Characterization of the DNA Binding Domains of Helix-Turn-Helix Proteins by Affinity Cleaving

Thesis by

Jumi A. Shin

In Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

California Institute of Technology

Pasadena, California

1992

(Defended December 20, 1991)

ACKNOWLEDGMENTS

It took me five years to write this thesis; they have been interesting years. Along the way, I've had the pleasure of knowing a lot of people who have taught me much about science and research, and a lot about myself and what I hope to get out of life. I have really changed, for the better I think, over the last few years, and I would like to thank those people who have been so important in my life.

Peter Dervan is the reason I came to Caltech. From my undergraduate experience, I knew that I wanted to work in the biological area of chemistry, and after meeting Peter, I figured "Why not?" I definitely made the right decision to work here, because I know that I have learned much more here in the last five years than I had learned in all my previous years; still, what I know now is just a drop in the bucket, and I'm incredibly excited about all the challenges that lie ahead. Peter's enthusiasm is really motivating and infectious. I feel tremendously lucky that I have been able to interact with Peter and with the fantastic group of people working in this lab.

There are many people that I have to thank for teaching me and being wonderful friends. I'd like to thank Jim Sluka and Dave Mack for all their help and expertise in the protein work, and Warren Wade for his boundless patience and friendship; my great friend Martha Oakley, a most generous and giving person with her time, kindness, and extensive knowledge of science; Mark Distefano, who taught me a lot about molecular biology and who has been a terrific (Scotch-drinking?) friend; Kevin Luebke, whom I've had the pleasure and the pain of sitting next to for nearly five whole years; Laura Kiessling, that crazy gal who wore wide belts for skirts; Scott Singleton, the All Omniscient God of Computers, Quantitative Affinity Cleaving, and Modeling

(and let's not forget Paul); Milan Mrksich, whose invaluable help with the modeling and photography can be seen in the color photos of this thesis; and George Best for his careful reading and helpful comments on this work. And I want to thank Bob for being a most supportive, fun, and precious friend; some of my happiest memories at Caltech are of you.

Finally, I have to thank my family. My parents have been two of the most supportive and loyal people a person could ever have; I always feel incredibly lucky that I was born to you. Dad, you've been a wonderful role model and have always encouraged me; I feel a lot more secure about my work, knowing you're behind me. And Mom, you are truly a treasure; you are so funny, wonderful, and beautiful; I can only hope I will become more like you as I get older. And Juyoung and Sungho, I wish you all of my love, and hope you enjoy your long lives ahead of you.

ABSTRACT

Protein recognition of specific DNA binding sites is critical for cellular processes; transcription, replication, and restriction are but a few of the biological activities regulated by DNA binding proteins. Although DNA can exist in any number of sequences and conformations, proteins have the exquisite capability of recognizing specific nucleotide sequences in an entire genome. **Chapter One** is a review that concentrates on sequence-specific DNA binding proteins primarily capable of regulating gene expression and also able to serve structural and catalytic roles in biological processes. The helix-turn-helix repressor proteins from phages λ and 434 and *lac* and *trp* repressors, leucine zipper proteins, zinc-finger proteins, the homeodomain, basic helix-loop-helix proteins, *arc* and *mnt* repressors, and double helix-turn-helix proteins are discussed.

Affinity cleaving studies on mutants of the DNA binding domain of Hin recombinase, Hin(139-190), are discussed in **Chapter Two**. Arg140 has been previously found to be extremely important for sequence-specific binding of Hin(139-190) to the *hix*L operator. Mutants in which Arg140 of Hin(139-190) is replaced with Lys, Ala, β-Ala, Glu, Gly, and Gln were studied; a mutant in which Arg142 is replaced by Lys is also discussed. The binding affinities and sequence specificities of these mutants are very different from that of Hin(139-190). Also Hin(139-184), in which the six carboxyl terminal residues are removed, was studied by affinity cleaving and shown to have a similar sequence specificity, albeit reduced binding affinity, compared to Hin(139-190). In **Chapter Three**, a procedure for quantitating thermodynamic parameters using the affinity cleaving technique is discussed. In this procedure, affinity cleaving reactions are run over a wide range of protein concentrations. A binding isotherm is generated and fit, using a least-squares program, and

reproducible values for binding constants can be obtained. Binding constants were quantitated for Hin(139-190), two of the mutant proteins, and Hin(139-184). This methodology was developed in order to obtain thermodynamic parameters of complexation between these synthesized proteins and their DNA binding sites. Binding constants were measured only for the *hix*L IRL and IRR sites for each protein, although in some cases, strong binding was also exhibited at the secondary and even tertiary sites. This method can resolve binding curves for individual sites in a cooperative system. Typical binding constants ranged between 10^5 M^{-1} and 10^7 M^{-1} at 20mM NaCl for the *hix*L site.

Affinity cleaving studies were performed on the DNA binding domain of *lac* repressor in **Chapter Four**. Based upon sequence homology analysis between the binding domains of *lac* and cro repressors, the DNA binding interaction between the *lac* repressor and operator was assigned as a helix-turn-helix motif. NMR studies on the DNA binding domain of *lac* repressor have shown that the recognition helix of *lac* binds the major groove of DNA in an orientation opposite that of cro. Affinity cleaving studies on the *lac* repressor DNA binding domain support the NMR results and establish the orientation of the recognition helix of the*lac* repressor.

In **Chapter Five** are discussed affinity cleaving studies on the*engrailed* homeodomain, for which a high-resolution cocrystal structure was recently published. Until now, affinity cleaving has been performed on proteins for which no high-resolution structural data exist. Because the *engrailed* homeodomain contains sequence elements very similar to Hin(139-190), it was chosen for affinity cleaving studies in order to compare and to interpret structural data obtained by affinity cleaving with a protein of known structure.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER ONE
Protein-DNA Recognition

CHAPTER TWO
Affinity Cleaving Studies on the DNA Binding Domain of Hin Recombinase

CHAPTER THREE
Quantitative Affinity Cleaving Studies on the DNA Binding Domain of Hin Recombinase

CHAPTER FOUR
Affinity Cleaving Studies on the *lac* Repressor DNA Binding Domain

CHAPTER FIVE
Affinity Cleaving Studies on the *engrailed* Homeodomain

APPENDIX

CHAPTER ONE
Protein-DNA Recognition

INTRODUCTION

The Lactose Paradigm

Forty years ago, the French geneticists Francois Jacob and Jacques Monod proposed an elegant yet simple model for prokaryotic gene regulation, which has remained the working paradigm in the field; this model was predominantly based upon biochemical and genetic studies conducted on *lac* repressor, a regulatory protein.[1,2] There is a cluster of three *lac* structural genes that comprise the *lac* operon: an operon codes for proteins required by the cell for enzymatic or structural functions. These proteins are called structural proteins, and they comprise the overwhelming majority of bacterial proteins in a cell. The *lac* genes are controlled by negative regulation: they are transcribed (i.e. mRNA transcripts are synthesized and will serve as the template for protein synthesis) unless turned off by a regulatory protein. Because the function of the regulatory protein is to prevent the expression of this cluster of structural genes, it is called a repressor protein, and this repressor regulates the activity of the operon. The repressor prevents transcription by binding to a sequence of DNA called the *lac* operator, which lies between the promoter and the operon. When the repressor binds at the operator, its presence prevents RNA polymerase from binding to the promoter and initiating transcription of the *lac* structural genes.

The repressor protein has a binding site for an inducer, which is a small molecule that causes production of enzymes able to metabolize it. In the *lac* system, the inducer is lactose; when the concentration of lactose rises

to a certain level, it binds to*lac* repressor and changes the conformation of protein such that it can no longer bind to DNA. Therefore, the repressor leaves the operator, and RNA polymerase can bind to the promoter and initiate transcription of the *lac* genes. This type of relationship in which the binding of a small molecule alters a protein's structural conformation is an example of allosteric control.

A Genetic Switch in a Bacterial Virus: the Lambda Phage System

Genes are regulated, or turned on and off, as an organism develops and as it adapts to different conditions.[3-8] The induction of a bacterial virus as the result of a transient change in the environment is an example of gene regulation. It was not until the 1960's, however, when repressors, including the λ repressor, were first detected biochemically and found to be proteins.[4] The λ virus, instead of multiplying in the host cell and destroying it, can switch off its genes and insert its DNA into the host cell's chromosome (Figure 1). The viral DNA is then replicated as a component of the bacterial host's chromosome and is passed on to its progeny. This dormant molecule of viral DNA is called a prophage, and the bacterium carrying it is a lysogen. A number of agents can induce the prophage to begin lytic growth; these inducing agents all damage host DNA and thus threaten the viability of the cell. Therefore, induction is effectively an escape mechanism for the virus.

The λ virus exhibits two modes of growth in which different sets of viral genes are expressed (Figure 2).[3,4] In lysogenic growth, only one protein is expressed: the λ repressor. The repressor turns off other genes, but stimulates transcription of its own gene; transcription activation is believed to occur when bound repressor interacts with RNA polymerase and assists in binding polymerase to the promoter and beginning transcription. In lytic growth, the regulatory protein is called λ cro, and it turns the repressor gene

off. (Cro was probably chosen to stand for "control of repressor and other things.")[4] Thus repressor turns off cro during lysogenic growth, and cro turns off repressor during lytic growth. Remarkably, both repressor and cro work by binding to the same region of λ DNA called the right operator, O_R (there similarly exists the left operator, which is much like the right operator). The operator includes three discrete sites designated O_{R1}, O_{R2}, and O_{R3}. Cro and repressor bind to each of these 17 base-pair sites as dimers, although these proteins do not dimerize in solution.

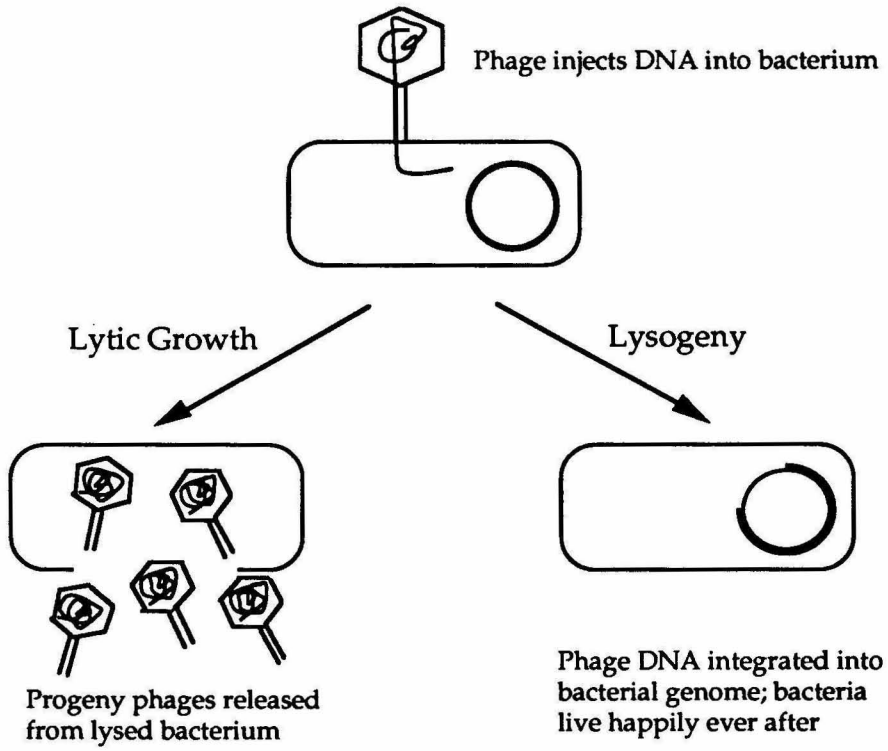


FIGURE 1. Lytic development involves the reproduction of phage particles with destruction of the host bacterium, but lysogenic existence allows the phage genome to be carried as part of the bacterial genetic information.

The repressor binds most strongly to O_{R1}; after binding at this site, O_{R2} becomes much more attractive to another repressor dimer and is filled immediately. This is an example of cooperativity in which one DNA binding molecule assists a second molecule to bind at another site. There is no

cooperative interaction between O_{R2} and O_{R3}; thus O_{R3} remains vacant until the concentration of repressor rises to a saturation level in which all three sites are occupied. The switch occurs when the host lysogen harboring the λ prophage is exposed to an inducing signal, typically ultraviolet radiation, which damages DNA.

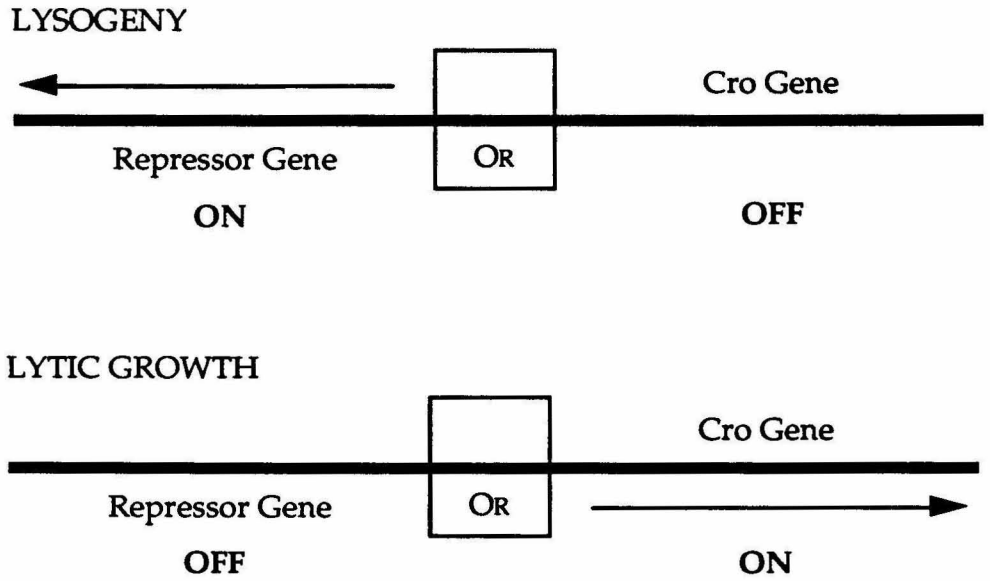


FIGURE 2. A molecular switch determines which of two sets of genes is turned on and thereby commits the virus either to the prophage (lysogenic) state or to lytic growth. The switch is in the region of viral DNA called the right operator (O_R), which is flanked by genes encoding two regulatory proteins. During lysogeny, the repressor gene is transcribed into mRNA; during lytic growth, cro is transcribed.

Upon irradiation, the cellular protein RecA, which ordinarily catalyzes DNA recombination, functions as a protease.[4] The protease cleaves λ repressor monomers into carboxyl-terminal and amino-terminal domains. The amino-terminal domain is responsible for DNA binding, but cannot dimerize efficiently without the carboxyl terminal and therefore cannot bind efficiently to the operator. The cell's concentration of functional repressor molecules drops, and both O_{R1} and O_{R2} are vacated. Now RNA polymerase

can bind to the rightward promoter and begin transcribing cro and other genes required for lytic growth. Unlike repressor dimers, cro dimers do not bind cooperatively to adjacent DNA sites, and cro does not facilitate the binding of RNA polymerase to the promoter. Cro also has reverse affinity from the repressor for operator binding sites, for cro binds most tightly to O_{R3}. As lytic growth continues, the concentration of cro increases until O_{R1} and O_{R2} are also occupied; this turns off transcription of the lytic genes. Thus the λ virus implements an elegant and efficient genetic switch between expression of two different sets of proteins. This mode of genetic control is regularly utilized in prokaryotes, including the P22 repressor and cro, and the 434 repressor and cro.[4]

Since the time of Jacob and Monod, the repressors and operators of several operons have been elucidated, and in some cases, the protein-DNA complex has been crystallized. These cocrystals include the repressor-operator complexes of λ repressor and λ cro, 434 repressor and 434 cro, catabolite gene activator protein (CAP), *trp* and *met* repressors, and zinc-finger proteins, which are regulatory proteins but not repressors. Although the *lac* operon has been the most extensively studied regulatory system, no crystal structure is yet available for the*lac* repressor or the cocrystal complex; solution-phase NMR structures of the*lac* repressor and the repressor-operator complex are available, however, and these studies have produced a wealth of detailed information about this system.

Remarkably, proteins utilize a fairly small number of structural motifs to recognize a large number of sequences of DNA. Many regulatory proteins exhibit one of these common structural motifs, the helix-turn-helix unit in which the first α-helix serves to anchor the second α-helix, which fits into the

major groove of DNA (Figure 3).[9-18] Regulatory proteins typically bind to DNA as cooperative dimers, and these cooperative interactions between proteins are essential for efficient DNA binding and transcriptional regulation. In Figure 3 are shown the protein crystal structures (in the absence of DNA) of λ repressor and cro, and CAP, which were some of the first regulatory proteins crystallized and which notably displayed a common structural unit, the helix-turn-helix, which was believed to interact with DNA.[9,10]

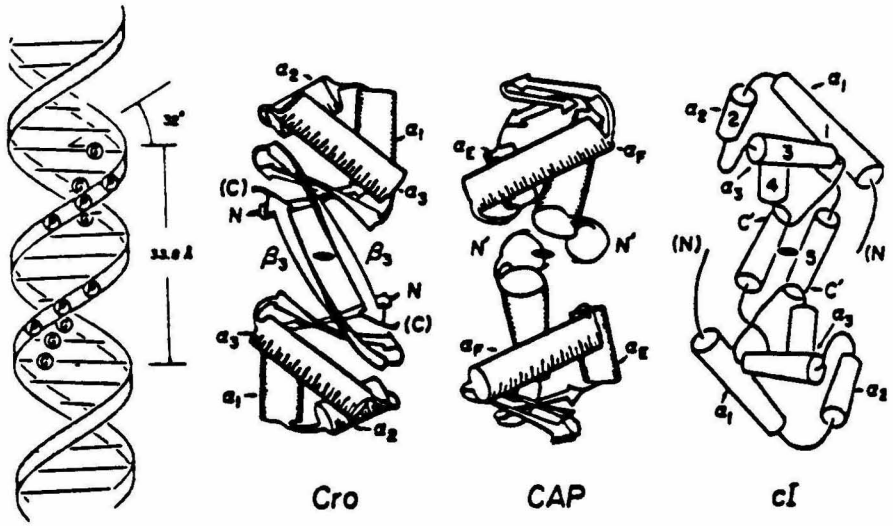


FIGURE 3. Schematic drawings of B-form DNA and the structures of λ cro, CAP, and λ repressor (cI).[9] The helix-turn-helix domains are α_2-α_3 for cro and repressor and α_E-α_F for CAP. The view of the proteins is of the face that binds to DNA. The DNA drawing shows the ethylated phosphates, which hinder binding of cro as well as the guanines (G), which are protected from dimethyl sulfate by cro in the major groove. The α-helices are represented as cylinders and the β-sheets as arrows.

PRINCIPLES OF RECOGNITION

That proteins can recognize particular DNA sequences was demonstrated by isolating the *lac* and λ repressors and showing that they bind to distinct segments of DNA. These proteins also bind to non-operator DNA

but with approximately 10^5 lower affinity. Structural, biochemical, and genetic studies of protein-DNA complexes have established two important sources of sequence-specific protein-DNA interactions: 1) Hydrogen bonding and van der Waals contacts between amino-acid side chains and exposed base pairs, primarily in the major groove of B-form DNA and to a lesser extent, the minor groove; 2) The sequence-dependent deformability of duplex DNA, which allows the operator site to exist in a particular conformation required for binding to a protein at a lower free energy than other sequences.[15]

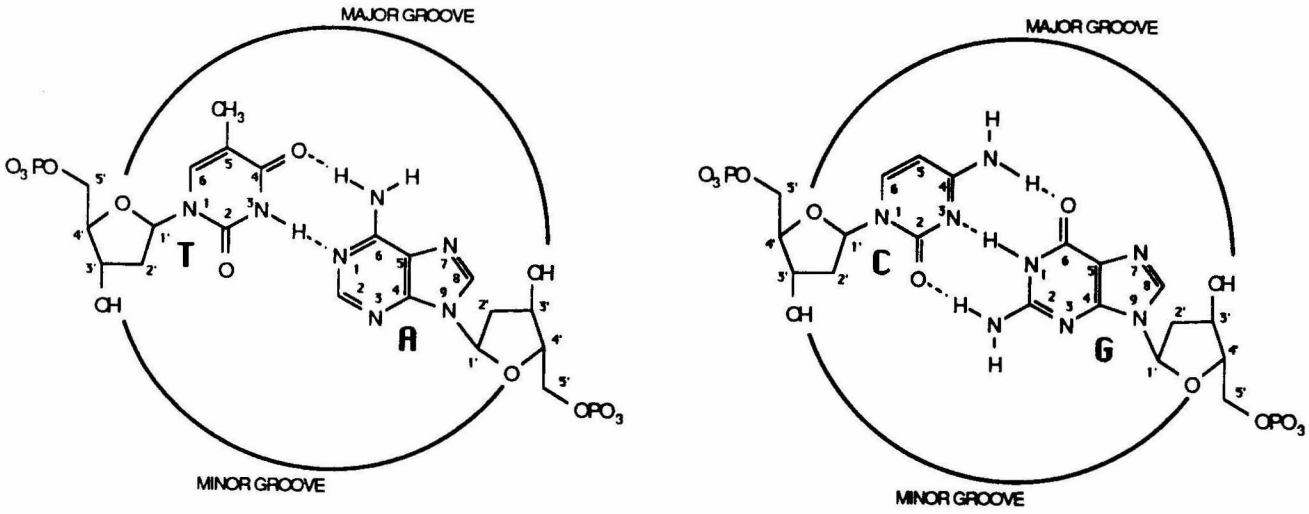


FIGURE 4. Schematic drawings of the TA and CG base pairs indicating the major and minor grooves.

The known structures of protein-DNA complexes rule out any simple way for particular amino acids to recognize specific base pairs,[19] although the structural variation among the four base pairs is very different as viewed from either the major or minor grooves (Figure 4). Although particular amino-acid side chains do not always recognize the same base pairs, there are some preferences; for example, the guanidinium group of arginine often makes a bidentate interaction with the N7 and O6 of guanine, as observed in the cocrystals for EcoR I and *trp* repressor and proposed for λ cro and CAP;

however, arginine is also observed to interact with the N7's of two adjacent adenines in EcoR I.[15,16] The major groove presents a distinctive hydrogen bond pattern for each of the four base pairs, as compared to the minor groove, in which only N2 of guanine distinguishes AT from GC. Furthermore, the minor groove of B-DNA is too narrow to accommodate an α-helix. Possibly, direct minor groove recognition can distinguish only a binary code (GC or CG vs. AT or TA), whereas major groove recognition can discriminate among all four base pairs, and perhaps this is a major consideration for sequence-specific protein recognition of DNA.[15,16]

B-DNA exhibits significant conformational variability; such observed distortions include changes in twist, groove width, and kinks.[15] GC base pairs generally have low propeller twist, because coplanarity is favored by the presence of three Watson-Crick hydrogen bonds. AT base pairs can display more twist, because their two Watson-Crick hydrogen bonds allow for more flexibility. AT-rich tracts favor narrowing of and bending toward the minor groove,[20] whereas GC-rich tracts favor kinks that narrow and bend toward the major groove.[15] The stiffness of DNA is sequence-dependent, and these elastic properties of DNA are of great importance to protein-DNA binding specificity.[21] The free energy cost for various nucleic-acid sequences to assume the conformation that is required for its binding to protein is not the same for different sequences. Structural studies have shown that proteins often bind a conformation of DNA that is altered from standard B-DNA.[15]

Water molecules have also been shown to be necessary for mediation of sequence-specific protein-DNA contacts.[15,16] In the *trp* repressor-operator cocrystal, there are no direct protein-DNA contacts; in this case, water molecules act as protein side chains by mediating hydrogen bond interactions. In the 434 repressor-operator cocrystal, water molecules mediate the

interaction of the guanidinium group on Arg43 with AT base pairs in the minor groove, and they orient the amide side chain of Gln33 toward the O4 of thymine in the major groove.[15,16]

THE HELIX-TURN-HELIX DNA BINDING DOMAIN

The helix-turn-helix (HTH) structural motif is the best characterized class of DNA binding domain[16,17]; the HTH is a recurring structure found in different, cooperatively interacting domains and is generally not a stably folded unit on its own. The HTH is found in both prokaryotic and eukaryotic regulatory proteins; it is composed of two α-helices (helices 2 and 3) separated by a short linker. From extensive biochemical, genetic, and structural studies, it has been found that helix 3 confers upon these proteins much of their DNA sequence specificity.[22,23] Such exquisite selectivity results from the ability of helix 3 to penetrate the major groove such that its amino-acid side chains can make direct or water-mediated contacts to specific base pairs and the phosphodiester backbone. Helix 3, often called the recognition helix, has essentially two surfaces: an inside, hydrophobic surface, which packs against the remainder of the protein, and an outside, hydrophilic surface, which interacts with the major groove. Helix 2 lies across the major groove above helix 3; the amino acids at the amino terminus of helix 2 make non-specific hydrogen bonds with the phosphodiester backbone, and favorable dipole-charge interactions between helix 2 and the major groove serve to anchor recognition helix 3 in the major groove. Regulatory proteins rely on the HTH motif, in large part, for recognition of operator sequences and transcriptional control.

A high degree of amino-acid sequence homology exists among these HTH domains, suggesting the fundamental importance of this evolutionarily

			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 [from J. P. Sluka, Ph. D. Thesis, California Institute of Technology, 1988]. These are all regulatory proteins except for HIN, γδ resolvase, and Tnp resolvase, which are recombinases. The sequences are shown N→C with hydrophobic residues in bold and charged residues marked appropriately. The columns marked N and C give the residue numbers for each protein. 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 occurrence). The brackets at the top mark the α_2 and α_3 helices of λ cro; α_3 is the DNA binding helix. The bottom five rows show tabulation of the occurrences of positive, negative, hydrophobic, and hydrophilic residues and the consensus for each column; bold-face numbers mark strong preferences for a given residue.

conserved structure (Table 1). Many proteins are suspected of being HTH DNA binding proteins on the basis of sequence homology[24]; already these predictions are being borne out as more structural studies are conducted on regulatory proteins. Comparison of known crystal structures reveals that conserved amino acids appear at points where helices 2 and 3 interact with each other and in the turn, which is instrumental in determining the angle between the two helices; these residues help form the structural skeleton of HTH units. Residues at positions 6, 10, 12, 17, and 20 are usually hydrophobic; they define the interior of the two-helix elbow and require additional hydrophobic residues from the domain to complete an enclosed hydrophobic core (Table 1). Residue 7 is typically Gly or Ala, and residue 9 is often Gly. These conserved residues generally do not contribute to DNA binding sequence specificity. The hydrophilic residues usually make direct and indirect contacts with the operator site.

A Structural Comparison of the λ repressor and the 434 repressor

Within the last few years, the cocrystal complexes of the λ repressor[25] and the 434 repressor[26] with their respective operators have been solved to high resolution. Comparison of the two complexes reveals that three conserved residues in the HTH domain make similar contacts.[27] These conserved residues and their interactions with phosphodiester oxygens help to establish a frame of reference within which other HTH residues make contacts that are critical for sequence-specific recognition, and these positioning contacts may be important features in HTH proteins. In contrast, the structural comparisons rule out any simple code for recognition between protein side chains and DNA base pairs.[27]

In the HTH regions of the λ repressor (residues 33-52) and the 434 repressor (residues 17-36), the polypeptide backbones of the two proteins are

virtually indistinguishable. The aligned sequences show a high degree of sequence homology; the conserved alanine and glycine (residues 37 and 41 in the λ repressor) are characteristic of the HTH unit and seem to stabilize folding (Figure 5). The conserved leucine (residue 50 in λ repressor) stabilizes packing of the HTH unit against the rest of the protein. The conserved lysine (residue 39 in the λ repressor) does not contact DNA, but may be favored because a positively charged amino acid tends to stabilize the carboxyl terminus of an α-helix, which is negatively charged because of the helix dipole effect.[28-30]

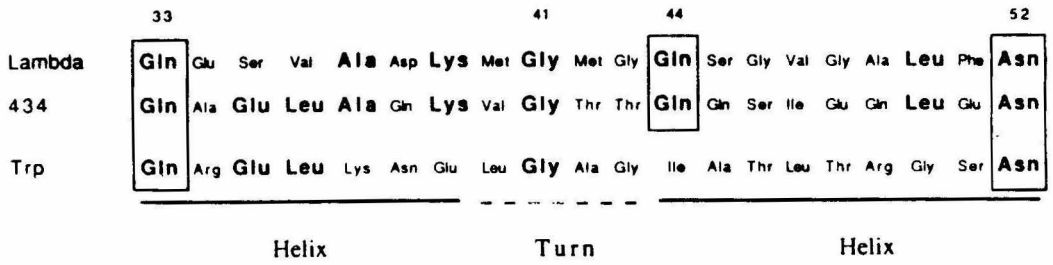


FIGURE 5. Alignment of the HTH units from the λ, 434, and *trp* repressors.[27] Conserved residues are shown in bold, and boxes enclose the conserved residues that contact DNA.

The first residues in the HTH unit, Gln_{λ}^{33} and Gln_{434}^{17} at the beginning of helix 2, make two hydrogen bonds with the DNA backbone; the peptide -NH contacts one phosphodiester oxygen near the outer edge of the operator, and the side chain $-NH_2$ bonds to an oxygen of a neighboring phosphate (Figure 6). The first residues in the second helix of the HTH unit are also glutamine: Gln_{λ}^{44} and Gln_{434}^{28} at the beginning of helix 3. These glutamines make a pair of hydrogen bonds with an adenine and allow for specific recognition of the AT base pair; this pair of hydrogen bonds can be formed only if adenine is present at this position. Gln_{λ}^{44} and Gln_{434}^{28} also make hydrogen bonds to Gln_{λ}^{33} and Gln_{434}^{17}. The glutamines at the

beginning of helices 2 and 3 therefore make critical contacts that help to anchor the helices with respect to each other and to anchor the protein at the operator site. The last amino acids in the HTH unit are Asn_{λ}^{52} and Asn_{434}^{36}; these asparagines form hydrogen bonds with the same phosphodiester oxygen that is contacted by the side-chain $-NH_2$ group of the first glutamine in the HTH. Each asparagine also forms a hydrogen bond to a carbonyl oxygen from the preceding turn of the α-helix.

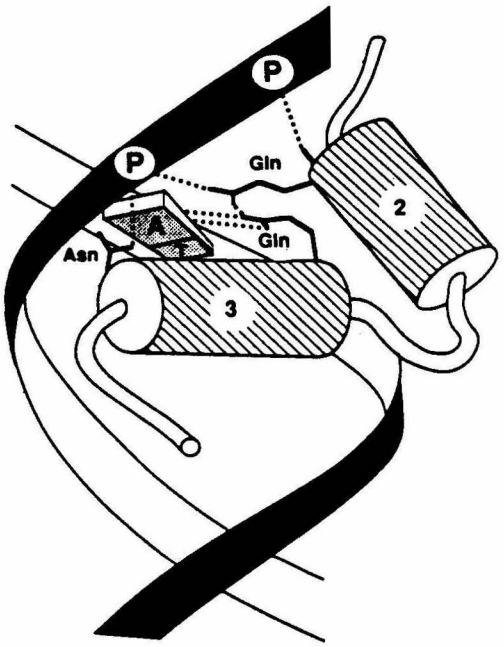


FIGURE 6. Sketch of the superimposed λ and 434 repressors half site.[27] Helices 2 and 3 are shown as cylinders; side chains are included for residues that make analogous contacts in the λ and 434 complexes; and a conserved AT base pair is shown. Hydrogen bonds are indicated by dotted lines. Highlighted residues correspond to Gln^{33}, Gln^{44}, and Asn^{52} of the λ repressor, and Gln^{17}, Gln^{28}, and Asn^{36} of the 434 repressor.

In each complex, the first, second, third, fifth, and sixth residues of the recognition helix (helix 3) interacts with base pairs in the operator site. The fourth residue, Val_{λ}^{47} and Ile_{434}^{31}, is buried in the back side of the α-helix and makes critical interactions with helix 2. Apparently, this orientation of helix 3 may be optimal for recognition, for any other orientation would bury two residues, leaving only four amino acids available for specific recognition of DNA.[27] Perhaps this helps to explain why the HTH motif is so precisely conserved.

SEQUENCE-SPECIFIC DNA BINDING PROTEINS

In the early 1980's, detailed structural insight into one of the mechanisms by which proteins carry out their transcription regulatory roles was provided by the X-ray crystal structures of the λ cro (66 amino acids)[31] and residues 1-92 of the λ repressor (236 amino acids)[32,33] from bacteriophage lambda, and *Escherichia coli* catabolite gene activator protein (CAP, 210 amino acids)[34,35] complexed with cyclic adenosine monophosphate (cAMP). From biochemical studies, it is known that these proteins bind to DNA operator sites as dimers; λ cro and CAP complexed with cAMP exist in solution as dimers, although most other known HTH transcription regulators exist in solution as monomers even at high protein concentrations.[9,10] Crystallographic studies on λ cro were performed on the monomeric protein, whereas crystallographic studies on CAP were performed on the intact dimer in complex with cAMP. Upon CAP's binding of cAMP, as well as *trp* repressor's binding of tryptophan and *lac* repressor's binding of lactose, changes in this protein's conformation enable it to bind to specific operator sites near certain promotors and to regulate transcription.[2]

All three protein dimers displayed a strikingly similar structural feature–a pair of α-helices linked by a turn (HTH) protruding from the protein's surface.[9,10] The helices were tilted and separated by approximately 34Å, equivalent to one full turn of B-DNA. All three proteins were found to have a high degree of sequence homology in the HTH region.[24] Model building studies on λ cro demonstrated that its recognition helix fit very well into the major groove of B-DNA,[36] and that protein-DNA contacts could be predicted; genetic studies on HTH proteins agree with the structural models proposed.[37] On the basis of model building studies on CAP, it was proposed that CAP might bind to a left-handed B-DNA major groove; that is, CAP's

recognition helix was tilted in an orientation different from that of λ cro and therefore, it was suspected that CAP binds to left-handed DNA.[9,10] Subsequent genetic and structural studies showed, however, that CAP binds to right-handed B-DNA, and that its recognition helix is tilted differently from that of λ cro.[38] Further insight was gained from the X-ray cocrystal structures of the 434 repressor (residues 1-69)[26,39,40] and the λ repressor (residues 1-92)[25] with their respective operators. Although the general features of the model for DNA binding were experimentally confirmed, it became evident that selective recognition of operator DNA is highly dependent on factors other than the amino-acid residues of the recognition helix.

Th 434 cro and the 434 repressor show a high degree of sequence homology (48%); the 434 repressor and the λ repressor also share significant sequence homology (26%). Thus it is not surprising that their three-dimensional structures are very similar, although the sequences of their operator sites differ remarkably. The real significance of the data generated by these cocrystal complexes lies not in the structure of the protein's recognition domain, but rather in the structures of the bound DNA and the protein-DNA interactions.

434 REPRESSOR

The 434 repressor-O_{R1} cocrystal complex, which was solved to 3.2Å resolution,[26] shows that the repressor undergoes little conformational change upon binding to its operator site (Figure 7). The repressor recognizes its operators by its complementarity to a particular DNA conformation as well as by direct interaction with base pairs in the major groove. In the major groove, non-polar contacts appear to be as important as hydrogen bonds.

Specific contacts are made by the side chains of Glu28, -29, and -33, which hydrogen-bond to base pairs 1, 2, and 3 (AT, GC, and AT), respectively, and by van der Waals contacts made by Thr27 and Glu29 with the methyl group of base pair 3 (AT).

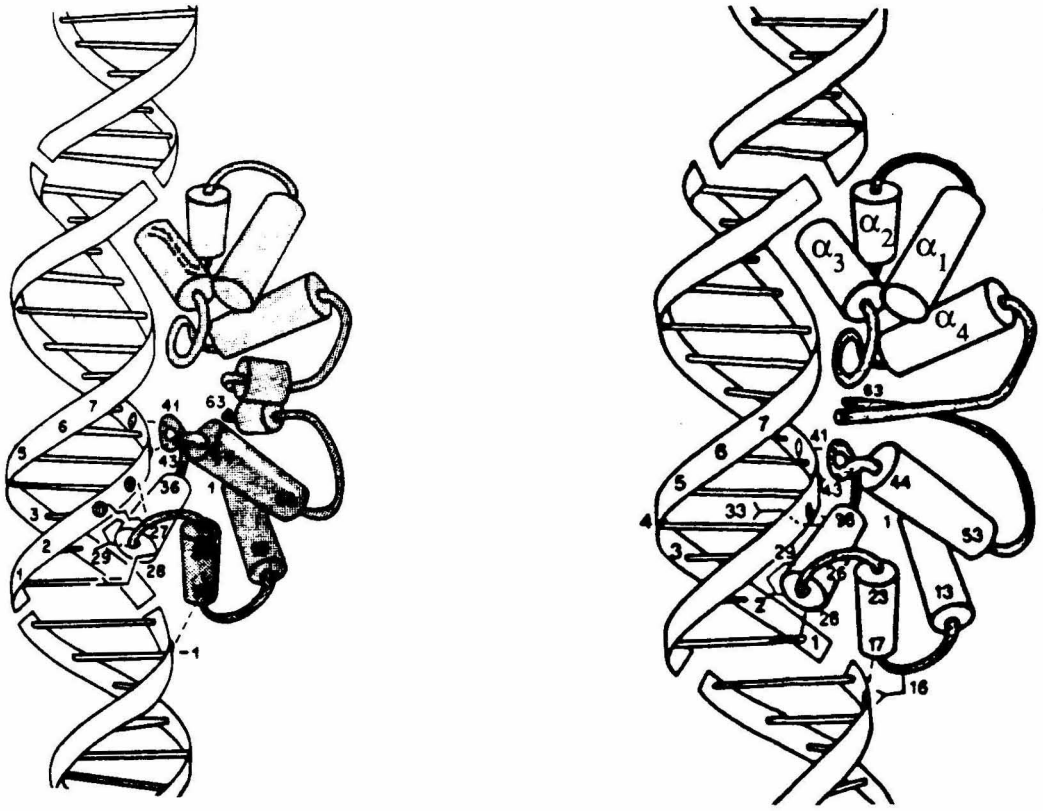


FIGURE 7. Representations of the 434 cro[42,43] and 434 repressor[26] cocrystal complexes. **Left.** 434 cro/14-mer complex. **Right.** 434 repressor-O_{R1} complex.

