipyrrole **NET**; *bis*(EDTA–netropsin–glycyl–GABA)phenoxazone (**BENP**).

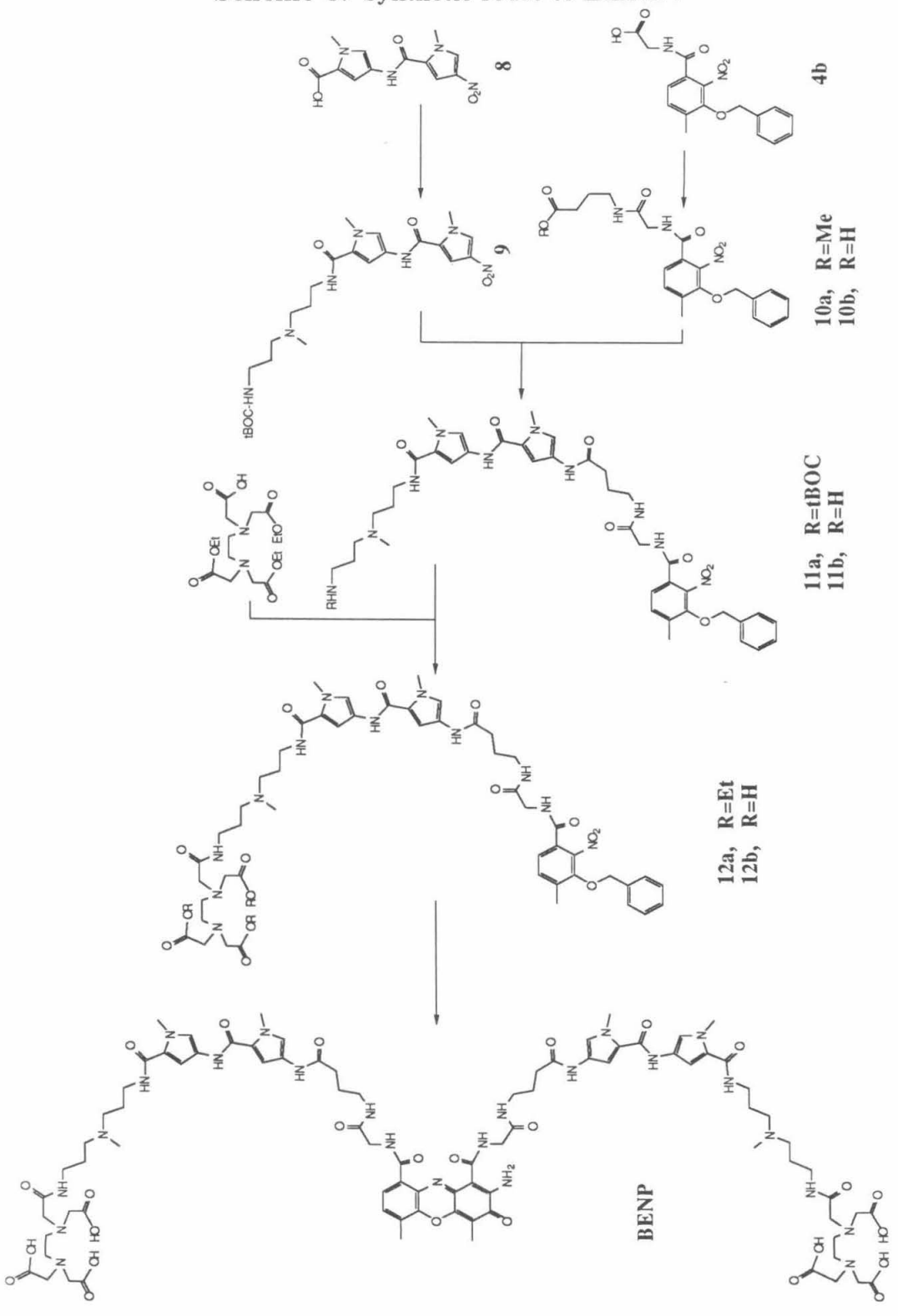
Synthesis of BENP

The synthetic route to *bis*(EDTA–netropsin–glycyl–GABA)phenoxazone is shown in Scheme 3. N,N′-Carbonyldiimidazole mediated coupling of the dipyrrole nitro acid **8**, with mono-*t*BOC protected 3,3′-diamino-N-methyldipropylamine (**7**) afforded **9** in 71% yield. Catalytic hydrogenation of **9** and subsequent condensation with the carbonyldiimidazole adduct of **10a** gave **11a** in 44% yield. Deprotection of **11a** to **11b** (95%) and condensation with excess carbonyldiimidazole activated EDTA-triethyl ester afforded **12a** in 78% yield. Saponification of **12a**, catalytic hydrogenation of **12b** and subsequent oxidative coupling with potassium ferricyanide in 1:1 methanol/phosphate buffer (pH 7.1), followed by chromatography on Chellex–100, gave the red-orange hexazwitterion of **BENP** in 71% yield.

DNA Cleavage with BENP•2Fe

The sequence specific DNA cleavage ability of **BENP**•2Fe was examined on a 517bp ^{32}P end-labelled restriction fragment from pBR322 as described for **BEDP**•2Fe. Figure 9 is the autoradiogram and figure 10 the histograms for the cleavage reaction. For comparison, **BEDF**•2Fe and the original hybrid, **BEDP**•2Fe, are included in figures 9 and 10. The binding sites for **BENP**•2Fe are different than those for both **BEDP**•2Fe and **BEDF**•2Fe. The left-hand-most **BEDP** site (5′-TATAGGTTAA, figure 10), the possible minor groove–intercalation–minor groove site, is lost with the redesigned **BENP**. Instead, **BENP** binds nearly the same site as **BEDF** in this region. The strong cleavage on the left hand end of this site indicates that one dipyrrole unit is tightly binding in the DNA minor groove, and is shifted right ~1bp relative to **BEDF**. The broad cleavage pattern on the right hand end of this site implies that the other

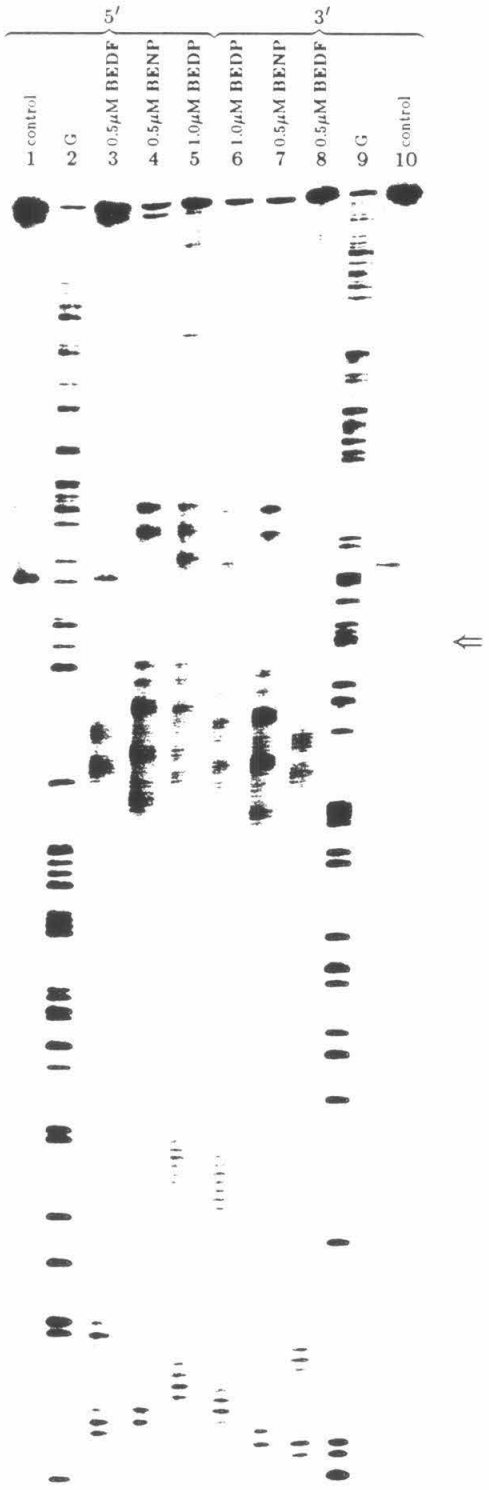
Scheme 3: Synthetic route to **BENP**.



dypyrrole unit is not binding tightly in the minor groove. **BEDP**•2Fe gives two ambigous cleavage loci (center of the histogram in figure 10). With **BENP**•2Fe, the left hand locus is also observed, though it appears to be less tightly bound, and the right hand one is lost.

The most intriguing binding/cleavage sites are the ones just below the arrow in figure 9 and on the right hand side of the histograms in figure 10. In this region **BEDF** binds the 8bp site 5′-TTATTTTT-3′. The original hybrid molecule, **BEDP**, apparently has overlapping binding sites in this region, giving at least four cleavage loci. The DNA cleavage patterns do not lend themselves to anaylsis of binding sites. It appears, however, that part of the **BEDP**•2Fe cleavage pattern in this region is similar to that of **BEDF**•2Fe. **BENP**•2Fe, on the other hand, gives a much less ambigous cleavage pattern in this region. It appears that there are two ajoining 10bp binding sites; 5′-ATTTGTTTAT-3′ and 5′-TTTTCTAAAT-3′, and that the very strong cleavage between the sites is due to the combined cleavage of two EDTA•Fe's, one from each binding site. It appears that the redesigned hybrid, **BENP**, is indeed more specific than the original hybrid molecule. However, **BENP** still appears to have monomeric and/or dimeric binding modes, though to a lesser extent. The two strong 10bp binding sites both include a single guanine residue at either the fifth or sixth base pair, i.e. stradling the center of the site. The inclusion of a single guanine at the center of the site, and the ± 4bp binding site shift relative to the fumaramide linked tripyrrole compound (**BEDF**), both suggest that at these two sites, binding may be minor groove–intercalation–minor groove as designed. It is also interesting to note that the cleavage pattern for **BEDP**•2Fe at this 20bp region is essentially the sum of the cleavage patterns for **BEDF**•2Fe and **BENP**•2Fe. This indicates that the original hybrid exhibits some of the DNA binding modes of both **BENP** and **BEDF**.

Figure 9: Autoradiogram of 5′ (lanes 1–5) and 3′ (lanes 6–10) ^{32}P end-labelled 517bp DNA restriction fragment on a high resolution denaturing wedge gel. Reaction conditions were end-labelled restriction fragment (~18,000cpm), 100μM bp calf thymus DNA, 20mM TRIS•acetate pH 7.9 and 5mM DTT. **BEDF**•2Fe, **BEDP**•2Fe and **BENP**•2Fe were equilibrated with the DNA for two hours at 37°C prior to the addition of DTT. Reactions were run two hours at 37°C and terminated by ethanol precipitaion. The ^{32}P end-labelled products were analyzed on an 8%, 1:20 crosslinked 50% urea denaturing polyacrylamide wedge gel (0.2mm thick at the top and 0.4mm thick at the bottom). Lanes 1 and 10 are intact DNA; lanes 2 and 9, Maxam-Gilbert G reaction; lanes 3 and 8 **BEDF**•2Fe(II), 0.5μM; lanes 4 and 7 **BENP**•2Fe(II), 0.5μM; lanes 5 and 6 **BEDP**•2Fe(II), 1μM. Bottom to the arrow is the sequence left to right in figure 10.


5′
3′
1 control
2 G
3 0.5μM BEDF
4 0.5μM BENP
5 1.0μM BEDP
6 1.0μM BEDP
7 0.5μM BENP
8 0.5μM BEDF
9 G
10 control

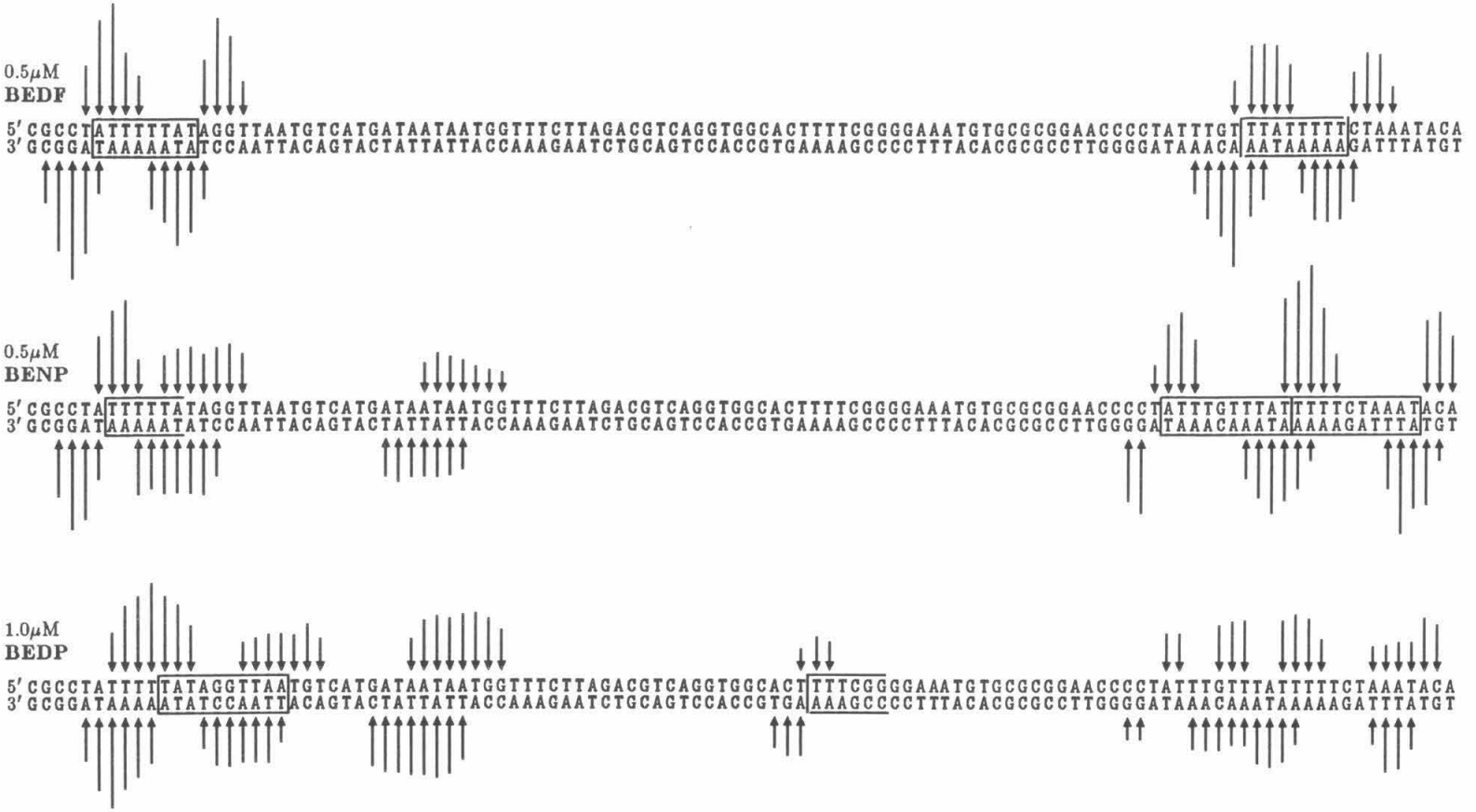


Figure 10: Histograms for the gel in figure 9. The sequences shown left to right correspond to the bottom of the gel to the arrow in figure 9. The height of the arrows represent extent of cleavage resulting in removal of the indicated base. Boxes represent binding sites assigned from the model in figure 6.

Does the Phenoxazone Intercalate?

The two strong cleavage sites observed for **BENP**, 5′-ATTTGTTTAT-3′ and 5′-TTTTCTAAAT-3′, *suggest* that both of the dipyrrolecarboxamide units are binding in the DNA minor groove and that the phenoxazone is intercalating. To determine if the phenoxazone is still competent at intercalation, the UV/VIS spectra of **BENP**•2Fe was examined in the presence and absence of calf thymus DNA. When bound (intercalated) to DNA, the **ACT** phenoxazone visible adsorption at 445nm undergoes a ~20nm red shift with 30% decrease in extinction.[2,65] Similarly, **NET** binding to DNA also induces a red shift and hypochromic effect in the **NET** UV adsorption at 296nm.

Figure 11 shows the UV/VIS spectra of 5μM **BENP**•2Fe(III) in the absence and presence of 5mM bp calf thymus DNA (1:1000). In the presence of calf thymus DNA, the phenoxazone adsorption undergoes a 25nm bathochromic shift and reduction in extinction. The adsorption for the dipyrroles at 296nm also undergoes a red shift (~14nm) and reduction in extinction. These two observations indicate that the phenoxazone is still able to intercalate, and that the dipyrrolecarboxamide units bind to the DNA in a manner similar to **NET**. Though the UV/VIS results do not specifically indicate *simultaneous* binding of all three domains of **BENP**, they do indicate that both the dipyrrole units and phenoxazone are still *competent* at their respective binding modes.

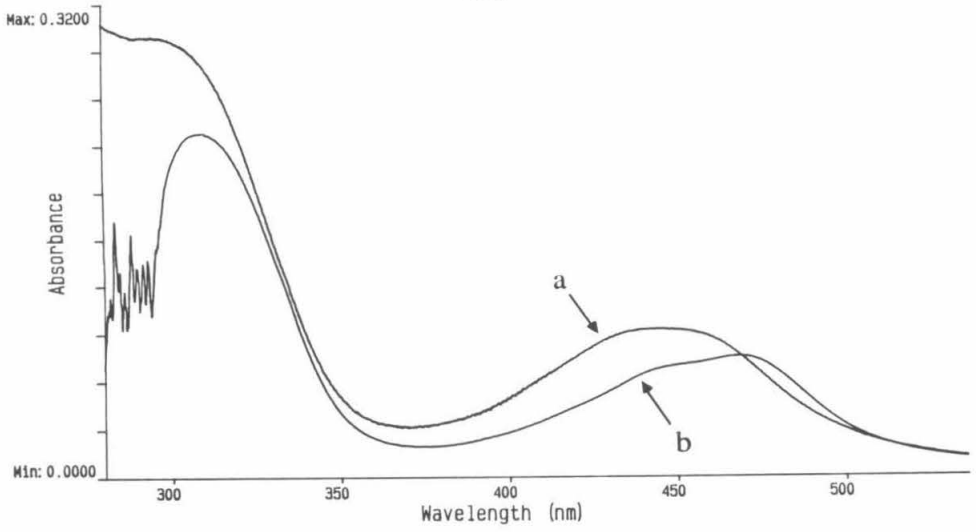


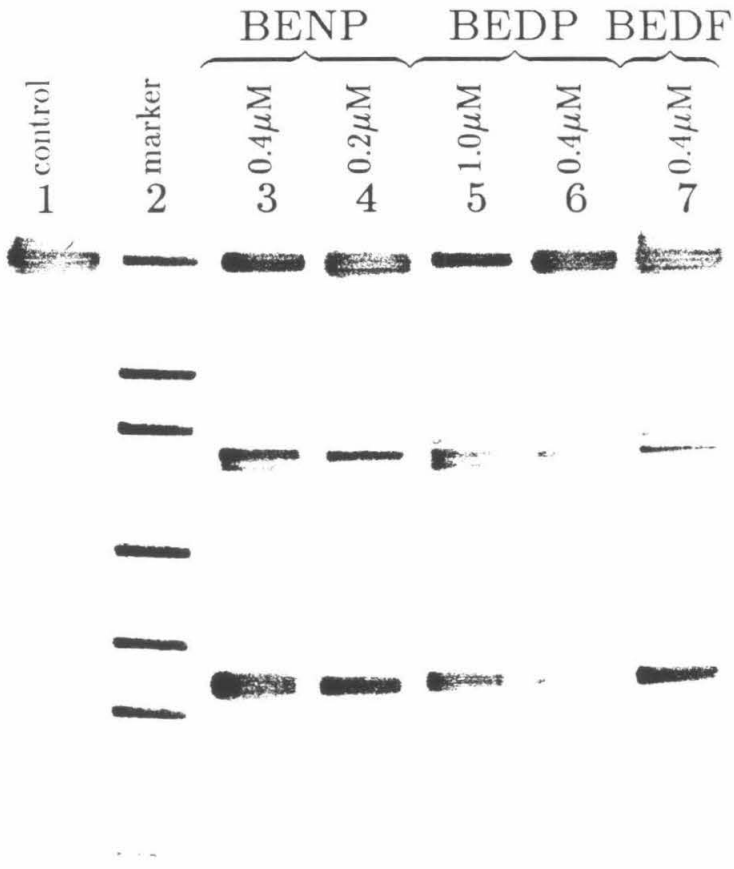
Figure 11: UV/VIS spectra of 5μM **BENP**•2Fe(III) in the absence (curve a) and presence of 5mM bp calf thymus DNA (curve b). The spectra of 5mM bp calf thymus DNA has been subtracted from curve b which is the source of the noise below 300nm. The adsorption around 450nm is due to the phenoxazone moeity and the one near 300nm is due to the dipyrroles.

Double Strand Cleavage with BENP•2Fe

Since there are 524,800 unique 10 base pair DNA sequences, it is possible that we have not examined the best binding sites for **BEDP** and **BENP**. However, since both hybrids contain two chelators, they should at least be reasonably competent at *double* strand cleavage. Double strand cleavage on larger DNA fragments should allow the searching of a large number of potential binding sites, as well as giving an indication of the overall specificity of the two hybrids. To examine their abilities to cause double strand cleavage, **BEDP**•2Fe and **BENP**•2Fe were allowed to react with ^{32}P end-labelled linearized pBR322 (4363 base pairs) under standard reaction conditions. The ^{32}P labelled reaction products were analyzed by non-denaturing polyacrylamide gel electrophoresis followed by autoradiography (figure 12).

The DNA double strand cleavage caused by 0.5μM **BEDF**•2Fe gives one strong

Figure 12: Non-denaturing agarose gel autoradiogram of DNA double strand cleavage. pBR322 was linearized with Sty I and ^{32}P end-labelled at only one $3'$ end. DNA cleavage conditions were the same as in figure 9. Lane 1 is a DNA control, lane 2 is a molecular weight marker (4364, 3371, 2994, 2368, 19998, 1768, 1372, 995 and 666 base pairs top to bottom), lanes 3 and 4 are **BENP**•2Fe cleavage reactions at 0.4 and 0.2μM, respectively. Lanes 6 and 7 are **BEDP**•2Fe cleavage reactions at 1.0 and 0.4μM, respectively. Lane 8 is 0.4μM **BEDF**•2Fe cleavage reaction.


BENP
BEDP
BEDF
control
marker
0.4μM
0.2μM
1.0μM
0.4μM
0.4μM
1
2
3
4
5
6
7

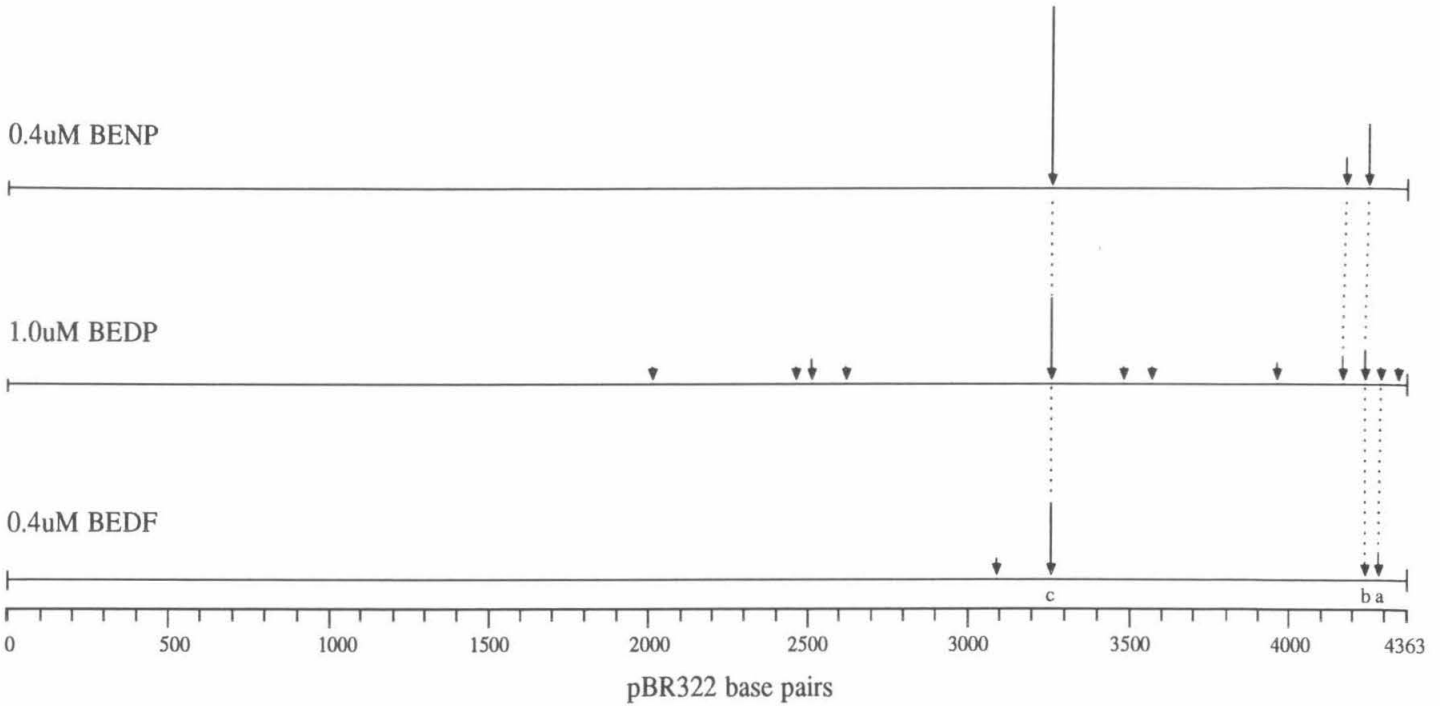


Figure 13: Histograms of the double strand cleavage data in figure 12. The height of the arrows represent the extent of cleavage at the indicated location in pBR322, +/-20bp. The sequence is numbered relative to the Eco RI site. The three major cleavage sites are; a, 4320 (low site on the 517bp fragment); b, 4240 (high site on the 517bp fragment) and c, 3235.

and three weaker sites.[63] The two cleavage sites near the top of the gel in figure 12 map to the two binding sites observed in the high resolution studies on the 517bp fragment (see figure 9). The strongest site in figure 13 maps to a region of pBR322 (3235 relative to Eco RI) which includes the longest run of A•T base pairs in the plasmid; 15bp, 5′-TTTTAAATTAAAAAT-3′. **BEDP**•2Fe (0.4μM) gives more than 12 detectable bands, as well significant background cleavage under the same conditions. The strongest **BEDP**•2Fe cleavage also maps to the 3235 site of pBR322. In contrast to **BEDP**•2Fe, 0.4μM **BENP**•2Fe gives only three cleavage bands, one of which is subtantially reduced by decreasing the **BENP**•2Fe concentration to 0.2μM. The higher site on the gel in figure 12 maps to the double binding site on the 517bp fragment observed in the high resolution studies (just below the arrow in figure 9 and right hand side of figure 10). The major double strand cleavage site for **BENP**•2Fe maps to the same region of pBR322, as do the major sites for **BEDP**•2Fe and **BEDF**•2Fe (3235). The broadness of the major cleavage band suggests that it may result from cleavage at more than one binding site in close ($\leq$40bp) proximity. It is clear from figure 12 that the redesigned hybrid (**BENP**) is significantly more specific than the original hybrid molecule **BEDP**.

BENP•DNA Binding Model

The observed facility of **BENP**•2Fe mediated double strand cleavage, in conjunction with the changes in the UV/VIS spectra in the presence of DNA, suggest that at least some binding events involve both minor groove binding of the two dipyrrolecarboxamides, and, simultaneously, intercalation of the phenoxazone moeity. The results of model building studies of this interaction are shown in figure 14. A model of **BENP**, without the metal chelates, was docked to a 10 base

pair B form DNA of sequence 5′-ATTTGTTTAT-3′. This sequence is one of the two strong binding sites observed in the high resolution DNA cleavage studies. The modeling studies indicate that the dipyrrolecarboxamide units are capable of interacting with a sequence of the form $(A/T)_4$ similar to that observed in the **NET**•dodecamer complex crystal structure. The modeling studies further indicate that the flexability of the GABA linker is required for proper orientation of the dipyrrolecarboxamide units relative to the intercalated phenoxazone. Interestingly, the model indicates that though **BENP** is nearly C_2 symmetric, the interactions of the phenoxazone carbonyls, glycyls and GABA portions of **BENP** with the DNA are highly asymmetric. This asymmetry appears to be caused by the asymmetry of the DNA binding site and leads to different conformations of the phenoxazone carbonyls and glycyl moeities. However, the remaining amides of both dipyrrolecarboxamide domains are still properly positioned in the minor groove.

CONCLUSION

The design, synthesis and testing of two DNA binding molecules that are *hybrids* of the natural products, distamycin/netropsin and actinomycin have been described. These molecules are hybrids in two other senses as well. They address the issue of where one can mix (and match) *groove binders* with *intercalators*, and they address the issue of whether one can couple A•T 'words' with G•C 'words'.

We have found that *bis*(distamycin–glycyl)phenoxazone will apparently bind 10 base pairs of DNA of general sequence $(A•T)_4(G•C)_2(A•T)_4$. However, not all sites allow simultaneous binding of the groove binders, suggesting that the extent and nature of the distortion of the sequences flanking the intercalator may be *sequence dependent.* Additionally, double strand cleavage of a 4363bp DNA fragment with

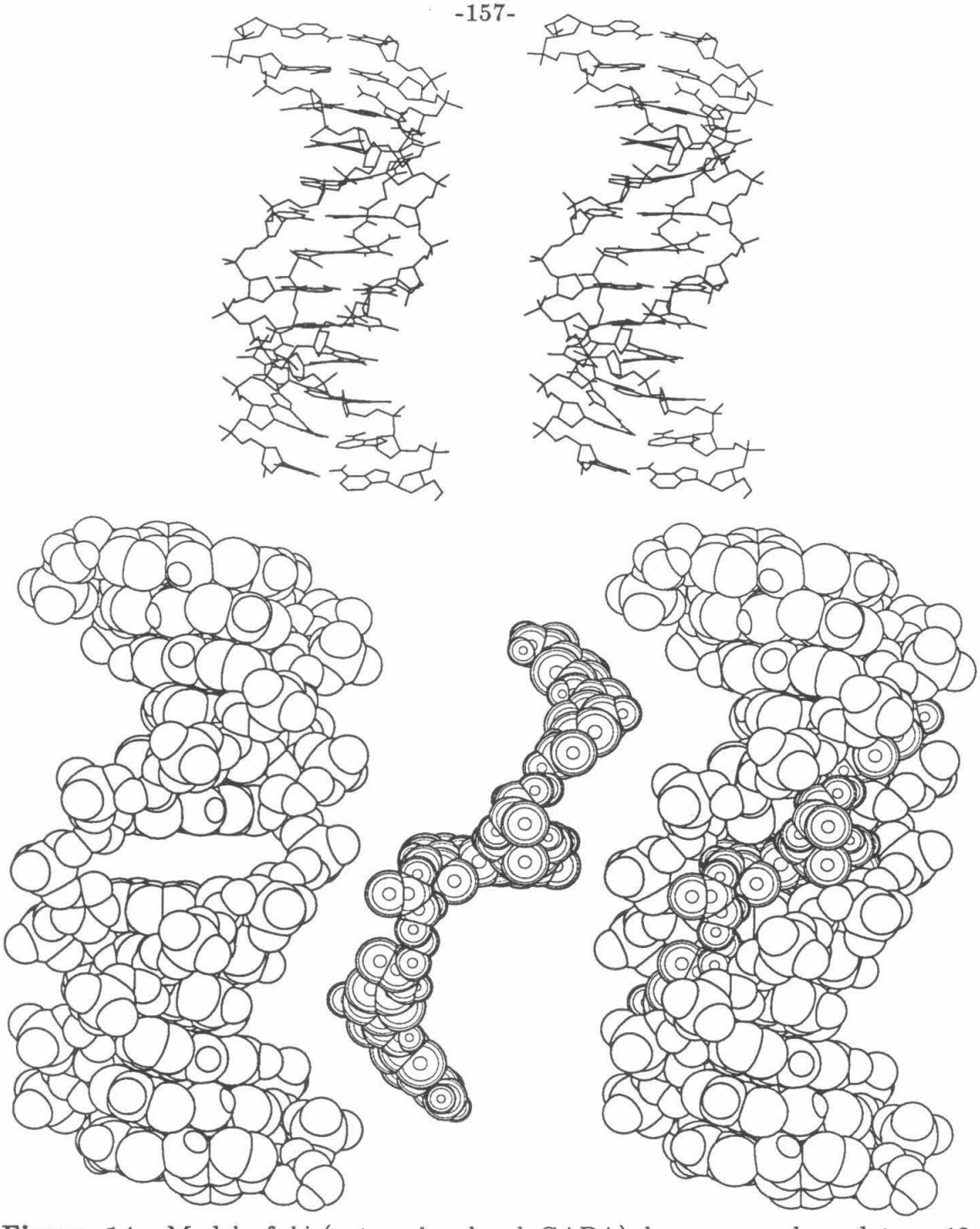


Figure 14: Model of *bis*(netropsin–glycyl–GABA)phenoxazone bound to a 10 base pair psuedo-B form DNA. (Top) Stereo drawing of the complex. (Bottom, left to right) Spacefilling models of the DNA with an empty intercalation pocket, **BENP** and the complex. In both models only heteroatom hydrogens are included.

BEDP shows a large number of weak cleavage sites, indicating that its sequence specific DNA binding is not optimized. A second generation hybrid, *bis*(netropsin–glycyl–GABA)phenoxazone, has shown improved sequence specific DNA binding ability and highly efficient and specific DNA double strand cleavage ability.

Experimental

See the experimental section for Chapter I for general comments and procedures used in the preparation of ^{32}P end-labelled restriction fragments. The following procedure for the preparation of nitro acid **1** is a modification of a published procedure[17,66] which is easier for large scale preps and gives a higher overall yield.

N-Methyl-4-nitropyrrole-2-carboxylic acid, 1a:[17,66] A solution of 100g (0.71mmol) N-methylpyrrole-2-carboxylic acid (Aldrich, 97+%) in 600ml acetic anhydride was cooled to -25°C in dry ice/acetone. A solution of 80ml concentrated nitric acid in 380ml acetic anhydride (prepared cold) was slowly added over 75 minutes with vigorous mechanical stirring and without allowing the temperature to rise above -20°C. After the addition was completed, the mixture was stirred an additional 3 hours at -25°C and one hour at ~-10°C. The resulting black solution was poured (carefully) into one liter of ice and stirred for two hours at room temperature. The solution (2.6l) was refrigerated (-20°C) overnight during which time brown crystals formed. The crystals were removed by filtration and washed with water until the washes were colorless. Drying under vacuum at 40°C gave 60.6g (50%) **1a** as light brown needles. NMR analysis indicated 97% 4-nitro and 3% 5-nitro products. Recrystallization from 300ml of 1:1 acetic acid/water at -20°C followed by filtration, water washing and vacuum drying gave 57.4g (48%) **1a** as a light tan material which was pure by NMR (NMR detection limit of 5-nitro acid and starting material is less than 1%): NMR (D_6-DMSO) δ 8.25 (d, 1H, j=2Hz), 7.25 (d, 1H, j=2Hz), 3.9 (s, 3H). The supernatants from the original reaction mixture were extracted with dichloromethane (3x600ml), dried (Na_2SO_4) and stripped, giving 50g of material, which was recrystalized as described above. NMR analysis of the tan colored product (28g, 23%) indicated 60% 5-nitro acid, 39% 4-nitro acid

and 1% starting material by weight. Further recrystallizations removed starting material, but the 6:2 ratio of 5- and 4-nitro acid products remained constant. Pure N-methyl-4-nitropyrrole-2-carboxylic acid methyl ester, **1b**, can be obtained from this mixture by esterification followed by chromatography on alumina.[17,66]

N-Methyl-4-nitropyrrole-2-carboxylic acid methyl ester, 1b:[17,66] To a cold solution of 17ml concentrated sulphuric acid and 175ml methanol were added 17.1g (0.10mmol) nitro acid. The solution was refluxed for 15 hours, then cooled to 0°C. The product was removed by filtration, washed with 5% sodium bicarbonate and water and dried *in vacuo*, giving 14.6g (79%) **1b** as off-white needles: NMR ($CDCl_3$) δ 7.1 (d, 1H, j=5Hz), 6.85 (d, 1H, j=5Hz),4.3 (s, 3H), 3.85 (s, 3H). Concentration of the supernatants afforded 1.5g (8.8%) recovery of starting material.

N-Methyl-4-(N-methyl-4-nitropyrrole-2-carboxamide)-pyrrole-2-carboxylic acid methyl ester, 1c:[17,66] In 200ml dry dimethyl formamide (DMF) was dissolved 14.6g (79.4mmol) nitro ester **1b**. Palladium on charcoal (1.5g) was added and the mixture hydrogenated under 1 atm hydrogen for 3 hours. The catalyst was removed by filtration through celite, giving a transparent black solution to which 20ml (0.2mol) triethylamine were added. N-methyl-pyrrole-2-carboxyl chloride (prepared by refluxing 14.8g (87mmol) nitro acid **1a** in 50ml thionyl chloride for two hours, followed by vacuum distillation of the excess thionyl chloride) in 50ml DMF was added to the filtered hydrogenation mixture and stirred overnight. Ice water (400ml) was added, the solution stirred for 20 minutes at room temperature and cooled to 0°C. The product was filtered, washed with 10% bicarbonate, water and cold methanol and dried *in vacuo*. Trituration with diethyl ether and redrying removed the last traces of DMF giving 22g (90%) of light yellow-green

nitro ester **1c**: NMR (D_6-DMSO) δ 10.3 (s, 1H), 8.2 (d, 1H, j=1.5Hz), 7.6 (d, 1H, j=2.0Hz), 7.5 (d, 1H, j=1.5Hz), 6.9 (d, 1H, j=1.5Hz), 4.0 (s, 3H), 3.85 (s, 3H), 3.75 (s, 3H).

N-Methyl-4-[N-methyl-4-(N-methyl-4-nitropyrrole-2-carboxamide) -pyrrole-2-carboxamide]-pyrrole-2-carboxylic acid methyl ester, 1d:[17,66] Nitro ester **1c** (21g, 69mmol) was dissolved in 200ml DMF with 1g palladium on charcoal and hydrogenated under 1 atm hydrogen for 24 hours. The catalyst was removed by filtration through celite and 26g triethylamine was added to the black solution. N-methyl-4-nitropyrrole-2-carboxyl chloride (from 12.5g, 73mmol, **1a**) in 75ml DMF was added and the solution stirred overnight. The product was removed by filtration and washed with 10% bicarbonate, water, cold methanol and ether followed by drying *in vacuo*. The light yellow-green solid was finely ground and digested in 1400ml boiling methanol for 15 minutes. Cooling to 0°C, filtering, washing (methanol and ether) and vacuum drying gave 22.5g (77%) of **1d** as yellow crystals: NMR (D_6-DMSO) δ10.35 (s, 1H), 10.00 (s, 1H), 8.2 (d, 1H, j=1.5Hz), 7.60 (d, 1H, j=1.5Hz),7.50 (d, 1H, j=1.5Hz), 7.05 (s, 1H, j=1.5Hz), 6.9 (d, 1H, j=1.5Hz), 4.0 (s, 3H), 3.83 (s, 3H), 3.75 (s, 3H).

N-Methyl-4-[N-methyl-4-(N-methyl-4-nitropyrrole-2-carboxamide) -pyrrole-2-carboxamide]-pyrrole-2-carboxylic acid, 1:[17,66] A solution of 1.6g (41mmol) sodium hydroxide in 70ml water was added to a suspension of 2.0g (4.7mmol) **1d** in 70ml ethanol. The mixture was refluxed 3 hours and the ethanol removed on a rotovap. The aqueous solution was diluted to 200ml with water and then slowly acidified with 6N hydrochloric acid to pH 1. The product was removed by filtration and washed with cold 1:1 methanol/water and acetone. Vacuum drying gave 1.8g (93%) of yellow solid **1**: NMR (D_6-DMSO) δ 3.84 (s, 3H),

3.87 (s, 3H), 3.97 (s, 3H), 6.85 (d, 1H, j=1.5Hz), 7.10 (d, 1H, j=1.5Hz), 7.25 (d, 1H, j=1.5Hz), 7.26 (d, 1H, j=1.5Hz), 7.45 (d, 1H, j=1.5Hz), 7.65 (d, 1H, j=1.5Hz), 8.2 (d, 1H, j=1.5Hz), 9.95 (s, 1H), 10.35 (s, 1H); UV/VIS (H_2O) 391 (ϵ=35,600), 236nm.

3-(*t*-Butylcarbamide)-3′-diamino-N-methyldipropylamine, 7: To 5g (34mmol) 3,3′-diamino-N-methyldipropylamine (Aldrich) were added 1.9g (8.7 mmol) di-*t*-butyldicarbonate (Fluka) and the mixture held at 50°C for one hour. The mixture was diluted with 30ml water and extracted with dichloromethane (3x30ml). The organic layer was stripped and the crude product chromatagraphed with 12% concentrated aqueous ammonium in methanol. Drying *in vacuo* gave 0.71g (34%) **7** as a colorless liquid which was pure by TLC: NMR ($CDCl_3$) δ 5.75 (broad singlet, 1H) 3.13 (p, 2H, j=6Hz), 2.73 (t, 2H, j=6Hz), 2.37 (t, 4H, j=6.8Hz), 2.17 (s, 3H), 1.7 (m, 6H), 1.49 (s, 9H).

Nitro-tripyrrolecarboxamide-linker-*t*-boc, 2: To a cold (0°C) solution of 1.16g (2.8 mmol) **1** and 0.91g (6.7mmol) N-hydroxybenzotriazole monohydrate in 50ml DMF were added 0.69g (3.4mmol) dicyclohexylcarbodiimide. The mixture was allowed to warm to room temperature overnight with stirring. A solution of 0.72g (2.9mmol) **7** in 5ml DMF was added and stirring continued for 15 hours. Dicyclohexylurea was removed by filtration followed by removal of DMF *in vacuo* at 40°C. After triturating with ether (3x), the crude product was chromatagraphed with 0.25% concentrated aqueous ammonia in methanol, giving 1.4g (79%) **2** as a yellow solid: NMR (D_6-DMSO) δ 10.25 (s, 1H), 9.90 (s, 1H), 8.14 (d, 1H, j=1.5Hz), 8.00 (m, 1H), 7.57 (d, 1H, j=1.5Hz), 7.22 (d, 1H, j=1.5Hz), 7.19 (d, 1H, j=1.5Hz), 7.02 (d, 1H, j=1.5Hz), 6.80 (d, 1H, j=1.5Hz), 6.7 (m, 1H), 3.92 (s, 3H), 3.83 (s,

3H), 3.78 (s, 3H), 3.2 (m, 4H), 3.25 (m, 4H), 2.12 (s, 3H), 1.6 (m, 4H), 1.37 (s, 9H).

2-Nitro-3-benzyloxy-4-methylbenzoic acid, 3:[67] A solution of 13.8g (0.25mol) freshly powdered pottassium hydroxide in 50ml DMSO was stirred for 5 minutes followed by the addition of 5g (25mmol) 2-nitro-3-hydroxy-4-methylbenzoic acid (Aldrich) and 12ml (96mmol) α-chlorotoluene (Aldrich). After stirring for three hours, 190ml of water was added and the mixture extracted with ether (2x100ml) to remove excess α-chlorotoluene. The aqueous layer was acidified to pH 1 with concentrated hydrocloric acid and extracted with ether (2x100ml). These ether extracts were washed with a 50% brine solution (4x50ml), dried (Na_2SO_4) and stripped. Drying in vacuum gave 7.1g (97%) **3** as a light yellow solid: NMR ($CDCl_3$) δ 8.2 (broad singlet, 1H), 8.8 (d, 1H, j=7.5Hz), 7.4 (m, 7H), 4.91 (s, 2H), 2.33 (s, 3H).

2-Nitro-3-benzyloxyl-4-methylbenzamide glycine methyl ester, 4a:[68] A solution of 7.0g (24mmol) **3** and 4.4g (27mmol) acydiimidazole in 60ml DMF was stirred for 24 hours under a dry atmosphere. N-ethylmorpholine (3.4g, 27mmol) and 3.4g (27mmol) glycine methyl ester hydrochloride were added and stirring continued for 12 hours. The reaction mixture was diluted with 600ml chloroform and extracted with 2.5% sodium bicarbonate (3x150ml), 5% hydrochloric acid (2x150ml), water (3x150ml) and saturated sodium chloride (150ml). The organic layer was dried (Na_2SO_4), and the chloroform removed on a rotovap. Drying *in vacuo* gave 8.0g (91%) **4a** as a cream colored solid which was pure by TLC: NMR ($CDCl_3$) δ 7.33 (m,7H), 6.74 (t, 1H, j=5Hz), 4.89 (s, 2H), 4.06 (d, 2H, j=5Hz), 3.74 (s, 3H), 2.31 (s, 3H).