What is most striking about this cocrystal structure is the extent to which the DNA is distorted from B-form DNA.[41] The middle four base pairs are overwound and compressed such that the phosphate-to-phosphate distance is reduced from 11.5Å for the canonical B-DNA minor groove to 8.8Å in the complex, and significant bending toward the minor groove occurs in the 434 repressor dimer-DNA complex.[16,26] The bending is not smooth: the operator is relatively straight at its center, but it bends symmetrically by

approximately 12° about three base pairs on either side of the center. The middle AT base pairs are twisted out of plane and participate in non-Watson-Crick bifurcated hydrogen bonds. Operators with AT base pairs at the center bind the repressor more tightly than those with GC base pairs.[16]

The cro and repressor proteins of bacteriophage 434 together regulate the switch between lytic and lysogenic growth of the phage. These proteins are far more similar in structure and sequence than the corresponding proteins of bacteriophage λ, which are congruent only in the HTH domains. Crystals of 434 cro (71 amino acids) complexed with the O_{R1} operator, which diffract to 3.2Å in one direction and 5.5Å in the other,[42] show that 434 cro similarly distorts the O_{R1} operator from B-DNA, although there exist major local differences between the DNA conformations of the repressor-O_{R1} complex and the cro-O_{R1} complex. Binding of the protein determines the precise conformation of the DNA operator in both the 434 repressor-DNA and the 434 cro-DNA complexes.[16]

λ REPRESSOR

The λ repressor (1-92) cocrystal with the 20-base pair O_{L1} fragment was resolved to 2.5Å resolution[25]; this complex also reveals that the protein itself undergoes little conformation change upon binding to DNA (Figure 8). The contacts made between the recognition helix and the major groove have been fairly well predicted by model-building studies[9,10]; a number of unanticipated interactions, however, were revealed by the crystal structure. The DNA is slightly less bent than in the 434 repressor and cro structures; twist and other helical parameters vary less and the base-pair planarity is more regular. No contacts are made between the repressor and the minor groove of DNA.

An additional feature of the λ repressor-O_{L1} complex is the contacts made by the amino-terminal arms, which wrap around the DNA and form hydrogen bonds with the major groove bases on the opposite side of the DNA helix (Figure 9).[25,32,44] A crystallographic study on the six-residue amino-terminal arms-O_{L1} complex and genetic and biochemical analysis on mutants of the amino-terminal arms was just published.[44] Previous crystallographic studies showed weak electron density for the arm, and it was believed that such flexible protein segments were unlikely to contribute significantly to DNA recognition. However, this recent study provides a high-resolution view (1.8Å resolution) of the λ repressor amino-terminal arms interacting with the O_{L1} site, and critical contacts are made by this flexible region. The crystal structure was determined at low temperature (-15°) to reduce thermal motion.

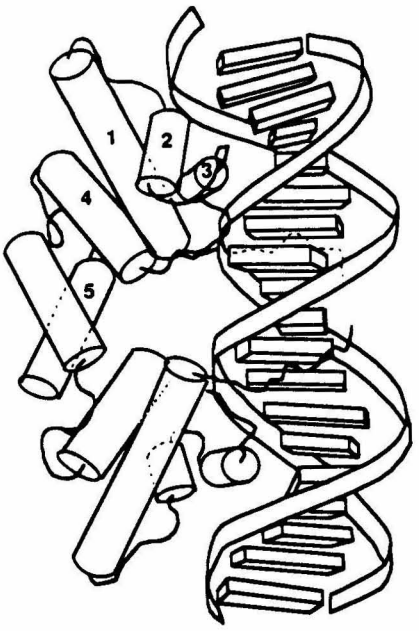


FIGURE 8. Sketch of the λ repressor-O_{L1} complex.[25] Helices in one monomer are numbered 1-5.

The amino-terminal arm of the λ repressor contains the sequence Ser1-Thr2-Lys3-Lys4-Lys5-Pro6. Random mutagenesis was conducted on each codon in the arm, and gel-shift assays were used to determine *in vitro*

binding activities. Residues 3, 4, and 5 are essential for repressor function; only Arg and Lys are acceptable at positions 3 and 5, and only Lys is functional at position 4. Residues 1 and 2 are not constrained to any particular conformation and do not make any important contacts. Lys3 and Lys4 each make hydrogen bonds to two guanines; Lys3 contacts N7 and O6 of two guanines, and Lys4 contacts two O6's of guanines. Interestingly, this amino-terminal recognition element, which plays a major role in sequence-specific DNA recognition, is outside the HTH domain, and deletion of this arm results in a greater than 8000-fold reduction in DNA binding affinity.[45]

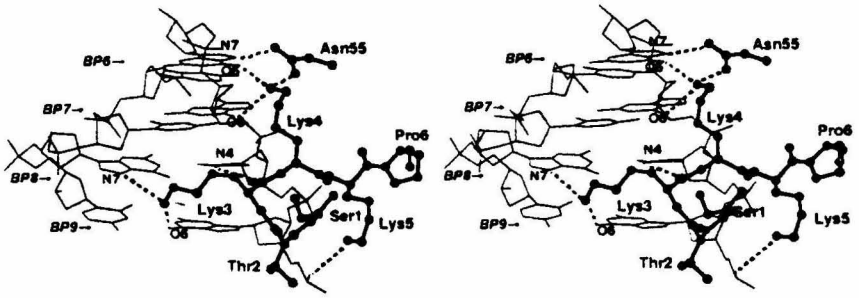


FIGURE 9. Stereo diagram showing the amino-terminal arm that contacts the consensus half site.[44] Lys3 and Lys4 contact guanines in the major groove; Lys5 contacts a phosphate group.

Both the λ cro and the λ repressor bind as dimers to the same regions of λ DNA called the left and right operators, O_L and O_R.[2,3,5,6,46,47] These two proteins comprise the regulatory λ switch between the two physiological states, lytic growth and lysogeny. Recently, the structure of the λ cro with a 17 base-pair consensus operator fragment has been solved to moderate resolution, and further work is currently in progress.[15,16] Like other repressor-operator complexes, the recognition helix of cro lies in the major groove, but preliminarily, it appears that there are significant differences in the structures of the complexed and uncomplexed crystals. The λ cro dimer is

held together by two β-strands from each monomer.[48,49] Upon binding DNA, the β-strands twist such that the monomers, including the HTH domains, rotate by more than 35° with respect to each other. Unlike the λ repressor (1-92)-O_{L1} complex, the λ cro-operator complex reveals that the DNA in the middle is overwound and the minor groove somewhat compressed. These changes were completely unanticipated, for the middle seven base pairs are less-flexible GC base pairs.

TRP REPRESSOR

The *trp* repressor is also a helix-turn-helix regulatory protein controlling the operon for synthesis of L-tryptophan in *E. coli* by a negative feedback loop[50]; in the absence of L-tryptophan, the repressor is inactive, the operon is switched on, and tryptophan is produced. As the concentration of tryptophan increases, it binds to the repressor and converts it to an active form that binds DNA and shuts off the gene. Comparison of the crystal structure of inactive unliganded *trp* aporepressor[51] with that of *trp* repressor[52-55] shows that tryptophan activates the dimer to bind to operator DNA a thousandfold by moving two symmetrical, flexible helical motifs. The dimerization domain is formed by an unusual arrangement of interlocking α-helices from both monomers in which five of each monomer's six helices are intertwined (Figure 10).

The structure of the *trp* repressor bound to an 18-base pair consensus sequence has been refined to 1.8Å resolution,[55] and this structure is quite surprising and unexpected. There are no direct hydrogen bonds or non-polar contacts to the bases that can explain the repressor's specificity for the operator sequence. Rather, the sequence appears to be recognized indirectly through water-mediated contacts and sequence effects on the geometry of the

phosphodiester backbone, which permits the formation of a stable interface. The recognition helix does not lie in the major groove, but rather the positively charged amino terminus of the helix protrudes almost perpendicularly into the major groove and makes dipole interactions with the negatively charged DNA backbone.

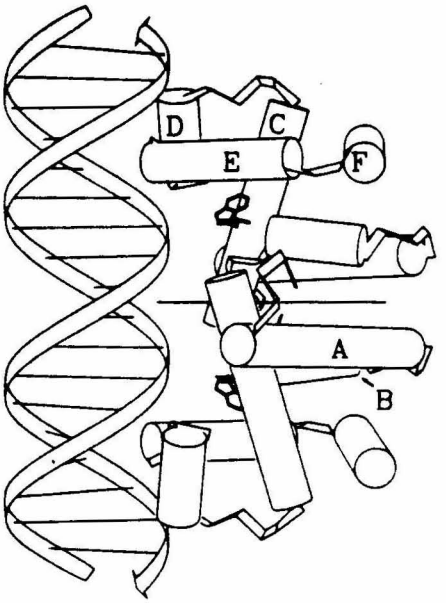


FIGURE 10. Schematic drawing of the *trp* repressor-consensus operator cocrystal structure.[52-55] Helix E is the recognition helix; its amino terminus protrudes perpendicularly into the major groove.

This astonishing mode for a regulatory protein to bind to its operator has caused the *trp* repressor cocrystal structure to be assailed; the conditions of crystallization (low salt and high alcohol) may favor formation of a non-specific complex.[17] Also the observed crystal complex appears to be at variance with *in vivo* genetic studies, which have shown that the oligonucleotide used in the crystallization of the *trp* repressor is not retarded in a gel-shift assay under conditions wherein a shorter oligonucleotide containing a different consensus sequence is retarded, and that methylation protection experiments on the full natural operator and the short consensus oligonucleotide give similar protection patterns.[56] Thus genetic studies appear to indicate that the *trp* repressor was crystallized with the wrong

operator sequence; this may explain why no base-specific contacts are made by protein, and that this cocrystal is simply a non-specific protein-DNA complex.

CATABOLITE GENE ACTIVATOR PROTEIN (CAP)

Very recently, the 3Å resolution cocrystal structure of CAP complexed with a 30-base pair operator showed that the DNA is bent by a whopping 90° (Figure 11).[57] This bend occurs almost exclusively from two 40° kinks that occur between base pairs 5 and 6 on either side of the dyad axis (Figures 11 and 12). These kinks, as well as smaller distortions in the DNA, derive from interactions between the protein and the DNA backbone and provide, in part, for specific binding through sequence-dependent distortability of the DNA. Additionally, sequence specificity is achieved through direct hydrogen bonding interactions between three side chains emanating from the recognition helix of CAP and the exposed edges of three base pairs in the major groove of the DNA helix. When CAP is complexed with its allosteric effector molecule cAMP, it activates transcription at more than 20 different promoters in *E. coli*. It is believed that this bend is an integral part of the mechanism for transcription activation, and that in addition to properly orienting CAP for possible interaction with RNA polymerase, wrapping of the DNA around CAP may result in upstream DNA contacts with RNA polymerase.

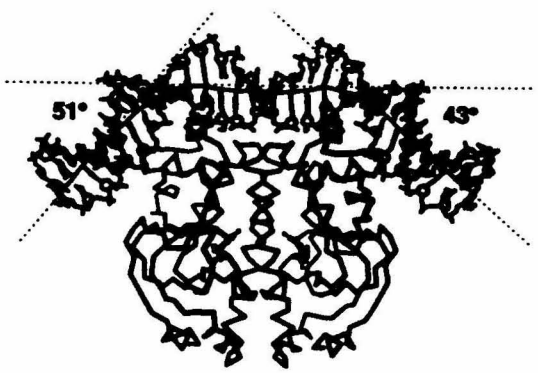


FIGURE 11. Structure of the CAP-DNA complex.[57] The protein is represented as an α-carbon backbone trace. The complex is positioned with the recognition helix perpendicular to the page and lines are extended through the axes of the central ten base pairs and the five terminal base pairs. The measured angles are as shown.

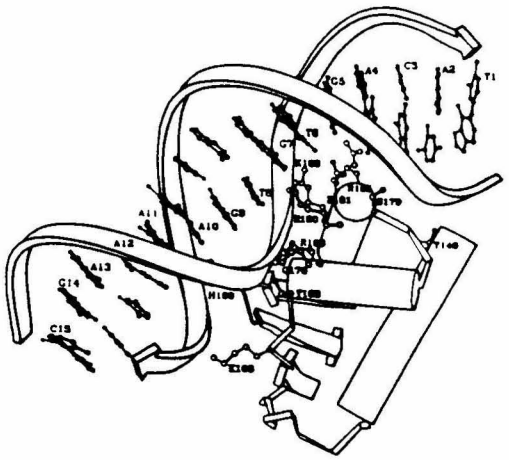


FIGURE 12. Drawing of the potential interactions between one DNA half site and a small domain of the CAP dimer.[57] Protein helices are represented as cylinders and β-sheets as arrows.

LAC REPRESSOR

Although most of the structural data for repressor-operator complexes have been generated by X-ray crystallography, ^{1}H and 2D NMR studies have recently been conducted on proteins and protein-DNA complexes.[58,59] These studies have not only provided a wealth of information, but as opposed to crystal structures, NMR provides solution-phase information on proteins. Recent NMR studies on the *lac* repressor headpiece (residues 1-51, 1-56, and 1-59)[60-62] have revealed dramatic variations on the HTH theme; in the case of *lac*, the orientation of the recognition helix is opposite that of other HTH proteins, namely that the amino terminus of the recognition helix is closer to the operator's center of symmetry which is reversed from other known HTH structures.[63-70] The tilt angle of the *lac* recognition helix is similar to that of other HTH proteins, whereas CAP's recognition helix has the same orientation as other HTH proteins, but its tilt angle is much different.

Before any structural data had become available on the *lac* repressor and its interaction with DNA, *lac* had been inferred to be similar to λ cro on the basis of sequence homology analysis[72]; NMR studies proved, however, that *lac* belongs to a different class of HTH proteins (Figure 13). It has been

suggested that several other regulatory proteins, including the *gal* and *deo* repressors, bind in a similar reverse manner and belong to the *lac* class of proteins[68,69]; the λ repressor and cro, *trp* repressor, 434 and P22 repressors, and CAP belong to the *cro* class of proteins.

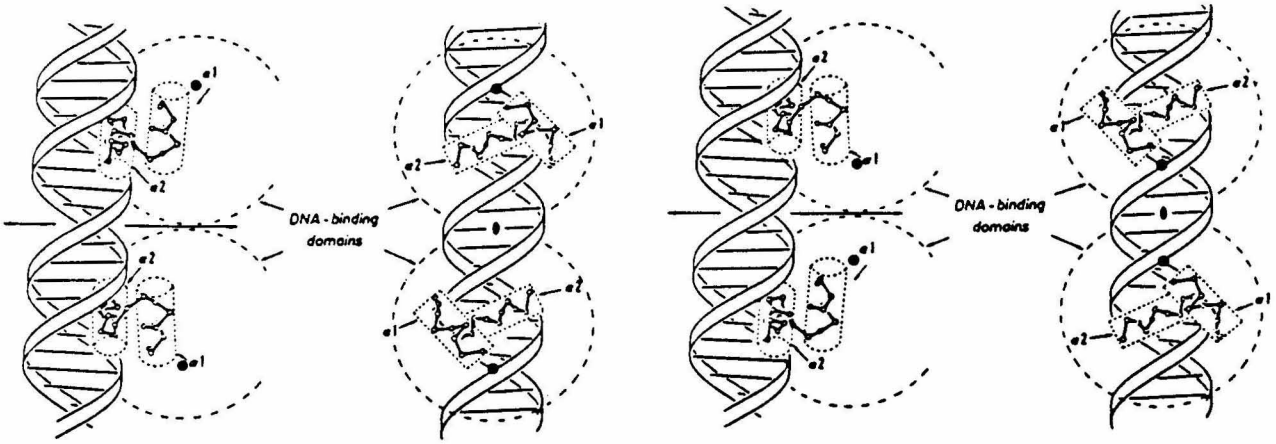


FIGURE 13. The two models for the structure of the *lac* repressor-DNA complex.[71] Heavy filled circles indicate the approximate position of the amino terminus of each subunit. **Left.** Model based on sequence homology analysis; this model is based on the experimentally documented models for the structures of the protein-DNA complexes of the λ repressor, λ cro, 434 repressor, 434 cro, and CAP. The helix-turn-helix motifs of two subunits of the *lac* tetramer are proposed to make equivalent, twofold related contacts with adjacent DNA major grooves. The second α-helix of each helix-turn-helix motif (the recognition helix) is proposed to interact within the major groove. **Right.** Model based on NMR studies; the orientation of the helix-turn-helix motif of each subunit is inverted relative to the orientation in the model on the left.

THE HOMEODOMAIN

The recent advances in our understanding of the genetic control of development are based upon the identification of master control genes that regulate physical development.[17,73-80] Regulatory genes were first identified in prokaryotes, and it became clear that they control the coordinate expression of a set of genes, as in the case of the *lac* repressor's governing its operon or the λ repressor/λ cro regulatory switch.[2] Drosophila geneticists identified a class of mutations called homeotic mutations, which lead to the replacement of one structure by another, resulting in mutated development, such as the

Antennapedia gene mutation in which a leg replaces an antenna in Drosophila. These homeotic genes were found to share a characteristic DNA segment, the homeobox, encoding a precisely defined protein domain of some 60 amino acids known as the homeodomain that functions as the DNA binding region of transcription factors and regulates development of the organism.[79] Homeobox-containing genes have been found in both insects and vertebrates; these genes are clustered and arranged in the same order along the chromosome as they are expressed along the anterior-posterior body axis.

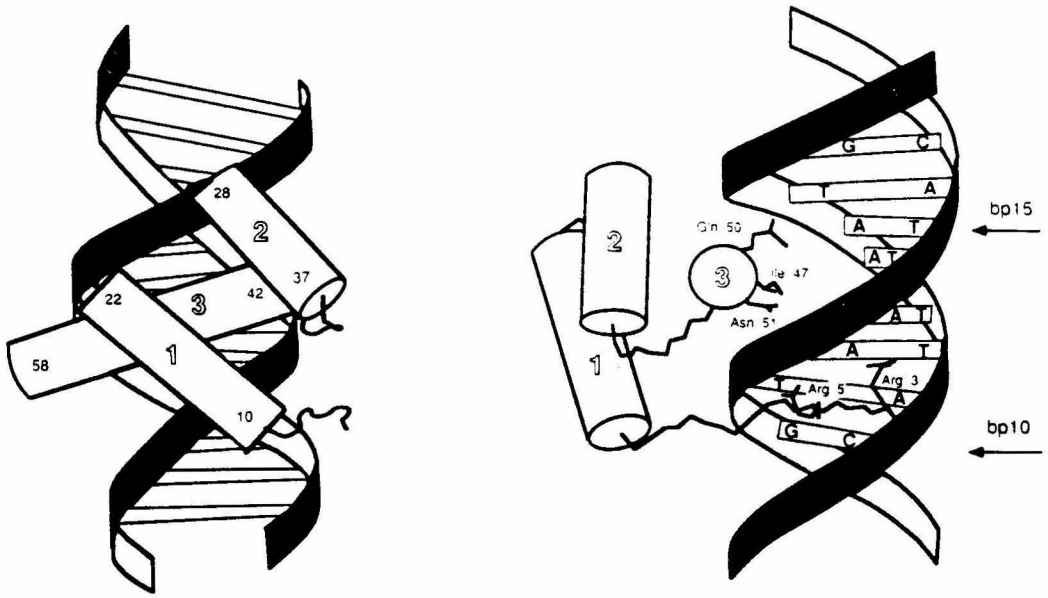


FIGURE 14. Sketch of the *engrailed* homeodomain-DNA complex.[81] **Left.** Front view of the complex; note the length of the recognition helix. **Right.** Side view of the complex; key contacts include side chains from Arg3 and Arg5 interacting with thymines 11 and 12 in the minor groove. Side chains from Ile47, Gln50, and Asn51 of the recognition helix contact base pairs in the major groove.

Homeodomain proteins had been suspected of possessing the HTH motif on the basis of sequence comparisons, although the sequence homology between homeodomains and prokaryotic HTH proteins is not significant.[17]

This HTH domain has been found widely distributed in prokaryotic organisms, and more recently, the HTH has also been discovered to be exploited by eukaryotes, first studied in Drosophila and more recently in mammalian transcription factors. As in the related prokaryotic proteins, the helix-turn-helix motif produces a scaffold which allows a protein to interact specifically with DNA.

The conformation of the homeodomain protein changes only minimally upon binding to operator DNA. The homeodomain binds to DNA with high affinity as a monomer,[81-88] in contrast to prokaryotic HTH proteins (K_D = 1-2 x 10^{-9} M for the *engrailed* homeodomain bound to a consensus operator in buffer containing 100mM KCl and 25mM HEPES, pH 7.6[81]; because as many as 10-15 ions may be displaced as a protein binds to DNA, the affinity of a protein for DNA can vary markedly with salt concentration[13]). The core recognition sequence is TAAT, which is conserved in virtually all homeodomain binding sites; furthermore, additional protein-DNA contacts are found outside the core sequence, and this does not occur in prokaryotic HTH proteins.

During the past year, NMR studies on a 68 amino-acid DNA binding fragment of the *Antennapedia* homeodomain from Drosophila (in the absence of DNA)[83,85] have revealed that its three-dimensional structure contains the HTH motif, which superimposes very well on the HTH unit of λ cro. The HTH unit on *Antp* is also the DNA binding and sequence-specific recognition structure of this eukaryotic protein. Also during the past year, the X-ray crystal structure of the 61 amino-acid *engrailed* homeodomain from Drosophila complexed with a 21 base-pair operator was resolved to 2.8Å resolution (Figure 14)[81]; the *engrailed* structure is the only existing homeodomain-DNA complex that has been solved. The binding site of *en*

has a B-DNA conformation and exhibits no bending or kinking, although there is slight base-pair distortion distal to the TAAT recognition site. Both *en* and *Antp* possess extremely long recognition helices of around 17 residues, as opposed to prokaryotic repressor recognition helices of around 9 residues. The *Antp* recognition helix is actually made up of one longer helix (helix 3) with a short 6 amino-acid helix (helix 4), which extends from helix 3 at a slight angle (Figure 15). Biochemical studies on the *Antennapedia* and thyroid transcription factor 1 homeodomains show that amino acids outside the recognition helix play an important role in determining DNA binding specificities of these two homeodomains.[87]

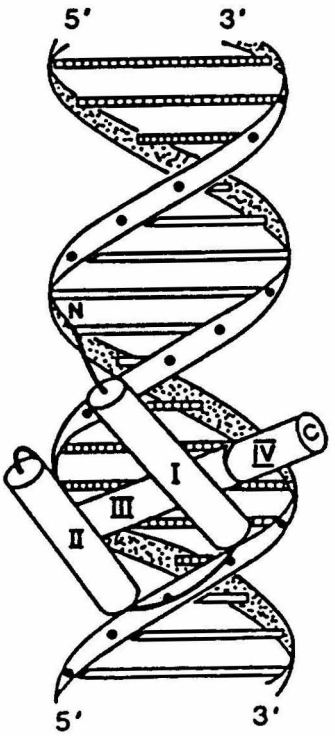


FIGURE 15. Schematic drawing of the *Antennapedia* homeodomain-DNA complex.[83,85] Note that Helix IV is angled away from Helix III.

Another novel aspect of homeodomain proteins is the direct contact made by the protein with the minor groove; this is in contrast to the prokaryotic repressors reviewed above, which make contacts only in the major groove. The amino-terminal arms on both *Antp* and *en* reach around the operator binding site into the minor groove behind; Arg5, a highly

conserved residue, makes a hydrogen bond to O2 of thymine 11 (the first T of the core recognition site), whereas Arg3 appears to contact O2 of thymine 12 (this residue's specific interaction is not well resolved). These minor-groove interactions strengthen the homeodomain's preference for AT versus GC tracts. These base-specific minor-groove contacts differ from prokaryotic HTH-DNA structures in which specific interactions are made only in the major groove.[81]

Recently the integration host factor (IHF), which assists during the integration of the DNA of bacteriophage λ into its host chromosome, was found to recognize DNA primarily through minor groove contacts, an unprecedented mode of recognition for a sequence-specific DNA binding protein.[89] Most sequence-specific DNA binding proteins have been shown to recognize predominantly, if not solely, the major groove, and this mode of interaction was believed to be required for site-specific recognition. The *engrailed* cocrystal structure demonstrates the significance of sequence-specific minor-groove contacts, and IHF shows that a protein can recognize its operator site through predominantly, if not solely, minor groove interactions.

THE LEUCINE ZIPPER

The leucine zipper provides the dimerization domain for a number of eukaryotic regulatory proteins, including the yeast transcription factor GCN4, the mammalian transcription factor C/EBP, and the oncogene products *fos*, *jun*, and *myc*, which all act as transcription factors.[90,91] In all five proteins there is a region of about 30 amino acids in which every seventh residue is leucine; because the periodic repeat of an α-helix is 3.4 amino acids, this region folds into a helix whose leucine residues form a ridge positioned on every other turn on one side of the helix (Figure 16). The leucine zipper

dimer is built up from two parallel helices in which the leucines face each other and intertwine to form a coiled-coil structure.[92-99] The DNA binding residues are immediately amino-terminal of the leucine-zipper dimer; the DNA binding residues are basic and form α-helices of about 20 amino acids, which track along each half of the DNA recognition site in the major groove.[100] Because these arms are so long, their helices are suggested to be kinked to allow them to follow the path of the major groove. This model has been called the "scissors grip" to indicate a Y-shaped molecule in which the arms extend about the DNA (Figure 17).[101]

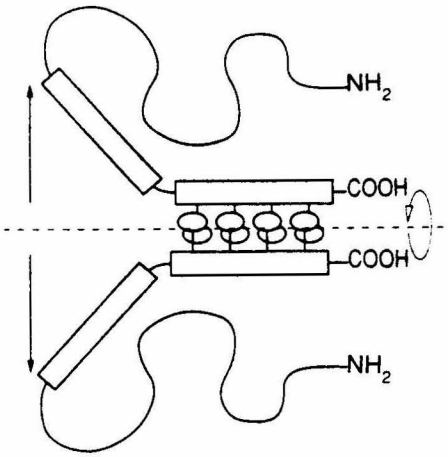


FIGURE 16. Schematic model of the DNA binding domain of C/EBP.[93] Two polypeptide chains are shown in a parallel dimeric conformation generated by specific interactions between the leucine repeat region of each unit. Leucine-repeat helices are indicated as rectangles with protruding leucine side chains. Angled rectangles adjacent to the leucine zipper correspond to the basic region.

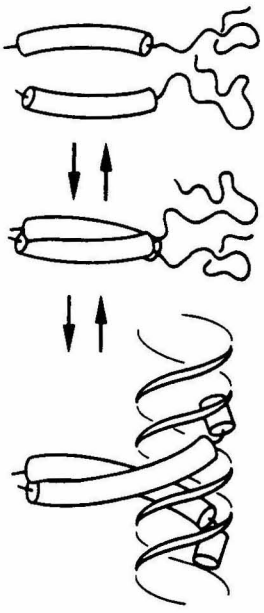


FIGURE 17. Binding reaction between C/EBP and DNA.[98] In the absence of DNA, C/EBP exists in equilibrium between the monomeric and dimeric states. Dimerization is mediated by the leucine zipper. When bound to DNA, the basic regions are presumed to become α-helices.

THE BASIC HELIX-LOOP-HELIX

The basic helix-loop-helix (bHLH) structure is common to a number of proteins involved in cell-type determination or transcriptional regulation.[102,103] This motif has been identified in proteins such as c-*myc*,[102] the muscle determination protein *MyoD*,[104] the Drosophila *achaete-scute* complex involved in neural determination,[74] and the Drosophila cell-type determination protein *daughterless*.[102] The helix-loop-helix structure in the bHLH motif serves as the protein dimerization domain; the conserved hydrophobic residues lie on one side of the helix. Each of these proteins contains homologous regions, which are believed to form two amphipathic α-helices separated by an intervening loop (Figure 18). This motif exhibits a remarkably stringent conservation of hydrophobic residues; in particular, leucines are found every seventh residue in each helix, similar to the leucine zipper structure. The basic region consists of about fifteen positively charged amino acids and is essential for specific binding to DNA. A proposed model of the bHLH dimerization is shown in Figure 18. Unlike the leucine zipper, the helices are proposed to dimerize in an antiparallel fashion. This dimerization domain is the key to transcription regulation.

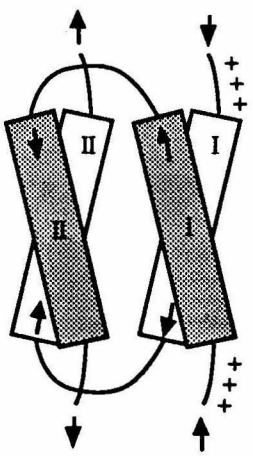


FIGURE 18. Model of the basic helix-loop-helix domains of two protens dimerizing to form a structure capable of bindng to DNA. The basic region is indicated by (+).

Recently the protein Id was isolated and found to be a negative regulator of *MyoD*.[105] Id is missing the basic region that is required for specific DNA binding; thus Id is believed to dimerize with *MyoD* to form a non-functional heterodimeric complex incapable of binding to DNA. The protein NFB has also been found to be a negative regulator of myogenic regulators.[104] On the other hand, Max is a bHLH protein that associates only with *myc* proteins, and this heterodimer binds DNA specifically; thus Max is a positive regulator of transcription.[106] Only the Myc-Max complex bound specifically to DNA, whereas only Max or only Myc did not exhibit appreciable binding under conditions in which the heterodimer did bind.

THE ZINC FINGER

Zinc fingers have been found in hundreds of proteins, and most of these interact with DNA through this motif.[107-109] This motif has been found solely in eukaryotic transcription factors. The zinc finger was first described upon analysis of the sequence of transcription factor TFIIIA from *Xenopus laevis* in 1987.[107] This factor, which regulates transcription of ribosomal 5S RNA, is a 344 amino-acid protein containing nine repeated sequences of about 30 residues each. The repeats have non-identical sequences, but each contains two cysteine residues at the amino end and two histidine residues at the carboxyl end; the last cysteine and first histidine are separated by twelve amino acids, including three hydrophobic residues, all at conserved positions (Figure 19). The cysteines and histidines complex one zinc(II) atom in a tetrahedral conformation, and this Zn(II) is necessary for proper folding of the zinc-finger structure. Zinc exhibits no redox chemistry, unlike copper and iron; redox reactions might prove harmful to DNA, RNA, and even the protein chain by catalyzing hydrolysis.[107] Various metals have been tested for

their ability to restore DNA binding activity to zinc fingers, and only zinc was found to be able to do so, although these peptides can fold into fingers in the presence of various metal ions including Co^{II}, Cd^{II}, and Hg^{II}.[107,108] The solution structure of individual zinc fingers has been studied by NMR.[110-112] Each finger can be described as a minoglobular protein with a close-packed, predominantly hydrophobic core and polar side chains on the surface. These isolated single fingers bind in a non-specific but zinc-dependent manner to DNA.

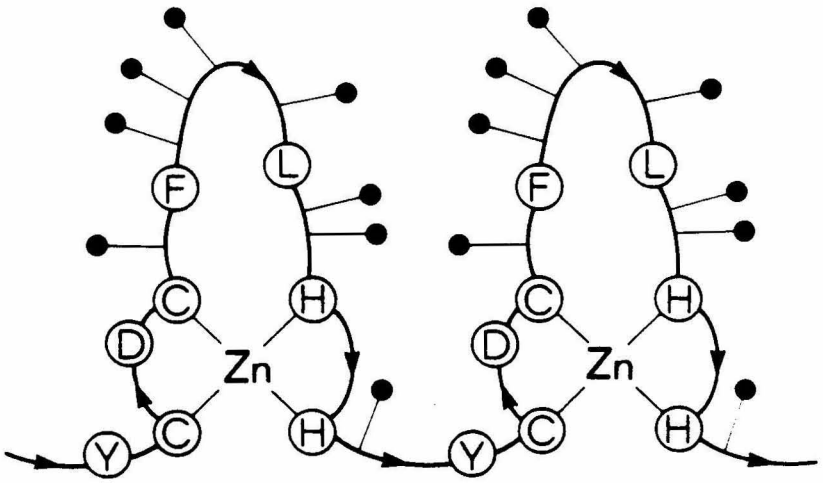


FIGURE 19. Drawing of two individual zinc fingers, each tetrahedrally coordinated to zinc.[107] Circled residues are the conserved amino acids including 2 Cys, 2 His, and a negatively charged Asp, and three hydrophobic residues that form a structural core. Black circles mark possible DNA binding side chains.

A model of the structure of a single zinc-finger domain was put forth based on analysis of sequences of suspected zinc-finger proteins and preliminary studies on a peptide corresponding to a single-finger domain.[113] In early 1988, Jeremy Berg proposed the structure shown below for a single domain (Figure 20)[113]; this was accomplished without the benefit of any three-dimensional structural information. More recent crystallographic and NMR studies have proved the Berg model for a single zinc-finger domain to be

virtually correct, although the several proposed models for the zinc finger complexed with DNA have not been borne out by the recently published cocrystal structure of a three-zinc-finger protein complexed to DNA.

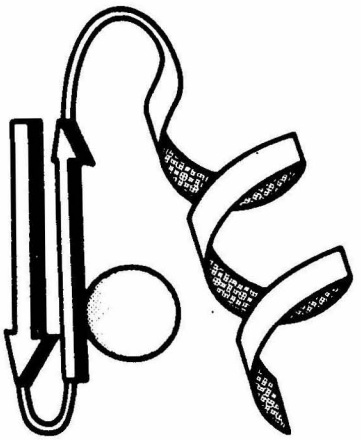


FIGURE 20. Schematic drawing of the proposed zinc finger structure.[113] Arrows on the left indicate the β-sheet structure, and the α-helix indicated by the curled ribbon is on the right. The circle in the middle indicates the zinc atom.

Recently, the cocrystal structure of the three zinc fingers (residues 349-421) of Zif268, a mouse immediate early protein, and a consensus DNA binding site was resolved to 2.1Å resolution (Figure 21).[114] In this complex, the zinc fingers wrap around the major groove of B-DNA. Each finger makes its primary contacts with a three base-pair subsite on one strand of the DNA, which is guanine-rich. This Zif268-zinc-finger peptide binds strongly and specifically to the guanine-rich operator site with $K_D = 6 \times 10^{-9}$. The crystal structure shows each zinc finger to be comprised of a small antiparallel β-sheet containing the two cysteines at the amino terminus packed against a short α-helix containing the two histidines at the carboxyl terminus. Each finger makes nearly equivalent, but not exact, contacts with the DNA. The amino terminus of each helix lies at the bottom of the major groove and is

held there by hydrogen bonds involving the Arg-Ser-Asp sequence at the beginning of each helix; the Arg guanidinium group contacts a guanine and is supported by hydrogen bonds to the carboxylate of Asp. These arginine-guanine contacts appear to be very important in the Zif complex.

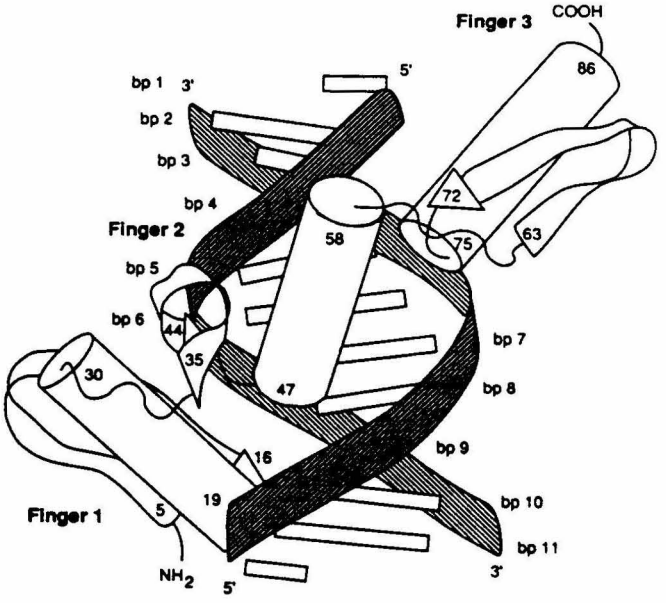


FIGURE 21. Sketch of the Zif268-DNA complex.[114] Helices are shown as cylinders and β-sheets as arrows. Protein-DNA recognition occurs with one strand of the DNA (the bottom strand in the major groove).