Glycine-2-nitro-3-benzyloxy-4-methylbenzamide, 4b: To a solution of 7.5g (21mmol) **4a** in 250ml methanol were added 100ml of 1N lithium hydroxide and the mixture was stirred at room temperature for 2 hours. The methanol was removed on a rotovap followed by dilution with 150ml water and slow acidification with 6N hydrochloric acid to pH 1. The product was removed by filtration and washed with 0.1N hydrochloric acid and ether, followed by vacuum drying. The 6.8g of crude product were digested in 250ml boiling chloroform then allowed to stand at 5°C overnight. Filtration followed by vacuum drying gave 6.5g (90%) **4b** as marble white crystals: NMR (D_6-DMSO) δ 12.7 (s, 1H), 9.12 (t, 1H, j=6Hz), 7.59 (s, 2H), 7.41 (s, 5H), 4.94 (s, 2H), 3.86 (d, 2H, j=6Hz), 2.37 (s, 3H).

1/2-Phenox-tripyrrolecarboxamide-linker-*t*-boc, 5a: A solution of 1.16g (1.81mmol) **2** and 0.2g palladium on charcoal in 3ml DMF was hydrogenated under one atmosphere hydrogen for 20 hours, followed by removal of the catalyst by filtration through celite. A solution of 0.65g (1.9mmol) **4b** and 0.34g (2.1mmol) acydiimidazole in 0.7ml DMF, after stirring for one hour, was added to the filtered hydrogenation mixture and stirred overnight. DMF was removed *in vacuo* at room temperature and the crude product triturated with ether (3x). The product was chromatographed twice with 30% methanol in dichloromethane (at ~5ml/min from a 40x150mm column), giving 0.82g (48%) **5a** as a bright yellow solid which was pure by TLC: NMR (D_6-DMSO) δ 10.01 (s, 1H), 9.96 (s, 1H), 9.91 (s, 1H), 9.12 (t, 1H, j=5.8Hz), 8.06 (t, 1H, j=5.6Hz),7.65 (s, 2H), 7.45 (m, 5H), 7.27 (d, 1H, j=1.0Hz), 7.20 (s, 2H), 7.06 (d, 1H, j=1.1hz), 6.96 (d, 1H, 1.2Hz), 6.86 (d, 1H, j=1.3Hz), 6.82 (t, 1H, j=4.8Hz), 5.00 (s, 2H), 4.01 (d,2H, j=4.3Hz), 3.87 (s, 6H), 3.82 (s, 3H), 3.21 (q, 2H, j=5.8Hz), 2.96 (q, 2H, j=5.8Hz), 2.53 (s, 3H), 2.32 (m, 4H), 2.16 (s, 3H), 1.64 (p, 2H, j=6.8Hz), 1.54 (p, 2H, j=6.7Hz), 1.39 (s, 9H).

1/2 - Phenox - tripyrrolecarboxamide - linker, 5b: To a solution of 725mg (0.77mmol) **5a** in 15 ml dichloromethane was added 10ml trifluoracetic acid. After 10 minutes, the mixture was slowly diluted with 75ml of ether which precipitated the di-trifluoracetic acid salt of **5b**. The crude product was chromatographed with 12% concentrated ammonia in methanol giving 605mg (93%) of **5b** as a light yellow solid: NMR (D_6-DMSO) δ 10.21 (s, 1H), 9.95 (s, 1H), 9.90 (s, 1H), 9.2 (broad singlet, 1H), 8.05 (t, 1H, j=6Hz), 7.65 (d, 1H, j=9Hz), 7.60 (d, 1H, j=9Hz), 7.42,(m, 5H), 7.25 (d, 1H, 1.0Hz), 7.19 (s, 2H), 7.05 (d, 1H, j=1Hz), 6.94 (d, 1H, j=1Hz), 6.83 (d, 1H, j=1Hz), 4.97 (s, 2H), 3.98 (s, 2H), 3.86 (s, 6H), 3.80 (s, 3H), 3.18 (m, 2H), 2.56 (broad singlet, 2H), 2.43 (s, 3H), 2.31 (m, 4H), 2.13 (s, 3H), 1.62 (p, 2H, j=8Hz), 1.48 (p, 2H, j=7Hz).

1/2-Phenox-tripyrrolecarboxamide-EDTA triethyl ester, 6a: A solution of 900mg (2.4 mmol) ethylendiamine tetraacetic acid triethyl ester[69] and 410mg (2.5mmol) acydiimidazole in 3ml DMF was stirred under a dry atmosphere for one hour, followed by the addition of 466mg (0.56mmol) **5b** in 0.5ml DMF. After stirring for 15 hours, the DMF was removed *in vacuo* at room temperature and the product triturated with ether (4x). Column chromatography with 1% concentrated aqueous ammonia in methanol afforded 630mg (95%) **6a** as a yellow solid: NMR (D_6-DMSO) δ 9.97 (s, 1H), 9.90 (s, 1H), 9.87 (s, 1H), 9.08 (t, 1H, j=5Hz), 8.00 (t, 1H, j=4Hz), 7.95 (t, 1H, j=5Hz), 7.64 (d, 1H, j=10Hz), 7.62 (d, 1H, j=10Hz), 7.41 (m, 5H), 7.24 (d, 1H, j=1.5Hz), 7.16 (s, 2H), 7.03 (d, 1H, j=1.5Hz), 6.93 (d, 1H, j=1.5Hz), 7.84 (d, 1H, j=1.5Hz), 4.96 (s, 2H), 4.05 (q, 6H, j=7Hz), 3.97 (s, 2H), 3.85 (s, 6H), 3.79 (s, 3H), 3.50 (s, 6H), 3.12 (q, 2H, j=6.7Hz), 2.72 (t, 2H, j=6.3Hz), 2.67 (t, 2H, j=6.3Hz), 2.42 (s, 3H), 2.31 (t, 2H, j=8.3Hz), 2.29 (t, 2H, j=8.3Hz), 2.13 (s, 3H), 1.62 (p, 2H, j=7Hz), 1.56 (p, 2H, j=7Hz), 1.17 (t, 9H, j=7Hz).

1/2-Phenox-tripyrrolecarboxamide-EDTA, 6b: To a solution of 200mg (167μmol) **6a** in 5ml of methanol was added 1.5ml of 1M lithium hydroxide and the mixture was allowed to stir in the dark at room temperature for 4 hours. The solvents were removed under reduced pressure and the product chromatagraphed on silica gel with 2% concentrated aqueous ammonia in methanol affording 170mg (91%) **6b** as a light yellow solid. NMR (1:10 trifluoroacetic acid/D_6-DMSO) δ 10.00 (s, 1H), 9.96 (s, 1H), 9.94 (s, 1H), 9.42 (bs, 1H), 9.13 (t, 1H, j=6Hz), 8.53 (t, 1H, j=6Hz), 8.22 (bs, 1H), 7.67 (d, 1H, j=8Hz), 7.61 (d, 1H, j=8Hz), 7.44 (m, 4H), 7.41 (m, 1H), 7.25 (d, 1H, j=1.3Hz), 7.20 (s, 2H), 7.12 (d, 1H, j=1.3Hz), 7.01 (d, 1H, j=1.7Hz), 6.99 (d, 1H, j=1.3Hz), 4.99 (s, 2H), 4.07 (s, 2H), 4.01 (d, 2H, j=5Hz), 3.98 (s, 2H), 3.94 (s, 2H), 3.88 (s, 9H), 3.85 (s, 4H), 3.36 (m, 2H), 3.30 (m, 4H), 3.25 (q, 2H, j=6Hz), 3.08 (m, 2H), 2.81 (d, 3H, j=4Hz), 2.45 (s, 3H), 1.90 (m, 4H).

BEDP: A mixture of 120mg (108μmol) **6b** and 30mg 5% palladium on carbon in 10ml methanol was hydrogenated under one atmosphere hydrogen in the dark for three hours. Under an argon atmosphere the catalyst was removed by filtration through celite and the celite was washed with 2ml methanol. The solution was diluted with 12ml of 67mM phosphate buffer (pH 7.1) and 89mg (270μmol, 2.5eq) potassium ferricyanide in 1ml phosphate buffer was added. The solution was stirred for two minutes during which time it turned deep red–orange characteristic of the phenoxazone moeity. The solution was allowed to stand for two hours, followed by removal of solvents under reduced pressure. The crude product was dissolved in minimal water and loaded on an Amberlite XAD–2 (Kodak) column. The column was washed with water (150ml), 3% EDTA–Na_2 (500ml) and water (1L). **BEDP** was eluted with 50% aqueous methanol. The solvents were removed *in vacuo*, the product was triturated with 2ml methanol and dried;

67mg (68%) **BEDP** as a red-orange solid. UV/VIS (H_2O) 438nm (phenoxazone), 305nm (pyrroles), 235nm; $\epsilon_{305}/\epsilon_{438}$=3.0, expected $\epsilon_{305}/\epsilon_{438}$=3.0. NMR (1:10 triflouroacetic acid/D_6-DMSO) δ 10.02 (s, 1H), 9.93 (s, 3H), 9.90 (s, 1H), 9.83 (s, 1H), 9.44 (bs, 2H), 9.08 (t, 1H, j=6Hz), 8.59 (t, 2H, j=5Hz), 8.24 (bs, 1H), 8.10 (bs, 2H), 7.72 (d, 1H, j=8Hz), 7.46 (d, 1H, j=8Hz), 7.25 (s, 1H), 7.23 (s, 1H), 7.20 (s, 4H), 7.17 (s, 1H), 7.15 (s, 1H), 7.07 (s, 2H), 7.02 (s, 2H), 4.44 (s, 8H), 4.03 (s, 4H), 4.00 (s, 8H), 3.88 (m, 18H), 3.57 (s, 3H), 3.52 (q, 6H, j=7Hz), 3.48 (m, 6H), 3.34 (bs, 4H), 3.29 (q, 6H, j=7Hz), 3.21 (m, 4H), 3.14 (m, 4H), 2.84 (d, 6H, j=4Hz), 2.56 (s, 3H), 2.20 (s, 3H), 1.95 (m, 8H).

N-Methyl-4-(N-methyl-4-nitropyrrole-2-carboxamide)-pyrrole-2-carboxylic acid, 8: To a suspension of 2g (6.5mmol) **1e** in 80ml ethanol was added a solution of 1.5g (37.5mmol) sodium hydroxide in 80ml water and the mixture refluxed for 3.5hr. The solution was cooled and concentrated to 100ml under vaccum. After acidification with 6N HCl, the product was removed by filtration, washed with water and dried to give 1.99g (104%) **8** as a yellow-green solid. NMR (D_6–DMSO) δ10.24 (s, 1H), 8.16 (d, j=2Hz, 1H), 7.56 (d, j=2Hz, 1H), 7.41 (s, 1H), 6.83 (d, j=2Hz, 1H), 3.94 (s, 3H), 3.83 (s, 3H).

9: To a cold (0°C) solution of 300mg (1.03mmol) **8**, 277mg (1.1mmol, 1.1eq) **7** and 303mg (2.3mmol, 2.2eq) N-hyroxybenzotriazole monohydrate in 10ml DMF was added 254mg (1.2mmol, 1.2eq) N,N′-dicyclohexylcarbodiimide. The solution was allowed to warm to room temperature over 16hrs. Dicyclohexylurea was removed by filtration and the solvents removed under vacuum at 30–40°C. The crude product was triturated with diethyl ether (3x) and dried. Chromatography with 0.3% concentrated aqueous ammonium hydroxide, 15% methanol in dichloromethane followed by re-chromatography of the mixed fractions with

0.5% concentrated aqueous ammonium hydroxide in methanol, afforded 381mg (71%) **9** as a yellow solid which was pure by TLC. NMR (D_6–DMSO) δ10.25 (s, 1H), 8.18 (d, j=1.7Hz, 1H), 8.13 (bs, 1H), 7.59 (d, j=2.2Hz, 1H), 7.21 (d, j=1.7Hz, 1H), 6.85 (bs, 2H), 3.96 (s, 3H), 3.82 (s, 3H), 3.35 (m, 4H), 3.22 (m, 2H), 2.95 (t, j=5.6Hz, 2H), 1.71 (bs, 2H), 1.62 (bs, 2H), 1.37 (s, 9H).

γ-Aminobutyric acid methyl ester hydrochloride: γ-Aminobutyric acid (5g, 49mmol) was suspended in 75ml methanol and hydrogen chloride was bubbled through the solution for three hours. The methanol was removed on a rotovap followed by 12 hours *in vacuo*. The crude product was recrystalized from 1:2 methanol/ether at -20°C. The extremely hydroscopic product was removed by filtration, washed with ether, and quickly transferred to a flask for vacuum drying, which gave 6.3g (86%) as white crystals which were pure by TLC: NMR (D_6-DMSO) δ 8.22 (broad singlet, 3H), 3.60 (s, 3H), 2.75 (broad multiplet, 2H), 2.42 (t, 2H, j=7Hz), 1.81 (p, 2H, j=7Hz).

1/2-Phenox-GLY-GABA-methyl ester, 10a: In 100ml dimethoxyethane were placed 6g (17.4mmol) **4b** and 2.55g (19.2mmol) N-hydroxysuccinamide monohydrate. The solution was cooled to 0°C, 4.3g(21mmol) dicyclohexylcarbodiimide was added, and the solution allowed to warm to room temperature overnight with stirring. Dicyclohexyl urea was removed by filtration 2.7g (18mmol) γ-aminobutyric acid methyl ester hydrochloride and 6ml N-ethylmorpholine was added. After stirring for four hours, the reaction mixture was diluted with 200ml water and extracted with chloroform (2x50ml). The organic layer was washed with 2.5% bicarbonate (2x50ml), brine (1x50ml) and dried (Na_2SO_4). Removal of the chloroform *in vacuo* followed by chromatography with 4% methanol in dichloromethane afforded 3.8g (48%) **10a** which was pure by TLC: NMR ($CDCl_3$) δ 7.59 (t, 1H,

j=5.8Hz), 7.39 (s, 7H), 7.01 (t, 1H, j=6Hz), 4.98 (s, 2H), 4.09 (d, 2H, j=5.8Hz), 3.64 (s, 3H), 3.31 (q, 2H, j=6Hz), 2.4 (m, 5H), 1.84 (p, 2H, j=7Hz).

1/2-Phenox-GLY-GABA acid, 10b: A solution of 3.7g(8.4mmol) **10a** and 72ml of 1N lithium hydroxide in 200ml methanol was stirred in the dark at room temperature for two hours. The methanol was removed on a rotovap, 100ml water were added and the solution slowly acidified with 6N hydrochloric acid to pH 1. The aqueous solution was extracted with ethyl acetate (3x50ml). The combined organic layers were washed with 0.2N hydrochloric acid (50ml) and brine (50ml). The ethyl acetate was removed *in vacuo* giving 3.2g (90%) **10b** as a cotton candy like solid: NMR (DMOS-D_6) δ 8.89 (t, 1H, j=6Hz), 7.83 (t, 1H, j=6Hz), 7.56 (s, 2H), 7.40 (3, 5H), 4.92 (s, 2H), 3.77 (d, 2H, j=6Hz), 3.09 (q, 2H, j=7Hz), 2.40 (s, 3H), 2.18 (t, 2H, j=7Hz), 1.61 (p, 2H, j=7Hz).

11a: A solution of 300mg (580μmol) **9** and 30mg 5% palladium on carbon in 3ml DMF was hydrogenated under 1Atm hydrogen in the dark for 48hrs at 10°C. The catalyst was removed by filtration through celite and the celite washed with DMF (3x1ml). A solution of 372mg (870μmol, 1.5eq) **10a** and 154mg (950μmol, 1.1eq relative to **10a**) carbonyldiimidazole in 5ml DMF was stirred 6hr at room temperature, filtered, and added to the filtered hydrogenation mixture. The resulting solution was stirred overnight in the dark followed by removal of solvent under reduced pressure at 35°C. The crude product was chromatographed twice with 0.25% concentrated aqueous ammonium hydroxide in methanol and once with 30% methanol in dichloromethane giving 232mg (44%) **11a** as a yellow solid which was pure by TLC. NMR (D_6–DMSO) δ9.84 (s, 1H), 9.81 (s, 1H), 8.98 (t, j=5.6Hz, 1H), 8.02 (t, j=5.0Hz, 1H), 7.98 (t, j=5.1Hz, 1H), 7.62 (d, j=8.1Hz, 1H), 7.59 (d, j=8.1Hz, 1H), 7.39 (m, 5H), 7.16 (s, 1H), 7.14 (s, 1H), 6.85 (s, 1H), 6.82 (s, 1H),

6.80 (m, 1H), 4.95 (s, 2H), 3.81 (s, 3H), 3.79 (s, 3H), 3.31 (s, 3H), 3.17 (m, 2H), 3.13 (m, 2H), 2.94 (q, j=5.2Hz, 2H), 2.41 (s, 3H), 2.25 (m, 6H), 2.13 (bs, 2H), 1.71 (p, j=6.8Hz, 2H), 1.61 (m, 2H), 1.52 (m, 2H), 1.36 (s, 9H).

11b: To a solution of 150mg (167μmol) **11a** in 3ml dichloromethane was added 2ml trifluoroacetic acid with vigorous stirring. The solution was stirred in the dark at room temperature for 10min followed by dilution with 15ml diethyl ether. The trifluoroacetate salt of the crude product was removed by filtration and chromatographed with 12% concentrated aqueous ammonium hydroxide in methanol affording 127mg (95%) **11b** as a light yellow solid which was pure by TLC. NMR (D_6–DMSO) δ9.87 (bs, 1H), 9.85 (s, 1H), 9.08 (bs, 2H), 8.50 (bs, 1H), 7.64 (d, j=7.8Hz, 1H), 7.58 (d, j=7.9Hz, 1H), 7.41 (m, 5H), 7.18 (s, 1H), 7.15 (s, 1H), 6.85 (s, 1H), 6.80 (s, 1H), 4.95 (s, 3H), 3.80 (m, 6H), 3.18 (m, 2H), 3.12 (m, 2H), 2.41 (s, 2H), 2.28 (m, 4H), 2.12 (s, 3H), 1.72 (p, j=7.3Hz, 2H), 1.61 (p, j=6.7Hz, 2H), 1.53 (p, j=6.8Hz, 2H).

12a: A solution of 117mg (312μmol, 5eq) EDTA triethyl ester[47] and 51mg (315μmol) acydiimidazole in 0.5ml DMF was stirred for 2hr at room temperature. A solution of 50mg (62μmol) **11b** in 2ml DMF was added and the resulting solution stirred overnight in the dark. DMF was removed under reduced pressure and the crude product triturated with diethyl ether (3x). Chromatography with 1% concentrated aqueous ammonium hydroxide in methanol afforded 56mg (78%) **12a** as a yellow glass which was pure by TLC. NMR ($CDCl_3$) δ9.34 (bs, 1H), 8.66 (bs, 1H), 8.33 (m, 1H), 7.81 (bs, 1H), 7.71 (bs, 1H), 7.36 (m, 5H), 7.28 (m, 2H), 7.16 (s, 1H), 6.72 (s, 1H), 6.58 (s, 1H), 4.91 (s, 3H), 4.12 (m, 6H), 4.03 (bs, 2H), 3.89 (s, 3H), 3.85 (s, 3H), 3.49 (s, 3H), 3.30 (m, 8H), 2.74 (m, 4H), 2.48 (m, 2H), 2.39 (m, 8H), 2.31 (s, 3H), 2.19 (s, 2H), 1.89 (m, 2H), 1.71 (m, 4H), 1.24 (m, 9H),

12b: To a solution of 56mg (48μmol) **12a** in 1ml methanol was added 0.435ml 1N (9eq) lithium hydroxide and the resulting solution was stirred 4hrs at room temperature in the dark. The solvents were removed under reduced pressure and the crude product chromatographed with 2% concentrated aqueous ammonium hydroxide in methanol giving 35mg (67%) **12b** as a yellow solid which was pure by TLC. NMR (10% trifluoroacetic acid/D_6–DMSO) δ9.82 (s, 1H), 9.76 (s, 1H), 9.25 (bs, 1H), 8.94 (t, j=5 6. Hz, 1H), 8.43 (t, j=5 2. Hz, 1H), 8.12 (bs, 1H), 7.93 (t, j=5.6Hz, 1H), 7.57 (d, j=8.0Hz, 1H), 7.52 (d, j=7.8Hz, 1H), 7.35 (m, 5H), 7.10 (s, 1H), 7.08 (s, 1H), 6.89 (s, 1H), 6.82 (s, 1H), 4.90 (s, 2H), 3.98 (s, 4H), 3.87 (s, 2H), 3.83 (s, 4H), 3.75 (s, 6H), 3.10 (m, 12H), 2.71 (bs, 4H), 2.35 (s, 3H), 2.19 (m, 2H), 1.79 (m, 4H), 1.66 (m, 2H),

BENP: A solution of 35mg (33μmol) **12b**, 10mg 5% palladium on carbon and 15μl triethylamine in 4ml methanol was hydrogenated under 1atm hydrogen in the dark for 3hrs. The solution was filtered through a 0.45μm cellulose acetate filter and the filter washed with 2ml methanol. To this solution was added 27mg (82μmol) potassium ferricyanide in 5ml 67mM phosphate buffer, pH 7.12. The solution was stirred two minutes in the dark during which time it turned deep red–orange characteristic of the phenoxazone moeity. After standing an additional 30min in the dark, the methanol was removed under reduced pressure and the remaining aqueous solution frozen and lyophilized. The crude product was dissolved in water and passed through a Chelex-100 (sodium form, Biorad) column, the eluent was frozen and lyphilized affording 22mg (71%) **BENP** as a red–orange solid which was homogenous by TLC. UV/VIS (H_2O) 448nm (phenoxazone), 300nm (pyrroles), and 220nm; $\epsilon_{300}/\epsilon_{448}$=2.65. NMR (10% trifluoroacetic acid/D_6–DMSO) δ10.80 (t, j=2.0Hz, 1H), 9.94 (d, j=6.0Hz, 2H), 9.83 (d, j=6.0Hz, 2H), 9.35 (m, 4H), 8.98 (t,

j=4.7Hz, 1H), 8.50 (m, 2H), 8.19 (m, 2H), 8.09 (m, 1H), 7.92 (m, 1H), 7.66 (d, j=7.9Hz, 1H), 7.44 (d, j=7.7Hz, 1H), 7.19 (s, 2H), 7.13 (s, 1H), 7.11 (s, 1H), 6.96 (s, 2H), 6.91 (s, 1H), 6.89 (s, 1H), 4.11 (s, 4H), 3.98 (s, 4H), 3.93 (s, 8H), 3.85 (m, 12H), 3.39 (m, 4H), 3.25 (m, 16H), 3.17 (m, 8H), 3.05 (m, 4H), 2.79 (m, 6H), 2.30 (m, 4H), 2.16 (s, 3H), 1.87 (m, 12H), 1.79 (m, 8H), 1.28 (s, 2H).