Recognition relies quite heavily on specific base contacts in the zinc-finger-DNA structure, much more so than in other known protein-DNA structures; these base contacts seem to play an important role in orienting the fingers, and this observation indicates that zinc fingers may be more flexible and adaptable than other motifs. The helices climb the wall of the major groove such that there is a phosphodiester backbone contact from the carboxyl terminus of the helix. The β-sheet helps to anchor the protein to one side of the major groove. There are almost no contacts between adjacent fingers in the Zif268-DNA complex, so the orientation of adjacent fingers with respect to each other is largely determined by the DNA; this observation is corroborated by NMR studies on a two-finger peptide in which it was discovered that the linker is flexible, and that adjacent finger domains do not interact in the

absence of DNA. Because zinc-finger domains lack much of the scaffolding seen in larger DNA binding domains, they are less likely to be sterically inhibited in their interactions and are likely to have a wider repertoire of possible DNA interactions than seen in the Zif268-DNA complex.[115-118] Indeed, other zinc fingers are known to bind to AT-rich sequences, and footprinting data on other zinc finger proteins, such as TFIIIA and SP1, cannot be reconciled with the geometry seen in the Zif complex.

Steroid Receptor Zinc Fingers

Many hormone receptors have also been found to contain zinc fingers, which use four cysteine atoms to bind zinc rather than the 2 Cys-2 His motif discussed above.[119-121] Members of the zinc-finger hormone-receptor family include the glucocorticoid, thyroid hormone, retinoic acid, and vitamin D3 receptors; these proteins contain a highly conserved DNA binding domain consisting of about 70 residues that bind to transcription-activating sequences called hormone response elements. Another type of zinc finger, including the retrovirus from the viral *gag* gene, Rauscher murine leukemia virus, and human immunodeficiency virus,[115] contains zinc fingers using three cysteines and one histidine to bind zinc. This 3 Cys-1 His motif is not completely limited to retroviruses; a human gene was recently cloned that is involved in recognition of a DNA sequence that had been implicated in sterol-mediated gene repression. This protein contains seven 3 Cys-1 His sequences and specifically binds to only single-stranded binding sites; these results provide further evidence that this motif is utilized for recognition of single-stranded nucleic acids. Clearly, the ubiquitous zinc finger motif can accommodate much variation in sequence and structure and quite possibly, in how it interacts with DNA.

The DNA binding domain of nuclear receptors is characterized by a pattern of eight cysteines which, for the glucocorticoid receptor, coordinate two zinc ions with tetrahedral geometry. NMR studies and the recently published cocrystal structure of the glucocorticoid receptor DNA binding domain-consensus operator complex resolved to 2.9Å shows that these zinc fingers are quite different from the 2 Cys-2 His motif discussed above (Figure 22).[121] Those discussed above act as independent, conformationally stable units each contributing to DNA binding, whereas zinc fingers of the receptor fold together as part of a larger, unified, globular domain. The glucocorticoid receptor DNA binding domain is an 86 amino-acid fragment encompassing residues 440-525, and it binds as a cooperative dimer in successive major grooves of operator DNA; the receptor fragment does not dimerize in solution.

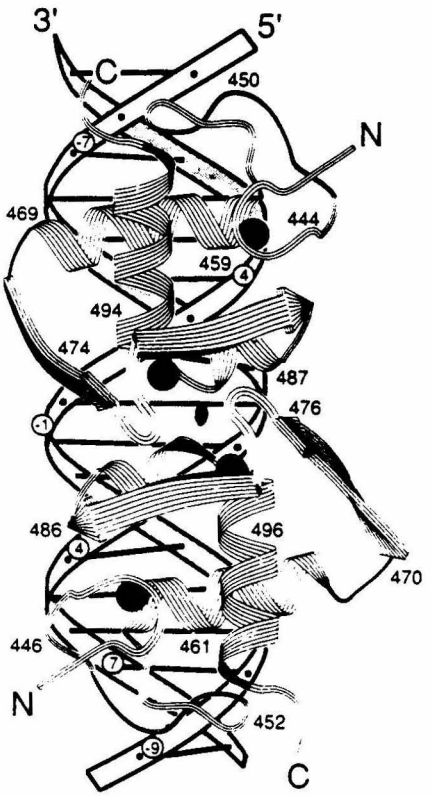


FIGURE 22. Schematic representation of the glucocorticoid receptor-DNA dimer complex.[121] Zinc ions are indicated as filled circles, helices as curled ribbons, and β-sheets as arrows. The dimerization domain lies above the central minor groove, but no protein-DNA interactions occur. There is a zinc atom assisting folding in each monomer's dimerization region and DNA binding region.

Like HTH proteins, the glucocorticoid receptor uses an α-helix to penetrate the major groove of DNA; this recognition helix is stabilized and oriented by a well conserved, local constellation of side chain from other parts of the protein. In this case the zinc center forms the core, whereas the hydrophobic brace forms the core in the HTH motif. Also similar to the HTH, the receptor uses an α-helix from each monomer in adjacent major grooves to recognize DNA. Zinc coordination stabilizes not only the helices in the DNA recognition region, but also the fold that is essential for dimerization. Many transcription factors use zinc (and other metals) to nucleate folding in substructures; this may have evolved from the structural economy and versatility provided by metal chelation.

ECOR I ENDONUCLEASE

The crystal structure of EcoR I endonuclease (276 amino acids) bound to a cognate thirteen base-pair oligonucleotide has been resolved to 3Å (Figure 23).[122,123] Kinetic data show that during the normal catalytic cycle, EcoR I is bound to non-specific DNA, which is not hydrolyzed, for a much larger fraction of time than it is bound specifically to cognate DNA, which is hydrolyzed. EcoR I is not a regulatory protein but a restriction endonuclease that cleaves duplex DNA at the site 5'-GAATTC-3', and it binds DNA as a dimer, each monomer containing a five-stranded β-sheet surrounded by six α-helices. The EcoR I binding site has been crystallized and studied and has served as a classic example of B-DNA.[41] This dimer significantly distorts its binding site; the central base pairs are unwound by 25°, widening the major groove by approximately 3.5Å and allowing the four parallel-α-helix bundle to wedge itself into the major groove.

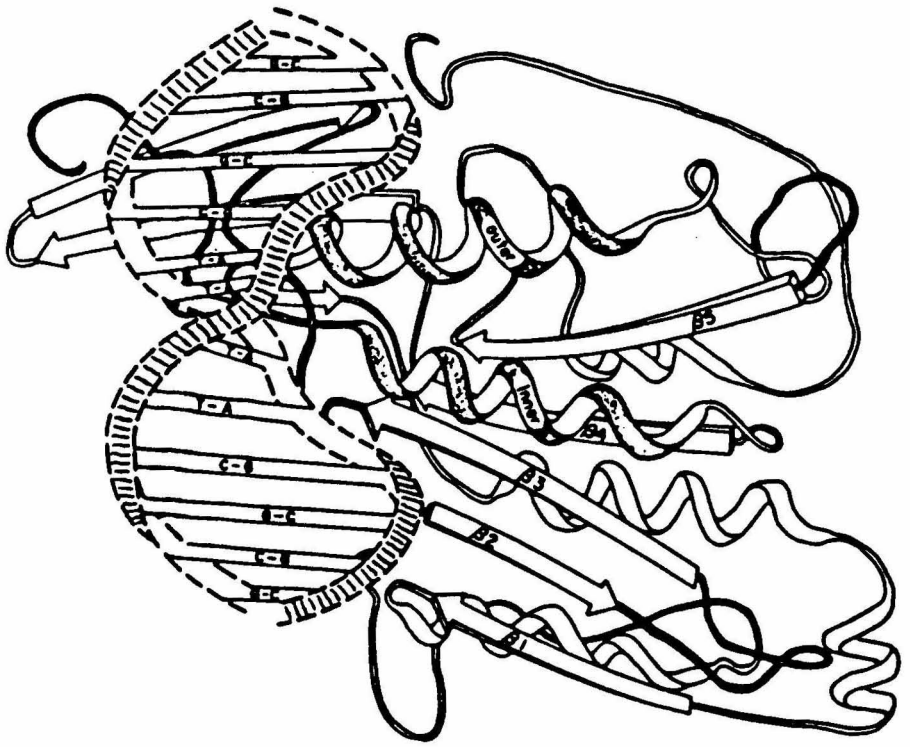


FIGURE 23. Schematic drawing of one subunit from the dimeric EcoR I-DNA cocrystal.[122,123] The other subunit would be rotated 180° about a horizontal line through the center of the first subunit in the plane of the page. Note the extensive kinking and unwinding of the DNA.

Two symmetrically placed kinks at positions +3 and -3 from the binding-site center increase the phosphate-phosphate distance of the backbone. Although the cocrystal structures of regulatory proteins show distortion in the bound DNA, the amount in the EcoR I complex is generally far greater; the CAP-DNA complex, however, shows the greatest amount of DNA distortion among known protein-DNA complexes. These differences may reflect the specific requirements of the two systems. Regulatory proteins act as on-off switches between two sets of genes, and a misreading of the DNA sequence is unlikely to harm the host organism and can easily be remedied. An error incurred by DNA cleaving proteins like EcoR I cannot be corrected and could prove to be very detrimental to the organism. An enzyme that

covalently modifies DNA must be capable of achieving a much higher level of sequence discrimination than other DNA binding proteins. The distortion of DNA by EcoR I also probably serves to catalyze the DNA cleaving reaction; that is, the strain in the phospohodiester backbone is likely to aid the hydrolysis reaction.

ARC AND *MET* REPRESSORS

Over four decades ago, the regulatory system in *E. coli* that controls the biosynthesis of methionine was discovered. Methionine is the precursor to S-adenosylmethionine (SAM), which is the chief donor of methyl groups in a variety of biochemical pathways. The *met* repressor uses SAM as a corepressor to control its own gene as well as genes for enzymes involved in the syntheses of methionine and SAM. The *met* repressor is comprised of 104 amino acids and forms stable dimers in solution.[124,125] Two dimers bind in adjacent major grooves to form a tetrameric complex. The *met* repressor binds operator sequences in tandem arrays such that repression depends not only on the affinity of the DNA protein interaction, but also on protein-protein contacts along the tandem array.[124]

The *met* repressor-SAM corepressor-18 base-pair DNA complex has been determined to 2.8Å resolution.[124] The dimeric *met* repressor is formed by two highly intertwined monomers; each monomer contains three α-helices and one β-strand. Upon dimerization, these β-strands form a two-stranded antiparallel β-sheet, which protrudes from the dimer surface. It is this β-sheet that recognizes and binds to a specific DNA site. NMR studies have deduced a similar structure for the *arc* repressor, which is involved in the switch from lysis to lysogeny in *Salmonella* bacteriophage P22.[126-130] The NMR structure shows a strongly intertwined dimer in which residues 8-14 of

each monomer form an antiparallel β-sheet. In the *arc* repressor study, two *arc* dimers bind in successive major grooves on one side of the DNA helix, similar to the *met* repressor.[130] Thus *arc* and *met* repressors are members of the same family of proteins that use an antiparallel β-sheet DNA binding motif.

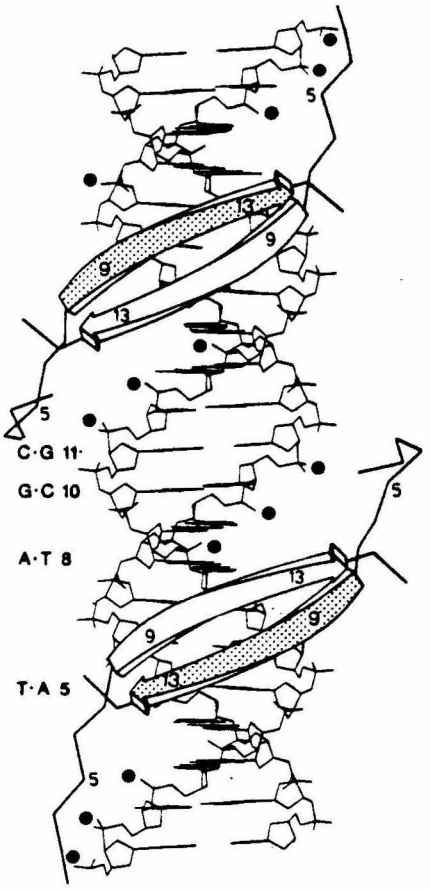


FIGURE 24. Representation of the *arc* repressor-operator complex.[124] Four monomers are shown with two antiparallel β-sheet structures lying in adjacent major grooves.

THE DOUBLE HELIX-TURN-HELIX

C-*myb*, the normal cellular homolog of the retroviral transforming gene v-*myb*, encodes a nuclear transcriptional regulatory protein, $p75^{c\text{-}myb}$, which is involved in regulating mammalian hematopoiesis.[131,132] C-*myb* protein functions in expression of *mim-1*, c-*myc*, *cdc2*, and DNA polymerase

α, and it also activates transcription from HIV-1 long terminal repeat.[133,134] Oncogenic activation of c-*myb* can occur when truncated versions of c-Myb are expressed that give rise to proteins that lack either an amino-terminal phosphorylation site that regulates specific DNA binding or a carboxyl-terminal trans-repressor domain.[135] This amino-terminal phosphorylation site serves as a negative regulatory element; when phosphorylated, sequence-specific DNA binding is inhibited. V-Myb does not possess this phosphorylation site, and is therefore oncogenic, for its activity cannot be controlled.

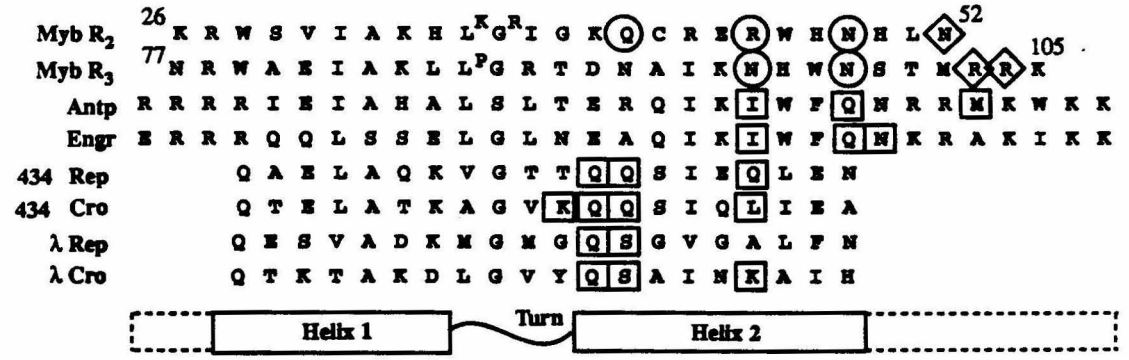


FIGURE 25. Sequence of the recombinant chicken Myb R_2 and R_3 domains aligned with HTH proteins. Amino acids making direct base-pair contacts are boxed. Positions in Myb R_2 and R_3 that strongly affect specific DNA binding when mutated are circled. Positions with moderate effect on DNA binding are marked as diamonds.

The DNA binding domain is located near the amino terminus and is composed of three highly conserved, imperfect 51- or 52-residue repeats designated R_1, R_2, and R_3.[132] Each repeat is believed to contain three α-helices, and the second and third helices are believed to be similar to the HTH motif found in prokaryotic repressors and eukaryotic homeodomains (Figure 25). Only R_2 and R_3 are required for sequence-specific DNA binding. Sequence alignment shows significant similarity between R_2R_3 and the HTH motif. The R_2R_3 fragment is believed to contain two consecutive HTH motifs; the two HTH structures are modeled to bind adjacent major grooves of DNA. A single α-helix, as in most HTH structures, is able to interact with

only 4-6 base pairs because of curvature of the major groove. C-Myb binds DNA as a monomer, but may achieve an extended contact surface with a double HTH motif in which there are two recognition helices lying in consecutive major grooves, thus doubling the number of sequence-specific interactions.[132]

CONCLUSION

Many regulatory proteins can directly recognize specific DNA operator sites by use of the helix-turn-helix structure, in which one helix lies in the major groove of DNA and makes sequence-specific interactions with base pairs and the DNA backbone; the other helix lies on top of this recognition helix and makes interactions with the DNA backbone, thus anchoring the recognition helix in the major groove. Nucleic acid sequences can also be recognized by virtue of the variable distortability among different DNA sequences. Overall, most operator sites show some distortion, which can be very dramatic as in the cases of CAP and EcoR I, and not as dramatic, as in the 434 repressor. Eukaryotic proteins, including the *engrailed* and *Antennapedia* homeodomains, *myc*, *MyoD*, and c-Myb, also utilize helix-turn-helix structures to recognize specific DNA sites; therefore, it appears that a wide array of proteins from all different species use evolutionarily conserved protein motifs to regulate gene expression.

Techniques for studying protein structure, including X-ray crystallography and NMR, are becoming more refined and advanced and are producing a wealth of knowledge about proteins and their complexes with DNA. *De novo* design of proteins much simpler than their natural counterparts, yet containing sufficient information in their sequences to specify a given function (for example, folding in aqueous solution or

membranes, formation of ion channels) have been designed.[136-138] Still, there exists no simple code for recognition between protein side chains and backbone and DNA base pairs and backbone, but gradually and rapidly, much is being learned about the complex relationships between proteins and DNA.

The question of how the amino-acid sequence of a protein specifies its three-dimensional structure has yet to be answered, but gradually this puzzle is being solved. How specific amino acids recognize different sequences of DNA is another question that is gradually being answered, and so far there is no simple code for protein-DNA recognition. Proteins are extremely complex, often large molecules, and it is difficult to discern those features in their sequences responsible for function, structural stability, and binding. Proteins that provide simple model systems have been designed and analyzed in order to understand these features and to provide answers to the questions above. However, much about proteins still remains unknown, although the last three decades have shown extraordinary advances in this area; the next several decades should be very interesting and informative.

CHAPTER TWO

Affinity Cleaving Studies on the DNA Binding Domain of Hin Recombinase

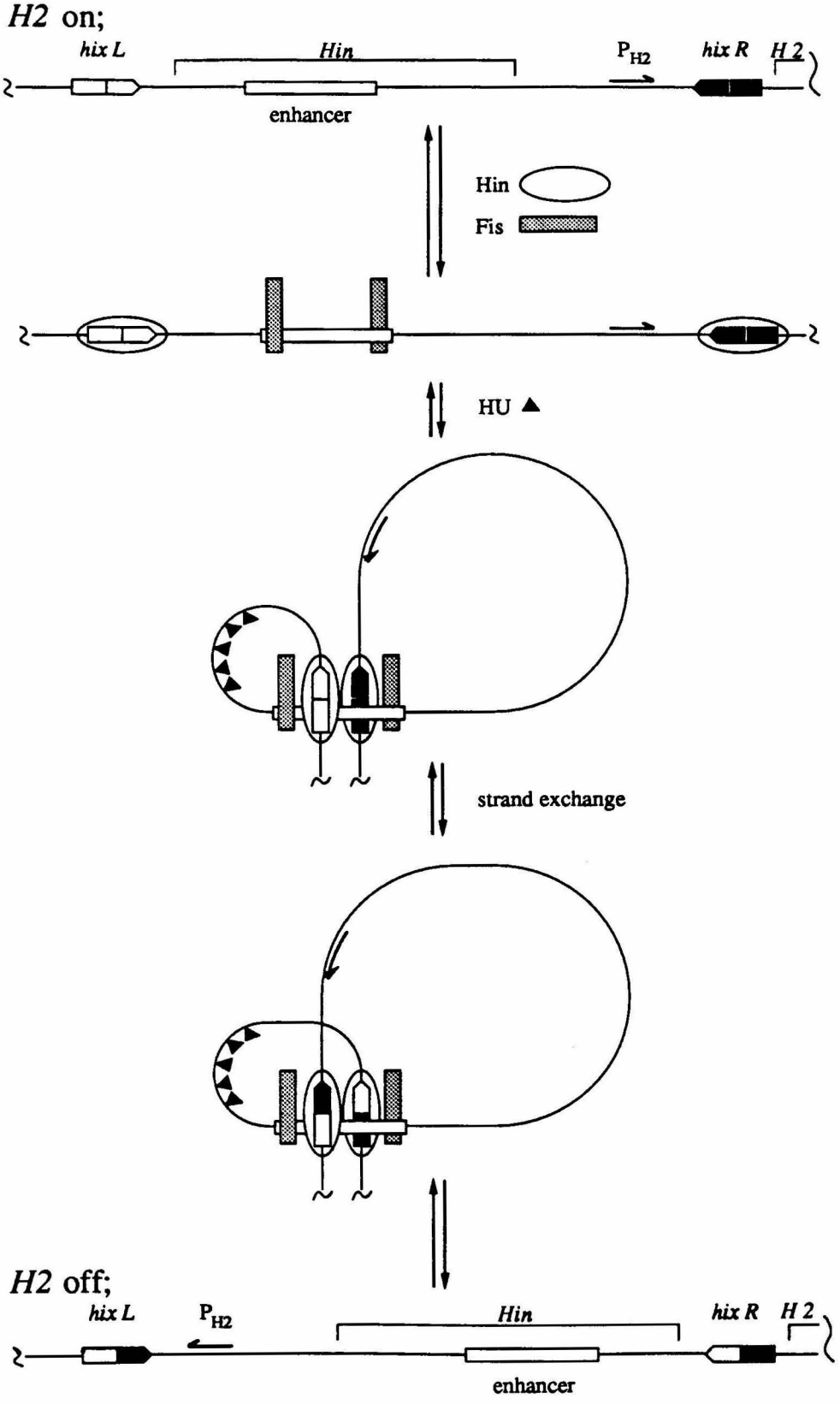
INTRODUCTION

Many sequence-specific DNA binding proteins have been found to consist of two separable domains, the smaller domain making DNA contacts and the larger domain performing catalytic and regulatory functions (discussed in Chapter 1). This allows the study of protein-DNA interactions using only the smaller DNA binding domain. These DNA binding domains still retain the sequence specificity of the intact protein; often, however, the binding affinity, or binding constant, is diminished by one to three orders of magnitude (discussed in Chapter 3). Examples of DNA binding domains that have been successfully studied include residues 1-69 of the 434 repressor,[26,39,40,139] residues 141-183 of γδ resolvase (183 amino acids),[140,141] residues 1-51, 1-56, and 1-59 of *lac* repressor (360 amino acids),[60-62,66,71,142-144] and the intact *engrailed* and *Antennapedia* homeodomains (60 amino acids).[81,83,85]

A few years ago, a synthetic 52 amino-acid peptide comprising residues 139-190 of Hin recombinase (190 amino acids, MW = 21K) was found to interact specifically with the 26 base-pair Hin recombination sites and to inhibit recombination (Figure 1).[81,83,85] Hin, a site-specific recombinase isolated from *Salmonella typhimurium,* switches expression between two different flagellar antigens, specified by the H1 and H2 genes, by inversion of a 996 base-pair segment of DNA.[145-147] Thus, Hin behaves much like a regulatory protein by serving as a switch between two flagellin types.[148] Hin

Figure 1

Scheme for Hin-mediated DNA inversion
(from J. P. Sluka, Ph. D. Thesis, 1988)



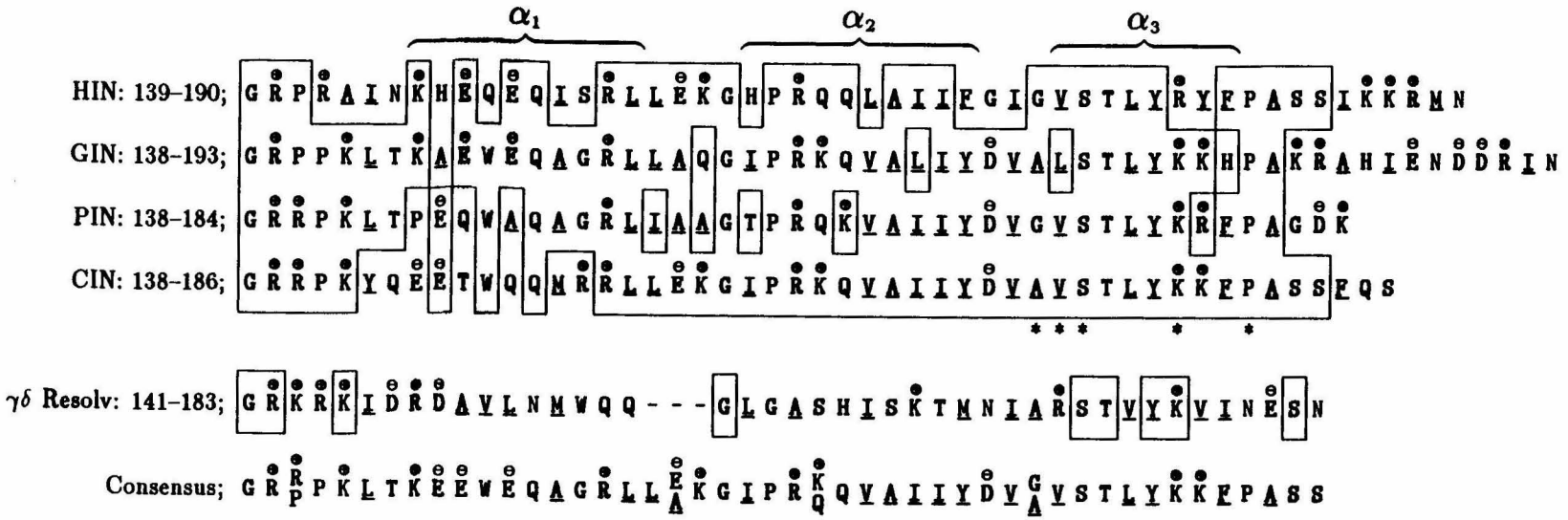


FIGURE 2. Protein sequence homology for the carboxyl-terminal domains of the HIN family of recombinases. Positively and negatively charged amino acids are indicated by circled + and − signs. Hydrophobic residues are underlined and positions that have been proposed to make DNA contacts for CAP, λ cro, and λ repressor are marked with an asterisk (*). As a comparison, the carboxyl-terminal 43-mer chymotrypsin cleavage product of γδ resolvase is included, as well as the consensus 46-mer for the resolvase and four recombinases. HIN shows 67% homology with the consensus sequence, GIN shows 85%, PIN shows 83%, CIN shows 85%, and γδ resolvase shows 20%. Net charges, excluding amino and carboxyl terminal charges are HIN +7, GIN +5, PIN +6, CIN +6, and γδ resolvase +5. (from J. P. Sluka, Ph. D. Thesis, California Institute of Technology, 1988.)

belongs to a family of recombinases including Gin, Cin, and Pin from phage Mu, phage P1, and e14 element of *E. coli*, respectively (Figure 2).

Recombination requires that the supercoiled DNA substrate contain the two 26 base-pair recombination sites *hix*L and *hix*R in inverted configuration and a 60 base-pair, *cis*-acting, enhancer sequence that increases the recombination rate 150-fold (Figure 1).[149] Two factors are required for efficient recombination: Fis (Factor II) is a 12K protein that binds to the enhancer site and assists in the proper alignment and topology for DNA inversion, and HU is a small, histonelike protein that may stabilize the recombination complex.[149-152] Hin binding sites have a nearly twofold symmetry, and the full Hin recombinase protein binds cooperatively. Because the half sites of *hix*L are not perfectly symmetric, Hin has a slightly higher affinity for one site over the other; the higher affinity site is the right-hand site referred to as *hix*L IRR (for Inverse Repeat Right), and the left site is *hix*L IRL. Hin binds to the higher affinity site first and then cooperatively assists a second molecule to bind to the other half site to form a dimer. Thus at low protein concentrations, binding is favored at *hix*L IRR, and as protein concentration increases, binding at both the left and right sites achieve parity as the sites become saturated. The dissociation constant for Hin recombinase binding to *hix*L has been estimated by quantitative DNase I footprinting to be approximately 4×10^{-10} M at 100mM NaCl[153] (the importance of salt concentration for DNA binding is discussed in Chapter 1).

THE AFFINITY CLEAVING TECHNIQUE

The technique most commonly used to examine the nature of protein-DNA contacts in solution is chemical or enzymatic footprinting.[154,155] The DNA bound molecule protects the DNA from chemical or enzymatic

degradation; these reaction products can be separated by gel electrophoresis, and those protected regions will appear as "footprints" or areas with little or no cleavage. Certain chemical reagents can actually determine the binding site to nucleotide resolution, although most footprinting reagents can only give more general information about the binding site and sequence. For instance, detailed information about particular amino acids or structural elements and their location with respect to DNA cannot be deduced from footprinting studies, so it is possible to determine only where a protein (or other molecule) binds to DNA, not how it binds. The affinity cleaving technique, however, has proven to be a powerful tool for more precise studies of molecular interactions with DNA.

Affinity cleaving was developed in the early 1980's in the Dervan group as a means of studying molecular recognition of DNA. By covalent attachment of the iron-chelating moiety ethylenediamine tetraacetic acid (EDTA), a DNA binding molecule could be transformed into a DNA cleaving molecule in the presence of Fe^{II}, reductant, and oxygen.[156,157] Under physiologically relevant pH, temperature, and salt conditions, a reducing agent such as dithiothreitol or sodium ascorbate can initiate the DNA cleavage reaction by the EDTA•Fe moiety. If EDTA•Fe is attached to a sequence-specific DNA binding molecule, cleavage at the molecule's binding sites will occur, and the extent of cleavage will be proportional to the molecule's binding affinity for that site; i.e., a more strongly bound site will exhibit more intense cleavage than that at a weaker site. This DNA cleaving moiety shows no preference for cleavage at particular base pairs or sequences,[158-160] and it imparts no DNA binding affinity to the molecule to which it is attached (discussed in Chapter 3). Therefore, the produced

cleavage is dependent solely on the intrinsic DNA binding ability of the molecule under investigation.

EDTA•Fe-equipped molecules cleave DNA by degradation of the deoxyribose backbone via a diffusible oxidant, presumably a hydroxyl radical.[155-158,160,161] The cleavage reaction is dependent on a reducing agent and molecular oxygen and typically extends over four to six base pairs on both strands of DNA.[158-160] These reaction products can be separated by high-resolution gel electrophoresis, and if the DNA is labelled with a radioactive marker such as ^{32}P, the electrophoresed products can be visualized by autoradiography (Figure 3). By conducting the gel electrophoresis with the cleavage reaction lanes running alongside DNA sequencing lanes, the precise nucleotide location of the DNA cleavage can be assigned. From affinity cleaving data, information about where the molecule binds to DNA can be ascertained, but even more specifically, the location of the structural element to which the EDTA•Fe DNA cleaving moiety has been attached can be deduced.

High-resolution assay of affinity cleaving

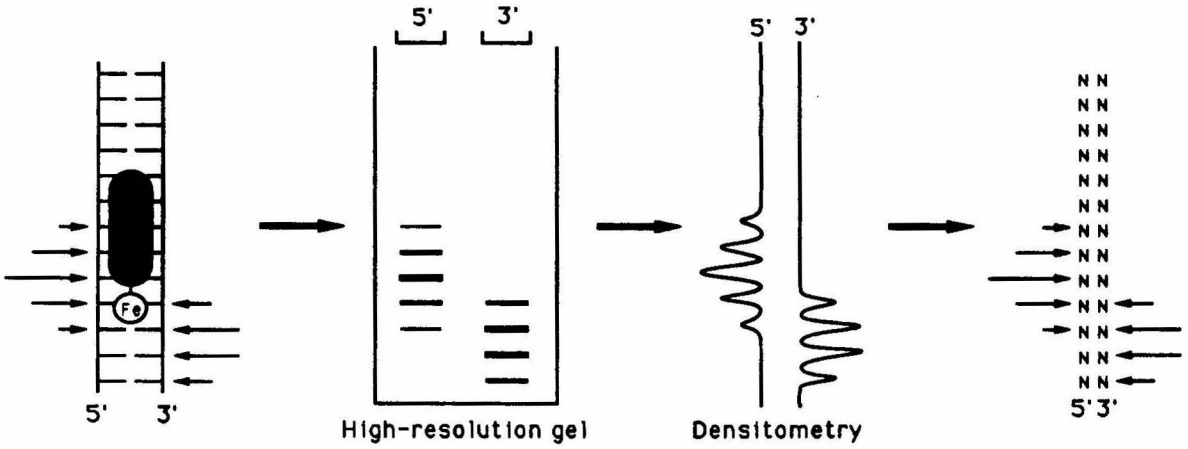


FIGURE 3

Because of the right-handed helical nature of duplex DNA, the groove in which the EDTA•Fe moiety is situated can be identified by complementary strand analysis of the cleavage pattern. An asymmetric cleavage pattern with the maximal cleavage loci shifted to to the 3' side on opposite strands corresponds to EDTA•Fe's being in or above the minor groove (Figure 4). However, the cleavage pattern exhibited when the DNA cleaving moiety is in the major groove is more complex. An asymmetric cleavage pattern with the maximal cleavage loci shifted to the 5' side on opposite strands occurs when EDTA•Fe is in or above the major groove; additionally, cleavage of lower efficiency appears on the distal strands of adjacent minor grooves (Figure 4). A pair of 3'-shifted asymmetric cleavage loci of unequal intensity on opposite strands results.

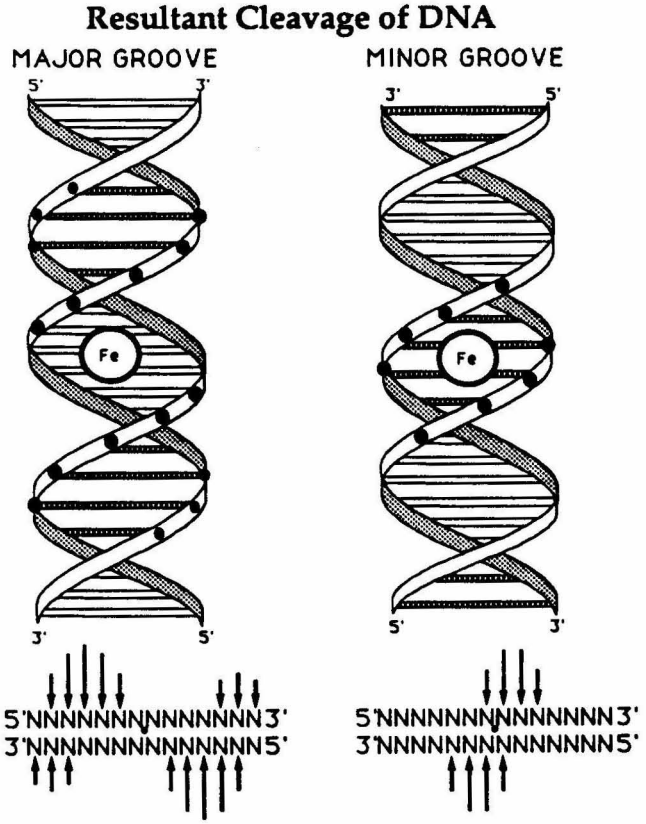


FIGURE 4

These patterns can be explained if the diffusible radical generated by EDTA•Fe reacts preferentially, although not necessarily exclusively, in the

minor groove; thus, in the case when EDTA•Fe lies in the major groove, radicals must diffuse into the two adjacent minor grooves in order to effect cleavage. In support of this view, a sinusoidal cleavage pattern is obtained when DNA bound to a precipitate of calcium phosphate is allowed to react with free EDTA•Fe, demonstrating the different reactivities the two grooves exhibit toward EDTA•Fe.[161]

The Dervan group has used affinity cleaving to study sequence-specific recognition of duplex DNA by naturally occurring antibiotics[159,160] and their designed analogues,[162-164] oligonucleotide triple-helix formation,[165-168] and proteins.[71,100,141,169,170] From these studies, much information has been gained about how these molecules interact with DNA and about the different cleavage patterns generated by EDTA•Fe; this data base assists in the interpretation of affinity cleaving results, which can become very complicated with proteins. Incorporation of EDTA•Fe at discrete amino acids of a protein should allow the location of these modified residues to be mapped to high resolution. The location of EDTA•Fe can be changed in order to examine structural elements of the protein-DNA complex; e.g., the location of the amino and/or carboxyl termini. Alternatively, EDTA•Fe can be kept in the same place and changes can be made in the protein; the influence of these changes on DNA binding can be ascertained by studying the affinity cleaving patterns.

SYNTHESIS AND ATTACHMENT OF EDTA DERIVATIVES TO PROTEINS

Proteins, including monoclonal antibodies, with covalently linked metal chelators have been prepared by "shotgun" approach: typically, metal chelators have been randomly coupled to available nucleophilic sites (e.g., the ε-amino on lysine[171] or the sulfhydryl on cysteine) on natural proteins. These

methods are somewhat problematic, for the number of chelators and their positions is variable and uncontrolled. Usually in these experiments, the chelators serve as molecular tags to allow monitoring of protein activity and location, so controlling the number and sites of chelation is not important. For structural studies of proteins, however, knowing the number and position of the site(s) of chelation is absolutely crucial. This research group has therefore developed methods for incorporating the metal chelator EDTA at specific sites on a protein and has performed structural studies on EDTA-derivatized proteins bound to DNA using the affinity cleaving method.

Synthetic route to tribenzyl-EDTA-GABA (BEG)

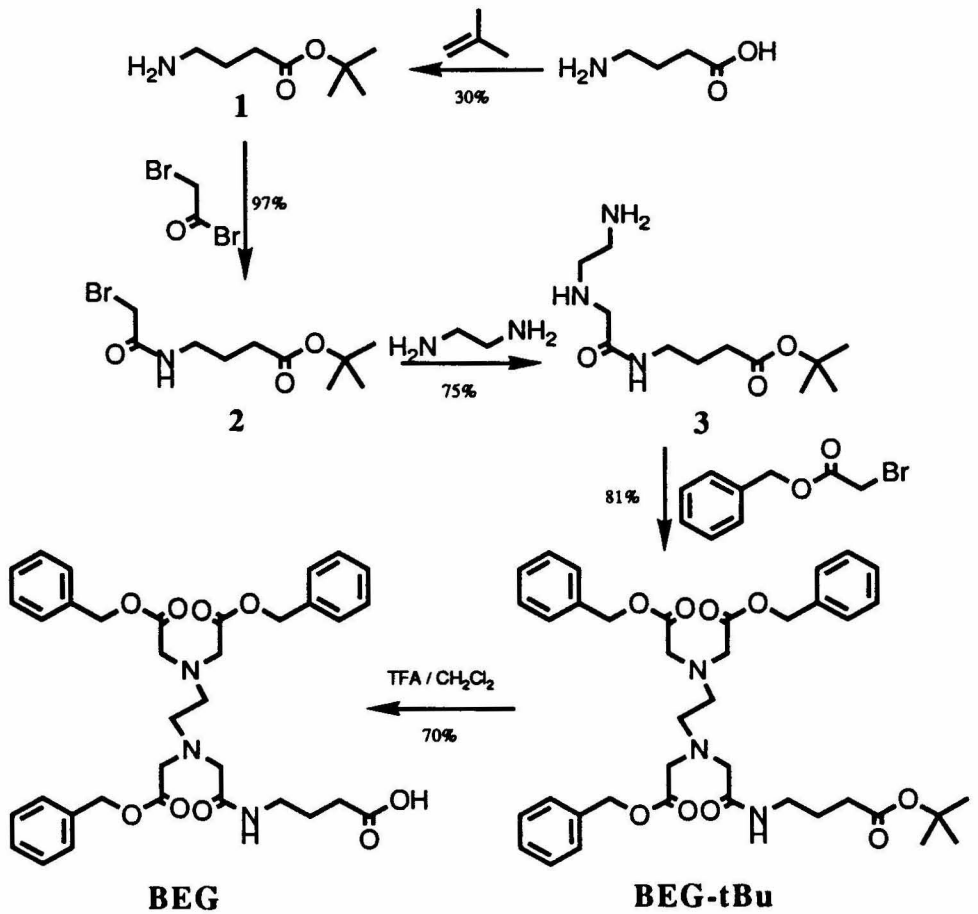


FIGURE 5

Coupling of BEG to Hin(139-190)

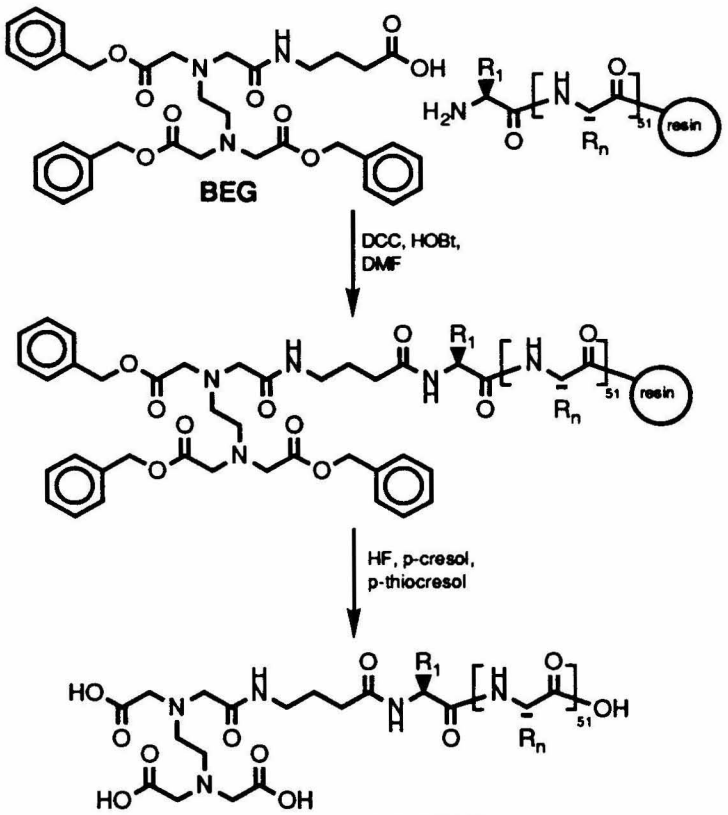


FIGURE 6

Initial affinity cleaving experiments were conducted with the DNA binding domain of Hin recombinase.[153,169] Jim Sluka attached a protected EDTA derivative compatible with Merrifield solid phase peptide synthesis to the amino terminus of the Hin DNA binding domain (carboxyl-terminal residues 139-190). Three of the four carboxylate arms of EDTA were protected as benzyl esters; the fourth carboxylate was coupled via an amide bond to a γ-aminobutyric acid (GABA) linker to minimize steric interference between the EDTA chelate and the Hin protein structure (Figure 5).[172] The resulting tribenzyl-EDTA-GABA (BEG) was suitable for coupling to the amino terminus of a protected, resin-bound, synthetic peptide; this coupling could be performed under the same conditions as that for coupling N-*t*-butoxycarbonyl (Boc) protected amino acids (Figure 6). BEG was coupled to the peptide as its

hydroxybenzotriazole (HOBt) ester in high yield, and cleavage of the peptide from resin by hydrofluoric acid yielded the EDTA-GABA-Hin protein. Because it does not possess a primary or secondary amine capable of further chain extension, BEG is suitable only for capping an amino terminus.

Coupling of Tricyclohexyl EDTA to Hin(139-190)

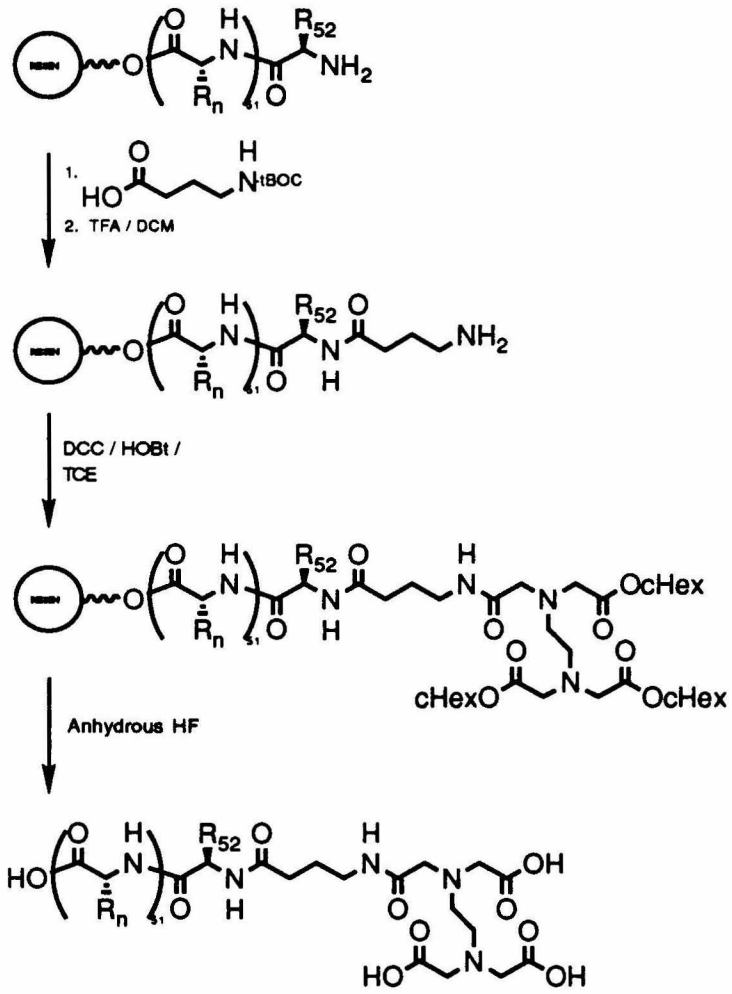


FIGURE 7

More recently, an EDTA derivative has been developed by John Griffin as an alternative to BEG; this tricyclohexyl ester of EDTA (TCE) offers more flexibility than BEG in that a variety of linkers, or none at all, can be used.[172]

The linker we have been using is Boc-protected γ-aminobutyric acid, so peptides made with BEG or TCE are identical in linker and metal chelator composition. TCE and Boc-GABA are, however, much simpler to synthesize than BEG (Figure 7).

PREVIOUS WORK ON THE HIN RECOMBINASE DNA BINDING DOMAIN

Studies on the Amino Terminus of HIN(139-190)

Previously, Jim Sluka performed initial studies on the Hin recombinase DNA binding domain.[153,169] In order to define the DNA binding domain of Hin, he synthesized two peptides of 31 and 52 amino acids from the carboxyl terminus (residues 160-190, 139-190, respectively; Figure 8). The Hin(139-190) 52-mer was chosen by comparison of the sequence of Hin with that of the related γδ resolvase of transposon γδ. Hin and γδ resolvase share 35% sequence homology; the 52 amino acids at the carboxyl terminus of Hin are homologous to the 43 amino-acid chymotrypsin cleavage product of γδ resolvase (Figure 2). The 31-mer was chosen for synthesis because it was believed to contain the helix-turn-helix motif of Hin (residues 160-190); this segment was identified by sequence homology analysis of Hin with regulatory proteins shown in Table 1 of Chapter 1.

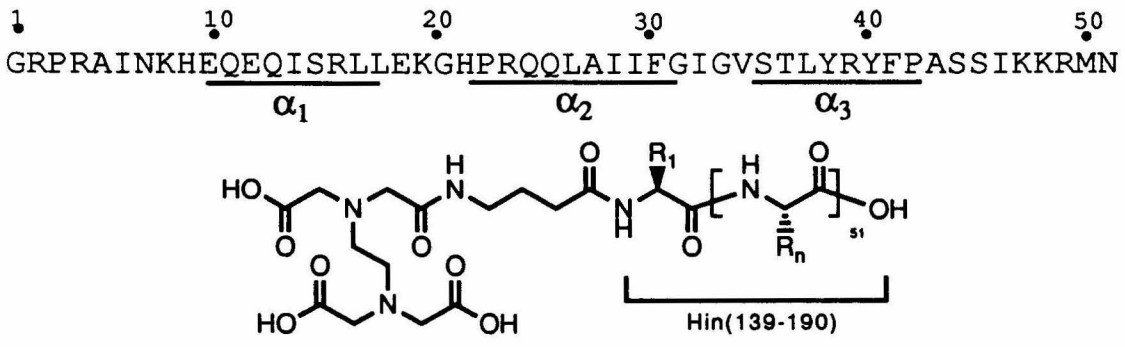


FIGURE 8. Top. Residues 139-190 comprising the DNA binding domain of Hin; α-helices are underlined. **Bottom.** The EDTA metal chelator and GABA linker attached to Hin(139-190).

The 31- and 52-mers were prepared by solid phase synthesis techniques.[173-175] In this method, synthesis begins with the carboxyl-terminal amino acid and builds up sequentially residue by residue. Each residue's reactive amino terminus is protected by the Boc group; the peptide synthesis cycle begins with deprotection of the amino terminus of the growing peptide chain by acid followed by a base-neutralization step. The next amino acid is then coupled to the free amino terminus of the peptide chain, and the cycle is repeated. After completion of the peptide syntheses, the 31- and 52-mers were cleaved from the resin bead using anhydrous hydrogen fluoride; the crude, fluffy peptide solid was purified by reverse phase high performance liquid chromatography (HPLC). The concentration of the purified peptide was assessed by measuring its ultraviolet absorbance at 275nm, using the extinction coefficient of 2810 $(M \bullet cm)^{-1}$, which is the extinction coefficient for both the 31- and 52-mer peptides containing two tyrosines [ε_{275} = 1405 $(M \bullet cm)^{-1}$ for tyrosine]. Nanomolar quantities of peptide were then packaged and stored at -20°.

The 31-mer did not bind DNA with sequence specificity or inhibit Hin-mediated recombination; just the helix-turn-helix domain, therefore, appeared to be insufficient for sequence-specific DNA recognition. More of the protein is necessary for the HTH structure to fold and pack properly. The 52-mer, however, was found to bind to Hin-specific sites on DNA and to inhibit recombination. The 52-mer bound to *hix* sites as did the full Hin protein with the exception of three base pairs at the center of the dimeric site; because the 52-mer was missing the Hin dimerization domain, this result was not surprising. In the DNA fragment used in Sluka's studies, the major 52-mer binding site is *hix*L. Another major cleavage site, called secondary Hin, is located just upstream from *hix*L, and although the site is not required for

Hin-mediated recombination, its location suggests that it is involved in autoregulation of Hin expression, for it is located just upstream from the initiation codon of the Hin gene. A third, very weak binding site called tertiary Hin lies between *hix*L and the secondary site; usually only one cleavage pattern is visible at this site (bound at the right half site), but at very high protein concentrations, light cleavage can also be seen at the left half site. DNA gel mobility retardation assays gave a binding constant of $7.1 \times 10^6\ M^{-1}$ (20mM NaCl, $\Delta G \approx -9.3$ kcal/mol) for the 52-mer at *hix*L, so the binding affinity of the DNA binding domain is about two to three orders of magnitude less than that of the full Hin protein. These results suggested that Hin, like many other DNA binding proteins, consists of discrete, separable domains, and that the carboxyl terminal domain of Hin is responsible for DNA recognition and binding.

Shorter and longer versions of the 52-mer were studied by Sluka; these included the 33, 37, 45, 49, 50, 51, 56, and 60 amino-acid, carboxyl-terminal sequences of Hin (Figure 9). The 33, 37, and 45 amino acid proteins showed no ability to bind DNA with sequence specificity. The 49, 50, 51, 56, and 60 amino-acid proteins could be footprinted and their EDTA•Fe derivatives cleaved DNA with sequence specificity; interestingly, gel retardation assays did not detect discrete binding of the 49-, 50-, and 51-mers at *hix*L. The 56- and 60-mers gave broad cleavage patterns that were difficult to decipher. The 51-mer, which differs from the 52-mer only in the removal of the amino-terminal glycine, showed footprinting and affinity cleaving patterns virtually identical to the 52-mer at 0.5μM protein concentrations; however, the fact that the 51-mer did not produce a gel shift (actually, it produced a smear) indicates that its binding constant must be lower than that for the 52-mer. The 49- and 50-mers, however, had completely lost binding affinity for *hix*L, even at high

Hin(160-190)	HPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	31-mer
[Fe•EDTA]Hin(158-190)	(EDTA-GABA)KGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	33-mer
[Fe•EDTA]Hin(154-190)	(EDTA-GABA)RLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	37-mer
[Fe•EDTA]Hin(146-190)	(EDTA-GABA)KHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	45-mer
[Fe•EDTA]Hin(142-190)	(EDTA-GABA)RAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	49-mer
[Fe•EDTA]Hin(141-190)	(EDTA-GABA)PRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	50-mer
[Fe•EDTA]Hin(140-190)	(EDTA-GABA)RPRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	51-mer
[Fe•EDTA]Hin(139-190)	(EDTA-GABA)GRPRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	52-mer
[Fe•EDTA]Hin(135-190)	(EDTA-GABA)GRLGGRPRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	56-mer
[Fe•EDTA]Hin(131-190)	(EDTA-GABA)ARAQGRLGGRPRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	60-mer
[Fe•EDTA]Hin(139-148)	(EDTA-GABA)GRPRAINKHE-CO_2^-	10-mer

FIGURE 9. Amino-acid sequences of synthetic proteins (proteins synthesized by J. P. Sluka). With the exception of Hin(160-190), each protein was prepared with and without EDTA-GABA at the amino terminus. On the right are listed the corresponding sizes of each protein.

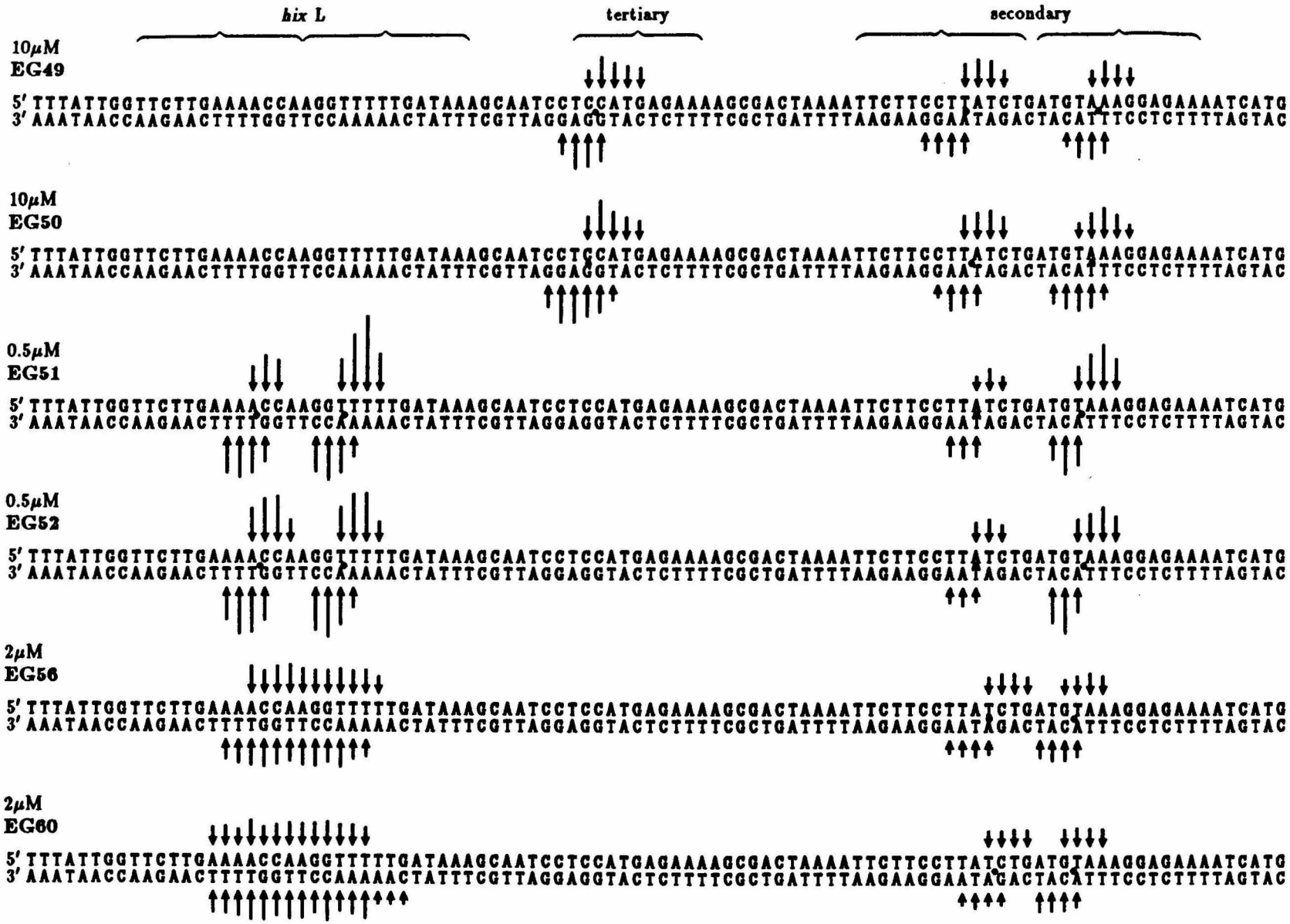


FIGURE 10. Histograms of synthetic proteins listed in Figure 9. Brackets mark *hix*L, tertiary, and secondary sites. Height of arrows indicates extent of cleavage.

protein concentrations (Figure 10). The 49- and 50-mers still retained binding specificity for the secondary Hin binding site, and at high concentrations, these proteins also picked up a tertiary site; the 51- and 52-mers also bound to the tertiary site at high concentrations.

The most interesting observation was that the only difference between the 50- and 51-mers is an arginine; without Arg140, the 50-mer, and for that matter the 49-mer, had a drastically reduced DNA binding affinity, and more significantly, the sequence specificity of these proteins was vastly different from that of the 51- and 52-mers, which reflect the sequence specificity of the full Hin recombinase. It must be noted that Arg140 is outside the helix-turn-helix domain of Hin, and affinity cleaving studies have shown that the amino-terminal residues of the 51- and 52-mers lie in or above the minor groove; these results are quite different from many other prokaryotic HTH repressors, which employ just the HTH domain to recognize the major groove and do not make contacts with the minor groove. Because of the importance of Arg140, a 10-mer consisting of the amino terminus of the 52-mer, residues 139-148, was prepared to study the amino-terminal interactions with DNA; the 10-mer did not bind with sequence specificity.

Studies on the Carboxyl Terminus of Hin(139-190)

Jim Sluka also incorporated the EDTA•Fe DNA cleaving moiety at the carboxyl terminus of Hin(139-190) in order to study whether Hin's carboxyl terminus (Asn190-Met189-Arg188-Lys187-Lys186-Ile-185), like the highly homologous amino-terminal arm of the λ repressor (Ser1-Thr2-Lys3-Lys4-Lys5-Pro6, discussed under the λ repressor section in Chapter 1), wraps around the DNA and binds in the major groove on the other side of the operator site. Met189 was replaced by lysine, and the amino protecting group FMOC on the lysine side chain was removed, and the free amine was capped

with BEG. Placement of EDTA-GABA at the second residue from the carboxyl terminus decreased the Hin(139-190) DNA binding affinity somewhat. This molecule did not give a clear affinity cleaving pattern, and it appeared that the carboxyl terminus is not held rigidly or close to the DNA. Although there is a high degree of sequence homology between the amino-terminal arm of the λ repressor and the carboxyl-terminus of Hin, Hin's carboxyl-terminal interactions with DNA do not appear to contribute to sequence-specific DNA binding as in the case of the λ repressor.[44] The carboxyl-terminal domains of Hin, Gin, Cin, and Pin, a highly homologous family of recombinases (Figure 2), are not conserved, corroborating the conclusion that the Hin carboxyl terminus does not interact closely with the binding site and is not an integral factor in sequence-specific binding.

THE HIN RECOMBINASE DNA BINDING DOMAIN

Amino-Terminal Mutants of Hin(139-190)

Because of the previous work done by Sluka on the amino-terminal residues of Hin(139-190), it was decided to pursue studies on the interactions between the Hin(139-190) 52-mer and the minor groove of DNA. Sluka's work demonstrated that Arg140 (the penultimate residue from the amino terminus of the 52-mer DNA binding domain), which interacts with the minor groove of DNA, was necessary for the sequence specificity of Hin(139-190) for *hix*L. Like the integration host factor protein (IHF) discussed under the section on homeodomains in Chapter 1, Hin recombinase is one of few known proteins in which interactions in the minor groove encode DNA binding specificity. The protein must therefore extend from the helix-turn-helix domain in the major groove over the phosphodiester backbone into the minor groove. The central six base pairs (CCAAGG) of *hix*L comprise the

recombination site and are not bound by Hin(139-190), and thus the binding site domains are separated by one turn of the DNA helix. Similarly, γδ resolvase DNA binding domains are separated by one turn of DNA, but other regulatory proteins, including the λ repressor and cro, 434 repressor, and CAP, bind to adjacent major grooves of DNA.

The five base pairs in the major grooves of the *hix*L, secondary, and tertiary Hin recognition half sites are all well conserved: these highly conserved base pairs in the major groove interact with the helix-turn-helix structure of Hin(139-190) and are important for sequence-specific recognition. *Hix*L IRL and IRR major grooves contain sequences 5′-TTCTT and 5′-TTATC, respectively; secondary IRL and IRR, 5′-TTCTT and 5′-TTCTC; and tertiary, 5′-TTCTC. Thus it may appear that all five half sites will have approximately equivalent affinities for the helix-turn-helix component of Hin(139-190). The following three-base pair sequences contained in the minor groove component of the recognition site appear to distinguish the relative binding affinities of the sites from each other: for *hix*L, the minor-groove sequence is 5′-AAA; for secondary, 5′-TTA; and for tertiary, 5′-GGA. These minor-groove sequences reflect the DNA binding affinities of Hin(139-190) for its binding sites: *hix*L (5′-AAA) > secondary (5′-TTA) >> tertiary (5′-GGA).

By shortening the 52-mer Hin(139-190) to the 51-, 50-, and 49-mers by removing glycine, arginine, and proline, respectively, from the amino terminus, Sluka found that the relative affinities for the five binding sites changed markedly from that of Hin(139-190), where *hix*L > secondary >>tertiary (5′-AAA > 5′-TTA >> 5′-GGA). The 49- and 50-mers also had greatly reduced affinity for DNA, requiring 20-fold higher concentration than that for the 51- and 52-mers for comparable cleavage to be observed. The 49- and 50-mers also have binding affinities reversed from that of the 52-mer in that

they prefer the secondary and tertiary sites over *hix*L: secondary (5'-TTA) ≥ tertiary (5'-GGA) >> *hix*L (5'-AAA). The 49- and 50-mers did not cleave at *hix*L; at 10μM 49- or 50-mer, cleavage at the secondary site was comparable to that of the 52-mer at only 0.5μM. Cleavage at the tertiary site was also strong with the 49- and 50-mers at 10μM concentration, nearly as strong as at the secondary site. These cleavage patterns for all the shortened proteins are 3'-shifted and remain in the same location as those for the 52-mer.

In order to study the Hin(139-190) amino terminus-minor groove interaction, several mutants of the Hin(139-190) 52-mer, in which the penultimate residue from the amino terminus (Arg140) was mutated, were synthesized by Suzanna Horvath at the Caltech Microchemical Facility (Figure 11). Benzyl-EDTA-GABA (Figure 5) was synthesized, covalently attached to the resin-bound proteins (Figure 6), both versions of each protein (with and without EDTA-GABA) cleaved from the resin with HF and purified by HPLC, and finally proteins were reacted with ^{32}P end-labelled restriction fragments in footprinting and affinity cleaving studies. These mutant proteins are listed in Figure 11. Proteins derivatized with the affinity cleaving moiety have the prefix "[Fe•EDTA]."

[Fe•EDTA]Hin(139-190)R140→E exhibits almost no ability to cleave DNA with sequence specificity; glutamic acid (designated as "E") possesses a negatively charged carboxylate side chain, which would be repelled by the negatively charged DNA phosphodiester backbone, whereas the wild-type arginine side chain is a positively charged guanidinium fork. [Fe•EDTA]Hin(139-190)R140→A, [Fe•EDTA]Hin(139-190)R140→βA, [Fe•EDTA]Hin(139-190)R140→G, and [Fe•EDTA]Hin(139-190)R140→Q (alanine is designated as "A," β-alanine as "βA," glycine as "G," and glutamine as "Q") all have neutrally charged side chains, and all exhibit

[Fe•EDTA]Hin(139-190)	(EDTA-GABA)GRPRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	52-mer
[Fe•EDTA]Hin(139-190)R140→A	(EDTA-GABA)G**A**PRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	52-mer
[Fe•EDTA]Hin(139-190)R140→βA	(EDTA-GABA)G**βA**PRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	52-mer
[Fe•EDTA]Hin(139-190)R140→K	(EDTA-GABA)G**K**PRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	52-mer
[Fe•EDTA]Hin(139-190)R140→E	(EDTA-GABA)G**E**PRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	52-mer
[Fe•EDTA]Hin(139-190)R140→G	(EDTA-GABA)G**G**PRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	52-mer
[Fe•EDTA]Hin(139-190)R140→Q	(EDTA-GABA)G**Q**PRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	52-mer
[Fe•EDTA]Hin(139-190)R142→K	(EDTA-GABA)GRP**K**AINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASSIKKRMN-CO_2^-	52-mer
[Fe•EDTA]Hin(139-184)	(EDTA-GABA)GRPRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASS-CO_2^-	46-mer
[Fe•EDTA]Hin(141-184)	(EDTA-GABA)PRAINKHEQEQISRLLEKGHPRQQLAIIFGIGVSTLYRYFPASS-CO_2^-	44-mer

FIGURE 11. Amino-acid sequences of synthetic proteins. Each protein was prepared with and without EDTA-GABA at the amino terminus. Mutated residues are shown in outline: R=Arg, A=Ala, βA=β-Ala, K=Lys, E=Glu, G=Gly, Q=Gln. On the right are listed the corresponding sizes of each protein.

sequence-specific DNA cleavage capabilities, albeit much weaker than the native 52-mer [Fe•EDTA]Hin(139-190). [Fe•EDTA]Hin(139-190)R140→K (lysine is designated as "K") has a positively charged amino side chain and is the mutant most similar to the native Hin(139-190); it also cleaves DNA with sequence specificity and gives the strongest cleavage pattern of all the mutants, and it is the only mutant whose binding could be detected by MPE footprinting at concentrations below 8μM. After preliminary studies showed that protein binding is significantly affected upon substitution of Arg140, a 52-mer protein with substitution of Arg142 by lysine was constructed, [Fe•EDTA]Hin(139-190)R142→K. [Fe•EDTA]Hin(139-190)R142→K also cleaves with sequence specificity.

The native 52-mer [Fe•EDTA]Hin(139-190) exhibits the cleaving preference *hix*L > secondary >> tertiary, whereas all of the mutants with substitutions at Arg140 give a reversed binding preference: secondary ≥ tertiary >> *hix*L. This reversed binding preference is identical to that for the 49- and 50-mers. [Fe•EDTA]Hin(139-190)R142→K gives a cleavage pattern quite similar to that of [Fe•EDTA]Hin(139-190), but its binding constant is over one order of magnitude weaker than that for [Fe•EDTA]Hin(139-190). All of the mutants give much weaker cleavage patterns than does [Fe•EDTA]Hin(139-190); therefore, Arg140 makes a significant binding contribution, both in binding affinity and in sequence specificity, whereas Arg142, as shown in [Fe•EDTA]Hin(139-190)R142→K, makes only a non-specific binding contribution.

The Arg140 residue in wild-type Hin(139-190) has a positively charged guanidinium side chain, which is a double-amino fork capable of donating hydrogen bonds. Replacing Arg140 with lysine in the [Fe•EDTA]Hin(139-190)R140→K mutant is a conservative change, for lysine's side chain is a

FIGURE 12. Autoradiogram of a high-resolution, denaturing polyacrylamide gel of affinity cleaving reactions on 3′ ^{32}P end-labelled restriction fragment from plasmid pMFB36. Reaction mixtures (10μl) contained ^{32}P end-labelled DNA fragment (~20,000 cpm), [Fe•EDTA]protein, 20 mM phosphate, pH 7.5, 20 mM NaCl, 0.1 mg/ml tRNA, and 5mM dithiothreitol (DTT). All components except DTT were incubated for 15 minutes at 22°C. Reactions were initiated by the addition of DTT (5 mM), and were allowed to proceed for 90 min at 22 °C followed by ethanol precipitation. All lanes contain 3′ end-labelled DNA. Lane 1, intact DNA control; lane 2, Maxam-Gilbert G reaction. Lane 3, 0.5μM [Fe•EDTA]Hin(139-190); lane 4, 4μM [Fe•EDTA]Hin(141-190); lane 5, 2μM [Fe•EDTA]Hin(139-184); lane 6, 2μM [Fe•EDTA]Hin(139-190)R140→A; lane 7, 2μM [Fe•EDTA]Hin(139-190)R140→βA; lane 8, 1μM [Fe•EDTA]Hin(139-190)R140→K; lane 9, 4μM [Fe•EDTA]Hin(139-190)R142→K; lane 10, [Fe•EDTA]Hin(139-190)R140→E; lane 11, 4μM [Fe•EDTA]Hin(139-190)R140→G; lane 12, 2μM [Fe•EDTA]Hin(139-190)R140→Q.