References and Notes

1. M. Record, T. Lohman, P. Haseth, *J. Mol. Biol.*, **107**, 145 (1976).
2. W. Muller, D. Crothers, *J. Mol. Biol.*, **35**, 251 (1968).
3. R. Dickerson, R. Drew, B. Conner, R. Wing, A. Fratini, M. Kopka, *Science*, **216**, 475 (1982).
4. W. Wilson, R. Jones, *Advances in Pharmacology and Chemotherapy*, **18**, 177 (1981).
5. L. Lerman, *J. Cell. Comp. Physiol.*, **64**, 1 (1964).
6. M. Waring in "*The Molecular Basis of Antibiotic Action*," E. Gail *et al.*, Eds., Wiley, New York, 1972, p.173.
7. W. Baur in "*Annual Reviews of Biophysics and Bioengineering*" L. Mullins, Ed., Annual Reviews, Palo Alto, California, 1978, p.287.
8. L. Wakelin, M. Waring, *Mol. Pharmacol.*, **10**, 544 (1974).
9. G. Quigley, A. Wang, G. Ughetto, J. van Brom, A. Rich, *Proc. Natl. Acad. Sci., USA*, **77**, 7204 (1980).
10. E. Gabbay, R. Scofield, C. Baxter, *J. Am. Chem. Soc.*, **95**, 7850 (1973).
11. H. Sobell, B. Reddy, K. Bhandray, S. Jain, T. Sakore, T. Seshadri, *Cold Springs Harbor Symposia on Quantitative Biology*, 87 (1977).
12. See for example T. Brennan, S. Sengupta, *J. Med. Chem.*, **26**, 448 (1983).
13. R. Pasternak, E. Gibss, R. Villafranca, *Biochemistry*, **22**, 2406 (1983).
14. C. Zimmer, *Prog. Nuc. Acid Res. Mol. Biol.*, **15**, 285 (1975).
15. F. Hahn in "*Antibiotics*", vol. III, J. Corcoran, F. Hahn, Eds., Springer-Verlag, New York, 1975, pp. 79–100.
16. M. Bialer, B. Yagen R. Mechoulam, *Tetrahedron*, **34**, 2389 (1978).
17. A. Zasedatelev *et al.*, *Stud. Biophys.*, **67**, 47 (1978).
18. H. Berman, S. Niedle, C. Zimmer, H. Thrum, *Biochem. Biophys. Acta.*, **561**, 124 (1979).

19. G. Luck, H. Triebel, M. Waring, C. Zimmer, *Nuc. Acid Res.*, **1**, 503 (1974).

20. C. Zimmer, G. Luck, *FEBS. Lett.*,**10**, 339 (1970).

21. A. Krey, *Progr. Mol. Subcell. Biol.*, **7**, 43 (1980).

22. M. Waring, *J. Mol. Biol.*, **54**, 247 (1970).

23. A. Kylov *et al.*, *Nuc. Acid Res.*, **6**, 289 (1979).

24. R. Wartell, J. Larson, R. Welss, *J. Biol. Chem.*, **250**, 2698 (1975).

25. A. Kolchinskii *et al.*, *Mol. Biol. (USSR)*, **9**, 14 (1975).

26. C. Zimmer, G. Luck, R. Nuske, *Nuc. Acid Res.*, **8**, 2999 (1980).

27. A. Krey, *FEBS Lett.*, **26**, 58 (1973).

28. D. Patel, L. Canuel, *Proc. Natl. Acad. Sci., USA*, **12**, 5207 (1977).

29. D. Patel, *Eur. J. Biochem.*, **99**, 369 (1979).

30. M. Kopka, C. Yoon, D. Goodsell, P. Pjura, R. Dickerson, *J. Mol. Biol.*, **183**, 553 (1985).

31. M. Kopka, C. Yoon, D. Goodsell, P. Pjura, R. Dickerson, *Proc. Natl. Acad. Sci., USA*, **82**, 1376 (1985).

32. M. Van Dyke, P. Dervan, *Nuc. Acid Res.*, **11**, 5555 (1983).

33. M. Van Dyke, P. Dervan, *Cold Springs Harbor Symposia on Quantitative Biology*, **47**, 347 (1983).

34. P. Schultz, J. Taylor, P. Dervan, *J. Am. Chem. Soc.*, **104**, 6861 (1982).

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

36. P. Schultz P. Dervan,*J. Biomol. Struct. Dyn.*, **1**, 1133 (1984).

37. H. Brockman, J. Lackner, *J. Chem. Ber.*, **101**, 1312 (1968).

38. J. Meienhofer, *J. Am. Chem. Soc.*, **92**, 3771 (1970).

39. T. Tanaka, K. Nakajima, K. Okawa, *Bull. Chem. Soc., JAPAN*, **53**, 1352 (1980).

40. A. Mauger, O. Stuart, J. Ferretti, J. Silverton, *J. Am. Chem. Soc.*, **107**, 7154 (1985).

41. A. Mayer *et al.*, *Top. Antibiotic Chem.*, **5**, 223 (1980).

42. W. Remers *et al.* in "*The Chemistry of Antitumor Antibiotics*," vol 1, Wiley, New York, 1979, pp. 63–132.

43. I. Goldberg, M. Hobinowitz, E. Reich, *Proc. Natl. Acad. Sci., USA*, **48**, 2094 (1962).

44. R. Wells, J. Larson, *J. Mol. Biol.*, **49**, 319 (1970).

45. R. Roberts, *Nucleic Acid Res.*, **11**, 1135 (1983)

46. F. Takusagawa, M. Dabrow, S. Neidle, H. Berman, *Nature*, **296**, 466 (1982).

47. S. Jain, H. Sobell, *J. Mol. Biol.*, **68**, 1 (1972).

48. T. Krugh, M. Young, *Nature (London)*, **269**, 627 (1977).

49. F. Quadrifoglio, V. Crescenzi, *Biophys. Chem.*, **2**, 64 (1974).

50. F. Quadrifoglio, A. Ciana, V. Crescenz, *Biopolymers*, **15**, 595 (1976).

51. H. Berg, K. Bachus, W. Jaekel, G. Luber, K. Teach, *Monatsberichte der deutschen Acadamie der Wissenschaften, Berlin*, **1**, 524 (1959).

52. L. Caclieri, R. Nemchin, *Biochim. Biophys. Acta.*, **87**, 641 (1964).

53. H. Devoe, I. Tinoco, *J. Mol. Biol.*, **4**, 518 (1962).

54. T. Krugh, E. Mooberry. Y-C. Chen Chiao, *Biochemistry*, **16**, 740 (1977).

55. Y-C. Chen Chiao, T. Krugh, *Biochemistry*, **16**, 747 (1977).

56. M. Van Dyke, Ph.D. Thesis, California Institute of Technology, Pasadena, California, 1984.

57. M. Lane, J. Drabrowiak, J. Vournakis, *Proc. Natl. Acad. Sci., USA*, **80**, 3260 (1983).

58. A. Scamrov, R. Beabealachvilli, *FEBS Lett.*, **164**, 97 (1983).

59. K. Fox, M. Waring, *Nuc. Acids Res.*, **12**, 9271 (1984).

60. *Occam's Razor:* (William of Occam, 1836) a scientific and philosophic rule that entities should not be multiplied unnecessarily which is interpreted as requiring that *the simplest of competing theories be preferred to the more complex* or that explanations of unknown phenomena be sought first in terms of known quantities.

61. M. Krivtsova, E. Moroskina, E. Glibin, E. Frisman, *Molekulyarnaya Biologiya*, **18**, 950 (1984).

62. J. P. Sluka, P. Dervan in "*New Synthetic Methodology and Functionally Interesting Compounds*," Z-I Yoshida, Ed., Elsevier, New York, 1986, pp. 307–322.

63. R. S. Youngquist, P. Dervan, *J. Am. Chem. Soc.*, **107**, 28(1985).

64. The entire sequence of pBR 322 has been determined; J. Sutcliffe, *Cold Springs Harbor Symposia on Quantitative Biology*, **43**, 77 (1979). However an error has been found in the original sequence: K. Peden, *Gene*, **22**, 277 (1983); An additional CG base pair at position 526 in the tetracycline resistance gene. This increases the size of pBR322 to 4363 base pairs.

65. S. Segupta, D. Schafer, *Biochim. Biophys. Acta*, **521**, 89 (1978).

66. P. Schultz; Ph.D. Thesis, California Institute of Technology, Pasadena, California, 1984.

67. H. Brockman, H. Muxfeldt, *Chem. Ber.*, **91**, 1242 (1958).

68. J. Taylor, P. Dervan, unpublished results.

69. R. Hay, K. Nolan, *J. Chem. Soc., Dalton*, 348 (1975).

CHAPTER III:

Bleomycin–Distamycin Hybrid Affinity Cleaving Molecules

As in the work described in Chapter II of this report, this chapter describes work aimed at the design, synthesis and characterization of hybrids of two naturally occurring sequence specific DNA binding compounds. In particular we will attempt to modify the A+T specificity of distamycin-like (**DST**) tripyrrolecarboxamides by incorporating the putative 5′-GC recognition domain of bleomycin (**BLM**).

Bleomycin

The bleomycins (figure 1) are a family of naturally occurring glycopeptide derived antibiotics that were first discovered in the 1960's by Umezawa and coworkers.[1–4] **BLM** has been used clinically against certain malignant lymphomas and squamous cell carcinomas[4] where its therapeutic activity is believed to be related to its ability to bind to and degrade DNA.[5] Numerous studies of **BLM** induced DNA degradation have established that DNA strand scission requires molecular oxygen and a metal ion, with Fe(II) (see figure 1b) being the most active and most extensively studied.[6–10]

Chemical degradation studies,[3] as well as total synthesis,[11,12] indicate that **BLM** consists of three functional domains, as indicated in figure 1a. The sugar

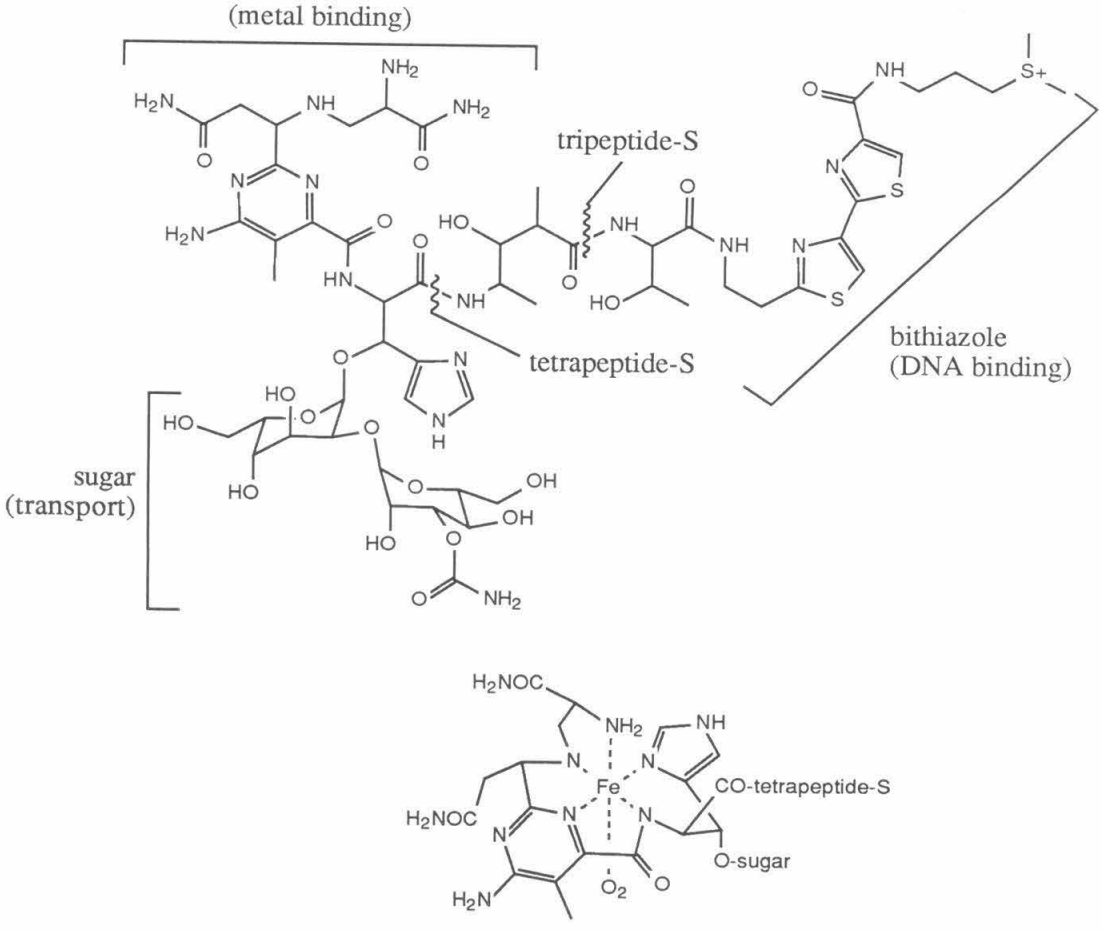


Figure 1: (Top) The three domains of bleomycin A_2. (Bottom) A proposed structure for the iron chelate of bleomycin.[5]

moeities are believed to be required for cellular transport,[13] the pyrimidine, β-aminoalanine and β-hydroxyimidazole for metal binding, and the bithiazole and terminal sulfonium domain for DNA recognition.[14,15]

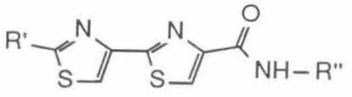
BLM preferentially cleaves the double stranded DNA sequence 5′-Pyr-G-C-Pur at the cysteine residue[16] by a specific oxidative degradation of the deoxyribose. Cleavage products are free nucleotide base and base propenal from the degraded nucleoside, and 5′ phosphate and 3′ phosphoglycolate termini for the flanking DNA

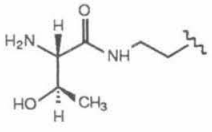
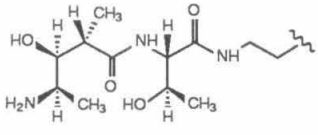
sequences.[10,17,18] The observed products are consistent with a mechanism involving abstraction of the deoxyribose C4′ hydrogen which is accessible via the DNA minor groove.[5] Hydroxyl radical and/or superoxide have been implicated as the hydrogen abstracting species.[19–21] However, the highly specific cleavage observed, i.e. only one deoxyribose per strand, indicates that the oxidant is not freely diffusable.

The sequence specific DNA recognition domain of **BLM** has been identified as the bithiazole containing tripeptide–S (figure 1a). Synthetic **BLM**'s which lack tripeptide–S are not competent at DNA cleavage. Furthermore, synthetic tripeptide–S efficiently inhibits DNA strand scission by wild type **BLM**, which indicates that the tripeptide–S is involved in DNA binding and not metal chelation. The association constants for tripeptide–S and **BLM** in the absence of metal ions are essentially identical, $1.2 \times 10^5 M^{-1}$ and $1.4 \times 10^5 M^{-1}$, respectively.[22,23] DNA binding of several tripeptide–S derivatives has been reported as shown in Table I.

Though there is little doubt about the major role of the bithiazole containing tripeptide–S in DNA binding, the *mode* of binding is still controversial.[5] The two major DNA binding models for tripeptide–S are intercalation and minor-groove binding. Evidence for intercalation of the bithiazole group includes a DNA unwinding angle of 12° (versus 26° for ethidium)[23] and an upfield shift in the ^{1}H–NMR bithiazole resonances in the presences of poly(dAdT).[22,25] However, the magnitude of both the unwinding angle and the ^{1}H–NMR shifts are significantly smaller than those observed for compounds which have been unambigously assigned as intercalators, i.e. ethidium, implying that the bithiazole does not fully intercalate. The minor groove model for the bithiazole•DNA complex is based on the observation that the cleavage chemistry occurs in the minor groove and on the structural similarities between the bithiazole and known minor groove binding molecules such

Table I: DNA interactions of bithiazole derivatives. Entries marked with "→" are the most directly related to the Bleomycins.



	R′	R″	M[a]	reference
→	$CH_3CONHCH_2CH_2$	$(CH_2)_3S^+(CH_3)_2$	$5x10^{-3}$	b, c
	$CH_3CONHCH_2CH_2$	$(CH_2)_4N^+H_3$	10^{-4} – $5x10^{-3}$	b
	C_6H_5	$(CH_2)_3S^+(CH_3)_2$	10^{-4}	b
	4-thiazolyl	$(CH_2)_3S^+(CH_3)_2$	10^{-4}	b
→		$(CH_2)_3S^+(CH_3)_2$	get	c
→		$(CH_2)_3S^+(CH_3)_2$	$3x10^{-5}$	c
→	"	$NH(CH_2)_4NHC(NH_2)_2^+$	$2x10^{-4}$	c
	$NH_2(CH_2)_2$	$NH(CH_2)_3NH(CH_2)_4NH_2$	$6x10^{-6}$	c

a. Concentration for 50% inhibition of Fe•**BLM**-mediated DNA cleavage.

b. From Oberley and Beuther.[21]

c. From Mayer and Que.[24]

as netropsin and Hoechst 33258[26] (shown in figure 2). Furthermore, phleomycin, which contains a reduced, and hence nonplanar, bithiazole ring, also binds and cleaves DNA with specificity similar to **BLM**.[27] DNA cleavage with phleomycin argues against a classical intercalation event.

There is also evidence that **BLM** may have more than one DNA binding mode. ^{1}H–NMR studies of tripeptide–S[25] and DNA–induced fluoresence quenching of **BLM**[28] both indicate *two* modes of binding. It is possible that **BLM** has

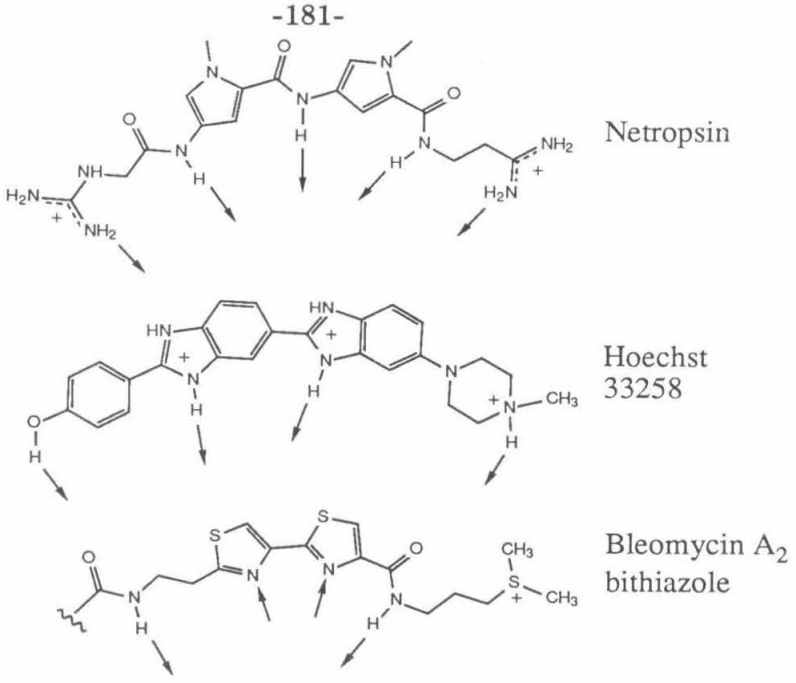


Figure 2: Comparison of the structures of netropsin, Hoechst 33258 and the DNA binding bithiazole portion of **BLM**. Arrows indicate potential hydrogen bond donors and acceptors. Note the common cresent shape of the three molecules with hydrogen bonding sites located on the concave side.[5]

mutually exclusive intercalation and minor groove binding modes for its bithiazole domain. Whether just one or both potential binding modes are the DNA *cleaving* modes is unclear.

Bleomycin–Distamycin Hybrids, Design and Synthesis

The conversion of the sequence specific DNA binding distamycin (**DST**) into a sequence specific DNA cleaving molecule by the incorporation of EDTA•Fe was described in the background section of Chapter II. We are concerned here with modifying the A+T specificity of the **DST**-like trypyrrolecarboxamides by incorporating the 5′-GC specific bithiazole domain of **BLM**. To this end I have prepared two hybrids of **DST** and **BLM**. The **DST** portions of the hybrids are tris-N-methylpyrrolecarboxamide units and the **BLM** portions are the bithiazole

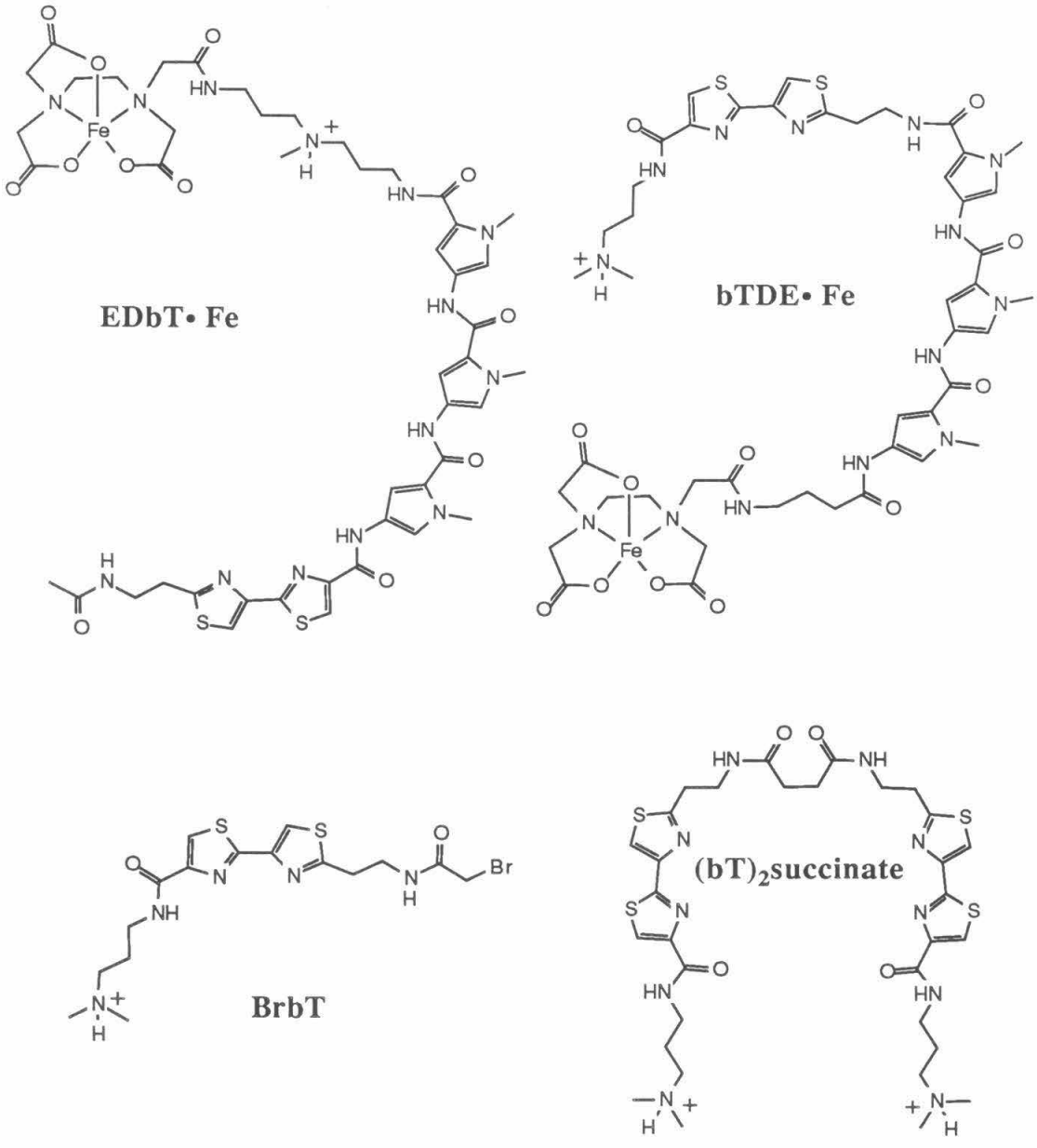


Figure 3: Iron chelates of the bleomycin-distamycin hybrids EDbT and bTDE and the bithiazole binding probes BrbT and $(bT)_2$succinate.

unit of the **BLM** tripeptide–S. Additionaly, I have incoporated the metal chelator EDTA to facilitate identification of DNA binding sites. In the first hybrid, EDTA–distamycin–bithiazole (**EDbT**, see figure 3), the bithiazole is coupled through its carboxy terminus directly to the amino terminus of the tripyrrolecarboxamide domain. The resulting **EDbT** is similar to a pyrrolecarboxamide homolog in which the cresent shape and extended π system are retained. The tripyrrolecarboxamide replaces the carboxy-terminal positive charge of **BLM**. In the second hybrid, bithiazole–distamycin–EDTA (**bTDE**), the amino terminus of the bithiazole is coupled to the carboxy terminus of the tripyrrolecarboxamide domain. The resulting **bTDE** has a two carbon linker between the two domains which interupts the extended π system and provides some independence in the positioning of the two domains. In **bTDE**, the carboxy terminus positive charge of **BLM** is retained. Both **EDbT** and **bTDE** were designed to bind to DNA sequences of the form 5′-(A•T)$_5$GC and 5′-GC(A•T)$_5$.

In addition to the two hybrid molecules, I have also prepared two bithiazole based molecules to try to gain insight into the *mode* of the **BLM** interaction with DNA. Previous work by Schultz and Dervan[29] has shown that an EDTA•Fe equipped bithiazole does not sequence-specificaly cleave DNA. Their results, however, do not rule out the possibility that the bithiazole unit is sequence specific. It may be that the specificity of the bithiazole is for a mono- or dinucleotide, i.e. G or GC and GT, and the EDTA•Fe cleavage over several base pairs has obscured the specific DNA binding. Furthermore, the EDTA•Fe moeity may be too large to be specifically delivered to the DNA binding site by the bithiazole. To circumvent these possible problems I have prepared a bromoacetyl equiped bithiazole similar to

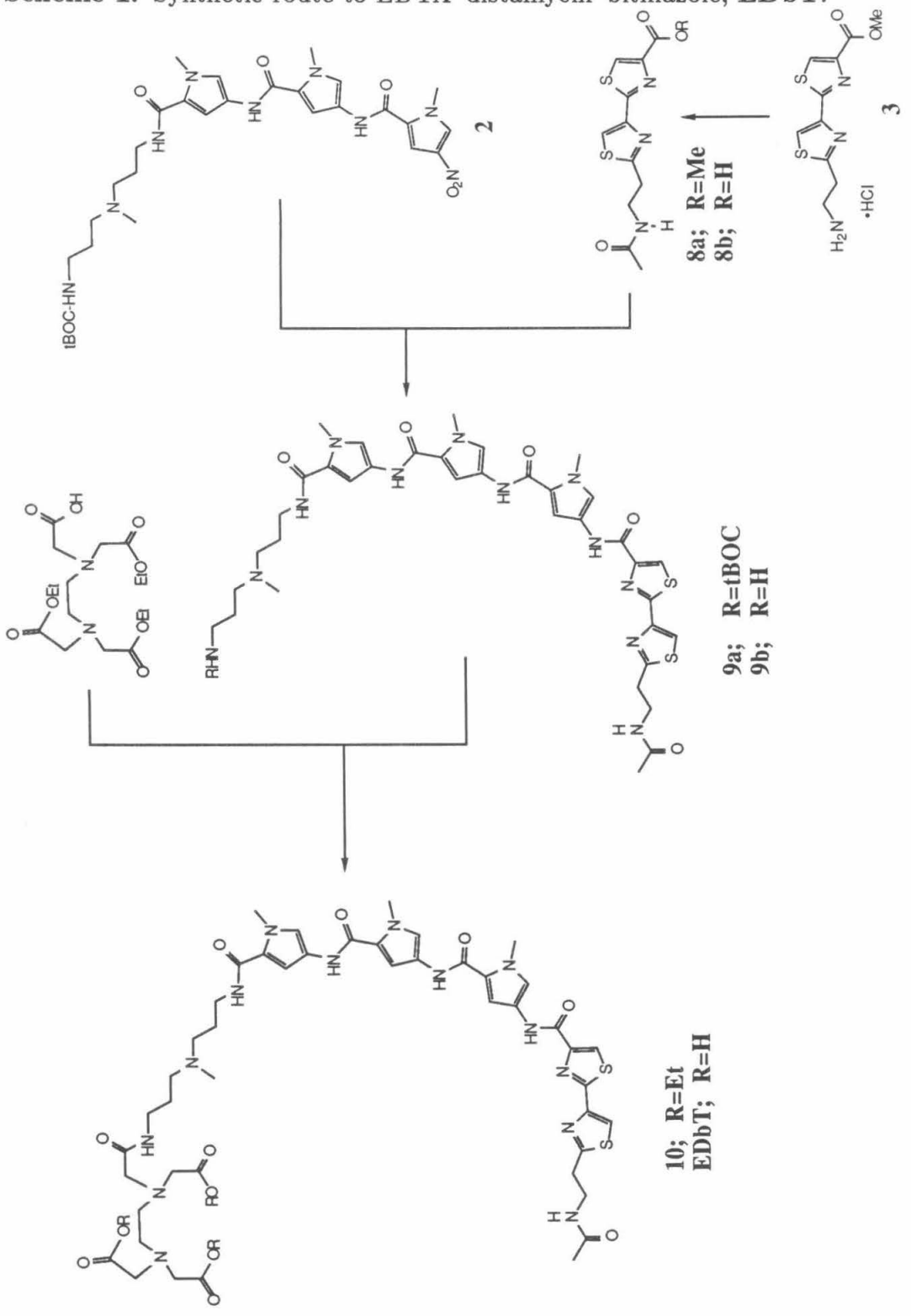
the bromoacetyl–tripyrrolecarboxamide of Baker and Dervan.[30] Bromoacetyl compounds alkylate the N3 of adenines (minor groove) and the N7 of guanines (major groove), which labilizes the nucleoside and cause DNA strand scission when heated in basic solution. The bromoacetyl function replaces the metal chelating and sugar domains of **BLM** (**BrbT**, figure 3). If the bithiazole unit is competent at sequence specific DNA binding, the bromoacetyl induced strand scission should provide information on both binding site location and groove identity.

A fourth compound was also prepared to examine the possibility that a bithiazole dimer may be more specific, and have a higher binding affinity, than the monomer. To investigate this possibility, the amino terminus of two bithiazole units was dimerized with a succinic acid linker to form **(bT)$_2$succinate**. The dimer retained both carboxy-terminal positive charges of **BLM**.

Synthesis of EDbT

The synthetic route to **EDbT** is shown in Scheme 1. The bithiazole amino acid methyl ester was prepared in nine steps using the published procedures from the **BLM** total synthesis.[31–35] The preparation of the tripyrrolecarboxamide starting material **2** was described in Chapter II of this report. The bithiazole methyl ester was acetylated with acetic anhydride (**8a**, 92%), followed by saponification to give the amino capped bithiazole acid **8b** in 87% yield. The tBOC-linker-nitropyrrole carboxamide, **2**, was hydrogenated to the aminopyrrole and coupled to N-hydroxybenzotriazole plus dicyclohexylcarbodiimide activated **8b** to afford **9a** in 70% yield. The tBOC protecting group was removed (**9b**, 93%), and EDTA triethyl ester attached via N-hydroxybenzotriazole plus dicyclohexylcarbodiimide mediated coupling to afford the triethyl ester of **EDbT** (**10**, 67%). Saponification of **10** produced **EDbT** in 76% yield.

Scheme 1: Synthetic route to EDTA–distamycin–bithiazole, **EDbT**.



Synthesis of bTDE

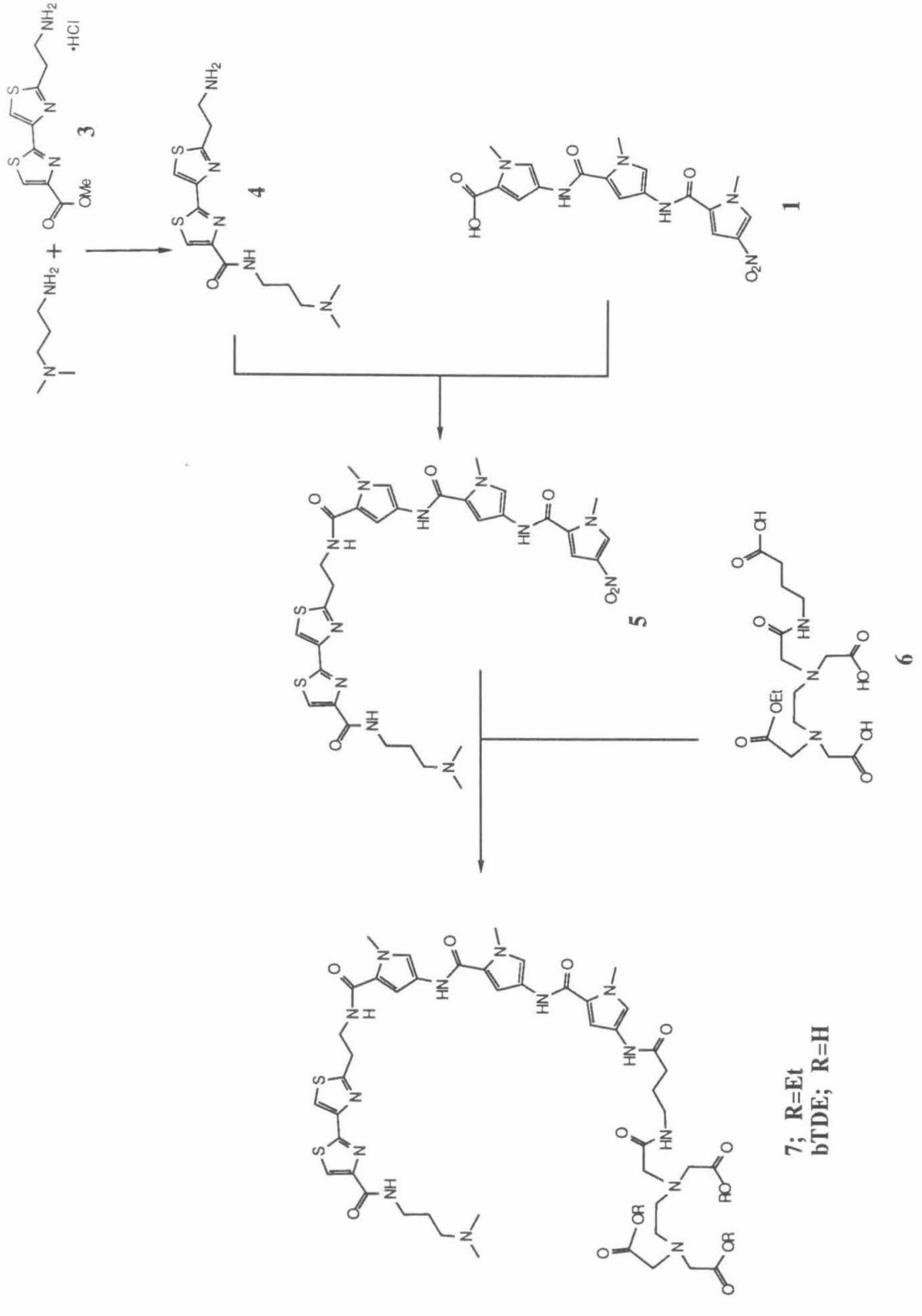
The synthetic route to **bTDE** is shown in Scheme 2. Heating the bithiazole methyl ester[31-33] with excess N,N-dimethylaminopropyldiamine produced the amino capped bithiazole **4** in 91% yield. N-hydroxybenzotriazole plus dicyclohexylcarbodiimide mediated coupling of **4** with nitrotripyrrolecarboxamide acid **1** afforded **5** in 61% yield. Catalytic hydrogenation of **5**, followed by reaction with acydiimidazole activated EDTA triethyl ester γ-aminobutyric acid amide (**6**), produced the triethyl ester of **bTDE** (**7**) in 17% yield. Saponification of **7** gave **bTDE** in 80% yield.

Synthesis of BrbT and $(bT)_2$succinate

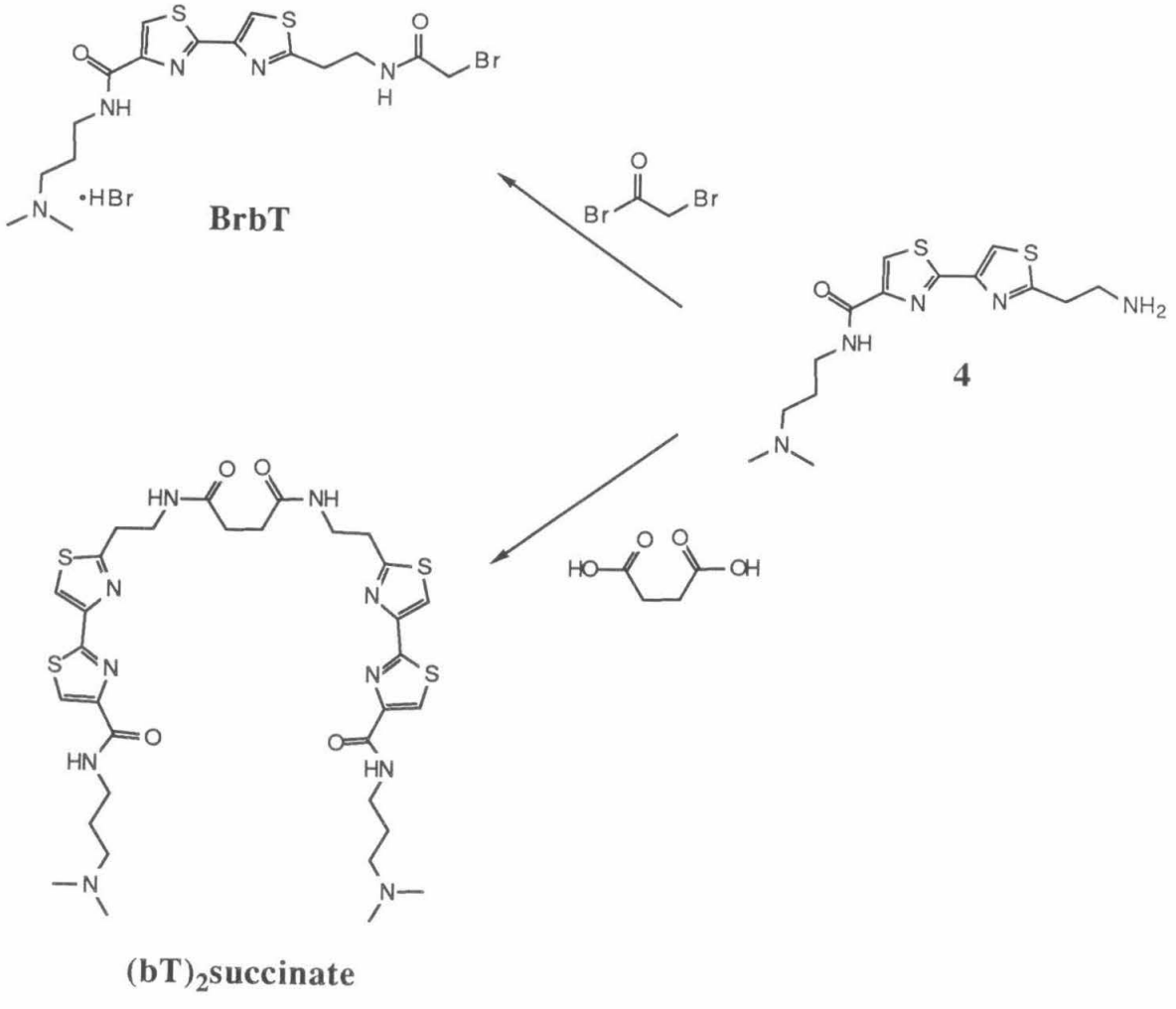
The synthetic routes to **BrbT** and **$(bT)_2$succinate** are shown in Scheme 3. The amino-capped bithiazole **4** was alkylated with bromacetyl bromide to afford **BrbT**•HBr in 70% yield.

Two equivalences of amino-capped bithiazole **4** were coupled to N-hydroxybenzotriazole plus dicyclohexylcarbodiimide activated succinic acid to afford **$(bT)_2$succinate** in 48% yield.

Scheme 2: Synthetic route to bithiazole–distamycin–EDTA, **bTDE**.



Scheme 3: Synthetic route to bromoacetyl–bithiazole, **BrbT**, and the dimer **$(bT)_2$succinate**.



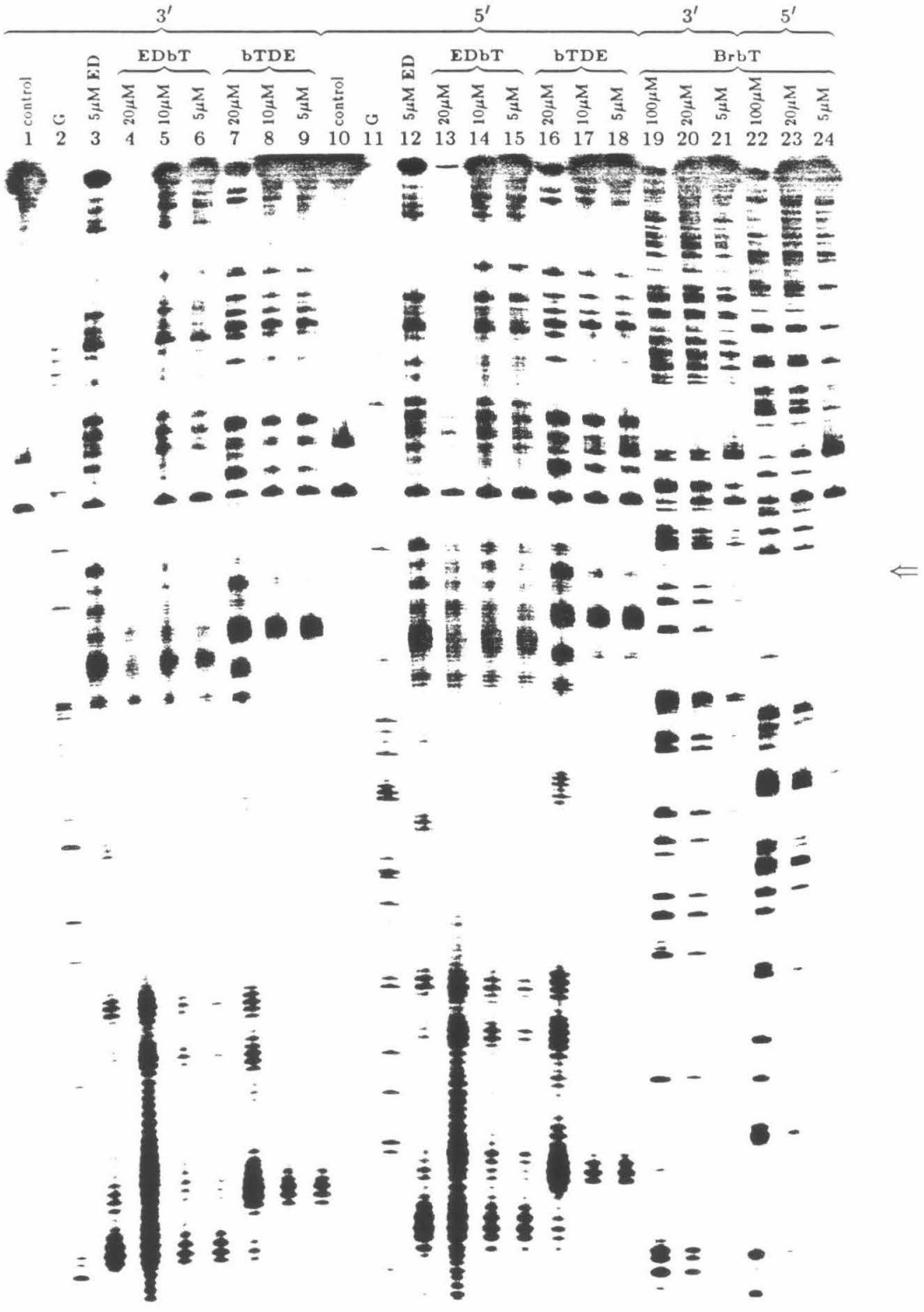
DNA Cleavage with EDbT•Fe and bTDE•Fe

The gel autoradiograms for **EDbT**•Fe and **bTDE**•Fe induced DNA cleavage on the 167 and 517 base pair fragments from pBR322 are shown in figures 4 and 6, respectively. For comparison EDTA-distamycin (**ED**) is included in both gels. Histograms for these cleavage reactions are shown in figures 5 and 7.