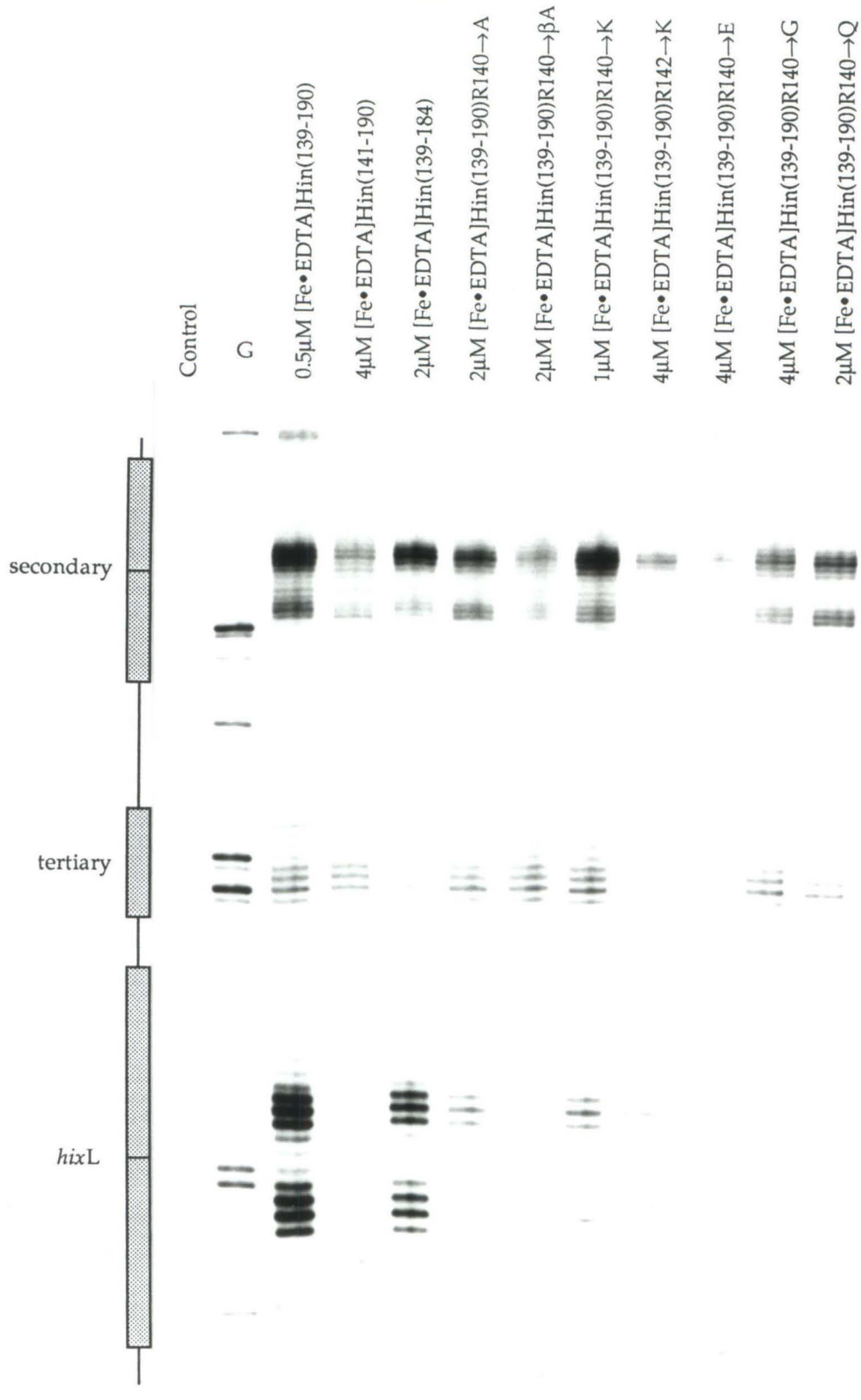
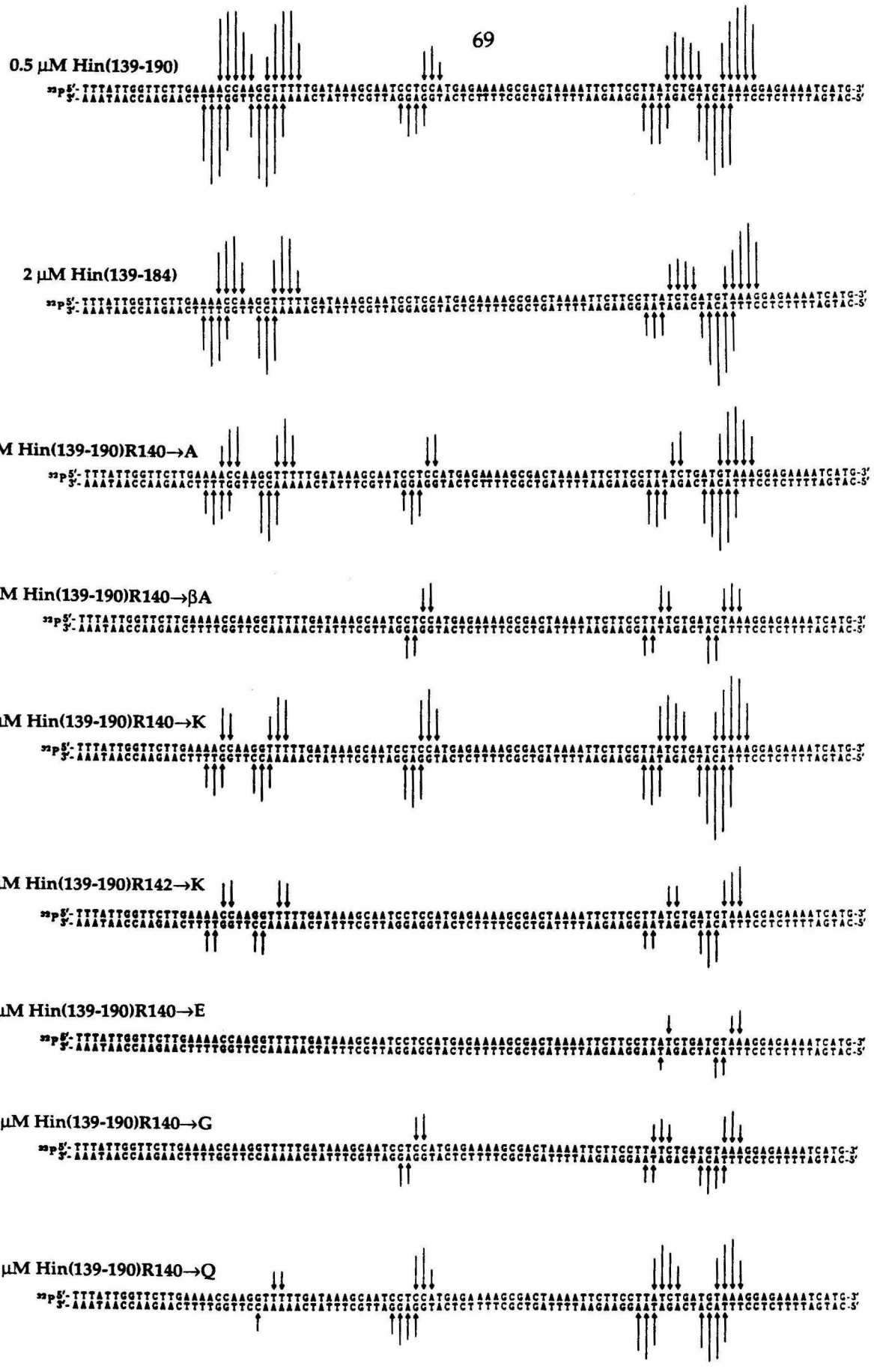

Control
G
0.5μM [Fe•EDTA]Hin(139-190)
4μM [Fe•EDTA]Hin(141-190)
2μM [Fe•EDTA]Hin(139-184)
2μM [Fe•EDTA]Hin(139-190)R140→A
2μM [Fe•EDTA]Hin(139-190)R140→βA
1μM [Fe•EDTA]Hin(139-190)R140→K
4μM [Fe•EDTA]Hin(139-190)R142→K
4μM [Fe•EDTA]Hin(139-190)R140→E
4μM [Fe•EDTA]Hin(139-190)R140→G
2μM [Fe•EDTA]Hin(139-190)R140→Q
secondary
tertiary
hixL

FIGURE 13. Histograms of affinity cleavage data in Figure 12. Arrow heights indicate relative extents of cleavage. Note that the affinity cleaving patterns for [Fe•EDTA]Hin(139-190), [Fe•EDTA]Hin(139-184), and [Fe•EDTA]Hin(139-190)R142→K show the same binding-site preferences for the *hix*L, secondary, and tertiary sites. The proteins that are mutated at Arg140–i.e. [Fe•EDTA]Hin(139-190)R140→A, [Fe•EDTA]Hin(139-190)R140→βA, [Fe•EDTA]Hin(139-190)R140→K, [Fe•EDTA]Hin(139-190)R140→E, [Fe•EDTA]Hin(139-190)R140→G, and [Fe•EDTA]Hin(139-190)R140→Q–show the same binding-site preferences, which are very different from that of [Fe•EDTA]Hin(139-190), [Fe•EDTA]Hin(139-184), and [Fe•EDTA]Hin(139-190)R142→K.


0.5 μM Hin(139-190)
2 μM Hin(139-184)
2 μM Hin(139-190)R140→A
2 μM Hin(139-190)R140→βA
1 μM Hin(139-190)R140→K
4 μM Hin(139-190)R142→K
4 μM Hin(139-190)R140→E
1 μM Hin(139-190)R140→G
2 μM Hin(139-190)R140→Q

positively charged single-amino group; [Fe•EDTA]Hin(139-190)R140→K displays both reduced DNA binding affinity and a binding preference reversed from that of the native 52-mer. It is suspected then that Arg140 does not merely interact non-specifically with the negatively charged phosphodiester backbone of DNA, but specifically recognizes bases in the minor groove or the sequence-dependent DNA conformation in the minor groove. Arg140 greatly prefers the 5'-AAA (*hix*L) sequence in the minor groove over 5'-TTA (secondary) and 5'-GGA (tertiary). In the case of the 434 repressor (discussed in Chapter 1), water molecules mediate the interaction between the guanidinium group of Arg43 and A,T base pairs in the minor groove; in the *engrailed* homeodomain cocrystal structure (Chapter 1), both Arg3 and Arg5 make thymine-specific contacts in the minor groove.

Poly(dA)•poly(dT) tracts are known to narrow the minor groove and facilitate DNA bending toward the minor groove (discussed under the Principles of Recognition section in Chapter 1). The guanidinium side chain of Arg140 may well be able to span the narrowed minor groove of the *hix*L binding site and interact with the phosphodiester backbones on either side, especially with the aid of water-mediated interactions as in the case of Arg43 of the 434 repressor; because the mutant proteins lack Arg140 and therefore cannot possibly make this type of dual interaction across the minor groove, the narrowed poly(dA)•poly(dT) tract may not be important for their interaction with DNA. Or Arg140 may be making specific interactions with bases in the minor groove; the fork on the arginine side chain allows it to make two hydrogen bonds, so more than one base may be recognized, and both adenine and thymine act as hydrogen bond acceptors in the minor groove. Both *hix*L IRR and IRL contain 5′-AAA in the minor groove; although the secondary Hin half sites contain 5′-TTA which should act as

hydrogen bond acceptors similar to *hix*L, the 5′-TTA sequence should still have a different conformation from 5′-AAA–differences may include propellor twist, narrowness of the minor groove, and other factors. We may conclude that Arg140 confers upon Hin(139-190) much of its high specificity for the 5'-AAA sequence in the *hix*L site. The conservative substitution of Arg140 with lysine changes the binding specificity of Hin(139-190) in favor of the secondary and tertiary sites, indicating that Arg140 does not merely make non-specific contacts at the *hix*L site; at the secondary and tertiary sites, however, Arg140 may make non-specific contacts, for binding of mutant proteins at these sites is not compromised to the great extent that it is at *hix*L.

The Arg-Pro-Arg motif appears to contribute significantly to the binding specificity of Hin, and it is essential to the protein's minor-groove interaction and binding affinity at the *hix*L site; arginines are well suited to binding in minor grooves containing A,T tracts, and proline provides unique and limited backbone torsional angles that fix the surrounding residues. Use of Arg-Pro-Arg as a DNA recognition element is not unique to Hin. Similar protein sequences can be found in the recombinases Cin, Gin, and Pin; these proteins also recognize a 5′-AAA tract in the minor groove of each half site. Cin and Pin contain the Arg-Arg-Pro sequence at the amino terminus; Gin contains Arg-Pro-Pro, which is rather different from the other recombinases (Figure 2).

Furthermore, several other helix-turn-helix proteins are known to make similar minor groove contacts. The 434 repressor contains Lys-Arg-Pro-Arg at the small loop between the α_3 and α_4 helices,[26] and this loop interacts with an A,T sequence in the minor groove (Chapter 1). In this case the minor groove is 2.5Å narrower than idealized B-DNA, and it appears that the Arg-Pro-Arg sequence is recognizing the DNA conformation rather than

contacting specific bases. Similarly, the γδ resolvase DNA binding domain (residues 141-183) contains an Arg-Lys-Arg-Lys sequence at its amino terminus that interacts with the minor groove. The *engrailed* homeodomain also possesses an Arg-Pro-Arg sequence in which Arg3 and Arg5 make thymine-specific contacts in the minor groove[81]; the *Antennapedia* homeodomain contains an Arg-Gly-Arg-Gly-Arg at its amino terminus, which is also believed to interact with an A,T sequence in the minor groove.[83,85] These positively charged sequences containing variants of Arg-Pro-Arg have been detected in a number of known DNA binding proteins[153]; it would seem that this sequence is particularly well suited for DNA recognition, in particular, recognition of a narrowed minor groove consisting of A,T bases that are capable of serving as hydrogen bond acceptors.

In the case of Hin(139-190), it can be argued that Arg140 is the major determinant of the protein's specificity for *hix*L over the secondary and tertiary Hin sites. Both Hin(142-190) and Hin(141-190) deletion proteins and all of the mutant Hin proteins with substitutions at Arg140 show virtually the same binding preference: secondary ≥ tertiary >> *hix*L. But in proteins that contain Arg140, Hin(140-190), Hin(139-190), and [Fe•EDTA]Hin(139-190)R142→K, the binding preference is more reflective of the naturally occurring Hin recombinase: *hix*L > secondary >> tertiary. Perhaps Hin(142-190), Hin(141-190), and the Hin proteins mutated at Arg140 reflect the "major-groove binding contribution" of Hin recombinase–that is, the sequence-specific binding contribution made by the helix-turn-helix structure, whereas the Arg-Pro-Arg amino terminus determines Hin's wild-type binding specificity.

Carboxyl-Terminal Studies on the Hin DNA Binding Domain

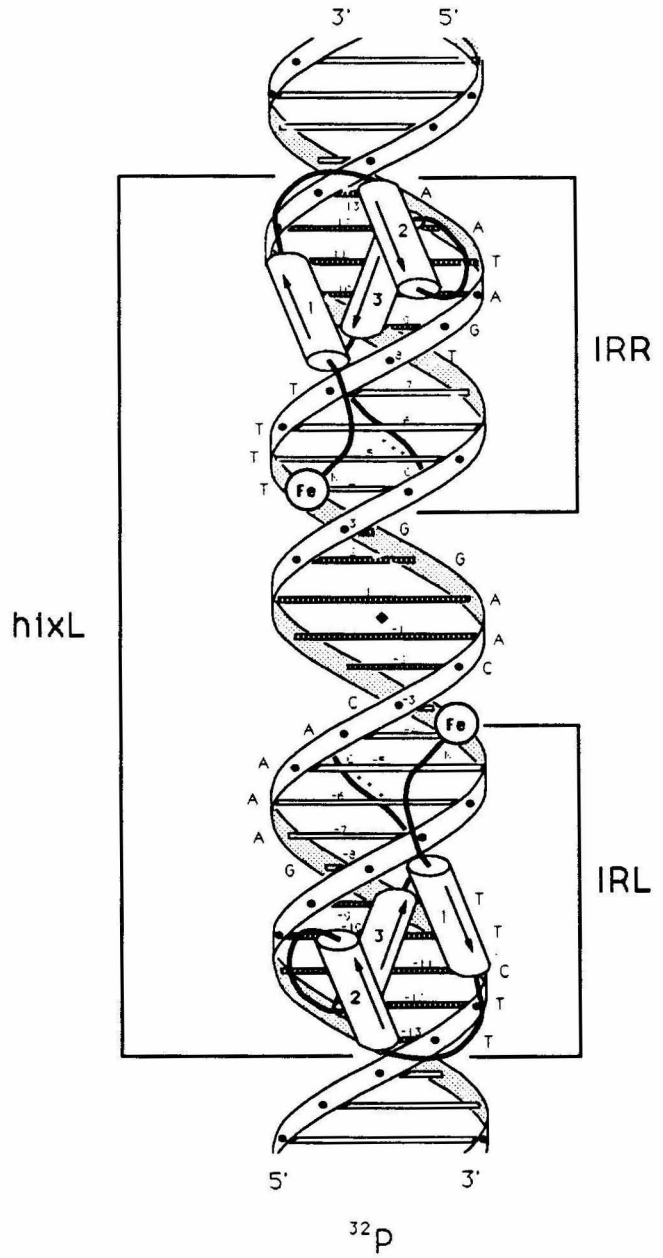
Two other mutants of Hin were synthesized at the Caltech Microchemical Facility; again two versions of each protein were prepared, with and without EDTA-GABA, which were cleaved from the resin by HF, purified by HPLC, and reacted with a ^{32}P end-labelled restriction fragment (Figure 11). Because of the lack of sequence homology at the carboxyl terminus amongst the recombinases Hin, Gin, Cin, and Pin (Figure 2), it was decided to remove the six non-conserved amino acids at the carboxyl terminus of Hin; it is significant to note that three of the six residues are positively charged (two lysines, one arginine).

Hin(139-184), a 46-mer, produces strong footprints similar to those of Hin(139-190) and gives a virtually identical cleavage pattern to that of [Fe•EDTA]Hin(139-190), although binding affinity is sacrificed by removal of these six residues. The binding constant of the 46-mer is about 50-fold less than that of the 52-mer. These six amino acids, therefore, make a significant non-specific binding contribution, but their removal does not compromise binding specificity. Hin(141-184) is missing the six carboxyl-terminal residues like the 46-mer Hin(139-184), but it is also missing the first two amino acids at the amino terminus (Gly 139 and Arg140)–i.e., Hin(141-190) is the Hin 50-mer minus the first six carboxyl-terminal residues. Site-specific binding by Hin(141-190) was not detected by footprinting, and affinity cleaving was not exhibited by [Fe•EDTA]Hin(141-190) at 8μM protein concentration; given that the Hin 50-mer binds only weakly at the secondary and tertiary Hin sites and that without the six carboxyl-terminal residues the binding constant is decreased by 50-fold, it is not surprising that Hin(141-190) fails to bind to the Hin binding sites.

Although there exists a high degree of sequence homology between the λ repressor's amino terminal arm, which contacts specific bases in the major groove (Chapter 1), and the carboxyl terminus of Hin, the Hin carboxyl residues are unnecessary for specific binding. Although the λ repressor arm and the Hin carboxyl terminus appear to be very similar structures and may be predicted to behave similarly, this is clearly not the case; this is another example that demonstrates the difficulty in predicting protein structure and specific interactions from sequence information.

FIGURE 14. Representation of the model of [Fe•EDTA]Hin(139-190) bound to *hix*L. The center of the dimeric operator site is indicated by the solid diamond. α-helices are shown as cylinders with arrows pointing from the amino to the carboxyl terminus.

Model of [Fe•EDTA]Hin(139-190) bound to *hix*L



CONCLUSION

A summary of the data gained from these experiments is listed below.

•[Fe•EDTA]Hin(139-190) shows the following DNA binding specificity:

*hix*L > secondary >> tertiary (5′-AAA > 5′-TTA >> 5′-GGA).

•The proteins in which Arg140 is mutated–[Fe•EDTA]Hin(139-190)R140→A, [Fe•EDTA]Hin(139-190)R140→βA, [Fe•EDTA]Hin(139-190)R140→K, [Fe•EDTA]Hin(139-190)R140→E, [Fe•EDTA]Hin(139-190)R140→G, and [Fe•EDTA]Hin(139-190)R140→Q–display a binding specificity reversed from that of [Fe•EDTA]Hin(139-190):

secondary ≥ tertiary >> *hix*L (5′-TTA ≥ 5′-GGA >> 5′-AAA).

•[Fe•EDTA]Hin(139-190)R142→K displays the same binding specificity as does [Fe•EDTA]Hin(139-190), although it binds with lower affinity.

••Therefore, Arg140 makes significant contributions to both DNA binding affinity and sequence specificity, whereas Arg142 contributes only to binding affinity. The amino-terminal interactions of Hin(139-190) with the *hix*L minor groove determines, in large part, the Hin protein's sequence specificity. Proteins lacking Arg140 are probably more reflective of the contribution made by the helix-turn-helix structure of Hin interacting with the major groove.

•[Fe•EDTA]Hin(139-184) retains the same binding specificity as does [Fe•EDTA]Hin(139-190), albeit its binding affinity is significantly reduced. [Fe•EDTA]Hin(141-184) shows no affinity cleaving activity.

••Therefore, the six carboxyl-terminal residues of the Hin DNA binding domain Hin(139-190) make non-specific contacts to DNA, most likely electrostatic interactions with the phosphodiester backbone, which contribute to binding affinity only and have no effect on sequence specificity.

MATERIALS AND METHODS

Materials. Protected amino-acid derivatives were purchased from Peninsula Laboratories; Boc-L-His(DNP) was obtained from Fluka. 4-Methylbenzhydrylamine (BHA) resin came from US Biochemical Corp. (Phenylacetamido)methyl (PAM) resin, dimethylformamide (DMF), diisopropylethylamine, dicyclohexylcarbodiimide in dichloromethane, N-hydroxybenzotriazole in DMF, and trifluoroacetic acid (TFA) were purchased from Applied Biosystems. Dichloromethane and methanol (HPLC grade) were obtained from Mallinckrodt, p-cresol and p-thiocresol from Aldrich, and diethyl ether (low peroxide content) from Baker. Doubly distilled water was further purified through the Milli Q filtration system from Millipore. tRNA (*E. coli* strain W, Type XX) came from Sigma and was dissolved in water and sterile-filtered. Enzymes were purchased from Boehringer Mannheim or New England Biolabs. UV-Vis spectra were recorded on a Perkin-Elmer Lambda 4C spectrophotometer. Laser densitometry on gel autoradiograms was done on an LKB Ultrascan XZ densitometer. Phosphorimaging was accomplished using a Molecular Dynamics 400S Phosphorimager and ImageQuant software. Storage phosphor screens were purchased from Molecular Dynamics.

Solid Phase Protein Synthesis

Synthesis. Manual protein syntheses were carried out by solid phase techniques in 20 ml vessels fitted with coarse glass frits, using synthetic protocols developed at the California Institute of Technology.[173-176] Automated protein syntheses were performed by Suzanna J. Horvath at the Caltech Microchemical Facility. Fully protected, resin-bound Hin(139-184) and Hin(141-184), referred to as the 46- and 44-mers, respectively, were synthesized on benzhydrylamine (BHA) resin using Boc protected amino

acids. BHA resin affords a neutrally charged amide terminus rather than negatively charged carboxylate terminus upon removal of the resin bead. The other proteins were synthesized on PAM resin, which gives a negatively charged carboxylate terminus, because these proteins contain the native carboxyl terminus. All of the EDTA-GABA derivatized proteins synthesized in this chapter were derivatized with BEG at the amino terminus. All of the Boc-protected amino acids and BEG were coupled by standard methods. Coupling yields were determined by quantitative ninhydrin monitoring with acceptable values being ≥99.7% near the beginning of the synthesis, falling off toward 99% at the end.

Deprotection and Purification. The histidine protecting group, dinitrophenyl (DNP), was removed by thiolysis at 25°, using 20% β-mercaptoethanol/10% diisopropylethylamine in DMF; this treatment was repeated twice for 30 minutes each time. Boc-protecting groups were removed with TFA and the resin dried; all other side-chain protecting groups as well as the proteins were cleaved from the resin and deprotected using anhydrous hydrofluoric acid in the presence of p-cresol and p-thiocresol radical scavengers for 60 minutes at 0°C. The HF was removed under vacuum, and the crude protein was precipitated by diethyl ether, dissolved in water and lyophilized. Before the crude proteins were subjected to HPLC purification, residual DNP groups were removed by treatment in 4M guanidine hydrochloride/50mM tris, pH 8.0/20% β-mercaptoethanol for 1 hour at 55°C. The synthesized proteins were then purified by reverse phase HPLC on a semipreparative C8 column (Vydac) with a linear gradient of acetonitrile/water with 0.1% trifluoroacetic acid (flow rate 3ml/min, 0 to 60% acetonitrile over 240 min). The proteins were homogeneous by the criteria of HPLC. Proteins were stored dry at -70 °C in 5 nmol aliquots (ε_{275}=2810 for two Tyr residues; ε_{275}=1405 for one tyrosine).

[Fe•EDTA]proteins were prepared by incubation of [EDTA]proteins with aqueous ferrous ammonium sulfate in a 1:1 molar mixture (30 min, 25 °C).

DNA Substrate

Radioactive Labelling of Restriction Fragment. Plasmid pMFB36 was linearized by digestion with Xba I. Linearized plasmid pMFB36 was 3'-end-labelled with [α^{32}P]-dATP and DNA polymerase I Klenow fragment.[177] Linearized plasmid pMFB36 was 5'-end-labeled by dephosphorylation, using calf intestinal alkaline phosphatase, followed by phosphorylation using [γ^{32}P]-ATP and T4 polynucleotide kinase.[177] Labelled linearized plasmid pMFB36 was digested with EcoR I, and the resulting labeled 557-base pair DNA fragment was isolated using non-denaturing polyacrylamide gel electrophoresis.

DNA Cleaving Experiments

Reaction Conditions. Reaction mixtures (10μl) contained ^{32}P end-labelled DNA fragment (20,000 cpm), [Fe•EDTA]protein, 20 mM phosphate, pH 7.5, 20 mM NaCl, 0.1 mg/ml tRNA (Sigma Chemical, Type XX), and 5mM dithiothreitol (DTT). All components except DTT were incubated for 15 minutes at 22°C. Reactions were initiated by the addition of DTT (5 mM), and were allowed to proceed for 90 min at 22 °C. Reactions were terminated by ethanol precipitation, dried, and resuspended in 100mM tris-borate-EDTA/80% formamide loading buffer. Reaction products were analyzed by electrophoresis on 8% polyacrylamide denaturing gels (1:20 crosslink, 7M urea). After electrophoresis, gels were dried and autoradiographed. Autoradiograms were analyzed by laser densitometry.

CHAPTER THREE

Quantitative Affinity Cleaving Studies on the DNA Binding Domain of Hin Recombinase

INTRODUCTION

The regulation of many prokaryotic and eukaryotic genes is dependent on specific, often cooperative, interactions that govern the binding of proteins to multiple DNA sites.[3] Although a qualitative understanding has been developed for the roles of various molecular entities in biological systems, little is known on a quantitative level. A quantitative characterization gives information on the strength of molecular interactions and the formulation of molecular mechanisms, and also enables prediction and modeling of a system's behavior. The measurement of binding constants, and by extension free energy values (ΔG's), of protein-DNA complexes represents a first step toward a full understanding of the thermodynamic profile of DNA complexation.

Most techniques used to study DNA protein interactions measure only macroscopic properties: i.e., these techniques are incapable of distinguishing binding interactions at individual sites in a multisite system, and therefore, only average properties over all sites can be measured. These classical binding techniques, including equilibrium dialysis, nitrocellulose filter binding, and gel retardation assays, measure binding of proteins (or other ligands) to DNA fragments. A short DNA fragment may contain only one binding site, but studies involving short fragments often exhibit end effects that are not observed when genomic DNA is investigated. In addition to increasing the number of potential protein binding sites, increasing the

length of the DNA fragment also increases the amount of non-specific protein-DNA interaction, and these techniques cannot separate the specific from non-specific binding contributions. Further complications arise because many interactions between regulatory proteins and operator sites are cooperative (discussed in Chapter 2). In cooperative systems, binding is dependent not only on the binding constant for one protein to an individual site, but also on interactions between proteins bound to neighboring sites. Classical binding techniques cannot resolve cooperative interactions, and therefore cannot give accurate thermodynamic values for the individual sites.[178]

QUANTITATIVE FOOTPRINTING

Quantitation of protein-DNA interactions was originally investigated using DNase I footprinting on the bacteriophage λ repressor in complex with the left and right operators, O_L and O_R (see the λ repressor section in Chapter 1).[179-181] This method, called quantitative footprinting,[178,182] resolves binding at individual sites, even in cooperative systems, and provides thermodynamically rigorous equilibrium binding curves. In a quantitative footprinting experiment, an end-labelled DNA duplex fragment containing specific binding sites is incubated with a series of known protein concentrations (most DNA binding ligands can be studied this way). After equilibration of the ligand-DNA complex, the mixture is exposed to DNase I, an endonuclease that introduces single-stranded nicks on DNA unprotected by bound ligand.[154] The DNA products are separated by gel electrophoresis and visualized by autoradiography. Typically, in the gel lanes with reactions at high ligand concentrations, a strong footprint is seen; as the ligand concentration gradually decreases, weaker footprints are exhibited, and

eventually no footprint is visualized at all. These data can be fit to a sigmoidal binding curve in which the high-ligand concentration data are at the top of the curve, and the low-ligand concentration data form the bottom.[178]

Experiments in which the equilibration time was varied from 30 minutes to 2 hours gave identical results, indicating that the association reaction between the λ repressor and the operator had reached equilibrium.[178] The equilibration time must be tested for each system. Experimental conditions were achieved such that DNase nicked only a fraction of the DNA fragments; under these conditions, each DNA fragment is cleaved only once, for the probability of multiple nicks on one fragment is vanishingly low. This condition of single-hit kinetics is necessary to ensure that the equilibrium of the ligand-DNA complex is unaffected by the DNase, the agent that is measuring the equilibrium, and that the population of full-length duplex available for protein binding is not significantly affected by DNase exposure.[182] Experiments were performed in which the DNase concentration was varied in order to ascertain whether DNase perturbs the binding equilibrium between the λ repressor-O_R complex. Over the range tested, virtually identical binding isotherms were obtained.[182] The thermodynamic values obtained from quantitative footprinting experiments on the λ repressor in complex with the single O_R1 site of the λ right operator (ΔG = -12.5±0.1 kcal/mol) were identical to those gained from filter-binding assays (ΔG = -12.5±0.2 kcal/mol).[182]

The polyacrylamide sequencing gels were cast using a specially designed comb; the 6mm lanes were separated by 6mm spacings in order to keep lanes distinct and to provide an accurate background reference level for

each lane.[178] Radioactive gels were exposed to X-ray film; it is necessary to ensure a linear film response with the best signal-to-noise ratio. In order to stay within the linear range of the film, preflashed X-ray film and an intensifying screen were used.[178,182] Therefore, measurement of the optical densities of all bands on an autoradiogram was exactly proportional to the radioactivity present in all bands on the gel. Autoradiograms were subjected to high-resolution two-dimensional scanning, and the data obtained were analyzed by a nonlinear least-squares fitting procedure.[178,182]

That quantitative footprinting yields thermodynamically valid individual-site isotherms is supported by several lines of evidence.[178,183] First, the only experimental variable that affects the degree of protection is the concentration of ligand. The relation of the density of any band, whether protected or not, to the total DNA in the equilibrium mixture is constant, except as modified by the variable ligand concentration. Rigorous procedures have been developed to ensure this by establishing conditions that do not perturb the ligand-DNA equilibrium, electrophoretic techniques that allow precise resolution of single bands, accurate quantitation of the density of measured bands, and methods that permit high resolution of binding isotherms.

QUANTITATIVE AFFINITY CLEAVING

Previously, the affinity cleaving method has been used to evaluate qualitative differences in binding strengths of EDTA•Fe conjugates of DNA binding ligands in complex with specific DNA sites. Like DNase footprinting, affinity cleaving can be extended to the quantitative measurement of equilibrium binding constants at individual sites. However, quantitative affinity cleaving has a number of advantages over the analogous footprinting

experiment. It eliminates the possibility of direct interactions between the complex and the DNA cleaving agent. Using the DNA cleaving functionality EDTA•Fe, which is virtually sequence independent, allows all binding sites to be examined with similar accuracy.

Affinity cleaving can be performed under a wide range of conditions; it has been performed between pH 5.5 and 10, at temperatures between 0°C and 45°C, and in the presence of a variety of cations and cation concentrations. Additionally, the presence of cleavage bands constitutes a positive signal that substantially increases the signal-to-noise ratio; in comparison, a footprint is a negative signal visualized by an absence of cleavage bands. Quantitative footprinting is limited by visualization of footprints; in practice, molecules with binding constants below 10^6-10^7 cannot be accurately studied by this technique.[183] The sole disadvantage to affinity cleaving is that the EDTA moiety must be covalently attached to the ligand under study. At neutral pH, the EDTA•Fe moiety carries a single negative charge, and this coupled with the steric bulk of EDTA•Fe may lower the binding affinity of the ligand-conjugate such that the binding constants obtained by quantitative affinity cleaving may be somewhat lower than that of the underivatized compound. This should not affect relative values between sites, however.

The experimental scheme for quantitative affinity cleaving is shown in Figure 1. The EDTA moiety is attached to the ligand; Fe(II) is loaded into the EDTA site, and in the presence of reducing agent and oxygen, EDTA•Fe produces localized backbone cleavage. By accurate measurement of the intensity of the cleavage pattern over a wide range of ligand concentrations, a best-fit binding curve can be generated. As in the footprinting experiment, the condition of single-hit kinetics must be achieved in quantitative affinity cleaving. The experiments discussed below are concerned only with protein

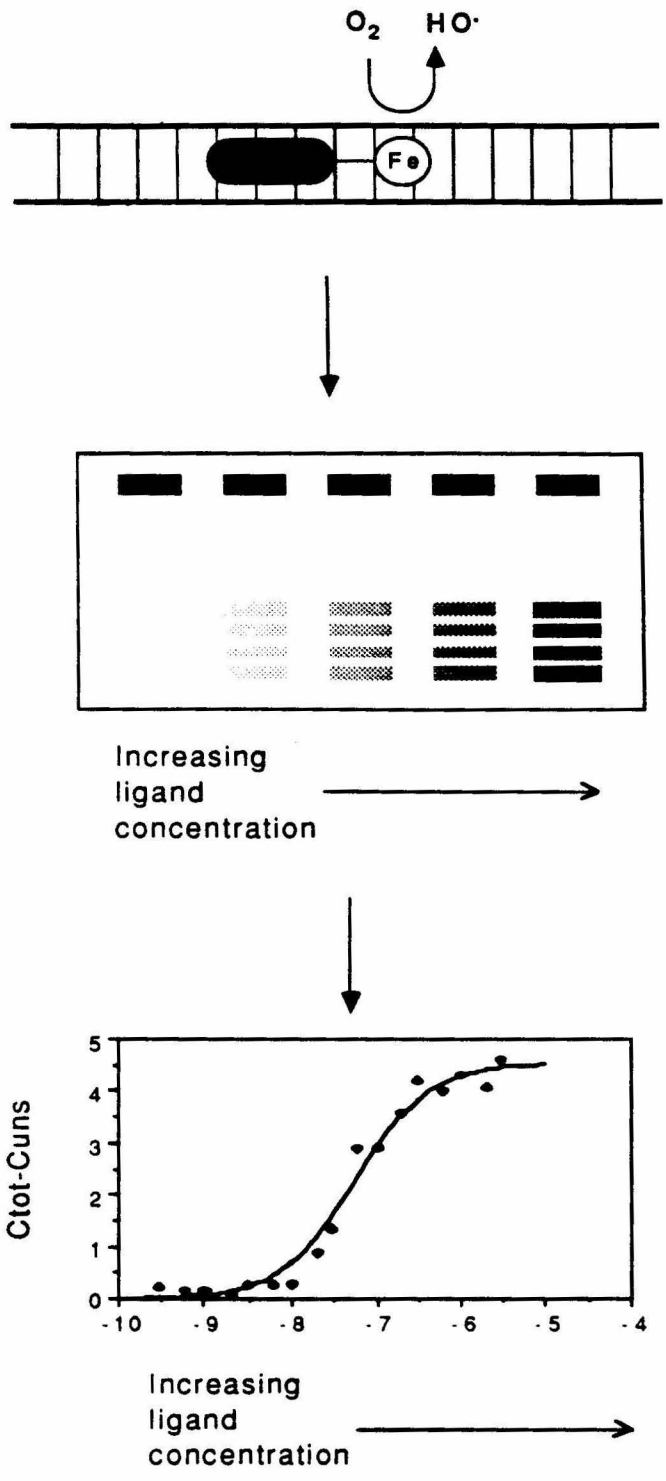


FIGURE 1. Protocol for quantitative affinity cleaving.

affinity cleavage, although extensive quantitative affinity cleaving studies have been conducted on triple-helical complexes. Synthetic proteins are easily derivatized with EDTA (see Chapter 2), and their relatively high sequence specificities and binding affinities make them an excellent choice for quantitative affinity cleaving studies. As an independent check for the accuracy of the quantitative affinity cleaving technique, measured binding constants can be compared to those obtained from DNase footprinting and gel retardation assays.

THEORY

The theoretical treatment for the quantitative affinity cleaving experiment follows directly from the quantitative footprinting formulation.[178,179] Using the definition of the fraction of DNA bound, accurate binding constants can be measured from footprinting experiments if only the ligand concentration and the fraction of DNA bound are known quantities. The ideas behind quantitative footprinting have been adapted to affinity cleaving, and the theory is discussed below.[163,178,182,184]

In the complexation reaction between the ligand, denoted L, and DNA

$$DNA + L \rightleftharpoons DNA \bullet L,$$

the binding constant is defined as

$$K_a = \frac{[DNA \bullet L]}{[DNA]_{free}[L]_{free}}.$$

This equation can be redefined using Y, the fraction of bound DNA.

$$Y = \frac{[DNA \bullet L]}{[DNA]_{total}}$$

$$K_a = \left(\frac{[DNA \bullet L]}{[DNA]_{total}}\right)\left(\frac{[DNA]_{total}}{[DNA]_{free}}\right)\left(\frac{1}{[L]_{free}}\right)$$

$$K_a = \left(\frac{Y}{1 - Y}\right)\left(\frac{1}{[L]_{free}}\right)$$

Theoretically, Y varies from 0 at very low ligand concentrations to 1 at saturation. The binding constant at a given site can then be determined by plotting the Y value versus ligand concentration over several orders of magnitude. At Y=0.5, the binding constant will be the reciprocal of the free ligand concentration ($K_a=1/[L]_{free}$).

These equations lead to two conclusions important to the footprinting and affinity cleaving experiments. First, measurement of the binding constant is independent of the concentration of DNA; this permits the use of radiolabelled DNA, because it is not necessary to know its specific activity. Second, accurate determination of the binding constant depends solely on accurate measurement of the free ligand concentration and the fraction of DNA bound. The fraction of DNA bound is directly proportional to the degree of protection from cleavage in the footprinting experiment; in affinity cleaving, the intensity of cleavage at a site is proportional to its occupancy. These quantities can be determined from measurement of the optical densities of bands on an autoradiogram or from the intensity of luminescence from photostimulable, storage phosphor-imaging screens.[185] $[L]_{free}$ is accurately measured by lowering the DNA concentration; when the concentration of DNA binding sites is <1% of the total ligand concentration, the approximation that the concentration of unbound ligand ($[L]_{free}$) is equal to the total concentration of ligand used in the reaction ($[L]_{total}$) is valid. This can be readily achieved by increasing the specific activity of the DNA. In the quantitative affinity cleaving studies, the lowest protein concentrations used

were in the nanomolar range, whereas calculations showed that the radioactive DNA concentrations were subpicomolar.

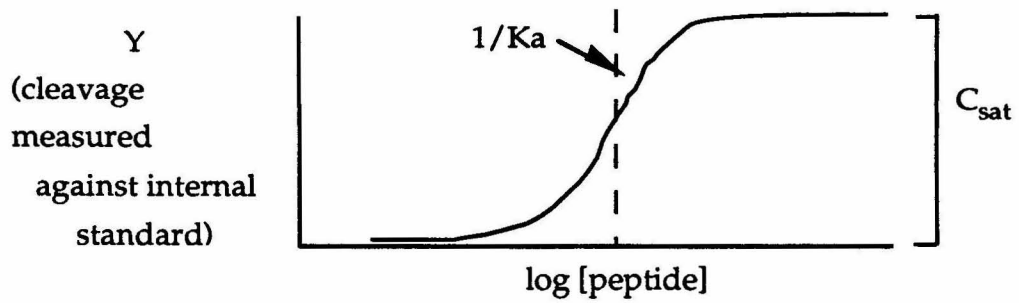


FIGURE 2. Schematic of a binding curve.