It is clear from the histograms that both **EDbT**•Fe and **bTDE**•Fe bind to the same DNA sequences as **ED**•Fe. The only significant difference observed between the bithiazole containing compounds and **ED** is the apparent binding site size as defined by the cleavage patterns. **ED**•Fe binds to five base pair A+T sequences, whereas both **EDbT** and **bTDE** apparently bind to six base pair sequences. Inspection of the binding site sequences does not appear to indicate altered specificity for the bithiazole containing compounds relative to **ED**. In particular there is no evidence for preferred binding at the type of DNA sequence for which the hybrid molecules were designed, i.e. 5′-GC$(A+T)_5$ or 5′-$(A+T)_5$GC. Furthermore, both **bTDE**•Fe and **EDbT**•Fe appear somewhat less efficient at DNA cleavage than **ED**•Fe.

The cause of the increased apparent binding site size is unclear. If the bithiazoles were not contributing to DNA binding, both **EDbT** and **bTDE** should have the same binding site size as **ED**. The observed increase in site size from 5 to 6 base pairs may reflect attempts by the hybrids to center themselves in A+T tracks of DNA in order to avoid steric clashes with the guanine 2-amino group. If this is the case, the hybrids are behaving somewhat like a higher homolog of **DST**. However the *decrease* in cleavage efficiency indicates that no significant specific interactions between the bithiazole and DNA occur.

Figure 4: Autoradiogram of the DNA cleavage products on the 517bp fragment from pBR322. Reaction conditions were 18,000cpm ^{32}P end-labelled DNA, 100μM bp calf thymus DNA, 40mM TRIS–acetate pH 7.9. The EDTA•Fe chelates were equilbrated with the DNA for 30 minutes at room temperature prior to the addition of DTT (5mM final concentration). EDTA•Fe cleavage reactions were run for one hour at room temperature. **BrbT** reactions were run for two hours at 45°C. All reactions were terminated by ethanol precipitation and analyzed on an 8%, 1:20 crosslinked, 50% urea denaturing polyacrylamide wedge gel. The gel was 0.2mm thick at the top and 0.4mm thick at the bottom. Lanes 1–9 and 19–21 contain 3′ labelled and lanes 10–18 and 22–24 5′ labelled restriction fragment. Lanes 1 and 10 are DNA controls and 2 and 11 Maxam-Gilbert G reactions. The remaining lanes contain **ED**•Fe, **EDbT**•Fe, **bTDE**•Fe and **BrbT** as indicated on the gel.


3′
5′
3′
5′
EDbT
bTDE
EDbT
bTDE
BrbT
control
G
5μM ED
20μM
10μM
5μM
control
G
5μM ED
20μM
10μM
5μM
20μM
10μM
5μM
100μM
20μM
5μM
100μM
20μM
5μM
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

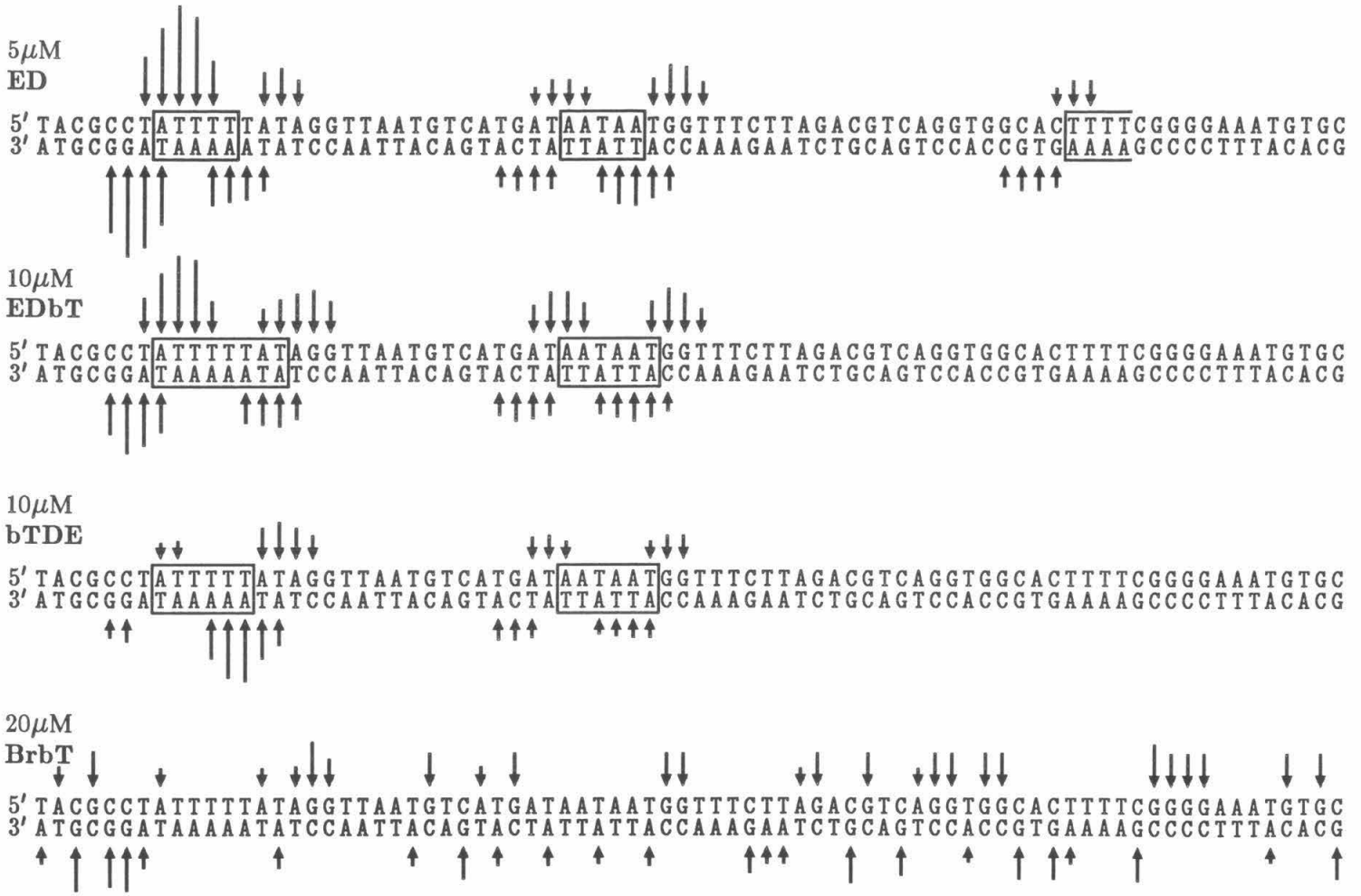
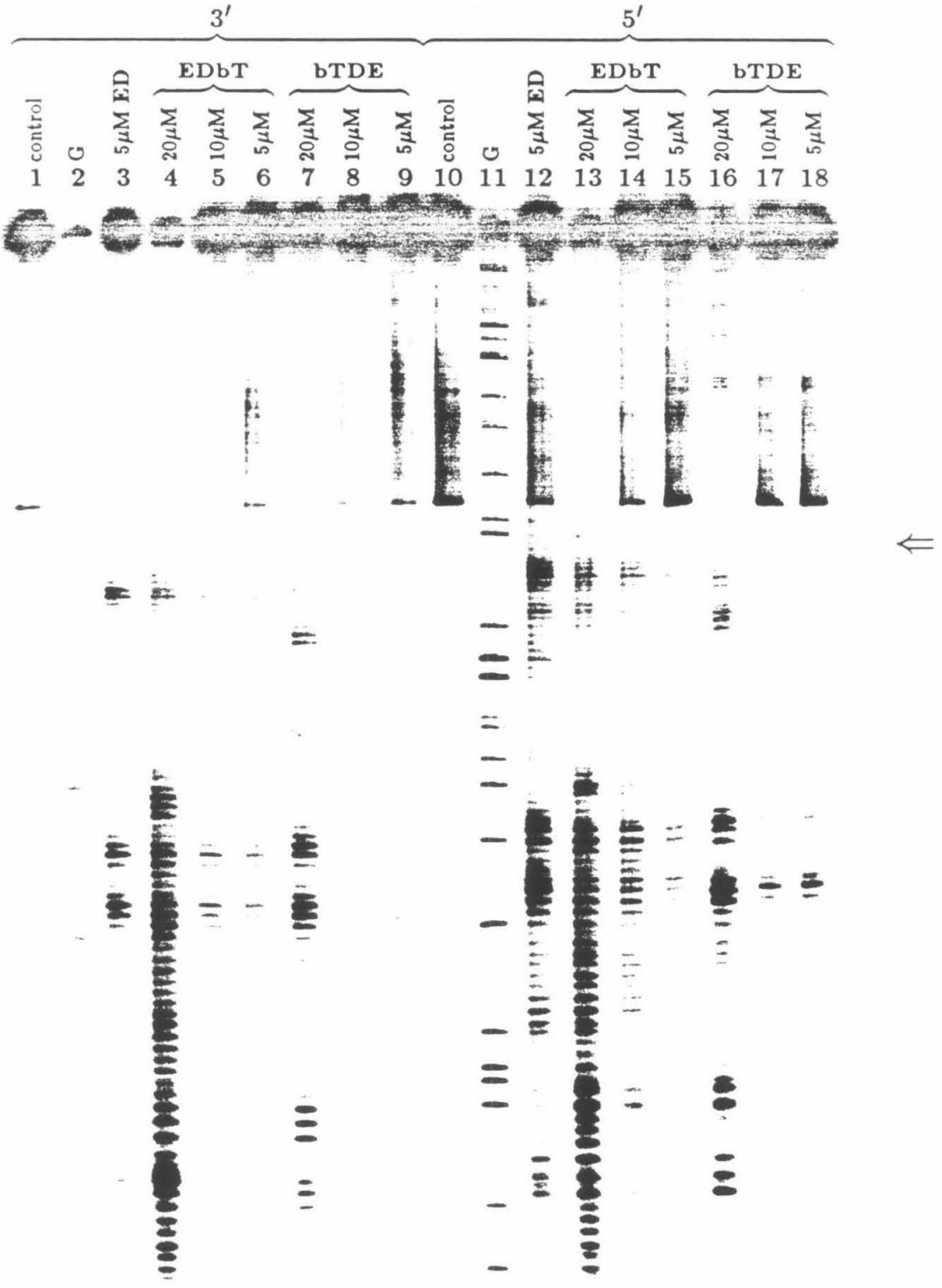


Figure 5: Histograms of the DNA cleavage on the 517bp (base pairs 3848 to 4362) restriction fragment from pBR322 from figure 4. The height of the arrows indicate extent of cleavage resulting in removal of the indicated base. Boxes define binding site location and size as described in Chapter II.

Figure 6: Autoradiogram of the DNA cleavage products on the 167bp fragment from pBR322. Reaction conditions were 18,000cpm ^{32}P end-labelled DNA, 100μM bp calf thymus DNA, 40mM TRIS–acetate pH 7.9. The EDTA•Fe chelates were equilbrated with the DNA for 30 minutes at room temperature prior to the addition of DTT (5mM final concentration). EDTA•Fe cleavage reactions were run for one hour at room temperature. All reactions were terminated by ethanol precipitation and analyzed on an 8%, 1:20 crosslinked, 50% urea denaturing polyacrylamide wedge gel. The gel was 0.2mm thick at the top and 0.4mm thick at the bottom. Lanes 1–9 contain 3′ labelled and lanes 10–18 5′ labelled restriction fragment. Lanes 1 and 10 are DNA controls and 2 and 11 Maxam-Gilbert G reactions. The remaining lanes contain **ED**•Fe, **EDbT**•Fe and **bTDE**•Fe as indicated on the gel.


3′
5′
EDbT
bTDE
EDbT
bTDE
control
G
5μM ED
20μM
10μM
5μM
control
G
5μM ED
20μM
10μM
5μM
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

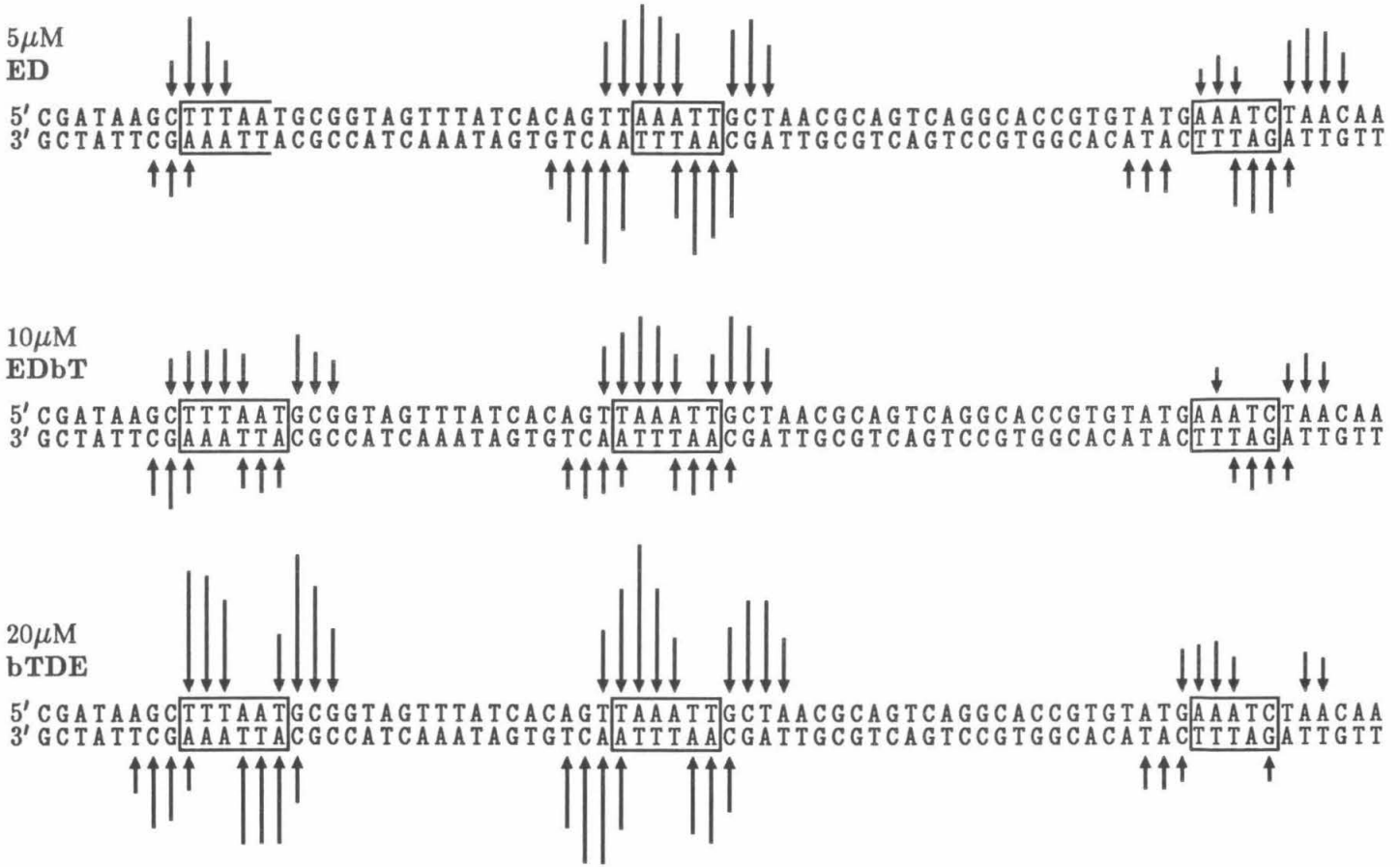


Figure 7: Histograms of the DNA cleavage on the 167bp restriction fragment from pBR322 from figure 6. The height of the arrows indicate extent of cleavage resulting in removal of the indicated base. Boxes define binding site location and size as described in Chapter II.

DNA Cleavage with BrbT

The ability of **BrbT** to cleave DNA via specific alkylation of adenines and guanines was investigated by incubation of **BrbT** with ^{32}P end-labelled DNA. It was found that reaction in sodium cacodylate or phosphate buffer, followed by heating in 10% aqueous piperidine, caused DNA cleavage at all guanine and some adenine residues. Additionaly, reaction in TRIS buffer at 45°C gave the same DNA cleavage pattern without the piperidine work-up. **BrbT** mediated DNA cleavage and cleavage histograms are shown in figures 4 and 5, respectively. **BrbT** gives strong cleavage at all G's and weakly cleaves a subset of the A's. If it is assumed that **BrbT** alkylates (cleaves) G's in the major groove and A's in the minor groove, as does dimethyl sulfate, then the observed cleavage pattern indicates that the bithiazole is not preferentially bound to either groove of the DNA. Cleavage at all G's could result from either specific, and equivalent, binding to, and alkylation of, all G's, or, alternatively, non-specific binding and *alkylation* of only G's. On the other hand, cleavage at only certain A's indicates specific binding of **BrbT** which results in enhanced cleavage of A's in/near the binding site. Inspection of the histogram in figure 5 allows several conclusions about the recognition sequence. First, A's in tracks of three or more A's, i.e. 5′-AAA or 5′-AAAA, are not cleaved, and double A's are only sometimes cleaved. Analysis of the flanking sequences around both the cleaved and uncleaved adenines indicates that cleavage most often occurs at 5′-pyrimidine-adenine sequences.

DNA binding by $(bT)_2$succinate

The sequence specific DNA binding of **$(bT)_2$succinate** was investigated by both MPE•Fe and Dnase I footprinting on the 167 and 517 base pair fragments from pBR322. **$(bT)_2$succinate** at 25, 50 and 100μM (100μM bp DNA) gave no discernable footprints with either reagent.

Discussion

The DNA cleavage with **BrbT** indicates that though the bithiazole does appear to have some sequence specific DNA recognition ability, it is not very great. Furthermore, the absence of significant differences in sequence specificity for the hybrid molecules relative to **ED**, as well as the lack of detectable binding by the bithiazole dimer, both strongly suggest that the bithiazole is not a DNA minor groove binding function in the same sense as is tripyrrolecarboxamide.

Conclusion

Hybrids of the DNA binding domains of Bleomycin and distamycin have been described. The sequence specific DNA binding abilities of these hybrids have been investigated by affinity cleaving methods which indicate that either the bithiazole group of bleomycin does not bind in the minor groove of DNA, or, alternatively, its binding mode is not compatible, and substantially weaker, than the tripyrrolecarboxamide of distamycin. The DNA binding ability of bithiazole itself was examined by including a bromoacetyl function which causes DNA strand scission via specific alkylation of adenine and guanine residues. The DNA cleaving ability of the bromoacetyl bithiazole indicates that the bithiazole has only modest preference for

5′-pyrymidine-adenine sequences in the minor groove and readily alkylates all guanines, presumably from the major groove. A bithiazole dimer was also investigated by footprinting methods, but no specific DNA binding was observed.

Experimental

See Chapter II for general procedures, the synthesis of nitrotripyrrolecarboxamide **1** and **7**, as well as the procedures used for the preparation of ^{32}P end-labelled DNA restriction fragments.

4: A solution of 140mg (0.52mmol) **3**[31–35] in 5ml (~50mmol, 100eq) N,N-dimethyldipropylamine was heated at 80°C for 15hrs, followed by removal of excess amine under reduced pressure. Chromatography with 5% concentrated aqueous ammonium hydroxide in methanol afforded 160mg (91%) **4** as a colorless liquid. TLC (4% NH_4OH/MeOH) R_f=0.27. ^1H–NMR ($CDCl_3$) δ 8.29 (t, j=6Hz, 1H), 8.11 (s, 1H), 7.85 (s, 1H), 3.53 (q, j=6Hz, 2H), 3.43 (s, 2H), 3.18 (s, 2H), 2.70 (bs, 2H), 2.43 (t, j=7Hz, 2H), 2.27 (s, 6H), 1.80 (p, j=7Hz, 2H).

5: A solution of 100mg (295μmol) **4**, 134mg (325μmol, 1.1eq) **1**, 99mg (650μmol, 2.2eq) N-hydroxybenzotriazole monohydrate and 67mg (325μmol, 1.1eq) dicyclohexylcarbodiimide in 3ml DMF was stirred overnight under a dry atmosphere. The DMF was removed under reduced pressure and the crude product triturated with diethyl ether (3x). Chromatography with 1:1 methanol–dichloromethane afforded 134mg (61%) **5** as a yellow solid. TLC (4% NH_4OH/dichloromethane) R_f=0.17. ^1H–NMR (D_6–DMSO) δ 10.29 (s, 1H), 9.97 (s, 1H), 8.65 (t, j=6Hz, 1H), 8.24 (m, 2H), 8.19 (d, j=2Hz, 1H), 8.15 (d, j=2Hz, 1H), 7.60 (d, j=2Hz, 1H), 7.28 (d, j=2Hz, 1H), 7.22 (d, j=2Hz, 1H), 7.04 (s, 1H), 6.89 (d, j=2Hz, 1H), 3.97 (s, 3H), 3.86 (s, 3H), 3.80 (s, 3H), 3.59 (q, j=6Hz, 2H), 3.32 (m, 4H), 2.28 (t, j=7Hz, 2H), 2.16 (s, 6H), 1.67 (p, j=7Hz, 2H).