In order to extend quantitative footprinting to quantitative affinity cleaving, a complicating factor must be considered. In an affinity cleaving experiment, the amount of cleavage at a band at the recognition site (C_{tot}) has a specific component (C), which is the contribution made by the affinity cleaving agent bound at the target site, and a non-specific component (C_{uns}) produced by unbound protein and unchelated iron in the reaction mixture (Figure 2).

$$C_{tot} = C + C_{uns}$$

In the affinity cleaving experiment, both components will vary with the concentration of the cleaving agent. Because Y is proportional to C, the equation for the binding constant as a function of Y can be rearranged as follows;

$$K_a = \left(\frac{C}{C_{sat} - C}\right)\left(\frac{1}{[L]_{free}}\right)$$

where C_{sat} is the amount of cleavage observed when Y=1. When $C=0.5C_{sat}$, $K_a=1/[L]_{free}$. C can be determined by rearranging the above equations, but C_{uns} is not directly measurable. It should, however, be proportional to the cleavage at a reference site (C_{ref}), removed from the target sequence and should have no specific cleavage component. Therefore, the specific component of cleavage will be given by

$$C = C_{tot} - kC_{ref},$$

where k is a scaling parameter that accounts for any intrinsic differences between the target and reference sites (e.g., the number of bands quantitated will likely differ between the two sites). Rearrangement of the above equations yields the final form of the theoretical curve to be fit to the experimental data points.

$$C_{fit} = \frac{C_{sat}(K_a[L])}{1 + (K_a[L])} + kC_{ref}$$

C_{tot} and C_{ref} are determined from the affinity cleaving data, and [L] is equal to the concentration of added cleaving reagent. The theoretical binding curve can be fit using k, C_{sat}, and K_a as adjustable parameters. Binding curves are then produced by graphing C_{tot}-kC_{ref} versus the ligand concentration as shown in Figure 1.

QUANTITATIVE AFFINITY CLEAVING ON EDTA•DERIVATIZED PROTEINS: [Fe•EDTA]Hin(139-190), [Fe•EDTA]Hin(139-184), [Fe•EDTA]Hin(139-190)R140→K, [Fe•EDTA]Hin(139-190)R140→A, and [Fe•EDTA]Hin(139-190)R142→K bound to *hix*L

The Quantitative Affinity Cleaving Experiment

A typical quantitative affinity cleaving experiment is conducted as follows. Each gel contains 21 data lanes, one G reaction sequencing lane, and

one control lane. The 6mm-wide lanes are separated by 6mm spacings using a comb of the same design as in the footprinting experiment. At three points per order of magnitude in protein concentration, this allows for five orders of magnitude in protein concentration to be studied per gel. A practical limit for protein concentration is 200μM; above this limit, the surfactant properties of proteins make accurate quantitation difficult. The DNA restriction fragment is approximately 200 base pairs in length and contains the *hix*L, secondary and tertiary Hin binding sites. In order to maximize the specific activity of the radiolabelled DNA, four radioactive deoxynucleotides are incorporated into each molecule. Thus, the final DNA concentration is subpicomolar, whereas the lowest protein concentration studied is nanomolar; these conditions allow use of the necessary approximation that $[L]_{total}$ be equal to $[L]_{free}$.

A solution containing EDTA-derivatized protein and Fe(II) was allowed to equilibrate and was then diluted appropriately for each of the 21 affinity cleaving reactions. Mechanized pipetmen were used to measure solutions with precision and accuracy. A stock solution containing radiolabelled DNA, buffer, and salts was prepared, distributed to each reaction tube, and allowed to equilibrate with the appropriately diluted protein at room temperature. Dithiothreitol (final DTT concentration=1mM) was added to each reaction, and the cleavage chemistry was allowed to proceed at room temperature for 10 minutes (for different proteins, this time varied, but was held constant for each experiment). The reaction was stopped and extracted with a 2:1 solution of phenol:chloroform to remove the protein, which at high concentrations has strong surfactant properties that make manipulations difficult. Reactions were then extracted with butanol to rid reactions of phenol, followed by ethanol precipitation. Reactions were taken up in formamide loading buffer containing 0.1% sodium dodecylsulfate,

which helps to mitigate the surfactant properties of any residual protein, and loaded on a denaturing polyacrylamide gel. All reactions were appropriately loaded on a gel such that each lane contained the same amount of radioactivity, as measured by scintillation counting. After loading, the tubes containing residuals from each reaction were again subjected to scintillation counting to ensure that all lanes had been loaded consistently. Lanes that were found to be loaded improperly were not used to obtain data.

Gels were run until at least three gels of high quality were obtained, and in many cases, four excellent gels were obtained. Data were collected for [Fe•EDTA]Hin(139-190) (the EDTA-GABA derivatized Hin recombinase 52 amino-acid DNA binding domain), [Fe•EDTA]Hin(139-184), and the mutant proteins [Fe•EDTA]Hin(139-190)R140→K, [Fe•EDTA]Hin(139-190)R140→A, and [Fe•EDTA]Hin(139-190)R142→K (see Chapter 2 for discussion of these affinity cleaving proteins). Data were analyzed only for the *hix*L binding site, although strong binding was also exhibited at the secondary Hin site. The proteins chosen for study were the strongest DNA binding proteins; the other mutants–[Fe•EDTA]Hin(139-190)R140→βA, [Fe•EDTA]Hin(139-190)R140→E, [Fe•EDTA]Hin(139-190)R140→G, [Fe•EDTA]Hin(139-190)R140→Q, and [Fe•EDTA]Hin(141-184)–bound only weakly, making accurate quantitation of their binding constants impossible. In the quantitative affinity cleaving experiment, binding constants as low as 10^5 M^{-1} were successfully measured.

Quantitation of Radioactive Gels

Gels were dried and exposed to photostimulable, storage phosphor-imaging plates.[185] A Molecular Dynamics 400S PhosphorImager was used to obtain data from these storage screens. Before the availability of two-dimensional analysis, X-ray film autoradiography was used for gel quantitation. Autoradiograms were scanned using an LKB UltroScan XL laser

densitometer, which has only one-dimensional scanning capabilities. By performing five one-dimensional scans separated by 1mm through each data lane and averaging the five scans, an approximation of a two-dimensional scan was obtained. The binding constants reported below are PhosphorImager data (except for gel retardation assays), although extensive work on the quantitative affinity cleaving technique was originally performed by laser densitometry.

Fitting Procedure

Each gel ideally provides 21 data points. All data points were included in analysis unless visual inspection revealed flaws at either the target or reference sites, if the cleavage intensity value C for a lane was greater than two standard deviations away from values in neighboring lanes, or a lane was found to have been loaded with an improper amount of radioactivity (discussed above). Most gels contained at least one rejected data point; a gel was rejected for analysis if more than three points were rejected. The site-specific cleavage for each lane was calculated using the equation

$$C = C_{tot} - kC_{ref},$$

where C_{tot} and C_{ref}, the cleavage intensities at the target and reference sites respectively, were determined from densitometry or phosphorimaging as described above. A theoretical binding curve[186,187] was fit from the experimental data using the apparent maximum cleavage C_{sat}, the binding constant K_a, and the scaling factor k as adjustable parameters with the following equation.

$$C_{fit} = \frac{C_{sat}(K_a[L])^n}{1 + (K_a[L])^n} + kC_{ref}$$

The exponent n is the Hill coefficient. When n=1, the system displays no cooperativity; when n=2, the system is dimerically cooperative. Because these quantitative affinity cleaving experiments were conducted on just the DNA binding domains of proteins (i.e., no functional or dimerization domains were present), no cooperativity was expected. However, upon fitting the data to the equation above, Hill coefficient values of around 2 were often obtained. This phenomenon was also seen in quantitative affinity cleaving experiments on triple-helical complexes; when the triplex equilibration time was greatly increased, however, Hill coefficients of around 1 were obtained. The obtained binding constant values were not affected by the changes in equilibration time. It is suspected that short equilibration times do not allow for proper equilibration of the ligand-DNA complex at lower ligand concentrations, and therefore, the resultant affinity cleavage drops off more quickly than it should, giving a binding curve with a steeper slope than expected.[187]

The difference between C_{fit} and C was minimized using a standard least-squares fitting algorithm (FORTRAN)[186,187] on an IBM PC/AT computer. The goodness of fit of the binding curve to the data points was judged by the reduced χ^2 criterion. Fits were judged acceptable for $\chi^2 \leq 1.0$, and most reduced χ^2 values were well below 0.5 (Table 1).

Error Analysis

All fits described in the text were performed with statistical weighting of the data points. The precision in the experiments was very high, indicating that sources of random error were well controlled. Because the bulky, negatively charged EDTA•Fe moiety is covalently attached to the protein, the binding constant values obtained were expected to be slightly lower than

those for the underivatized protein. For quantitative affinity cleaving and footprinting studies, the error in the observed cleavage intensity was assumed to be proportional to the error in the specific activity of the DNA sample loaded onto each lane. The largest contribution to the error in loaded radioactivity is the strong surfactant properties of proteins at high concentrations; protein concentrations above 10μM caused difficulties in manipulations of the reactions and gel loading. After ethanol precipitation and drying of each reaction, the precipitant was resuspended in loading buffer. High protein concentrations caused these solutions to stick to the plastic walls of the Eppendorf tubes and pipet tips; phenol:chloroform extractions removed most of the protein, but some of the surfactant problems persisted. Proteins with lower binding constants ($\leq 10^6$) were more difficult to measure because higher concentrations of protein had to be used in order to achieve saturation binding. Standard deviations among different measurements of K_a were typically 20-50% of the average K_a.

RESULTS

In Table 1 are listed the binding constants and their averages obtained from individual gels; each binding constant was assessed at 20mM NaCl. Binding constants for each *hix*L half site, IRL and IRR (inverted repeats left and right), are listed. A block of the four most intense cleavage bands for each half site and a reference block of 4-7 bands for each lane were quantitated by photostimulable, storage phosphor autoradiography,[185] and the optical densities for each block (in absorbance units x mm^2) are reported. The values shown in Table 1 are from proteins derivatized with the affinity-cleaving moiety EDTA-GABA. DNase I quantitative footprinting was performed on [Fe•EDTA]Hin(139-190) in order to correlate data from the footprinting and

TABLE 1.

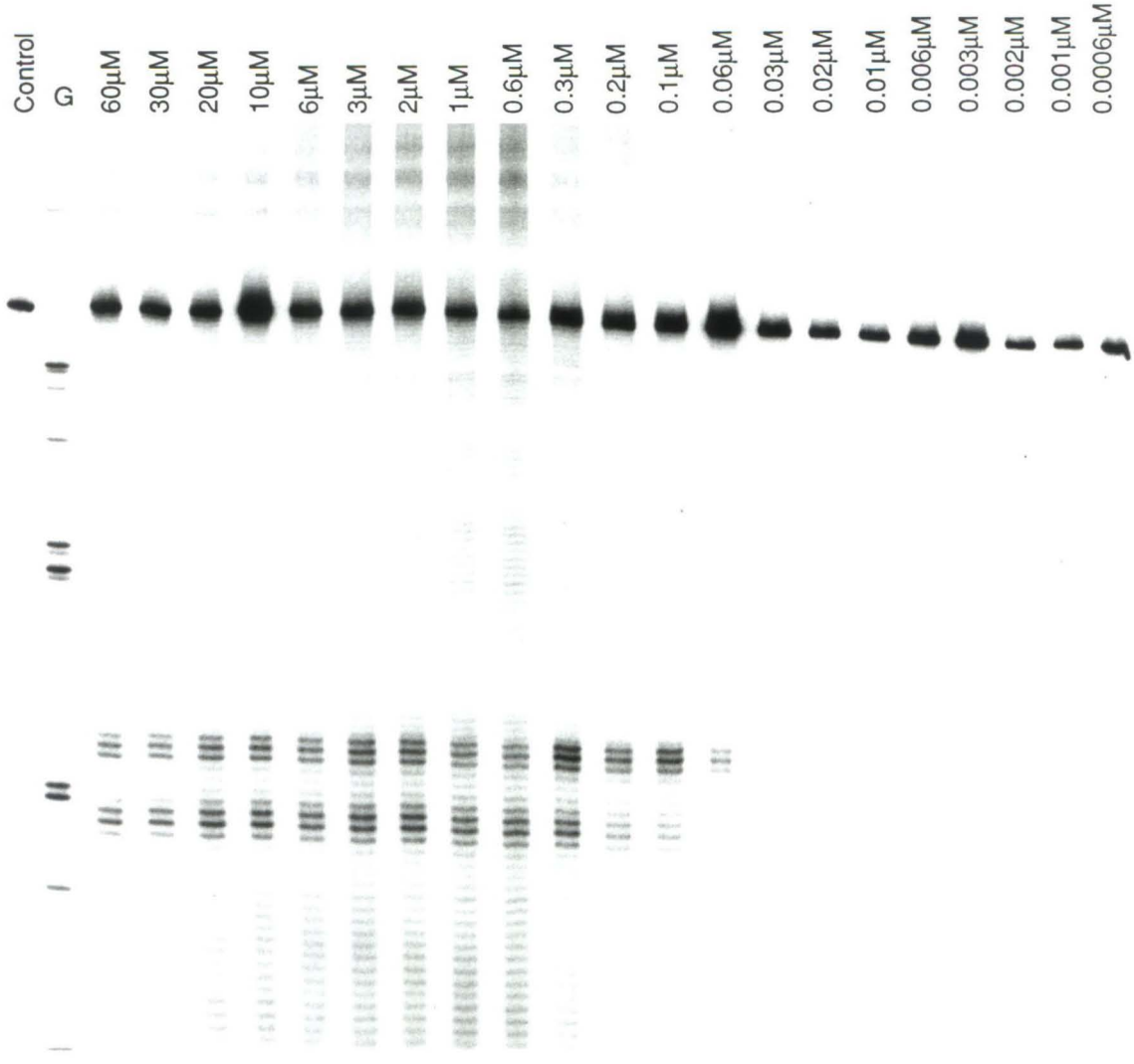
Gel	K_a (M^{-1}) [NaCl]=20mM	χ^2	Average K_a	Standard Deviation
[Fe•EDTA]Hin(139-190) IRL				
JAS III-34 IRL	4.44×10^6	0.5718	**4.69×10^6**	7.24×10^5
JAS III-37 IRL	4.13×10^6	0.2092		
JAS III-38 IRL	5.51×10^6	0.1905		
[Fe•EDTA]Hin(139-190) IRR				
JAS III-34 IRR	2.25×10^7	0.3184	**1.84×10^7**	4.90×10^6
JAS III-37 IRR	1.30×10^7	0.2142		
JAS III-38 IRR	1.98×10^7	0.2804		
[Fe•EDTA]Hin(139-184) IRL				
JAS III-49 IRL	6.30×10^5	0.0404	**7.32×10^5**	8.72×10^4
JAS III-50 IRL	7.14×10^5	0.2206		
JAS III-51 IRL	7.42×10^5	0.2216		
JAS III-52 IRL	8.41×10^5	0.1368		
[Fe•EDTA]Hin(139-184) IRR				
JAS III-49 IRR	6.92×10^5	0.0467	**8.17×10^5**	9.76×10^4
JAS III-50 IRR	7.91×10^5	0.1858		
JAS III-51 IRR	8.70×10^5	0.1543		
JAS III-52 IRR	9.14×10^5	0.1449		
[Fe•EDTA]Hin(139-190)R140→K IRL				
JAS III-53 IRL	2.30×10^6	0.0960	**1.84×10^6**	4.14×10^5
JAS III-55 IRL	1.50×10^6	0.5344		
JAS III-56 IRL	1.72×10^6	0.4617		
[Fe•EDTA]Hin(139-190)R140→K IRR				
JAS III-53 IRR	4.27×10^6	0.0879	**2.91×10^6**	1.19×10^6
JAS III-55 IRR	2.07×10^6	0.3484		
JAS III-56 IRR	2.31×10^6	0.2527		
[Fe•EDTA]Hin(139-190)R140→A IRL				
JAS III-88 IRL	1.39×10^5	0.5295	**3.50×10^5**	2.53×10^5
JAS III-90 IRL	4.94×10^5	0.6743		
JAS III-91 IRL	6.33×10^5	0.3167		
JAS III-92 IRL	1.33×10^5	0.2987		
[Fe•EDTA]Hin(139-190)R140→A IRR				
JAS III-88 IRR	9.09×10^4	0.1692	**2.00×10^5**	1.46×10^5
JAS III-90 IRR	2.96×10^5	0.3738		
JAS III-91 IRR	3.54×10^5	0.1209		
JAS III-92 IRR	6.11×10^4	0.1814		
[Fe•EDTA]Hin(139-190)R142→K IRL				
JAS III-93 IRL	9.57×10^5	0.4985	**9.21×10^5**	1.51×10^5
JAS III-95 IRL	1.07×10^6	0.3552		
JAS III-96 IRL	7.10×10^5	0.8901		
JAS III-97 IRL	9.50×10^5	1.3237		
[Fe•EDTA]Hin(139-190)R142→K IRR				
JAS III-93 IRR	1.34×10^6	0.1633	**1.94×10^6**	7.28×10^5
JAS III-95 IRR	2.90×10^6	0.4244		
JAS III-96 IRR	2.10×10^6	0.5663		
JAS III-97 IRR	1.41×10^6	0.6242		

affinity cleaving experiments. DNase I is a large enzyme that cleaves only selectively, and therefore, it tends to give large footprints[154]; in the case of Fe•EDTA]Hin(139-190) (and by extension, the other Hin proteins), DNase cannot resolve the two half sites of *hix*L, so any binding constant measurement could be only for the full site. Quantitative footprinting data (not shown) for Fe•EDTA]Hin(139-190) gave binding constants similar to those obtained from affinity cleaving, although resolution of the half sites is not possible by DNase footprinting. Clearly, footprinting is limited by the resolution capabilities of DNase I; affinity cleaving, however, has no such limits on resolution.

Gel retardation assays,[188] also known as Crothers' gels, were performed on Hin(139-190), the underivatized Hin DNA binding domain. The concept behind gel retardation is that protein binding to DNA will cause a decrease in mobility of the DNA during gel electrophoresis. The protein-DNA complex moves more slowly through a gel matrix, for not only is this complex much larger than the DNA alone, but also protein binding typically causes kinking at the DNA binding site, and both factors contribute to decreased mobility. It is important to note that proteins are often positively charged, whereas DNA is negatively charged; although the protein-DNA complex is overwhelmingly negatively charged, the protein-DNA complex "breathes," meaning that the protein and DNA separate and rebind according to the system's kinetics, and that the gel matrix provides a "cage" around the separated complex, helping to force it to rebind. It is possible that during those times that the protein is unbound to DNA, it will tend to migrate away from the DNA toward the opposite electrode. Thus, binding constants obtained by gel retardation may be on the low side, and it will be difficult to visualize a shift in DNA mobility when proteins with low binding constants are studied. Interestingly,

FIGURE 3. Autoradiogram of a quantitative affinity cleaving gel of [Fe•EDTA]Hin(139-190) bound to the 3′ end-labelled Xba I-Spe I fragment (218 base pairs) from pMFB36. This autoradiogram is from the gel labelled JAS III-38 (see JAS notebook III). Lane 1, control; lane 2, G sequencing reaction. Lanes 3-23 contain decreasing concentrations of protein as indicated. Each reaction contains 20mM phosphate buffer, pH 7.5, 20mM NaCl, and 1mM dithiothreitol. Each reaction proceeds at room temperature for 10 minutes. Reactions are stopped and extracted with 200μL of a 2:1 solution of phenol:chloroform, followed by butanol extraction and ethanol precipitation. Reactions are taken up in formamide loading buffer and loaded onto an 8% polyacrylamide denaturing gel.

Quantitative Affinity Cleaving Gel
[Fe•EDTA]Hin(139-190)



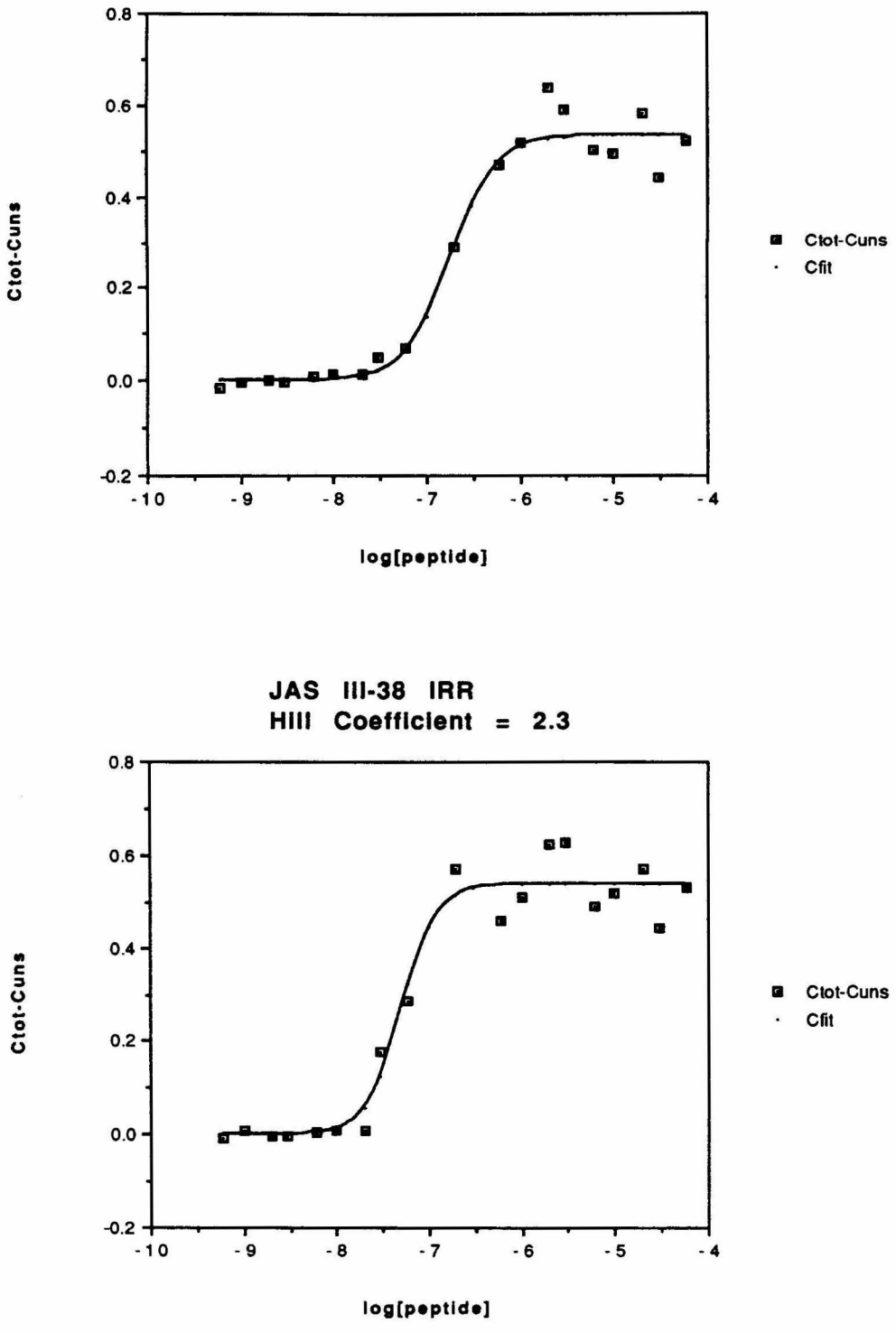


FIGURE 4. Binding isotherms for gel JAS III-38. **Top.** Isotherm for the *hix*L IRL binding site. **Bottom.** Isotherm for the *hix*L IRR binding site.

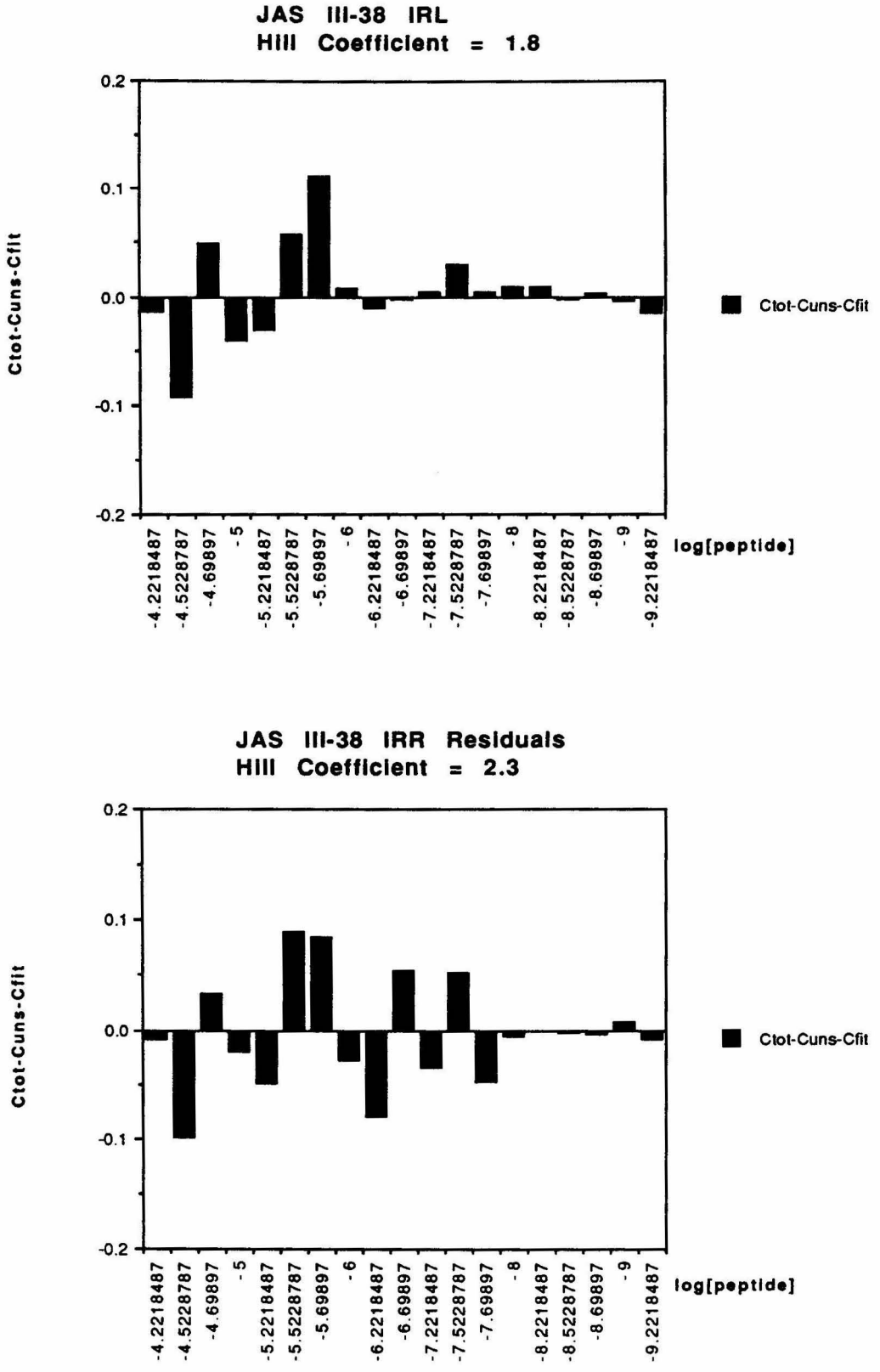


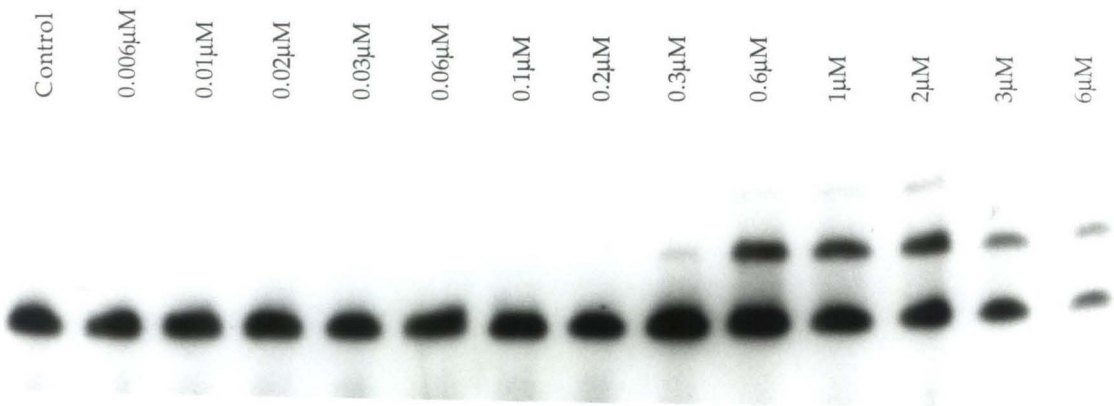
FIGURE 5. Residuals for data from JAS III-38 shown in Figure 4; residuals measure the difference between the obtained data points and the fit curve.

although binding constants could be extracted from gel retardation assays for the 52-amino acid Hin DNA binding domain Hin(139-190), the truncated proteins synthesized by Jim Sluka (discussed in Chapter 2) did not give gel shifts; the 51-mer Hin(140-190) gave a smear, and the 50- and 49-mers, Hin(141-190) and Hin(142-190) respectively, gave no gel shifts. By affinity cleaving, the 51-mer showed strong binding at *hix*L, and the 50- and 49-mers both showed detectable binding. Thus, affinity cleaving is a more sensitive technique for detecting protein-DNA binding than is gel retardation. A representative quantitative affinity cleaving gel of [Fe•EDTA]Hin(139-190) is shown in Figure 3 and the data in Figure 4; representative gels and data of [Fe•EDTA]Hin(139-184), [Fe•EDTA]Hin(139-190)R140→K, [Fe•EDTA]Hin(139-190)R140→A, and [Fe•EDTA]Hin(139-190)R142→K are shown in the Appendix.

Gel retardation assays were performed on Hin(139-190), and mobility shifts could be seen both for monomeric protein binding followed by dimeric binding (the DNA fragment used in these studies contained only the *hix*L site; see Figure 6); as another comparison to affinity cleaving, it cannot be known from gel retardation which half site has the higher binding constant and saturates first, whereas with affinity cleaving, the differences between half sites is clearly visualized. Gel retardation data (obtained by laser densitometry) are listed in Table 2. The thermodynamic values obtained for Hin(139-190) by gel retardation methodology are very similar to those obtained for [Fe•EDTA]Hin(139-190) by quantitative affinity cleaving. It was expected that the affinity cleaving values would be slightly lower because of attachment of the EDTA-GABA moiety at the amino terminus of [Fe•EDTA]Hin(139-190); however, Scott Singleton has shown that covalent attachment of EDTA-GABA at the 5′ terminus of an oligodeoxyribonucleotide

FIGURE 6. Autoradiogram of a gel retardation assay (8% polyacrylamide, 1:30 cross-link, 50% urea) on Hin(139-190). The gel is JAS II-80. Reaction concentrations are indicated in the figure. Each reaction was run in a total volume of 20μL. Each reaction contained 80mM NaCl, 20mM Tris-HCl buffer, pH 7.6, 1.2mM EDTA, 0.5 mg/ml bovine serum albumin, and 1mM dithiothreitol and was allowed to incubate with protein for ~20 minutes at room temperature. 5μL of ficoll loading buffer were added to each reaction; reactions were then loaded onto an 8%, 1:30 cross-link nondenaturing polyacrylamide gel using a 100μL Hamilton syringe. Each gel holds fourteen reactions that were loaded, starting with the reaction containing the lowest protein concentration. The syringe was flushed between each lane loading. The gel was loaded while it was running at 80V and turned up to 230V after loading for ~3 minutes to run samples into the gel. The voltage was then turned down to 90V and run for 1-2 hours. Gels were then dried and exposed to X-ray film.

Gel Retardation Assay
Hin(139-190)



had no measurable effect on the stability of the corresponding triple-helical complex.

TABLE 2.

Protein bound to *hix*L	[NaCl] mM	K_a (M^{-1})	Average K_a	Method
Hin (full protein)	100	2.5×10^7		DNase I footprinting
Hin(139-190)*	20	7.1×10^6		Gel retardation
Hin(139-190)				Gel retardation
JAS II-95	20	8.0×10^6	4.0×10^6	
JAS II-80	20	3.0×10^6		
JAS II-81	20	1.1×10^6		
Hin(139-190) IRL				Gel retardation
JAS II-95	20	5.3×10^6	2.3×10^6	
JAS II-80	20	1.4×10^6		
JAS II-81	20	2.2×10^5		
Hin(139-190) IRR				Gel retardation
JAS II-95	20	8.6×10^6	6.2×10^6	
JAS II-80	20	5.3×10^6		
JAS II-81	20	4.9×10^6		

*Reported by Anna Glasglow.

DISCUSSION

Examination of the binding constants in Table 1 reveals that quantitative affinity cleaving gives highly reproducible results; even studies on those proteins with weaker binding constants, [Fe•EDTA]Hin(139-190)R140→A and [Fe•EDTA]Hin(139-190)R142→K, give very consistent values. The lowest binding constants, ~10^5 M^{-1}, appear to be the limit for measurement by quantitative affinity cleaving studies on protein-DNA complexes because of the difficulty in manipulating the reactions. Thus, quantitative affinity cleaving can precisely measure binding constants one to

two orders of magnitude lower than those detected by quantitative footprinting and gel retardation assays. This is a significant improvement over previous techniques.

The binding constants for Hin(139-190) obtained from quantitative affinity cleaving and gel retardation, a proven technique for studying protein-DNA complexes,[188] are strongly corroborative; the values obtained from affinity cleaving are higher than those from gel retardation by a factor of two, but this is well within experimental error, and as was discussed above, gel retardation may give lower thermodynamic values. Additionally, the gel retardation assay requires use of short DNA fragments (≤ 100 base pairs); a 56 base-pair fragment containing the *hix*L site was used in these experiments. The *hix*L site is situated seven base pairs from one end of the DNA, and end effects may cause a decrease in binding. Importantly, the binding constants from quantitative affinity cleaving data for each half site are consistently twice those from gel retardation data. So quantitative affinity cleaving is not only a precise method producing internally consistent binding constants, but also these values compare favorably with those obtained by proven techniques. Therefore, quantitative affinity cleaving is a sensitive and valid method for studying thermodynamic properties of ligand-DNA complexes.

Further examination of Table 1 reveals several interesting points. The binding constant for [Fe•EDTA]Hin(139-190) binding at the *hix*L IRR half site is about 3-4 times higher than that for [Fe•EDTA]Hin(139-190) at *hix*L IRL. However, the proteins in which Arg140 is mutated, [Fe•EDTA]Hin(139-190)R140→K and [Fe•EDTA]Hin(139-190)R140→A, give binding constants that are virtually identical; as a matter of fact, the binding constant between [Fe•EDTA]Hin(139-190)R140→A and IRL is higher than that for IRR, the

reverse of [Fe•EDTA]Hin(139-190). Arg140 lies in the minor groove in a tract of A's, which are known to give a narrowed minor groove and appear to be a common recognition element for arginine (discussed in Chapter 1). At *hix*L IRR, this tract contains five A's, whereas IRL contains only four A's; thus, the IRR minor groove is likely to be slightly narrower than that of IRL. This may be an extremely important recognition element for Arg140; previously, Jim Sluka demonstrated the importance of Arg140 for the Hin DNA binding domain's sequence-specific recognition of DNA. In the mutant proteins [Fe•EDTA]Hin(139-190)R140→K and [Fe•EDTA]Hin(139-190)R140→A, there is no Arg140 to recognize the minor groove and to discriminate between tracts containing four A's versus five A's; note that no preferential binding at IRR is exhibited by these mutants. Also, the binding constants for [Fe•EDTA]Hin(139-190)R142→K, which contains Arg140 but not Arg142, show that [Fe•EDTA]Hin(139-190)R142→K prefers the IRR half site over IRL by over a factor of two, which is similar to that of [Fe•EDTA]Hin(139-190), though [Fe•EDTA]Hin(139-190) shows stronger discrimination between the two sites.

Seemingly at odds with the above explanation are the data from [Fe•EDTA]Hin(139-184), which contains Arg140 but is missing the six carboxyl-terminal residues; the IRL and IRR binding constants are almost equal. Removal of these six amino acids, three of which are positively charged, from [Fe•EDTA]Hin(139-190) to afford [Fe•EDTA]Hin(139-184) reduces binding by a factor of ~50. As was discussed in Chapter 2, comparison of the affinity cleaving patterns of [Fe•EDTA]Hin(139-184) and [Fe•EDTA]Hin(139-190) indicates little change in their DNA binding specificities, so these six residues make a significant, generally non-specific binding contribution. But these carboxyl-terminal residues make some

contribution toward sequence discrimination between *hix*L IRR and IRL. Extensive work developing the quantitative affinity cleaving method leaves little doubt that any of the results gained are spurious, so it is difficult to explain the data from [Fe•EDTA]Hin(139-184). Perhaps one reasonable explanation is that removal of the six carboxyl-terminal residues causes the Hin DNA binding domain to alter its conformation somewhat; because it is now missing three positively charged residues (2 Arg, 1 Lys), which most likely interact with the negatively charged DNA backbone, the protein may try to compensate for this loss. Another way to look at the problem is to note that the intact protein [Fe•EDTA]Hin(139-190) employs three recognition elements to bind to specific sequences of DNA: the helix-turn-helix structure in the major groove, the carboxyl-terminal residues in the major groove, and the amino terminal Arg-Pro-Arg in the minor groove (discussed in Chapter 2). The intact protein, therefore, adopts a conformation that maximizes all three elements' interactions with DNA. The truncated protein [Fe•EDTA]Hin(139-184) is missing the carboxyl-terminal element in the major groove; thus, the truncated protein adopts a conformation that maximizes DNA interactions with only two recognition elements. Although [Fe•EDTA]Hin(139-184) exhibits the same overall binding preference as does [Fe•EDTA]Hin(139-190) (*hix*L > secondary >> tertiary), there exist subtle differences in binding preference, and these differences may be partly due to the powerful abilities of proteins to force themselves and their DNA binding sites to adopt conformations that maximize protein-DNA interactions.