EDTA triethyl ester γ-aminobutyric acid amide, 6: To a solution of 5g (13mmol) EDTA triethyl ester[36] and 1.52g (13mmol) N-hydroxysuccinimide

in 100ml dioxane was added, with stirring, 2.7g (13mmol) dicyclohexycarbodiimide in 20ml dioxane. The solution was stirred for 12 hours at room temperature, filtered and the filtrate concentrated. The filtrate was then dissolved in 100ml dimethoxyethane and added with stirring to a solution of 2g (20mmol) γ-aminobutyric acid and 1.68g (20mmol) sodium bicarbonate in 100ml water. After stirring 12 hours, the solution was concentrated under vacuum and chromatographed with 10% methanol in dichloromethane to give 4g (65%) **6** as a yellow glass. ^{1}H–NMR (D_6–DMSO) δ 8.00 (t, j=5Hz, 1H), 4.08 (m, 6H), 3.53 (s, 4H), 3.45 (s, 2H), 3.19 (s, 2H), 3.10 (m, 2H), 2.70 (t, j=6Hz, 2H), 2.20 (t, j=6Hz, 2H), 1.63 (m, 2H), 1.19 (t, j=7Hz, 9H).

bTDE triethyl ester, 7: A solution of 100mg (136μmol) **5** and 20mg 5% palladium on carbon in 2ml DMF was hydrogenated under one atmosphere hydrogen overnight in the dark. The hydrogenation mixture was filtered through celite and the celite washed with 2ml DMF. To the resulting yellow solution was added 69mg (150μmol, 1.1eq) **6** and 31mg (190μmol, 1.4eq) acyldiimidazole. After stirring 15 hours, the DMF was removed under reduced pressure and the crude product triturated with diethyl ether (3x). Chromatography with 0.75% concentrated aqueous ammonium hydroxide in methanol afforded 27mg (17%) **7** as a yellow solid. ^{1}H–NMR (D_6–DMSO) δ 9.91 (s, 1H), 9.90 (s, 1H), 9.81 (s, 1H), 8.64 (t, j=7Hz, 1H), 8.31 (s, 1H), 8.24 (s, 1H), 8.14 (s, 1H), 7.23 (d, 1H), 7.20 (d, 1H), 7.16 (d, j=2Hz, 1H), 7.05 (d, j=2Hz, 1H), 6.90 (d, j=2Hz, 1H), 6.88 (d, j=2Hz, 1H), 4.10 (m, 6H), 3.85 (s, 3H), 3.83 (s, 3H), 3.80 (s, 3H), 3.60 (m, 6H), 3.50 (m, 4H), 3.20 (m, 4H), 2.75 (m, 4H), 2.30 (m, 6H), 2.22 (m, 4H), 1.70 (m, 6H), 1.20 (m, 9H).

bTDE: To a suspension of 22mg (19μmol) **7** in 0.5ml ethanol plus 0.25ml water was added 0.25ml 1N lithium hyroxide. After stirring for four hours in the dark,

the solvents were removed under reduced pressure at room temperature. The crude product was redissolved in 2ml water, filtered, frozen and lyophilized. Chromatography with 2% concentrated aqueous ammonium hydroxide in methanol afforded 16mg (80%) **bTDE** as a yellow solid. UV/VIS (H_2O) 212nm, 245nm, 298nm, (ϵ=46,000M^{-1}). ^{1}H–NMR (1:10 trifluroacetic acid – D_6–DMSO) δ 9.94 (s, 1H), 9.93 (s, 1H), 9.86 (s, 1H), 9.36 (bs, 2H), 8.66 (t, j=6Hz, 1H), 8.49 (t, j=6Hz, 1H), 8.30 (s, 1H), 8.28 (bs, 1H), 8.16 (s, 1H), 7.24 (d, j=2Hz, 1H), 7.21 (d, j=2Hz, 1H), 7.19 (d, j=2Hz, 1H), 7.10 (d, j=2Hz, 1H), 6.96 (d, j=2Hz, 1H), 6.91 (d, j=2Hz, 1H), 4.12 (s, 2H), 3.99 (bs, 4H), 3.90 (s, 3H), 3.86 (s, 3H), 3.83 (s, 3H), 3.63 (bs, 4H), 3.40 (m, 4H), 3.30 (m, 2H), 3.20 (m, 2H), 3.16 (m, 2H), 3.00 (m, 2H), 2.82 (m, 5H), 2.32 (t, j=7Hz, 2H), 1.96 (p, j=7Hz, 2H), 1.79 (p, j=7Hz, 2H),

8a: To a solution of 80mg (300μmol) **3** and 200μl (1.4mmol) triethylamine in 2ml DMF was added 72μl (760μmol, 2.6eq) acetic anhydride. After stirring for 14 hours at room temperature, the DMF was removed under vacuum at room temperature. Chromatography with 8% methanol in dichloromethane afforded 85mg (92%) **8a** as a white solid.

8b: To a solution of 81mg (270μmol) **8a** in 2ml 50% aqueous ethanol was added 100mg (2.5mmol) sodium hydroxide. The resulting solution was refluxed 2.5 hours, followed by removal of the ethanol under reduced pressure. The resulting solution was acidified with 1N HCl and the water removed under vacuum. The crude product was dissolved in 15ml hot ethanol, cooled to room temperature and NaCl removed by filtration. Concentration of the supernatants afforded 67mg (87%) **8b.**

9a: A solution of 174mg (270μmol) **2** and 25mg 5% palladium on carbon in 5ml DMF was hydrogenated under one atmosphere hydrogen for 15 hours in the dark.

The catalyst was removed by filtration through celite and to the reulting solution was added 67mg (230μmol) **8b**, 76mg (500μmol) N-hydrozybenzotriazole and 56mg (271μmol) dicyclohexylcarbodiimide. After stirring 14 hours in the dark, the DMF was removed under vacuum at room temperature and the crude product triturated with diethyl ether (3x). Chromatography with 0.5% concentrated aqueous ammonium hydroxide in methanol afforded 140mg (70%) **9a** as a yellow solid.

9b: To a solution of 120mg (135μmol) **9a** in 3ml dichloromethane was added 2ml trifluoroacetic acid. After stirring 10 minutes, the solvents were removed under reduced pressure. Chromatography with 12% concentrated aqueous ammonium hydroxide in methanol afforded 99mg (93%) **9b** as a yellow solid.

EDbT triethyl ester, 10: A solution of 90mg (114μmol) **9b**, 47mg (125μmol) EDTA triethyl ester[36], 38mg (250μmol) N-hydroxybenzotriazole and 26mg (125-μmol) dicyclohexylcarbodiimide in 5ml DMF was stirred 14 hours in the dark. Dicyclohexylurea was removed by filtration and the DMF under reduced pressure. The crude product was triturated with diethyl ether (3x), followed by chromatography with 1% concentrated aqueous ammonium hydroxide in methanol to give 89mg (67%) **10** as a yellow solid.

EDbT: To a solution of 60mg (52μmol) **10** in 1ml ethanol was added 1ml of 0.5N lithium hydroxide. After stirring for four hours at room temperature, the solvents were removed under reduced pressure. Chromatography with 2% concentrated aqueous ammonium hydroxide in methanol afforded 42mg (76%) **EDbT** as a yellow solid. The 500MHz ^{1}H–NMR was consistent with the assigned structure. UV/VIS (H_2O) 220nm, 248nm (ϵ=34,000M^{-1}), 304nm (ϵ=45,000M^{-1}).

BrbT: To a solution of 53mg (160μmol) **4** in 10ml dichloromethane was added 15μl (1.05eq) bromoacetyl bromide. After stirring at room temperature for 30 minutes, the solvents were removed under reduced pressure at room temperature. The crude product was chromatographed with 2% aqueous ammonium hydroxide in methanol. The combined fractions (40ml) were concentrated to 10ml and hydrogen bromide was bubbled through the solution, followed by final drying under reduced pressure to afford 62mg (70%) **BrbT** as a white solid. ^{1}H–NMR (D_6–DMSO) δ 8.91 (bs, 1H), 8.68 (bs, 1H), 8.31 (s, 1H), 8.18 (s, 1H), 5.76 (s, 2H), 4.12 (m, 2H), 3.54 (m, 2H), 3.32 (d, j=7Hz, 6H), 3.19 (m, 4H), 2.02 (m, 2H).

(bT)$_2$succinate: To a cold (0°C) solution of 50mg (147μmol) **4**, 47mg (310μmol) N-hydroxybenzotriazole and 8.1mg (70μmol) succinic acid in 5ml DMF was added 31mg (150μmol) dicyclohexylcarbodiimide. After stirring 14 hours at room temperature, the DMF was removed under reduced pressure and the crude product triturated with diethyl ether (3x). Chromatography with 4% concentrated aqueous ammonium hydroxide in methanol gave 25mg (48%) **(bT)$_2$succinate** as a white solid. The 400MHz ^{1}H–NMR was consistent with the assigned structure. UV/VIS (H_2O) 285nm (ϵ=28,000M^{-1}), 215nm.

References and Notes

1. H. Umezawa, K. Maeda, T. Takeuchi, Y. Okami, *J. Antibiot., Ser. A.*, **19**, 200 (1966).

2. H. Umezawa in *"Bleomycin: Current Status and New Developments"*, S. Curter, S. Crooke, H. Umezawa, Eds., Academic Press, New York, 1978, p.15.

3. H. Umezawa, *Pure Appl. Chem.*, **28**, 665 (1971).

4. H. Umezawa, *Biomedicine*, **18**, 459 (1973).

5. J. Stubbe, J. Kozarich, *Chem. Rev.*, **87**, 1107 (1987).

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

7. M. Takeshita, A. Grollman, E. Ohtsubo, H. Ohtsubo, *Proc. Natl. Acad. Sci., USA*, **75**, 7983 (1978).

8. A. D'Andrea, W. Haseltine, *Proc. Natl. Acad. Sci., USA*, **75**, 3608 (1978).

9. R. Burger, J. Peisach, S. Horowitz, *J. Biol. Chem.*, **256**, 11636 (1981).

10. J. Giloni, M. Takeshita, F. Johnson, A. Grollman, *J. Biol. Chem.*, **256**, 8608 (1981).

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

12. Y. Aoyage, H. Suguna, K. Katano, J. Primeau, L. Chang, S. Hecht, *J. Am. Chem. Soc.*, **104**, 5537 (1982).

13. L. DeReimer, C. Meares, D. Goodwin, C. Diamanti, *J. Med. Chem.*, **22**, 1619 (1979).

14. J. Kross, D. Henner, W. Haseltine, L. Rodriguez, M. Levin, S. Hecht, *Biochem.*, **21**, 3711 (1982).

15. L. Fisher, R. Kuroda, T. Sakai, *Biochem*, **24**, 3199 (1985).

16. C. Mirabelli, A. Ting, C. Huang, S. Mong, S. Crooke, *Cancer Res.*, **42**, 2779 (1982).

17. R. Burger, A. Berkowitz, J. Peisach, S. Horowitz, *J. Biol. Chem.*, **255**, 11832 (1980).

18. C. Haidle, K. Weiss, M. Kuo, *Mol. Pharmacol.*, **8**, 531 (1972).

19. R. Burger, J. Peisach, W. Blumberg, S. Horowitz, *J. Biol. Chem.*, **254**, 10906 (1979).

20. J. Lown, A. Joashua, *Biochem. Pharmacol.*, **29**, 251 (1980).

21. W. Oberley, G. Buether, *FEBS Lett.*, **97**, 47 (1979).

22. M. Chen, A. Grollman, S. Horowitz, *Biochem.*, **16**, 3641 (1977).

23. L. Povirk, M. Hogan, N. Dattagupta, *Biochem.*, **18**, 96 (1979).

24. R. Mayer, L. Que, *J. Biol. Chem.*, **259**, 10356 (1984).

25. T. Booth, T. Sakai, J. Glickson, *J. Biol. Chem.*, **258**, 4211 (1983).

26. R. Dickerson in *"Mechanism of DNA Damage and Repair: Implications for Carinogenesis and Risk Assesment in Basic Life Sciences,"* M. Smi, L. Grossman, Eds., Plenum, New York, vol. 38, 1986, p. 245.

27. J. Kross, W. Henner, S. Hecht, W. Haseltine, *Biochem.*, **21**, 4310 (1982).

28. C. Huang, L. Galuan, S. Crooke, *Biochem.*, **19**, 1761 (1980).

29. P. Schultz, P. Dervan, unpublished results.

30. B. Baker, P. Dervan, *J. Am. Chem. Soc.*, **107**, 8266 (1985).

31. K. Zee-Cheng, C. Cheng, *J. Heterocyclic Chem.*, **78**, 1439 (1970).

32. T. Takita *et al.*, *Tetrahedron Lett.*, **22**, 671 (1981).

33. J. Riordan, T. Sakai, *J. Heterocyclic Chem.*, **18**, 1213 (1981).

34. T. Sakai, J. Riordan, T. Booth, J. Glickson, *J. Med. Chem.*, **24**, 279 (1981).

35. E. Gilbert, E. Rumanowski, P. Newallis, *J. Chem. Eng. Data*, **13**, 130 (1968).

36. R. Hay, K. Nolan, *J. Chem. Soc., Dalton*, 1348 (1975).

CHAPTER IV:

Studies In Metalloporphyrin Mediated Affinity Cleaving

The general strategy of covalently linking a DNA-binding small molecule to a function capable of cleaving duplex DNA, i.e. "affinity cleaving," has been exploited by several groups including our own. Affinity cleaving reagents can be separated into two general classes; sequence specific and non-sequence specific. One of several examples of a non-sequence specific affinity cleaving molecule is methidiumpropyl-EDTA-Fe(II), **MPE**•Fe.[1,2] **MPE**•Fe consists of the DNA binding moiety methidium[3-6] covalently linked to an EDTA•Fe chelate known to cause DNA strand scission in the presence of a reducing agent and molecular oxygen.[1,2]

There are also examples of sequence specific affinity cleaving molecules. This class of compounds is perhaps best exemplified by **DE**•Fe(II), which consists of the sequence specific minor groove binding N-methylpyrrole tripeptide covalently linked to EDTA•Fe(II).[7] The DNA binding portions of both **MPE** and **DE** are modeled after compounds known to have high affinity for double stranded DNA. **MPE** is derived from ethidium bromide and **ED** from distamycin A, a naturally occurring antibiotic.

The mode of action of a different type of cleaving moiety, metalloporphyrins, has not been reported in detail. Several porphyrin containing affinity cleaving

molecules have appeared in the literature, but, to date, none have been sequence specific. We have undertaken the design, synthesis and characterization of a sequence specific affinity cleaving compound which contains a metalloporphyrin as the reactive moiety; porphyrin-distamycin–Fe(II), (**PD**•Fe(II)). Before examining **PD**•Fe we shall first look at some of the oxidative chemistry of metalloporphrins in general, as well as their interactions with double stranded DNA.

Metalloporphyrins as Oxidation Catalysts

Oxidation reactions catalyzed by metalloporphrins have received considerable attention of late as potentially useful organic oxidation systems, and as models for the catalytic cycle of the mono-oxygenase cytochrome P-450.[8–12] Reactions studied have included unactivated C-H to C-OH and C-X and epoxidation of olefins.

From a conceptual viewpoint, the oxidation reactions can be divided into two limiting classes: the generation of a diffusable reactive species and the generation of a bound metal-oxygen species. In the first class, the oxidation products are consistent with hydroxy radical attack (and selectivity), and are independent of the method of generation. Fenton type reactions[10] are in this class, including EDTA•Fe(II), tetraphenylporphinato-Fe(II) or P-450 with hydroperoxides or peracids.[13]

$$\text{Fe(II)} + \text{H}_2\text{O}_2 \rightarrow \text{Fe(III)} + \text{HO}^- + \text{HO}\bullet$$

$$\text{HO}\bullet + \text{substrate} \rightarrow \text{products}$$

The second class of reactions is one in which the method of generation of the active species is reflected in the product(s). If the active oxidizing species is a metallo-oxygen compound, the observed products should be sensitive to the environment, i.e. the metal chelator, of the active species.

metal(ligands) + oxygen source → active metal-oxo species

active metal-oxo species + substrate → metal(ligands) + products

A number of metalloporphrins have been studied which appear to belong to the second class of reactions. Studies have been carried out on the effect of different metals (Fe, Mn, Cr *etc.*)[8,9] and oxygen sources (iodosobenzene[9] and p-cyano-N,N-dimethylaniline-N-oxide[6]) on epoxidation.

Metalloporphrin Mediated DNA Cleavage

It has been proposed that hemin (ferriprotoporphyrin IX chloride) is important in gene expression *in vivo*, where it is possible that some of its effects are due to nicking of the DNA.[14] It has also been observed that hemin, in the presence of a reducing agent and molecular oxygen, cleaves supercoiled plasmid DNA *in vitro*.[14] Various synthetic porphyrins, such as iron(III)-meso-tetra(N-methylpyridyl)porphines, are believed to interact with duplex DNA in several ways:

1. Intercalate into G•C regions.[15]
2. Outside bind in A•T regions.[15]
3. Cause strand scission in the presence of dithiothrietol and oxygen.[16,17]

Several affinity cleaving molecules which use an iron porphrin to generate DNA strand scission have been studied. Covalent attachment of an iron porphrin and a DNA intercalator produces a family of compounds which, in the presence of 2-mercaptoethanol or sodium bisulfite and oxygen, cleave plasmid DNA[17,18] (Table I). To date, the sequence specificity (if any) and cleavage products have not been reported for these compounds. It seems reasonable however, to expect that sequence specificity is small, and since the cofactors are similar (O_2+reducing agent), the products should be similar to those produced by **MPE**•Fe.

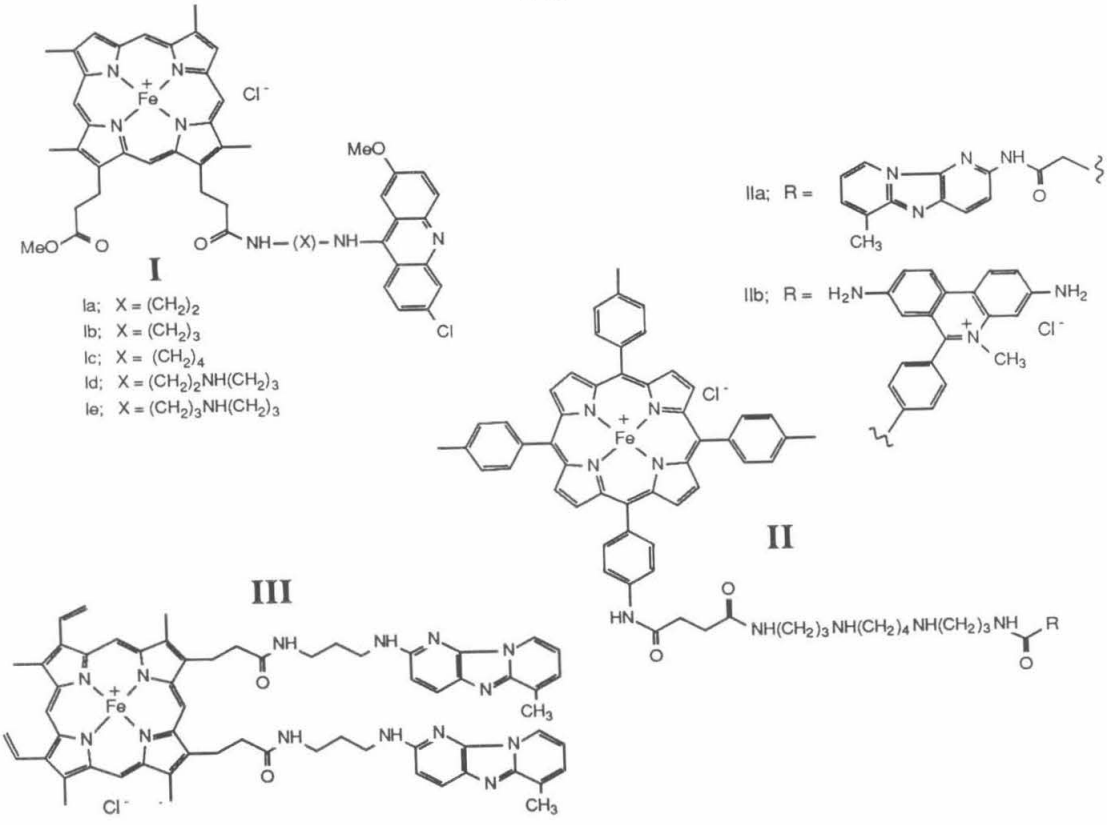


Table I: Cleavage of Closed Circular DNA by Iron Porphyrin Derivatives.

Small Molecule	Conc. (μM)	DNA (μM bp)	% Cleavage
Ia[a]	50	?	67
Ib	50	?	67
Ic	50	?	69
Id	50	?	70
Ie	25	?	88
IIa[b]	10	100	39
	100	100	45
IIb	50	100	76
	100	100	76
III[b]	100	100	52[c]

a. PM2-CCC-DNA, pH 7.1, 20mM 2-mercaptoethanol, 37°C, 20min.[18]

b. PM2-CCC-DNA, pH 7.8, 4mM $Na_2S_2O_4$, 50mM NaCl, 20°C, 1hr.[17]

c. Total recovery of DNA was low, reason unknown.

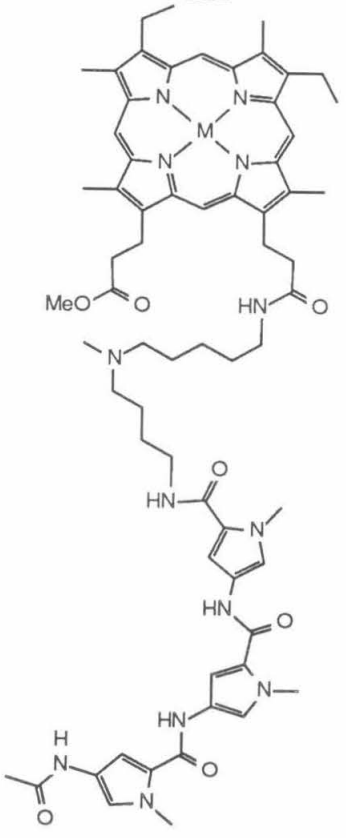
Porphyrin–Distamycin: PD

Since metalloporphrins are capable of generating strand scission in duplex DNA, we set out to design, synthesize and characterize a sequence specific affinity cleaving molecule with a metalloporphyrin as the reactive moiety. Sequence specific delivery of the metalloporphyrin should allow us to differentiate between a diffusable reactive species, such as hydroxy radical, and a metal-oxo mediated reaction where the binding site is better defined.

We choose a strategy similar to that used for **ED**; sequence specific minor groove binding by a moiety related to the natural product distamycin A, covalently linked to a function capable of causing strand scission of double stranded DNA. Porphyrin–distamycin, **PD**, uses mesoporphyrin IX as the iron chelator, which is linked through one of its own side chains to a short tether. This, in turn, is linked to the carboxy terminus of an N-methylpyrrole tripeptide, which is known to have a high affinity for A+T regions of duplex DNA (see chapter II for a more detailed description).