CONCLUSION

By extension of the affinity cleaving technique, quantitative affinity cleaving has been developed as a powerful tool for the measurement of

thermodynamic parameters of ligand-DNA complexes. The theory behind quantitative affinity cleaving is based on that for quantitative DNase I footprinting, a known technique for measurement of thermodynamic properties of protein-DNA complexes. Equilibrium binding constant values for protein-DNA complexes obtained by quantitative affinity cleaving were highly reproducible and compared very favorably with values gained from gel retardation assays, also a known technique. Binding constants were obtained for [Fe•EDTA]Hin(139-190), [Fe•EDTA]Hin(139-184), [Fe•EDTA]Hin(139-190)R140→K, [Fe•EDTA]Hin(139-190)R140→A, and [Fe•EDTA]Hin(139-190)R142→K at the *hix*L IRR and IRL half sites, and binding constants as low as ~10^5 could be extracted. These results provide a quantitative basis for understanding protein complexation with DNA.

MATERIALS AND METHODS

Proteins were synthesized and prepared as described in Chapter 2.

Radioactive Labelling of Restriction Fragments. The plasmids pMFB36 and pMS536 were generous gifts from Prof. Mel Simon, Department of Biology, California Institute of Technology.[189] Plasmid pMFB36, used in the quantitative affinity cleaving experiments, was linearized by digestion with Xba I. The linearized plasmid was 3′ end-labelled with [α^{32}P]-dATP, -dCTP, -dGTP, and -dTTP and DNA polymerase I Klenow fragment.[177] Labelled linearized plasmid pMFB36 was digested with Spe I, and the resulting labelled 218 base pair DNA fragment was isolated using non-denaturing gel electrophoresis. Plasmid pMS536, used in the gel retardation assays, was linearized by digestion with Hind III. The linearized plasmid was 3′ end-labelled with [α^{32}P]-dATP, -dCTP, -dGTP, and -dTTP and DNA polymerase I Klenow fragment.[177] Labelled linearized pMS536 was digested with EcoR I,

and the resulting labelled 56 base pair DNA fragment was isolated using non-denaturing gel electrophoresis.

Quantitative Affinity Cleaving Reaction Conditions. A concentrated solution of protein and Fe^{II} in a 1:1 mixture was allowed to equilibrate at room temperature ~10 minutes. The protein was then diluted into 21 separate Eppendorf tubes for each of the 21 affinity cleaving reactions. EDP2 Mechanized Pipetmen (Rainin) were used to measure solutions with precision and accuracy. Reactions were run in a total volume of 10μM. A stock solution containing radiolabelled DNA, buffer, and salts was prepared such that each reaction would contain 20mM phosphate buffer, pH 7.5, and 20mM NaCl. 6μL of this stock was distributed to each reaction tube and allowed to equilibrate with 2μL of the appropriately diluted protein at room temperature, ~15 minutes. 2μL of 5mM dithiothreitol (final DTT concentration=1mM) was added to each reaction and allowed to proceed at room temperature, 10 minutes (for different proteins, this time varies, but each gel is internally consistent). The reaction was stopped by addition of 50μL of water containing 2μL of 1mg/mL tRNA carrier. Each reaction was extracted with 200μL of a 2:1 solution of phenol:chloroform. Reactions were then extracted with butanol, followed by ethanol precipitation. Reactions were taken up in formamide loading buffer containing 0.1% sodium dodecylsulfate and loaded on denaturing polyacrylamide gels. All reactions were appropriately loaded on a gel such that each lane contained the same amount of radioactivity, which was determined by scintillation counting. After loading, the tubes containing residuals from each reaction were again subjected to scintillation counting to ensure that all lanes had been loaded consistently. Gels were then dried and exposed to storage phosphor-imaging plates.

Gel Retardation Assay Reaction Conditions. This protocol came from Anna Glasgow. Each reaction was run in a total volume of 20μL. Each reaction contained 80mM NaCl, 20mM Tris-HCl buffer, pH 7.6, 1.2mM EDTA, 0.5 mg/ml bovine serum albumin, and 1mM dithiothreitol. Each reaction was allowed to incubate with protein for ~20 minutes at room temperature. 5μL of ficoll loading buffer (1x reaction buffer, 5% ficoll, 0.1% bromophenol blue) was added to each reaction; reactions were then loaded onto an 8%, 1:30 cross-link, non-denaturing polyacrylamide gel (pre-electrophoresed at 80V for ~30 minutes in 1x TBE buffer), using a 100μL Hamilton syringe equipped with tygon tubing on the needle tip to prevent bubble formation. Each gel holds fourteen reactions which were loaded starting with the reaction containing the lowest protein concentration. The syringe was flushed with water between each lane loading. The gel was loaded while it was running at 80V and turned up to 230V after loading for ~3 minutes to run samples into the gel. The voltage was then turned down to 90V and run for 1-2 hours. Gels were then dried and exposed to X-ray film.

Quantitation of Radioactive Gels. The gel retardation assay gels were exposed to Kodak XRP film. Autoradiograms were scanned using an LKB UltraScan XL laser densitometer. The output report was directly sent to the LKB data analysis program GSXL on an IBM PC/AT computer. Data were then fit using a nonlinear least-squares fitting program (FORTRAN) on an IBM PC/AT. The quantitative affinity cleaving gels were quantitated using storage phosphor technology. Kodak Storage Phosphor Screens S0230 from Molecular Dynamics were pressed against dried gels and exposed. A Molecular Dynamics 400S PhosphorImager was used to obtain data from storage screens, and data were analyzed by performing volume integrations of the target and reference sites using the ImageQuant v.3.0 software running on

an AST Premium 386/33 computer. Data were fit using a nonlinear least-squares fitting program on an IBM PC/AT computer.

CHAPTER THREE

Affinity Cleaving Studies on the *lac* Repressor DNA Binding Domain

INTRODUCTION

Although the *lac* repressor-*lac* operator system of *Escherichia coli* serves as the classic model for genetic regulatory systems, the details of the repressor-operator interaction are poorly understood. Because the *lac* repressor protein sequence is highly homologous to helix-turn-helix proteins with known (co)crystal structures including the λ repressor,[25] λ cro,[46,190] 434 repressor,[26,39,40,139] 434 cro,[42,43] and CAP,[57] a model for the *lac* repressor-operator interaction inferred from these known structures has been proposed.[72] However, on the basis of proton NMR spectroscopic investigations, Kaptein and co-workers have proposed that the helix-turn-helix motif of the *lac* repressor binds to DNA in an orientation opposite that of the helix-turn-helix motifs of the λ repressor, λ cro, 434 repressor, 434 cro, and CAP.[66] The orientation of the helix-turn-helix motif of *lac* repressor in the *lac* repressor-DNA complex has been determined by the affinity cleaving method. The DNA cleaving moiety EDTA•Fe was attached to the amino terminus of a 56-residue synthetic protein corresponding to the DNA binding domain of the *lac* repressor. The affinity cleaving pattern resulting from the complex of the EDTA•Fe-derivatized *lac* repressor DNA binding domain and the *lac* operator site strongly supports the proposal of Kaptein and co-workers; although *lac* repressor is a helix-turn-helix protein, it does not belong to the same class of helix-turn-helix motif as that of the λ repressor and cro, 434 repressor and cro, and CAP.

THE *LAC* PARADIGM

From a genetic and biochemical standpoint, one of the best-characterized protein-DNA complexes studied is the *lac* repressor-*lac* operator system.[1,2] *Lac* repressor is a transcriptional regulatory protein that governs the expression of three enzymes involved in the metabolism of lactose: *lacZ* codes for the enzyme β-galactosidase, *lacY* codes for β-galactoside permease, and *lacA* codes for β-galactoside transacetylase (Figure 1).[2] These three genes comprise the *lacZYA* gene cluster; this gene cluster and those elements that regulate its expression comprise a unit of gene expression commonly referred to as an operon, and the *lac* operon serves as the most extensively studied paradigm of gene expression.

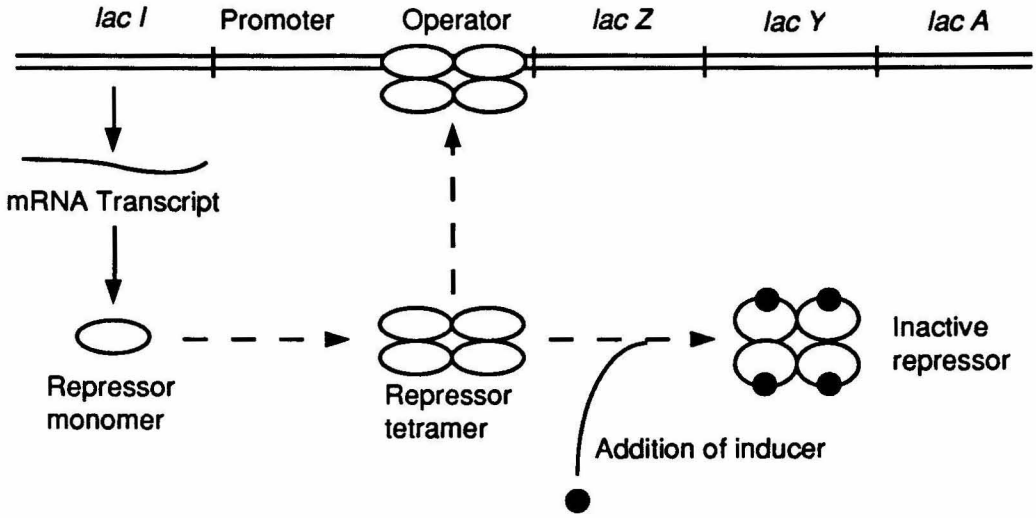


FIGURE 1. The *lac I* gene synthesizes a repressor whose tetramer binds to the operator and prevents transcription of the structural gene cluster *lacZYA*. Addition of inducer converts the repressor into an inactive form that cannot bind to the operator; transcription can now start at the promoter, and the three metabolic enzymes are synthesized.

The *lac* repressor functions by binding to a specific DNA site, *lac* operator, located adjacent to the *lac* promoter. Transcription of the *lacZYA* gene cluster is controlled by negative regulation, meaning that it is

transcribed unless repressed by the regulatory protein *lac* repressor. *Lac* repressor binds to its operator and prevents transcription of *lacZYA*, but when the repressor binds the molecule lactose, repressor is converted to an inactive form that cannot bind to the operator. Therefore, transcription starts at the promoter, the three enzymes are produced that will then metabolize the lactose, and regulation of the cellular concentration of lactose is achieved. Lactose is an inducer molecule, for it causes the production of enzymes capable of metabolizing it. When the inducer binds at one site of the *lac* repressor, it changes the protein's conformation at another site; in this case, lactose binding changes the protein conformation at its DNA binding site, thus affecting the ability of the *lac* repressor to binding to its operator. This is an example of allosteric control.

THE *LAC* REPRESSOR DNA BINDING DOMAIN

Lac repressor and its specific DNA site have been subjected to extensive structural, biophysical, biochemical, and genetic analyses. *Lac* repressor is a tetramer of four chemically identical subunits, each consisting of 360 amino acids[1]; the *lac* operator exhibits twofold symmetry and binds two subunits of the tetrameric repressor (Figure 1). The DNA binding domain, or headpiece, consists of the first 50-60 residues of the *lac* repressor, while the remaining ~300 residues make up the core of the *lac* repressor. Digestion of intact repressor with trypsin yields amino-terminal headpieces 1-51 and 1-59, plus a tetrameric core with full inducer binding activity; chymotryptic digestion gives residues 1-56 and the core.[60-62] All three headpieces fold independently, bind to the *lac* operator in a similar fashion,[2,69] and protect the same bases from methylation as does intact repressor.[62] Although no X-ray crystal structures for the *lac* repressor or its DNA complex exist because of the lack of

suitable crystals, the three-dimensional structures of the DNA binding domain of the *lac* repressor and the *lac* repressor-DNA complex have been determined by proton NMR spectroscopic investigations.[63,65,66,69,143,191,192] The consensus DNA binding site for the *lac* repressor is 5'-AATTGTGAGCGCTCACAATT-3'[193,194]; the site is 20 base pairs in length and exhibits perfect twofold sequence symmetry. The left DNA half site for the *lac* repressor is 5'-AATTGTGAGC-3'.

Lac repressor is a member of a class of at least 150 sequence-specific DNA binding proteins–prokaryotic and eukaryotic–that utilize a highly conserved α-helix-turn-α-helix motif to mediate interaction with DNA.[9,10,46,191,195,196] The helix-turn-helix is the best characterized protein recognition motif for double helical DNA because of the large number of high-resolution crystal structures. Recent cocrystal structures of helix-turn-helix proteins bound to their DNA operators have been determined for the DNA binding domains of the 434 repressor (1-69),[26] the λ repressor (1-92),[25] and the *engrailed* homeodomain from Drosophila.[81] These proteins bind DNA as symmetrical dimers, except for the *engrailed* homeodomain, which binds as a monomer. Each monomer contains a helix-turn-helix unit. One of the helices, commonly referred to as the recognition helix, fits into the major groove and makes base-specific contacts; the other helix packs against the recognition helix, making mostly contacts with the phosphodiester backbone (see Chapter 1). In the case of the *lac* repressor, amino acids 5 through 25 constitute the helix-turn-helix motif.

Two models have been proposed for the structure of the *lac* repressor-DNA complex. The models differ in the orientation of the helix-turn-helix motif of each subunit with respect to the DNA site (Figure 2). Five members of the helix-turn-helix class of proteins, including the λ repressor, 434

repressor, 434 cro, and CAP, have been documented to interact with DNA as illustrated schematically in Figure 2B. Based on amino-acid sequence and three-dimensional structural homology considerations, it was proposed that the helix-turn-helix motif-DNA orientation in the case of *lac* repressor is similar or identical to the orientation documented in the cases of λ repressor and cro, 434 repressor and cro, and CAP.[72,197] In this model, the amino-terminus of the helix-turn-helix motif of *lac* repressor is near base-pair 4 of the DNA half site, and the carboxyl terminus of the helix-turn-helix motif of *lac* repressor is near to base pair 10 of the DNA half site.

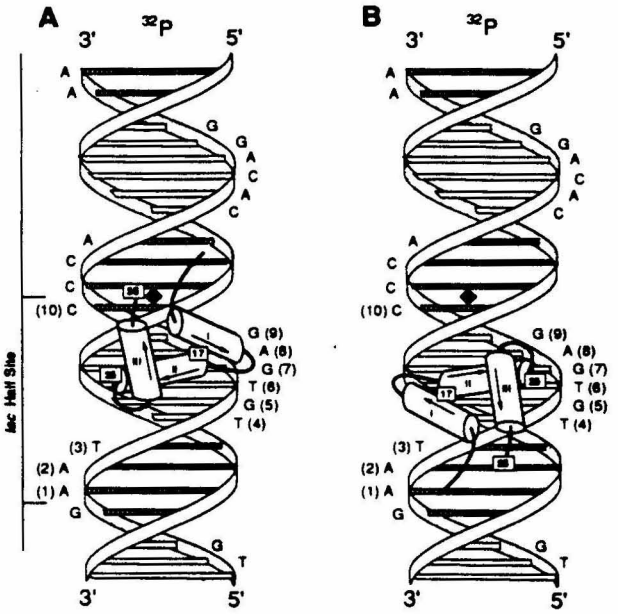


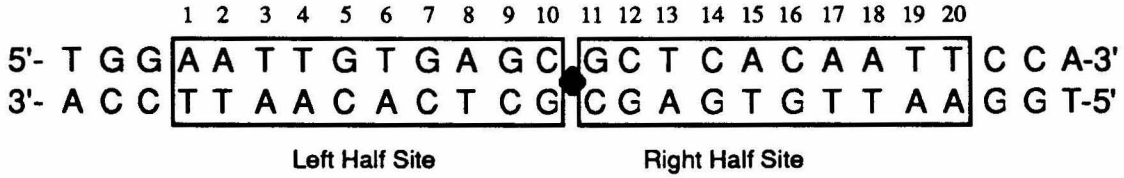
FIGURE 2. The two proposed models for the structure of the complex between [Fe•EDTA]*lac*(1-56) and the left DNA half site. α-Helices are illustrated as cylinders with arrows pointing from the amino terminus to the carboxyl terminus. Black diamond indicates axis of symmetry. **A.** Model of Kaptein and co-workers; the orientation of the helix-turn-helix motif of each subunit is inverted relative to the orientation in B. **B.** Model of Matthews and co-workers; this model is based on the experimentally documented models for the structures of the protein-DNA complexes of λ repressor, λ cro, 434 repressor, 434 cro, and CAP.

In contrast, given the proton NMR spectroscopic investigations, Kaptein and co-workers[66,69,144,198] have proposed that the helix-turn-helix

motif-DNA orientation in the case of *lac* repressor is opposite–180° different from–the orientation in the cases of the other proteins. In this model, the amino terminus of the helix-turn-helix motif of *lac* repressor is near base-pair 10 of the DNA half site, and the carboxyl-terminus of the helix-turn-helix motif of the *lac* repressor is near base pair 4 of the DNA half site (Figure 2A).

Recent genetic[68] and photocrosslinking[199] results provide strong support for the model of Kaptein and co-workers. Genetic results suggesting that amino-acid 18 of *lac* repressor contacts base-pair 7 of the DNA half site[197,200,201] are consistent with both models. However, additional genetic results suggesting that amino-acid 22 of *lac* repressor contacts base-pair 5 of the DNA half site are consistent only with the model of Kaptein and co-workers.[68] Photocrosslinking results suggest that amino-acid 17 of *lac* repressor is close to base-pair 8 of the DNA half site, and that amino-acid 29 of repressor is close to base pairs 3 and 4 of the DNA half site.[199] The numbering scheme for the full consensus *lac* operator (black diamond indicates axis of symmetry) is shown below. Affinity cleaving results discussed below strongly support the model of Kaptein and co-workers.

Consensus *lac* Operator



AFFINITY CLEAVING STUDIES WITH [Fe•EDTA]*lac*(1-56)

The DNA cleaving moiety, ethylenediamine tetraacetic acid complexed with ferrous iron (EDTA•Fe^{II}), can be incorporated at discrete amino-acid

residues in a protein, thus allowing the positions of these residues in a DNA complex to be mapped to nucleotide resolution by high-resolution gel electrophoresis.[100,141,158,160,169,170,202] The affinity cleaving method has been used to determine the location of the amino terminus of *lac* repressor within the *lac* repressor-DNA complex. EDTA was attached, via a four-carbon linker, to the amino terminus of the chemically synthesized 56-residue *lac* repressor DNA binding domain, *lac* repressor(1-56), to construct [EDTA]*lac*(1-56), which was then incubated in solution with Fe^{II} to form [Fe•EDTA]*lac*(1-56) (Figure 3). After formation of the [Fe•EDTA]*lac*(1-56)-DNA complex, DNA cleavage was initiated by addition of the reducing agent sodium ascorbate. DNA cleavage occurs just beyond base-pair 10 of the DNA half site.

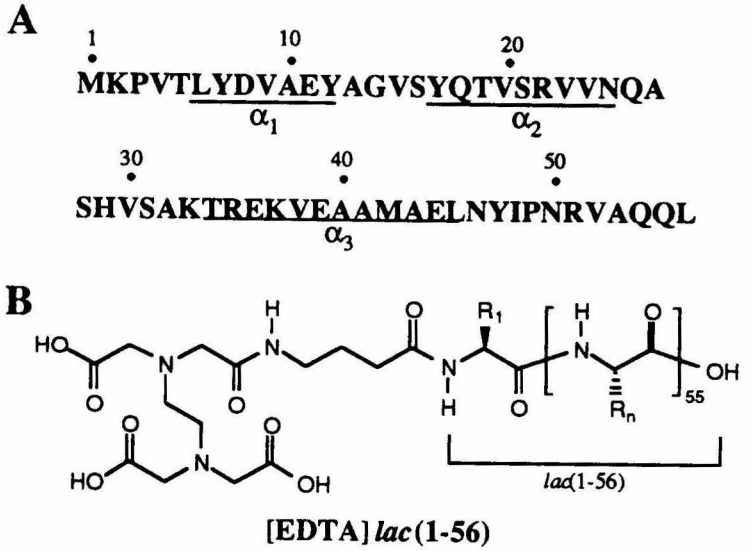


FIGURE 3. A. The sequence of amino acids 1 to 56 of *lac* repressor. **B.** Synthetic protein [EDTA]*lac* (1-56).

The *lac* repressor DNA binding domain containing the amino-terminal residues 1-56 was manually synthesized by standard Merrifield solid phase synthesis methods.[173,174] Circular dichroism shows that the 56-mer possesses significant helical content and gives a spectrum similarly found in

previous headpiece studies.[60] Two proteins were prepared: one protein contains the amino-terminal residues of *lac* repressor, *lac*(1-56), and the other protein is *lac*(1-56) derivatized with the EDTA chelating agent attached to the protein's amino terminus via a four-carbon linker, [EDTA]*lac*(1-56). The protein was purified by high performance liquid chromatography (HPLC), and shown to be homogeneous by mass spectrometry and sequencing analysis.

Originally, two DNA plasmids were sent from R. H. Ebright at Rutgers University. One plasmid, pOE442, contained an idealized, consensus full *lac* operator,[193,203] whereas the other plasmid, pRZ446, contained the wild-type full operator. Working with these plasmids proved to be very difficult: pOE442 did not appear to be the DNA it should have been, and both pOE442 and pRZ446 gave unsatisfactory and ambiguous data. New operators were therefore designed, and originally machine-synthesized oligodeoxyribonucleotides, rather than restriction fragments from plasmid DNA, were used.

Oligonucleotide Synthesis of Palindromic *lac* Operator Sequences

Initial studies on the *lac* DNA binding domain were started on two palindromic oligonucleotides. An oligonucleotide 56-mer containing the full consensus *lac* operator site and a 62-mer containing the full consensus operator but with a six base-pair separation of half sites were prepared (Figure 4). In the 62-mer, the two half sites comprising the full consensus operator are separated by one half-turn of B-DNA (six palindromic base pairs) in order to keep the two half sites and the affinity cleaving patterns distinct and unambiguous. On each side of the operator are idealized *lac* operator flanking sequences.[193,203] Because both the 56-mer and 62-mer are self-complementary, palindromic duplex DNA was expected to form. With this experimental design, both ends of the duplex operator would be radioactively labelled. Normally, only one end of the DNA should be labelled, so that

upon DNA cleavage initiated by [Fe•EDTA]*lac*(1-56), the DNA fragments can be separated and visualized unambiguously by gel electrophoresis. In this case, however, there is a twofold axis of symmetry that removes this restriction.

FIGURE 4. Top. 56-mer palindromic consensus full *lac* operator site (outlined) surrounded by idealized flanking sequences. **Bottom.** 62-mer palindromic consensus full *lac* operator site (outlined) with six base-pair insert (one half turn of B-DNA).

The data generated in this experiments were inconclusive. Control experiments using the single-stranded DNA cleaving enzyme S1 nuclease showed heavy, periodic cleavage of the DNA. This suggested the presence, despite slow annealing conditions intended to favor formation of the thermodynamically favored duplex, of hairpin structures made possible by the use of palindromic sequences. Thus, another experimental approach was required.

Oligonucleotide Synthesis of the *lac* Operator Half Site

In order to circumvent the problems described above, two complementary oligonucleotides, a 42-mer and 38-mer, were synthesized (Figure 5). Only the left consensus half site was used in order to achieve the simplest affinity cleaving data; this site was surrounded by the wild-type flanking sequences. Only one end of the complementary duplex has a four-base pair overhang in which radioactive nucleotides can be incorporated by DNA polymerase; the other end is blunt. Thus, the left end can be labelled by

incorporation of radioactive nucleotides at the 3'-end of the lower strand (38-mer); a 5'-radioactive labelling can be achieved by using T4 kinase to transfer a radioactive phosphate from an ATP molecule to the 5'-end of the upper strand (42-mer), followed by annealing to the 38-mer to form the duplex.

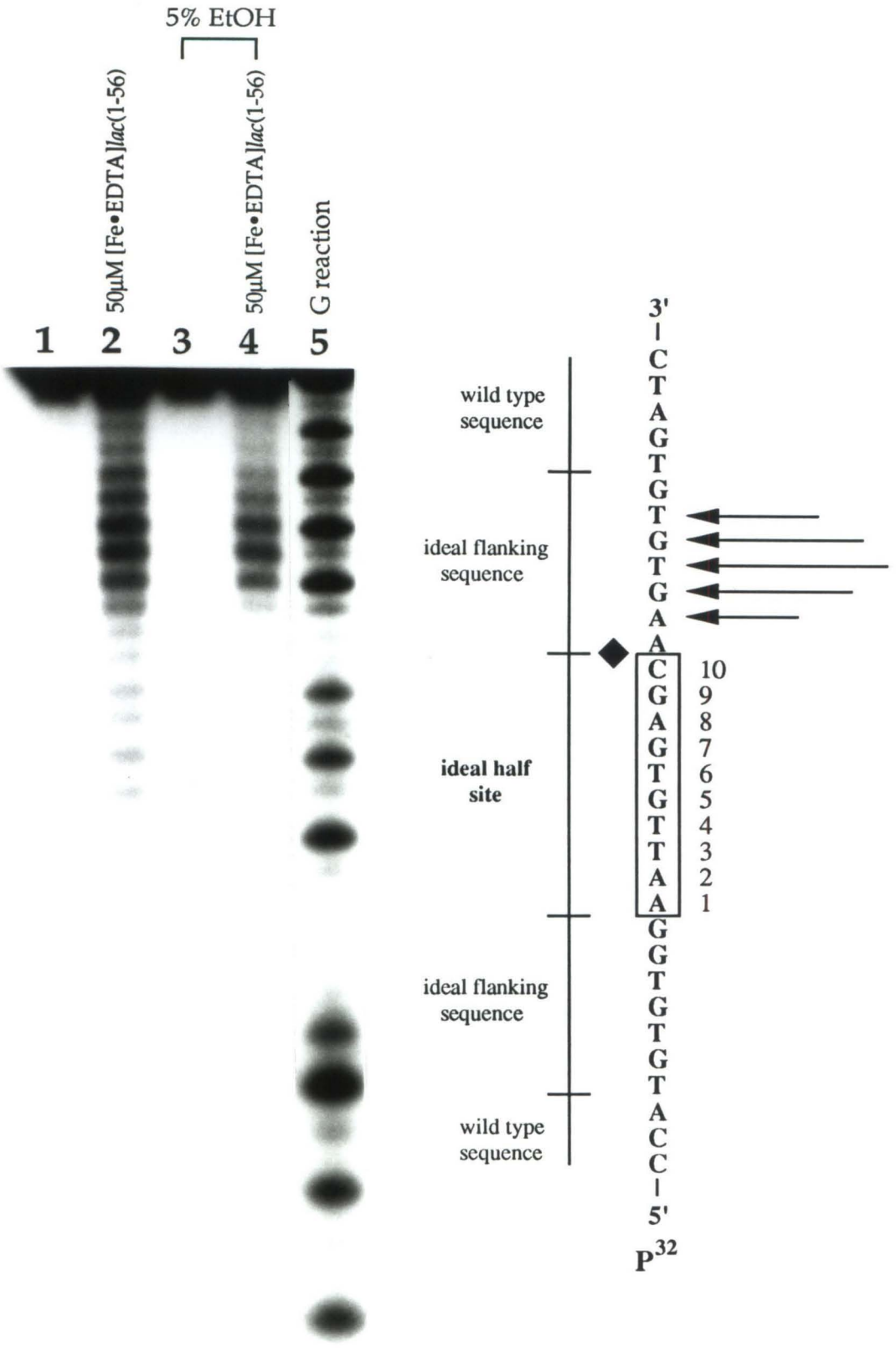
5'-TATACGAATTCCATGTGTGG**AATTGTGAGC**AAGTGTGTGATC-3'
3'-GCTTAAGGTACACACC**TTAACACTCG**TTCACACACTAG-5'

FIGURE 5. Duplex left half site of *lac* operator surrounded by wild-type flanking sequences. The top strand is the 42-mer, and the bottom strand is the 38-mer. Only the left end of the duplex is labelled with radioactive nucleotides by DNA polymerase I.

Affinity cleaving results clearly showed an affinity cleaving pattern in the location expected for the NMR model (Figure 6), and thus these results corroborate Kaptein and coworkers' orientation of the *lac* repressor recognition helix. Many unsuccessful attempts were made to footprint *lac*(1-56) using a number of footprinting agents: DNase, dimethyl sulfate (DMS), methidium propyl EDTA (MPE), and diethyl pyrocarbonate. Although definitive affinity cleaving data were gathered from this experiment, it was necessary to demonstrate that this synthetic *lac* repressor DNA headpiece, the first synthesis of any *lac* DNA binding protein, was capable of binding to the *lac* operator, and that this interaction could be detected by known footprinting techniques.

Also, the affinity cleaving data, although conclusive, was not particularly clean; visualization of the gel-separated products of the [Fe•EDTA]*lac*(1-56)-operator reaction showed a high level of background cleavage (Figure 6). Because these oligonucleotides are quite short, it is possible that end effects interfered with protein-DNA binding, and *lac*(1-56)

FIGURE 6. Autoradiogram of a high-resolution denaturing polyacrylamide gel (12% acrylamide, 1:20 cross-link, 50% urea) showing oligonucleotides Lac38 and Lac42. All lanes are 5'-^{32}P end-labelled. Lanes 1 and 3, intact DNA control; lane 5, Maxam-Gilbert G reaction. Lanes 2 and 4, affinity cleaving reaction with [Fe•EDTA]*lac*(1-56) at 50μM, tris-acetate buffer, pH 7.8, 20mM NaCl, and 10μM calf thymus; cleavage reaction was initiated with 1mM sodium ascorbate. Lane 2, 30 minutes reaction, room temperature. Lane 4, one-hour reaction (contains 5% ethanol to reduce background cleavage), room temperature. At right is the histogram data for just the 5'-^{32}P end-labelled strand (the 42-mer); arrows represent the extent of cleavage for [Fe•EDTA]*lac*(1-56). The consensus half site is boxed and numbered; the solid black diamond indicates where the axis of symmetry would be for the full site.


5% EtOH
50μM [Fe•EDTA]lac(1-56)
50μM [Fe•EDTA]lac(1-56)
G reaction
1
2
3
4
5
wild type sequence
ideal flanking sequence
ideal half site
ideal flanking sequence
wild type sequence
3'
5'
P32
10
9
8
7
6
5
4
3
2
1

would bind more strongly to its operator site if it were on a larger piece of DNA more closely mimicking naturally occurring genomic DNA. This stronger binding would perhaps give cleaner affinity cleaving and footprinting results. Therefore, plasmids containing both the full and half sites were constructed; this allowed the use of restriction fragments containing more than 300 base pairs.

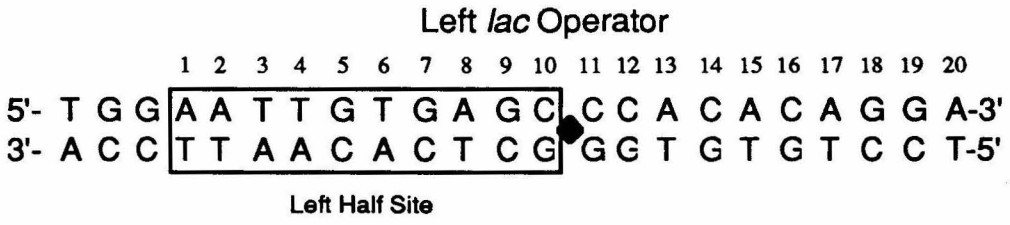
Plasmids Containing the Consensus Full and Half *lac* Operator Sites

Two pairs of oligonucleotides were synthesized; one pair consisted of the consensus *lac* operator half site surrounded by idealized flanking sequences (66- and 76-mers), and the other pair consisted of the consensus full site (76- and 86-mers) similar to the sequences shown in Figures 4 and 5. Each pair of oligonucleotides was annealed and ligated into the multiple cloning site polylinker region of cloning vectors pUC18 and pUC19.[204] A pUC18 derivative containing the *lac* half site (pJS18H) and a pUC19 derivative containing the full site (pJS19F) were obtained and large quantities of each plasmid were produced.

Restriction fragments (>300 base pairs in length) from these plasmids were reacted with [Fe•EDTA]*lac*(1-56) and the products separated by gel electrophoresis; autoradiogram visualization of these gels showed that, as in the previous experiments, a high level of background cleavage occurs when [Fe•EDTA]*lac*(1-56) is used. This suggests that the binding constant for the *lac* DNA binding domain to its operator site is quite low–on the order of 10^5–and NMR studies with the *lac* repressor headpiece give a similar number.[64] In comparison, the 52-mer DNA binding domain of Hin recombinase exhibits a much stronger binding constant of ~10^7 (see Chapter 3). Given the low degree of sequence specificity of the *lac* DNA binding domain for its operator, the high amount of background cleavage is not unexpected. This low degree of

specificity is also consistent with the difficulty in obtaining footprints of the *lac* DNA binding domain with the restriction fragments; many unsuccessful trials with different footprinting reagents were tried and failed. Because of the high level of background cleavage, data obtained from the reaction of [Fe•EDTA]*lac*(1-56) and the full operator site were somewhat ambiguous (Figure 7). A cleavage pattern resulting from [Fe•EDTA]*lac*(1-56) with a single half site was expected to result in the least ambiguous data with regard to orientation assignment; consistent with this expectation, the cleanest and simplest data were obtained from the restriction fragment containing just the half site (Figure 8). Parallel DNA cleavage experiments with the full twofold symmetric consensus DNA site for *lac* repressor yield results consistent with those seen for the single half site (Figure 7).

DNA cleavage experiments were performed with [Fe•EDTA]*lac*(1-56) and a 317 base-pair DNA fragment containing the left DNA half site in the presence of the reducing agent sodium ascorbate. DNA cleavage was observed at five adjacent nucleotides on one DNA strand and at four adjacent nucleotides on the opposite DNA strand (Figure 8). These nucleotide positions are located immediately beyond base-pair 10 of the left DNA half site: i.e., at base pairs 11, 12, 13, 14, 15 and 16 (numbering scheme shown below; black diamond indicates location of the axis of symmetry for the full site). These results establish that the EDTA•Fe moiety of [Fe•EDTA]*lac*(1-56) is near to base pairs 11 to 16 of the left DNA half site in the [Fe•EDTA]*lac*(1-56)-DNA complex. These results cannot be easily reconciled with the model of Matthews and co-workers. In contrast, these results are in good agreement with the model of Kaptein and co-workers (Figure 9).



The five adjacent nucleotides at which DNA cleavage was observed on one DNA strand are shifted approximately one nucleotide in the 3' direction relative to the four adjacent nucleotides at which cleavage was observed on the opposite DNA strand. A 3'-shift in the DNA cleavage pattern is diagnostic of DNA cleavage by EDTA•Fe located in or near the DNA minor groove. Therefore, this result suggests that the EDTA•Fe moiety of [Fe•EDTA]*lac*(1-56) is near the DNA minor groove in the [Fe•EDTA]*lac*(1-56)-DNA complex. This is in agreement with the model of Kaptein and co-workers. To conclude, as proposed by Kaptein, the helix-turn-helix motif of *lac* repressor binds to DNA in an orientation opposite that of the helix-turn-helix motifs of λ repressor, λ cro, 434 repressor, 434 cro, and CAP.