Synthesis

The synthesis of **PD** is given in Scheme I. Mesoporphyrin IX was choosen for two reasons; its high stability in solution (protoporphyrin IX, the natural product, is notoriously light sensitve) and ease of preparation from readily available hemin chloride. Catalytic hydrogenation of hemin chloride reduces the vinyl side chains and removes the iron, giving mesoporphyrin IX diacid.[19,20] Catalytic hydrogenation of **1** (see chapter II), followed by reaction with acetyl chloride gave **2** in 86% yield. Saponification of **2** (**3**, 85%), followed by diimidazole mediated coupling with N,N-diaminopropyl-N-methyl amine, afforded **4** in 31% yield. The



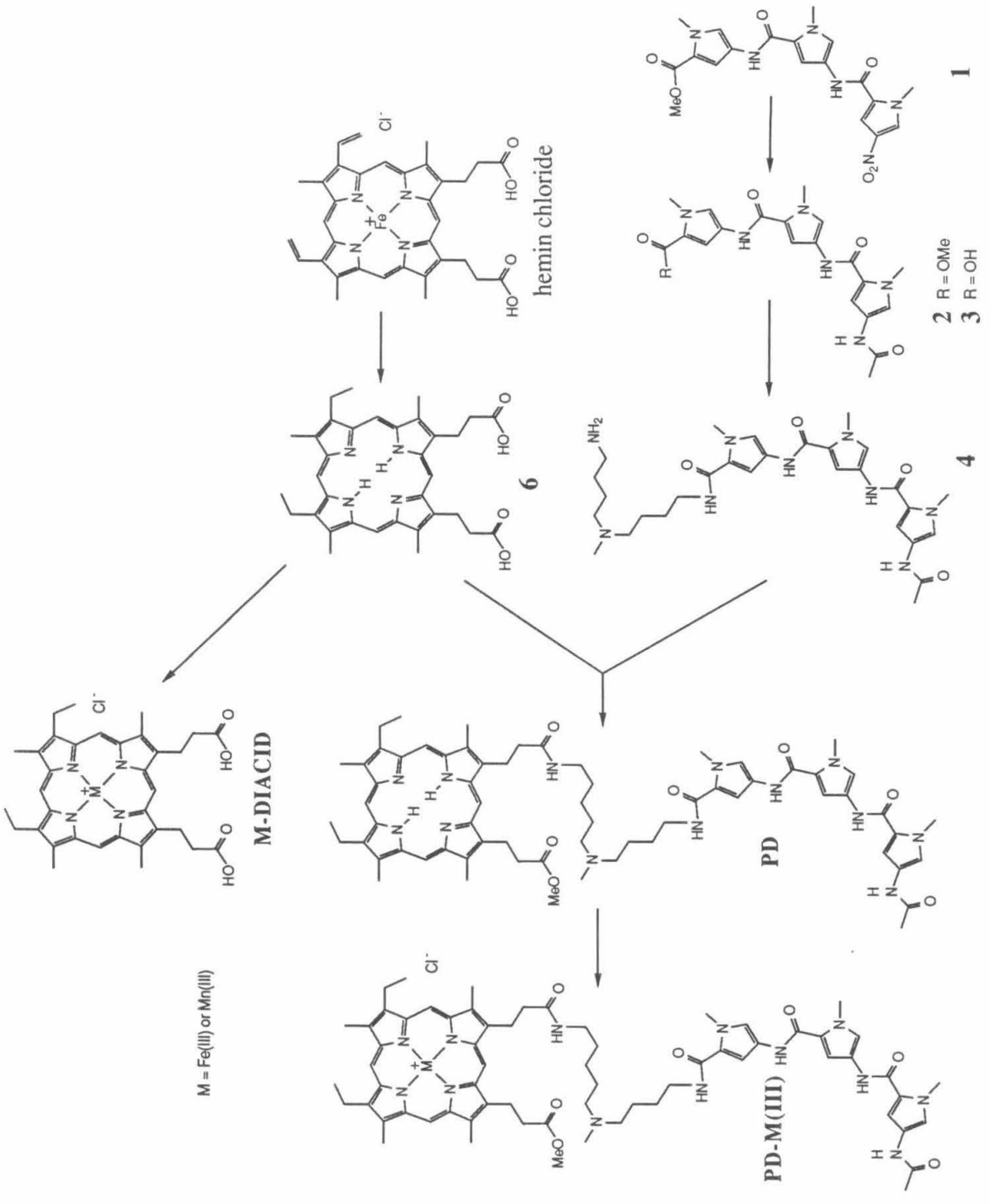
PD
M = Fe or Mn

porphyrin diacid was activated with diimidazole, coupled with **4** and worked up in methanol, giving the porphrin free base **PD** (34%). Iron and manganese were inserted by reaction with the appropriate metal (II) chloride in refluxing DMF.[21]

Results

The cleavage of duplex DNA by **PD**•Fe was determined by monitoring the conversion of supercoiled (form I) pBR322 plasmid DNA[22] to open circular (II) and linear forms (III)[23] by non-denaturing 1% agrose gel electrophoresis followed by ethidium bromide staining. The introduction of one single strand break converts form I to II. The generation of a second single strand break within ~16 base pairs of an existing break converts II to III.[24] The results for **PD**•Fe and diacid•Fe are shown in Table II. **ED**•Fe is also included for comparison.[7]

Scheme III: Synthetic route to **PD**.



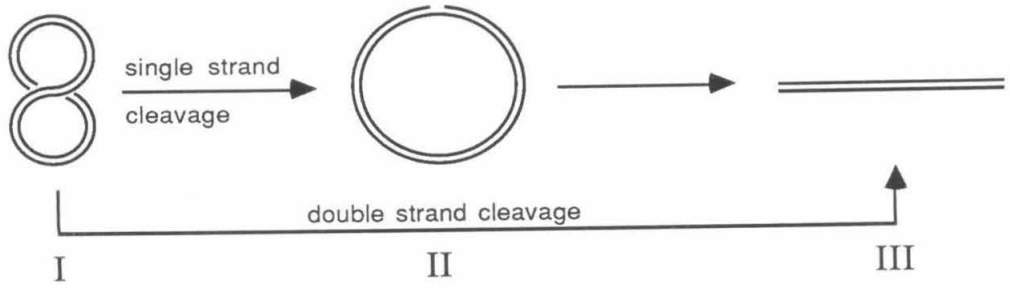


Table II: Cleavage of pBR322 plasmid DNA.[a]

Reagent	Conc μM	% form[b] I	II	III	% DNA recov.[c]
none	-	97	3	0	100
PD•Fe	100	0	0	0	0
PD•Fe	80	46	54	0	5
PD•Fe	40	59	41	0	60
PD•Fe	20	90	10	0	100
PD•Fe	1	98	2	0	100
DIACID•Fe[d]	100	3	92	4	100
DIACID•Fe[d]	10	40	60	0	100
ED•Fe[e]	1	26	76	1	100
ED•Fe[e]	0.1	74	26	0	100

a. Form I pBR322, reagent, buffer (40mM TRIS, 5mM NaOAc, pH 7.9) and DTT (1mM) were allowed to react at 25°C for 2h in the dark.

b. Corrected for the decreased stainability of form I.

c. Total recovery of DNA as determined by densitometry; I + II + III, ±10%.

d. Agrees well with Mueller *et al.*[14]

e. From Taylor *et al.*[7]

The data in Table II clearly shows that **PD**•Fe does not cause DNA strand scission at levels comparable to **ED**•Fe(II). At 1μM **ED**•Fe appears to be at least 40 times more efficient than **PD**•Fe. **PD**•Fe is not even as efficient as the diacid•Fe.[14]

The last column in Table II lists the approximate total recovery of DNA calculated by summing the form I, II and III bands and comparing to the DNA standard

lanes. **PD**•Fe at concentrations greater than about 40μM gives very low total recovery of DNA. The missing DNA is not even apparent as a smear the length of the gel, which implies that the low recovery is not due to gross over-digestion of the DNA (giving a population of linear DNA's of various lengths). At high **PD**•Fe concentrations, a band was frequently observed at the beginning of the gel, i.e. did not migrate at all, which suggests that the low recovery may be due to **PD**•Fe induced precipitation of the DNA. Low recovery of DNA in this type of assay has also been reported for a hemin-bis-intercalator (compound **III** in Table I).

In an attempt to increase the efficiency of **PD**•Fe cleavage, different reducing agents and oxygen sources were tested as cofactors under the same reaction conditions. Sodium borohydride, thioethanol and sodium ascorbate (at 1mM), in place of DTT as the reducing agent, all gave similar or worse results. Different oxygen sources (iodosobenzene, m-chloroperbenzoic acid and peracetic acid at 1mM) also did not improve **PD**•Fe cleavage.

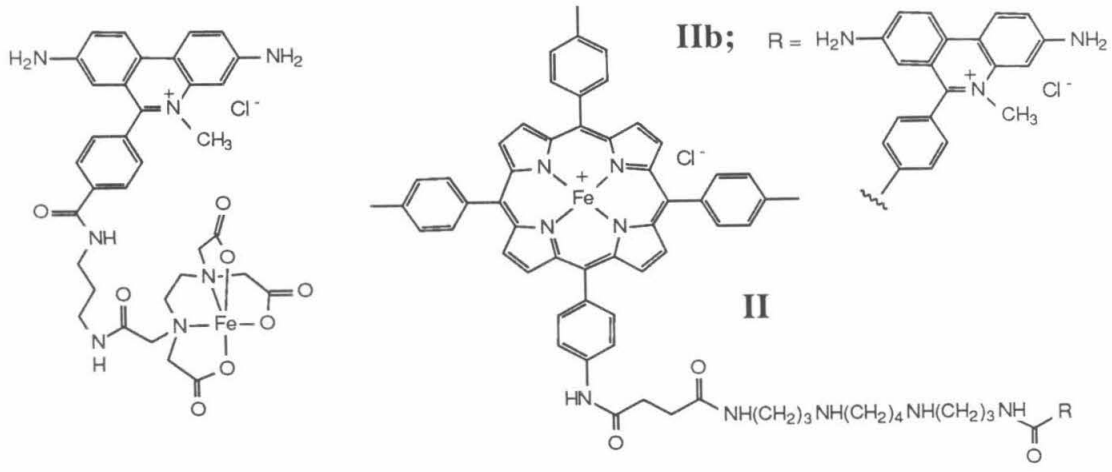
To determine whether **PD**•Fe was exhibiting any sequence specificity, even at its low efficency, strand cleavage was examined on 3′- and 5′-^{32}P labelled 516 base pair restriction fragments from pBR322.[7] The 516 fragment is known to have at least four strong **DE**•Fe cleavage sites. With 20μM **PD**•Fe at standard assay conditions (100μM base pair, 2mR/hr•lane), no DNA cleavage was detected. At higher activity levels (4mR/hr•lane) and long autoradiogram exposures (three times normal), faint but even cutting was observed, and was indistinguishable from the background obtained without **PD**•Fe. It appears then, that **PD**•Fe does not sequence specifically cleave, and therefore bind, duplex DNA.

Discussion

Since **PD**•Fe is not an affinity cleaving molecule, the obvious question is why. Iron porphyrins do cause DNA cleavage (Table I), and covalent attachment of a DNA cleaving function to distamycin gives an efficient sequence specific DNA cleaving molecule, i.e. **ED**. **PD**•Fe is not only not sequence specific, it is also less efficient at causing strand scission than diacid•Fe. It appears that the iron porphyrin is interfering with the binding of the pyrrole tripeptide *and* the pyrrole tripeptide is interfering with the iron porphyrin's ability to cleave DNA.

The exact nature of the interactions which prevent **PD**•Fe from behaving as expected are unknown. One possible cause of low cleavage/binding affinity is aggregation of **PD** due to hydrophobic effects and/or μ-oxo dimer formation.[25,26] Supporting evidence for aggregation effects is found in the low solubility of **PD**•Fe in the aqueous reaction mixture (saturated solution of **PD**<0.5mM). **ED**•Fe is at least an order of magnitude more soluble than **PD**•Fe. An alternative cause of the low binding affinity is the stearic bulk of the porphyrin ring system, which may prevent the tripyrrole from binding in the minor groove of DNA. A third possible cause of low cleavage efficiency is intramolecular chelation of the iron-porphyrin by the tripyrrole unit. Intramolecular chelation could block both proper DNA binding by the tripyrrole and inhibit the redox chemistry of the metalloporphyrin.

The difference in cleavage efficiency of two molecules with the same DNA binding unit, methidium, but different cleaving functions (EDTA•Fe versus TPP•Fe), is of interest (Figure 1). It appears that the EDTA version is effective at reagent concentrations that are two to three orders of magnitude lower than for the porphyrin version. This particular porphyrin even has the advantage of the spermine



[**MPE**•Fe(II)] = 10^{-7}M
[DNA] = 10^{-5}Mbp
Form II = 91%
(25°C, 30min, 1mM DTT)

[IIb] = 10^{-4}M
[DNA] = 10^{-4}Mbp
Form II = 76%
(20°C, 1h, 4mM $Na_2S_2O_4$)

Figure 1: Comparson of the efficiences of two Fe-chelate equiped ethidium based DNA cleaving molecules.

linker whose positive charges (the two amines at pH 7.8) should increase the overall binding affinity electrostatically. If the same pattern is followed in **DE** and **PD**, and since **DE** cuts efficiently at 10^{-6}M, **PD** concentrations of about 10^{-4}M should be needed. Unfortunately, 10^{-4}M **PD** is too high a concentration for meaningful analysis (Table II), since it is apparently near the solubility limit and the reagent to base pair ratio is excessively high (10 **PD**•Fe:bp).

Conclusion

The design, synthesis and DNA binding/cleaving ability of **PD**•Fe was described. **PD**•Fe consists of iron-mesoporphyrin IX coupled to the sequence specific DNA binding N-methylpyrrole tripeptide. **PD**•Fe was found to be very inefficient at cleaving duplex DNA and to exhibit no detectable sequence specificity.

Experimental

See the experimental sections for Chapter I, II and III for general comments.

N-methyl-4-[N-methyl-4-(N-methyl-4-(acetamide)pyrrole-2-carboxamide)-pyrrole-2-carboxamide]-pyrrole-2-carboxylic acid methyl ester, 2: A solution of 8.0g (19mmol) nitro ester **1** (Chapter II) and 1g palladium on charcoal in 100ml DMF was hydrogenated under 1atm hydrogen for 15 hours. The catalyst was removed by filtration through celite and a solution of 9.4g (93mmol) triethylamine and 4.4g (56mmol) acetyl chloride in 10ml DMF was added. The solution was stirred over night. The reaction mixture was concentrated under reduced pressure to 50ml and diluted with 100ml cold water. The product was removed by filtration, washed with water, cold methanol and ether followed by vacuum drying. Recrystalization from 2:3 methanol/water gave 7.0g (86%) **2** as a yellow solid: NMR (D_6–DMSO) δ 9.1 (s, 2H), 9.80 (s, 1H), 7.45 (d, 1H, j=2Hz), 7.21 (d, 1H, j=2Hz), 7.13 (d, 1H, j=2Hz), 7.05 (d, 1H, j=2Hz), 6.90 (d, 1H, j=2Hz), 6.84 (d, 1H, j=2Hz), 3.88 (s, 9H), 3.74 (s, 3H), 1.97 (s, 3H).

N-Methyl-4-[N-methyl-4-(N-methyl-4-(acetamide)-pyrrole-2-carboxamide)-pyrrole-2-carboxamide]-pyrrole -2-carboxylic acid, 3: A solution of 2.6g (5.9mmol) **2** and 2.5g (64mmol) sodium hydroxide in 1:1 ethanol/water was refluxed for two hours. The ethanol was removed on a rotovap and the aqueous solution slowly acidified with 6N hydrochloric acid. The product was removed by filtration, washed with 0.1N hydrochloric acid and cold methanol, and dried under vacuum giving 2.1g (86%) **3**: NMR (D_6–DMSO) δ 9.90 (s, 2H), 9.82 (s, 1H), 7.40 (d, 1H, j=1.5Hz), 7.21 (d, 1H, j=1.5Hz), 7.14 (d, 1H, j=1.5Hz), 7.05 (d, 1H, j=1.5Hz), 6.83 (d, 2H, j=1, 5Hz), 3.82 (s, 9H), 1.93 (s, 3H).

P3-linker, 4: A solution of 2.5g (15mmol) carbonyl diimidazole in 10ml DMF was added to 3.3g (7.8mmol) **3** in 60ml DMF. After stirring for five hours, 11g (78mmol) N, N-di(3-amino-propyl)-N-methyl amine were added and stirring continued overnight. DMF was removed under reduced pressure and the product triturated three times with ether. Chromatography with 10% concentrated aqueous ammonia in methanol gave 1.4g (31%) **4** as a bright yellow solid: NMR (D_6–DMSO) δ 9.88 (s, 1H), 9.86 (s, 1H), 9.78 (s, 1H), 8.02 (t, 1H, j=5Hz), 7.25 (s, 1H), 7.19 (s, 1H), 7.15 (s, 1H), 7.05 (s, 1H), 6.88 (s, 1H), 6.84 (s, 1H), 3.86 (s, 3H), 3.84 (s, 3H), 3.80 (s, 3H), 3.20 (m, 2H), 3.08 (broad singlet, 2H), 2.68 (m, 6H), 2.15 (s, 3H), 1.96 (s, 3H), 1.62 (p, 2H, j=8Hz), 2.50 (p, 2H, j=8Hz); UV/VIS (H_2O) 302nm (ϵ=32,000), 233nm (ϵ=27,000).

PD: In 4ml DMF was placed 100mg (0.16mmol) mesoporphyrin IX dihydrochloride[19,20] and 80mg (0.52mmol) acydiimidazole. After stirring for five hours, a solution of 86mg (0.16mmol) **4** in 1ml DMF was added and the solution stirred overnight. Another 30mg (0.16mmol) acydiimidazole were added, the solution stirred two hours, and 4ml methanol added. After stirring overnight, the solvents were removed under reduced pressure at 35°C and the product was triturated three times with ether. Chromatography with 3% concentrated aqueous ammonia in methanol afforded 60mg (34%) **PD** as a purple solid: NMR (D_6–DMSO) δ 10.27 (s, 1H), 10.14 (s, 1H), 9.91 (s, 1H), 9.88 (s, 1H), 9.83 (s, 1H), 7.98 (broad singlet, 1H), 7.88 (broad singlet, 1H), 7.26 (s, 1H), 7.18 (s, 1H), 7.15 (s, 1H), 7.08 (s, 1H), 6.89 (s, 1H), 6.80 (s, 1H), 4.35 (broad singlet, 4H), 4.08 (broad singlet, 4H), 3.80 (s, 9H), 3.73 (s, 3H), 3.62 (s, 6H), 3.59 (s, 6H), 3.11 (broad singlet, 4H), 3.02 (broad singlet, 6H), 2.03 (s, 3H), 1.84 (s, 3H), 1.80 (broad singlet, 6H), 1.68 (m, 2H), 1.33

(broad singlet, 2H), 1.24 (broad singlet, 2H), –4.04 (s, 2H); UV/VIS (methanol) 615, 565, 530, 496, 390, 305, 235, 210nm.

Metal Insertion: Iron and manganese were inserted by adding 1.1 equivalence of the appropriate metal (II) chloride to a refluxing solution of the porphyrin in DMF.[21] The solution was heated until a small aliquot was no longer bright red under long wavelength UV on a TLC plate. The DMF was removed under reduced pressure and the product triturated three times with ether. Metal insertion was confirmed by changes in the adsorption spectra characteristic of iron (III) and manganese (III) mesoporphyrins.[27,28]

References and Notes

1. Hertzberg, R. P., Dervan, P. B., *J. Am. Chem. Soc.*, **104**, 313 (1982).
2. Hertzberg, R. P., Dervan, P. B., *Biochemistry*, **23**, 3934 (1984).
3. Bresloff, J. L., Crothers, D. M., *J. Mol. Biol.*, **95**, 103 (1975).
4. Bauer, W., Vinograd, J., *J. Mol. Biol.*, **33**, 141 (1968).
5. LePecq, J. B., Paoletti, C., *J. Mol. Biol.*, **27**, 87 (1967).
6. Waring, M. J. in "The Molecular Basis of Antibiotic Action" Wiley: New York, 1981, p. 274–306.
7. Taylor, J. S., Schultz, P. G., Dervan, P. B., *Tetrahedron*, **40**, 457 (1984).
8. Groves, J. T., Haushalter, R. C., Nakamura, M., Nemo, T. E., Evans, B. J., *J. Am. Chem. Soc.*, **103**, 2884 (1981).
9. Groves, J. T., Nemo, T. E., *J. Am. Chem. Soc.*, **105**, 5786 (1983).
10. Tabushi, I., Yazaki, A., *J. Am. Chem. Soc.*, **103**, 7371 (1981).
11. Nee, M. W., Bruice, T. C., *J. Am. Chem. Soc.*, **104**, 6123 (1982).
12. Chank, C. K., Kuo, M. S., *J. Am. Chem. Soc.*, **101**, 3413 (1979).
13. Walling, C., *Accts. Chem. Res.* , **8**, 125 (1975).
14. Mueller, G. C., Aft, R. L., *J. Biological Chem.*, **19**, 12069 (1983).
15. Pasternak, R. F., Gibbs, E. J., Villafranca, R. J., *Biochemistry*, **22**, 2406 (1983).
16. Byrnes, R. W., Fiel, E. J. in "Conversation in Biomolecular Stereodynamics III, Collected Abstracts", State University of New York, New York, 1983, p. 126.
17. Hashimoto, Y., Lee, C. -S., Shudo, K., Okamato, T., *Tetrahedron Lett.*, **24**, 1523 (1983).
18. Lown, J. W., Joshua, A. V., *J. Chem. Soc., Chem. Commun.*, 1298 (1982).
19. Corwin, A. H., Erdman, J. G., *J. Am. Chem. Soc.*, **68**, 2473 (1946).
20. Caughey, T. *et al.*, *J. Org. Chem.*, **31**, 2631 (1966).

21. Adler, A. D., Longo, F. R., Kamps, F., Kim, J., *J. Inorg. Nucl. Chem.*, **32**, 2443 (1970).

22. The entire sequence of pBR322 has been determined: Sutcliffe, J. G. *Cold Spring Harbor Symposium Quant. Biol.*, (1982). However an error has been found in the original sequence: Peden, K. W. C., *Gene*, **22**, 277, (1983).

23. Johnson, P. J., Grossman, L. I., *Biochemistry*, **16**, 4217 (1977).

24. Friefelder, D., *Biopolymers*, **7**, 681 (1969).

25. Fleischer, E. B., Srivastava, T. S., *J. Am. Chem. Soc.*, **91**, 2403 (1969).

26. Cohen, I. A., *J. Am. Chem. Soc.*, **91**, 1980 (1969).

27. Collman, J. P., Gagne, R. R., Reed, C. A., Halbert, T. R., Lang, G., Robinson, W. T., *J. Am. Chem. Soc.*, **97**, 1427 (1975).

28. "The Porphyrins" vol. I and II, Dolphin, D., Ed., Academic Press, New York, 1978.