^{31}P NMR studies performed on the 56-residue headpiece also provide evidence that supports the 2D NMR studies[70]; by studying the protein interactions with the phosphodiester backbone, specific phosphate ^{31}P perturbations were observed, which were consistent with NMR, genetic, and biochemical data. Affinity cleaving data show that the amino terminus of the *lac* headpiece lies closer to the 3', rather than the 5', end-labelled strand of DNA (the strand on the right in Figure 8; heights of arrows indicate relative extent of cleavage, and the 3' end-labelled strand shows significantly stronger cleavage compared to the 5' end-labelled strand on the left). The ^{31}P NMR data also show that the *lac* repressor headpiece amino terminus preferentially recognizes the strand on the right in Figure 8.[70]

Intact *lac* repressor is believed to bind to its operator as a dimer with an approximate 2-fold symmetry axis at the center of the binding site. An important question then is whether these headpiece-operator complexes are representative of the whole *lac* repressor-operator complex. Genetic experiments provide support for the validity of the headpiece-operator studies. Mutagenesis experiments have been conducted in which the first two amino acids of the *lac* recognition helix were substituted with those for *gal* repressor (Tyr 17→Val, Gln 18→Ala).[67] This mutant repressor had high affinity for the *gal* operator, which differs from the *lac* operator at positions 4 and 2. Also, a *lac* repressor variant with Arg22 replaced by Asn abolishes repressor binding to the ideal operator; this mutant, however, binds to a variant of the ideal *lac* operator in which G•C at position six has been changed to T•A.[69] 2D NMR also predicts a contact between Arg22 and G•C base-pair 5.[66] All of the experimental results thus far–affinity cleaving, ^{1}H and ^{31}P NMR, and biochemical and genetic experiments–have unambiguously fixed the orientation of the *lac* repressor recognition helix as opposite to that of λ repressor and cro, 434 repressor and cro, and CAP. It appears that *lac* is not alone in this respect; both *gal* and *deo* may also belong to the *lac* class of repressors.[69]

FIGURE 7. Autoradiogram of a high-resolution denaturing polyacrylamide gel (8% acrylamide, 1:20 cross-link, 50% urea) showing the full consensus *lac* operator. All lanes are 3' ^{32}P end-labelled. Lane 1, intact DNA control; lane 2, Maxam-Gilbert G reaction. Lanes 3-10, affinity cleaving reactions with [Fe•EDTA]*lac*(1-56) at 50 μM, tris-acetate buffer, 20 mM NaCl, and 100 μM calf thymus; the cleavage reaction was initiated with 1 mM sodium ascorbate. Lanes 3 and 7, pH 6.2; lanes 4 and 8, pH 6.6; lanes 5 and 9, pH 7.0; lanes 6 and 10, pH 7.4. Lanes 7-10 contain 5% ethanol in an attempt to reduce background cleavage. At right is the histogram data (the 5' end-labelled data were obtained from an autoradiogram not shown); arrows represent the extent of cleavage for [Fe•EDTA]*lac*(1-56). The consensus full site is boxed; the black diamond indicates the axis of symmetry.

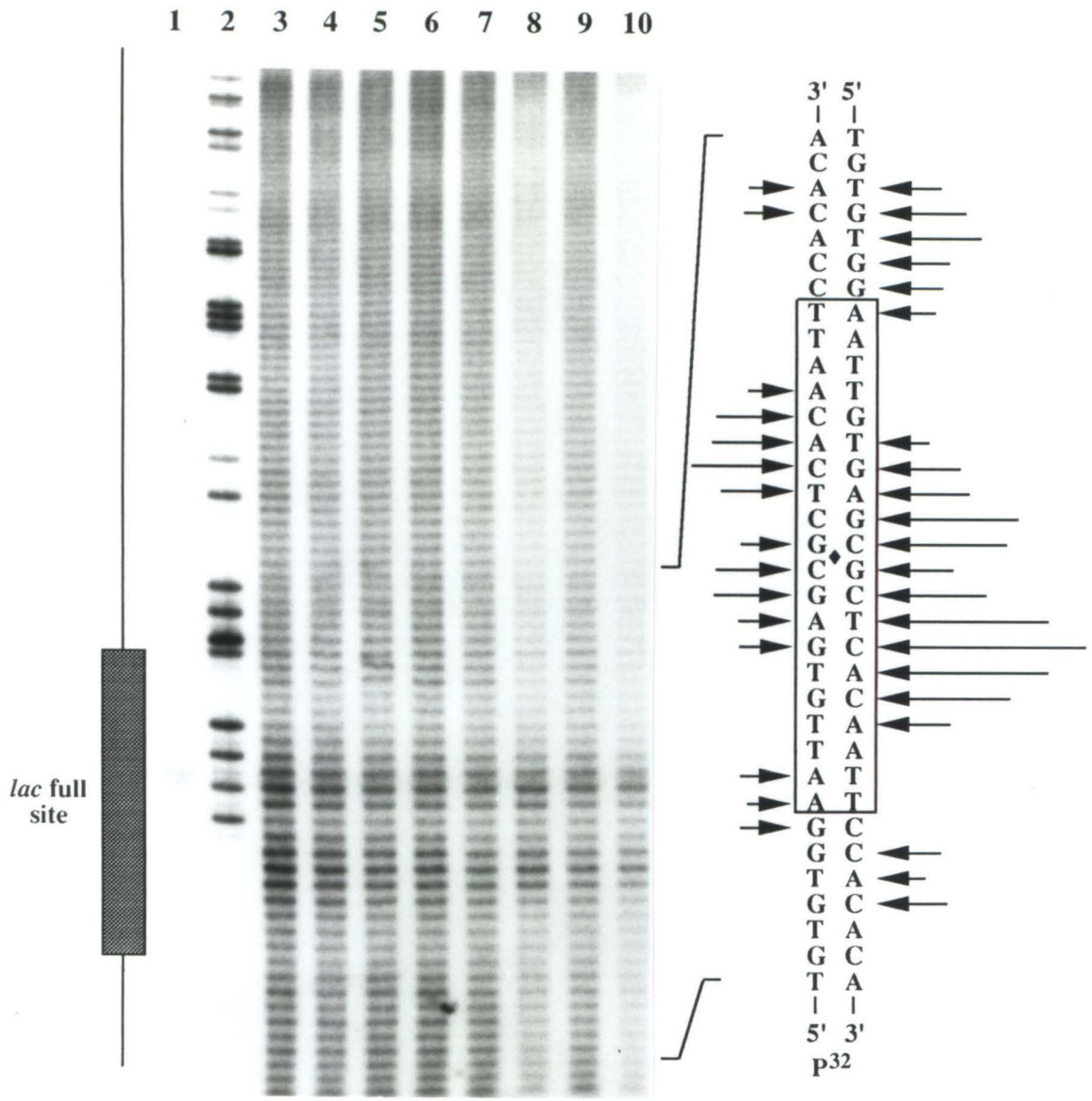

1 2 3 4 5 6 7 8 9 10
lac full site
3' 5'
ACACACCCTTAACACACTCGCGAGTGTTAAGGTGTGT 5'
TGTGTGTGGGAATTGTGAGCGCTCACAATTCCACACA 3'
p32

FIGURE 8. Left. Autoradiogram of a high-resolution denaturing polyacrylamide gel (8% acrylamide, 1:20 cross-link, 50% urea) showing the left consensus *lac* half site. Lanes 1-4 present data for DNA 3' end-labelled on the right DNA strand. Lanes 5-8 present data for DNA 5' end-labelled on the left DNA strand. Lanes 1 and 5, intact DNA; lanes 2 and 6, chemical-sequencing A reaction; lanes 3 and 7, chemical-sequencing G reaction; lanes 4 and 8, affinity cleaving reactions with [Fe•EDTA]*lac*(1-56) at 50 μM, tris-acetate buffer, 20 mM NaCl, and 100 μM calf thymus; the cleavage reaction was initiated with 1 mM sodium ascorbate. **Right.** [Fe•EDTA]*lac*(1-56) affinity cleavage. The figure illustrates the 10 base pair left DNA half site for *lac* repressor. Arrows indicate nucleotides at which DNA cleavage is observed; length of the arrows indicates the relative extent of cleavage. The black diamond indicates where the axis of symmetry would be for the full site.

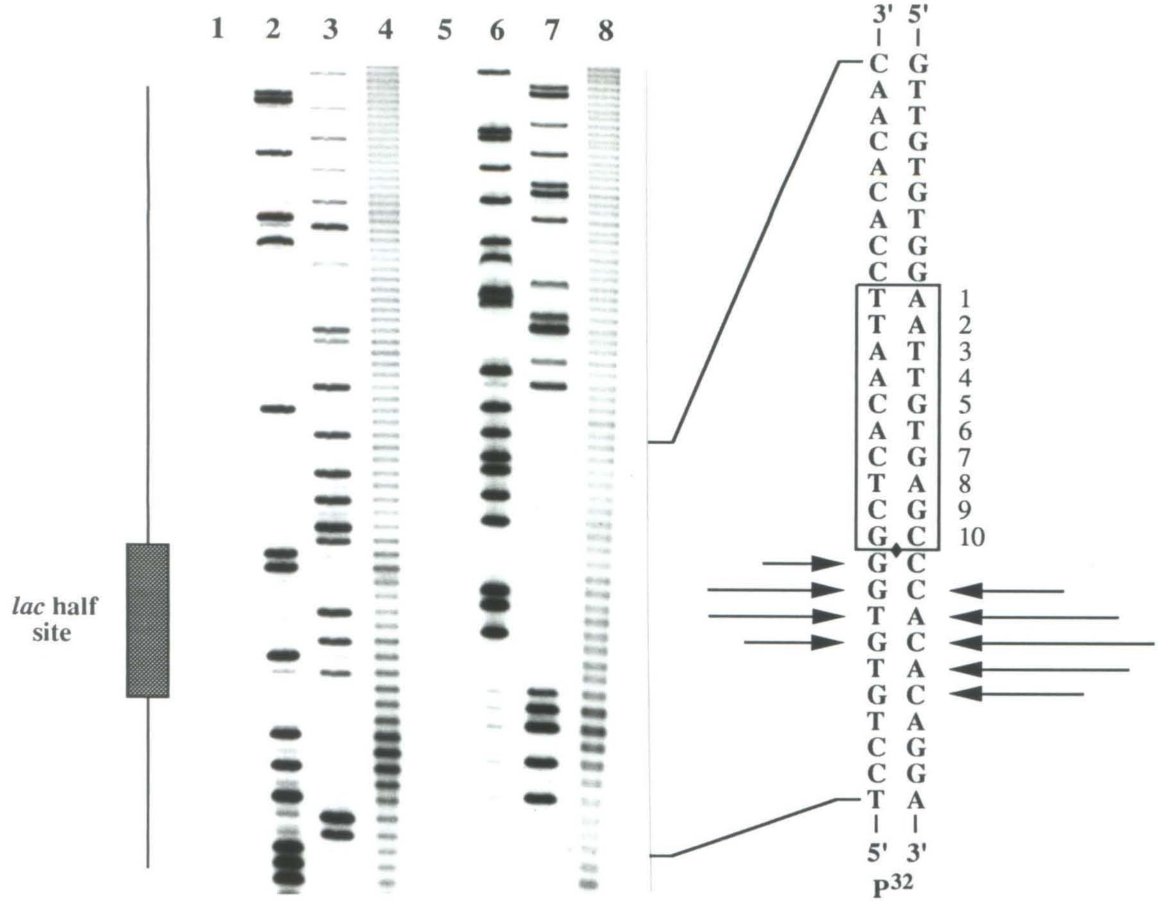
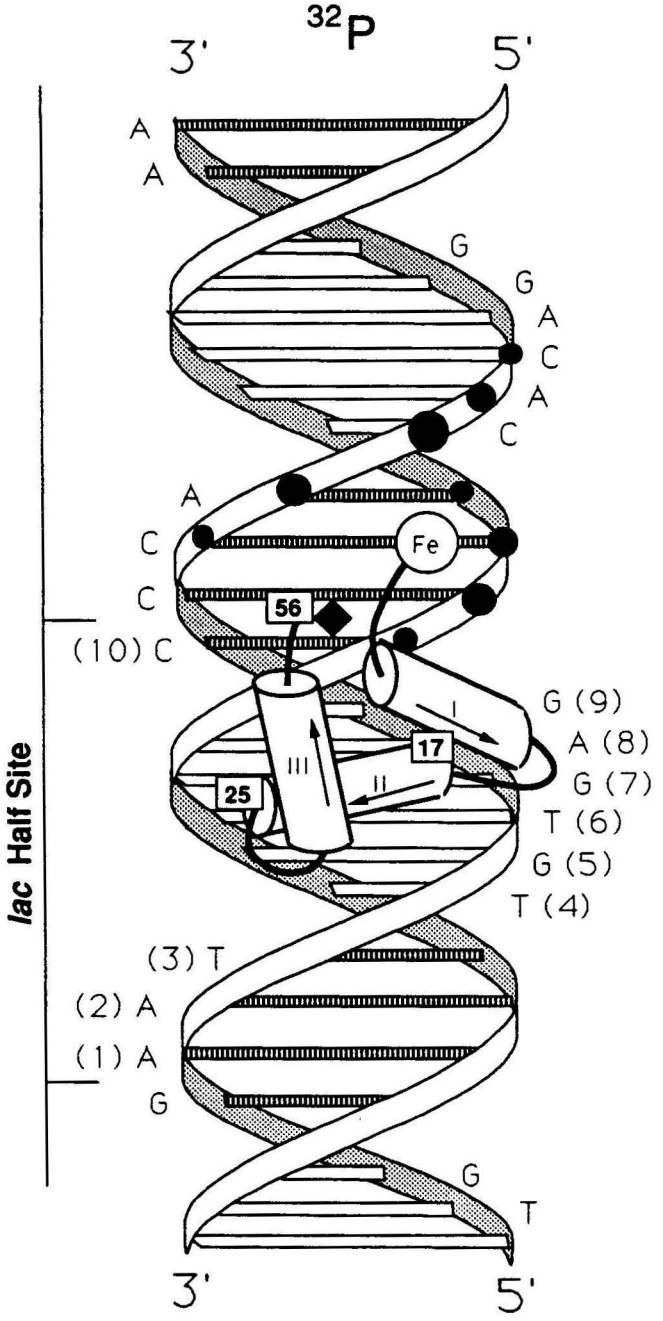

1 2 3 4 5 6 7 8
lac half site
3' 5'
C G
A T
A T
C G
A T
C G
A T
C G
C G
T A 1
T A 2
A T 3
A T 4
C G 5
A T 6
C G 7
T A 8
C G 9
G C 10
G C
G C
T A
G C
T A
G C
T A
C G
C G
T A
5' 3'
P32

FIGURE 9. Model for the structure of the complex between [Fe•EDTA]*lac*(1-56) and the left DNA half site. α-Helices are illustrated as cylinders with arrows pointing from the amino terminus to the carboxyl terminus. Black diamond indicates axis of symmetry. Filled circles indicate nucleotides at which DNA cleavage is observed; the diameter of the circles indicates the relative extent of cleavage. The model is in agreement with the model of Kaptein and co-workers (Figure 2A).


32P
3'
5'
A
A
G
G
A
C
A
C
A
C
Fe
C
56
C
(10) C
G (9)
A (8)
G (7)
T (6)
G (5)
T (4)
I
17
III
II
25
(3) T
(2) A
(1) A
G
G
T
3'
5'
lac Half Site

MATERIALS AND METHODS

[EDTA]*lac*(1-56)

Materials. Protected amino-acid derivatives were purchased from Peninsula Laboratories; Boc-L-His(DNP) was obtained from Fluka. 4-Methylbenzhydrylamine (BHA) resin came from US Biochemical Corp. Dimethylformamide (DMF), diisopropylethylamine, dicyclohexylcarbodiimide in dichloromethane, N-hydroxybenzotriazole in DMF, and trifluoroacetic acid (TFA) were purchased from Applied Biosystems. Dichloromethane and methanol (HPLC grade) were obtained from Mallinckrodt, p-cresol and p-thiocresol from Aldrich, and diethyl ether (low peroxide content) from Baker. Doubly distilled water was further purified through the Milli Q filtration system from Millipore. tRNA (*E. coli* strain W, Type XX) came from Sigma and was dissolved in water and sterile-filtered. Enzymes were purchased from Boehringer Mannheim or New England Biolabs.

Synthesis. Manual peptide synthesis was carried out by solid phase techniques in 20 ml vessels fitted with coarse glass frits, using synthetic protocols developed at the California Institute of Technology.[175,176] Fully protected, resin-bound *lac*(1-56) was synthesized on benzhydrylamine resin using t-Boc protected amino acids. γ-aminobutyric acid was incorporated at the deprotected amino-terminus of resin-bound *lac*(1-56) using t-butoxycarbonyl-γ-aminobutyric acid. EDTA was incorporated at the deprotected amino terminus of resin-bound [γ-aminobutyric acid]*lac*(1-56) using tricyclohexyl-EDTA and standard coupling chemistry. [EDTA-γ-aminobutyryl]*lac*(1-56), referred to as [EDTA]*lac*(1-56), was cleaved from the resin and deprotected using anhydrous hydrofluoric acid in the presence of p-cresol and p-thiocresol radical scavengers.

[EDTA]*lac*(1-56) was purified by reverse phase HPLC on a semipreparative C8 column (Vydac) with a linear gradient of acetonitrile-water with 0.1% trifluoroacetic acid (flow rate 3ml/min, 0 to 60% acetonitrile over 240 min). The peptide was homogeneous by the criteria of HPLC, amino-acid analysis, and amino-acid sequencing. Mass spectrometric analysis: calculated, 6584.5; found 6587±13. [EDTA]*lac*(1-56) was stored dry at -70 °C in 5 nmol aliquots (ε_{275}=5620 for four Tyr residues). [Fe•EDTA]*lac*(1-56) was prepared by incubation of [EDTA]*lac*(1-56) with aqueous ferrous ammonium sulfate in a 1:1 molar mixture (30 min, 25 °C).

DNA Substrate

Construction of Plasmids pJS18H and pJS19F. Standard techniques were used in plasmid construction.[205] Oligodeoxyribonucleotides were designed to place Xma I and Pst I restriction endonuclease sites at the 5' and 3' ends of the insert, respectively. Automated oligonucleotide synthesis was performed on an Applied Biosystems 380B DNA Synthesizer using β-cyanoethyl-phosphoramidite chemistry. All chemicals were purchased from Applied Biosystems. Removal of the oligonucleotide from the support and deprotection was accomplished by treatment with ammonium hydroxide. Oligonucleotides were purified on 20% acrylamide gels (1:20 cross-link, 50% urea). The complementary oligonucleotides were allowed to anneal, and a phosphate group was added to each 5' end with T4 polynucleotide kinase and ATP (New England Biolabs and Calbiochem, respectively). pUC18 and pUC19 vectors[204] (Life Technology Research Laboratories) were cut with Xma I and Pst I, and the synthesized insert was ligated with T4 DNA ligase into the pUC18/pUC19 multiple cloning site polylinker region (New England Biolabs). The transformation was conducted on competent cells purchased from Bethesda Research Laboratories; through α-complementation, these

transformed DH5α competent cells were capable of producing blue/white colonies suitable for screening on agar plates containing Bluo-gal and IPTG. Colonies were chosen and grown in overnight cell cultures (5 ml); the cells were collected, lysed, and the isolated plasmid was sequenced (Pharmacia T7 and T7 Deaza Sequencing Kits). The overnight cell culture yielding plasmid of the desired sequence was inoculated into large overnight cultures (500 ml). Cells were harvested, lysed, and the recovered plasmid was purified by cesium chloride density gradient.[205]

Radioactive Labelling of Restriction Fragment. Plasmid pJS18H was linearized by digestion with EcoR I. Linearized plasmid pJS18H was 3'-end-labelled with [α^{32}P]-dATP and DNA polymerase I Klenow fragment.[177] Linearized plasmid pJS18H was 5'-end-labeled by dephosphorylation using calf intestinal alkaline phosphatase, followed by phosphorylation using [γ^{32}P]-ATP and T4 polynucleotide kinase.[177] Labelled linearized plasmid pJS18H was digested with Nde I, and the resulting labeled 317-base pair DNA fragment was isolated using non-denaturing polyacrylamide gel electrophoresis. A similar labelling procedure was used to label pJS19F; instead of linearizing the plasmid with EcoR I, Hind III was used.

DNA Cleaving Experiments

Reaction Conditions. Reaction mixtures (10μl) contained ^{32}P end-labelled DNA fragment (20,000 cpm), 50 μM [Fe•EDTA]*lac*(1-56), 1 mM sodium ascorbate, 50 mM tris-acetate, pH 7.0, 20 mM NaCl, 0.1 mg/ml tRNA (Sigma Chemical, Type XX), and 0.5 mg/ml bovine serum albumin. All components except sodium ascorbate were incubated for 30 min at 37 °C. Reactions were initiated by the addition of sodium ascorbate (1 mM), and were allowed to proceed for 30 min at 22 °C. Reactions were terminated by desalting, using Sephadex G-50 spin columns. Reaction products were analyzed by denaturing

electrophoresis on 8% polyacrylamide gels (1:20 crosslink, 7M urea). After electrophoresis, gels were dried and autoradiographed. Autoradiograms were analyzed by laser densitometry.

CHAPTER FIVE

Affinity Cleaving Studies on the *engrailed* Homeodomain

INTRODUCTION

Expression of Homeobox Genes during Development

Organisms develop according to a precise developmental program that specifies their body plan in great detail and determines the sequence and duration of developmental events. This information is stored on DNA in homeotic genes; how organisms express this information is a fundamental problem in developmental biology.[73] Normal development requires an organism to govern the expression of thousands of structural genes in a spatially and temporally controlled fashion. The effects of mutations in homeotic genes in Drosophila were identified as early as 1915; these mutations transform certain parts or entire body segments into different structures[73]–for example, one identified mutation causes a leg instead of an antenna to protrude from a fly's head. Clearly, the complex machinery controlling development must be precisely regulated in order to achieve a normal, functional organism.

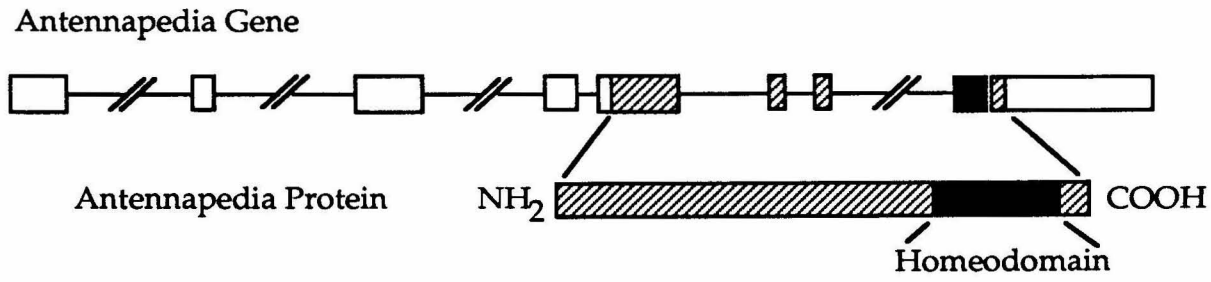


FIGURE 1. Structural organization of the *Antennapedia* gene and protein. Exons 1-8 are separated by introns, which are not drawn to scale. The homeobox is located in exon 8 near the carboxyl terminus of the protein. (adapted from W. J. Gehring , *Science* **1987**, *236*, 1245-1252.)

The homeobox was first identified in the early 1980's as a region of homology common to several homeotic genes in Drosophila, including the *Antennapedia* gene (Figure 1) and the segmentation gene *fushi tarazu* (*ftz*).[76] The homeobox is a small DNA segment of approximately 180 base pairs that codes for the 60 amino-acid homeodomain protein, which is near the carboxyl terminus of the full homeotic protein and is especially conserved.[73] Since its initial discovery, the homeobox has been found in many of the genes that commit cells to follow specific pathways of development in the early embryo.[75,77,78] Over forty such genes exist in Drosophila, and over half (15 of 25) of those that have been cloned and characterized contain the homeobox sequence.[76] Homeodomain proteins are important factors for governing transcription in a wide range of species, from yeast to man.[206-209] Homeobox expression has been studied in the development of other eukaryotic organisms, including worms, frogs, and mice. In both insects and vertebrates, homeotic genes are clustered, and the individual genes of these clusters are arranged in the same order along the chromosome as they are expressed along the anterior-posterior body axis.[79]

The amino-acid homologies between three Drosophila homeodomains and a *Xenopus laevis* (toad) homeodomain start and end abruptly; each homeodomain contains 60 residues and is very rich in arginines and lysines.[210] The extreme conservation found in the amino-acid sequences between the Drosophila and Xenopus homeodomains suggests that the domain has a vital function in the control of early development. It appears that very selective pressure has been applied to keep this region constant.

Previous genetic and biochemical studies suggested that the homeodomain mediates the sequence-specific DNA binding activities of these proteins. Mutation studies in this region indicated that this region

plays a critical role in the control of gene expression *in vivo*; more definitive *in vitro* assays have shown that the homeodomain is necessary for DNA binding, whereas other regions are not as important. Point mutations and in-frame deletions within the homeobox can completely abolish binding activity in the expressed homeodomain.[76] Furthermore, homeobox "swap" experiments, in which homeobox sequences from different homeotic genes are replaced by other homeoboxes, show that DNA binding specificity is a property of the homeodomain.[76] Although a given homeodomain can recognize a range of DNA sequences, these data are consistent with homeodomain mediation of DNA recognition and binding.

Searches through protein sequence banks have also shown that these homeodomains are highly homologous with yeast mating-type regulatory proteins, MATα1 and α2; these proteins control the expression of many unlinked mating and sporulation genes, which control a form of cell differentiation.[211] Like the yeast MAT proteins, homeodomains are critical for normal development in eukaryotes, and these proteins act as transcription factors. In Drosophila, homeodomains have been found that govern segmentation development in the embryo. The Drosophila homeodomains *Ultrabithorax*, *fushi tarazu* (*ftz*), and *Antennapedia* (*Antp*) have the same amino-acid sequence in the putative recognition helix and bind to the same DNA sequence. The *engrailed* homeodomain from Drosophila varies at one position from *Antp* and *ftz*, but these closely related homeodomains bind to similar sequences, and this competitive binding may play a role in the transcription of developmental genes.[82]

SIMILARITY OF THE HOMEODOMAIN TO HELIX-TURN-HELIX PROTEINS

Since the early 1980's, researchers had speculated that homeodomain proteins might contain a helix-turn-helix structure.[76,79,209] Sequence comparisons among the homeodomains from Drosophila, sea urchin, and humans and the MAT proteins from yeast show a weak similarity to the helix-turn-helix motif found in prokaryotic gene regulatory proteins. The *Antp* homeodomain has been found to represent the consensus sequence of the homeodomains most closely[73,79,84]; among homeodomains, the most highly conserved region corresponds to the putative recognition helix.[73] This area contains four invariant residues, three of which are also found in the yeast MAT proteins. This is in marked constrast to the prokaryotic regulatory proteins, which show significant variation in their recognition helices, reflecting the fact that each prokaryotic protein binds to a different DNA operator sequence. Additionally, a single amino-acid substitution of the ninth residue of the recognition helix appears to be sufficient to switch the DNA binding specificities of different homeodomains.[80] *Bicoid*, which controls anterior development in Drosophila, contains Lys at position 9, whereas *Antp* uses Gln; when Lys9 in the *Bicoid* recognition helix is replaced by Gln, the mutant protein no longer recognizes *Bicoid* binding sites but instead recognizes *Antp* sites.[80] The homeodomain recognition helix is also much longer than that for prokaryotic regulatory proteins; the homedomain recognition helix is ~17 amino acids, whereas that for the prokaryotic helix-turn-helix is ~10 residues.[81,83,85]

Another significant difference between homeodomain and prokaryotic regulatory proteins is their modes of DNA binding[84]; prokaryotic helix-turn-helix proteins bind to operators as dimers, whereas homeodomains bind to monomeric binding sites. Although homeodomain proteins do not use

cooperativity to enhance sequence specificity and binding affinity for DNA as prokaryotic regulatory proteins do, homeodomains exhibit extremely high binding constants; homeodomains, which are typically only 60 residues in length, bind to operators with binding constants around 10^9-10^{11},[81,88] whereas the dimerically bound DNA binding domains of prokaryotic helix-turn-helix proteins (around 50-90 amino acids in length) give binding constants of 10^6-10^7 (see Chapter 3). In studies with purified 434 and P22 repressor DNA binding domains and the corresponding half-operator sites, only trace amounts of protein-DNA complexes formed at protein concentration $\geq 10^{-5}$ M. Kinetic studies on the *Antp* homeodomain suggest that the dramatic difference in the operator affinities between *Antp* and prokaryotic DNA binding domains is due to the low dissociation rate of the *Antp*-DNA complex.[212]

Homeodomain proteins bind to very similar operator sites; the core consensus sequence is 5′-TAAT. Genetic studies conducted on the *Antp* and *bicoid* homeodomains demonstrate that the TAAT motif, which is conserved in virtually all homeodomain binding sites, is not used to discriminate between the *Antp* and *bicoid* sites. The data showed that mutation of bases 3′ to the TAAT motif (especially the base lying adjacently 3′ to the TAAT–i.e., 5′-TAATN) affects the binding specificity between the *Antp* and *bicoid* homeodomains.[80] Similar results have been found in studying the optimal DNA binding sequence of the *Ultrabithorax* homeodomain from Drosophila. The central TAAT bases play a primary role in determining the binding affinity with significant secondary contributions derived from flanking bases.[88] It is likely that the TAAT is necessary to distinguish the operator sites from non-specific DNA, and thus the homeodomain binding site consists of

two subsites–a common TAAT core element and and the specificity-determining bases that lie on the 3′ side.

Comparison of the high-resolution cocrystal structures of 434 repressor[26] (discussed in Chapter 1) and the *engrailed* homeodomain[81] and the NMR structure of *Antp*[83,85] (discussed below) shows significant differences in their modes of DNA binding. The positioning of the homeodomain is quite different from the prokaryotic regulatory proteins; the homeodomain binding motif is shifted and reoriented with respect to DNA, and the amino-terminal arm of the homeodomain crosses the phosphodiester backbone to make important specific interactions in the minor groove.[84] Although recent structural studies on the *engrailed* (*en*) and *Antp* homeodomains show that both possess helix-turn-helix DNA binding structures, their modes of interaction with DNA differ significantly from that of prokaryotic helix-turn-helix proteins.

AFFINITY CLEAVING STUDIES ON THE *ENGRAILED* HOMEODOMAIN

X-ray crystallography has been the technique most predominantly used for gaining high-resolution structural data on proteins and protein-DNA complexes. More recently, NMR has proven to be an extremely powerful technique for studying these complicated structures.[58,59] NMR has yielded much high-resolution information about protein complexes that have not been amenable to crystal formation; for example, the *lac* repressor and its DNA complex, on which numerous crystallization attempts have failed, have been successfully studied by solution NMR (Chapter 4). Affinity cleaving is also a solution-phase technique that can yield structural data. For example, affinity cleaving has revealed that the amino terminus of the DNA binding domain of Hin recombinase, Hin(139-190), lies in or above the minor groove,

and has fixed the orientation of the recognition helix (helix 3) with respect to the operator site.[153,213] The affinity cleaving method has never been compared with a known method for obtaining structural data, such as X-ray crystallography; such a comparison would likely provide valuable information on design and interpretation of affinity cleaving experiments. A lot of affinity cleaving data has been generated in studies on the Hin DNA binding domain, and it was decided to conduct an affinity cleaving study on a protein, similar to Hin, with a high-resolution crystal structure in order to correlate affinity cleaving data with crystallographic data.

Originally, 434 repressor was chosen for affinity cleaving studies. In 1988, a high-resolution X-ray crystal structure of the DNA binding domain of 434 repressor-operator complex was determined.[26] Although there is no known structure of the Hin recombinase DNA binding domain, sequence homology analysis strongly supports the assertion that Hin is a helix-turn-helix (HTH) protein[84]; affinity cleaving studies show that the putative helices of Hin(139-190) fold in a similar orientation as that of 434 repressor when complexed to DNA.[153,213] Like Hin, 434 repressor contains an Arg-Pro-Arg sequence which makes contacts with the minor groove. The Arg-Pro-Arg in 434 repressor, residues 41-43, is located at the carboxyl terminal of the HTH motif in a loop between helices 3 and 4, where helix 3 is the recognition helix (Chapter 1); however, the Arg-Pro-Arg in Hin(139-190), residues 140-142, is located at the amino terminus of the HTH motif before the start of helix 1.

Before any experiments on 434 repressor were started, the crystal structure of the *engrailed* homeodomain-operator complex was determined[81]; *engrailed* also uses an HTH unit, which folds similarly to 434 repressor, to bind to DNA. Like Hin, *engrailed* contains an Arg-Pro-Arg sequence at the amino terminus before the start of helix 1. It is interesting to note that Hin

and 434 repressor are prokaryotic proteins, whereas *engrailed* is eukaryotic; thus, there appears to be a high degree of structural conservation among a wide range of species. It is also of interest that Hin has been proposed to be more similar in structure and binding to the eukaryotic homeodomain than to prokaryotic regulatory proteins (discussed below).[84] Because *engrailed* and Hin share the common amino-terminal sequence, which lies in the minor groove, it was decided to synthesize the *engrailed* homeodomain derivatized with the affinity cleaving moiety EDTA•Fe. Studies were also conducted on the *engrailed* homeodomain derivatized with the metal chelator Gly-Gly-His at the amino terminus.

Structural Studies on the *engrailed* and *Antennapedia* Homeodomains

Recently, the structures of two homeodomain-operator complexes were determined: the *Antp* homeodomain complexed with its DNA operator was determined by NMR spectroscopy,[83,85] and the *en* homeodomain-operator cocrystal was determined to 2.8Å resolution.[81] Both proteins contain a helix-turn-helix structure, which binds to the operator, although the homeodomain helix-turn-helix and its interactions with DNA differ significantly from that of prokaryotic regulatory proteins. This hardly seems surprising considering that the binding constant for the *Antp* homeodomain monomer is four orders of magnitude higher than that of either the 434 or P22 repressors' binding as monomers to their half sites.[84]

In the *Antp* structure, a 68 amino-acid protein fragment corresponding to residues 297-363 of the *Antp* protein formed a 1:1 complex with a 14 base pair DNA fragment; the *Antp* homeodomain binds to this site with a binding affinity of ~10^9 M^{-1} and half-life of ~90 minutes.[85] The NMR solution structures show that the protein undergoes little conformational change upon binding to its operator. With one exception, all observed protein-DNA

contacts occur in the major groove. Arg5, a highly conserved amino acid in known homeodomains, reaches into the minor groove to interact with a sugar proton; interestingly, in the NMR structure of the free homeodomain, no defined spatial structure could be found for the apparently flexible amino-terminal segment comprising residues 0-6. Arg5, which is outside of the helix-turn-helix structure, appears to make a significant binding contribution, for methylation and ethylation interference is observed when bases in the minor groove are modified. Base-specific contacts occur between Ile47, Gln50, and Met54 in helices III and IV; it is worth noting that Gln50 occupies position 9 of helix III (see discussion above), and that it shows an NOE with H5 of C7 in the major groove. All other intermolecular NOE's are with sugar protons. In *Antp*, the recognition helix is not smoothly continuous. A kink separates helices III and IV such that the short helix IV is at a slight angle to helix III. Helix III has the largest number of contacts with the major groove, and they are all on the same side of the helix.

The cocrystal structure of a 61 amino-acid peptide containing the *en* homeodomain complexed to a 21 base-pair DNA duplex shows structural features and protein-DNA interactions similar to those for *Antp*.[81] The DNA duplex in the cocrystal is a relatively straight segment of B-DNA; the only distortion is shown in the major groove which is several angstroms wider than normal in the region where helix 3 binds.

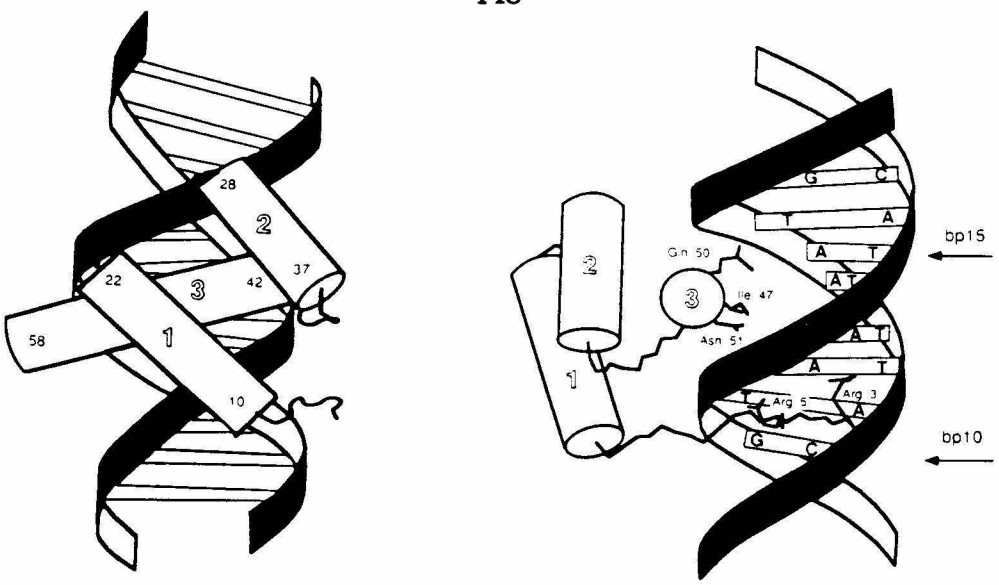


FIGURE 2. Sketch of the *engrailed*-DNA complex. α-Helices are indicated as cylinders, and critical contacts are shown on the right.

Although careful sequence comparisons allow homeodomains to be grouped into subfamilies, it seems likely that all homeodomains will have similar structures and use similar modes of DNA recognition. Like *Antp, en* uses a helix-turn-helix motif to bind to its operator, but unlike *Antp, en* contains a long, continuous recognition helix with no kinks (Figure 2). Gel retardation experiments confirm that the *en* homeodomain binds tightly to the operator used in crystallization (K_D=1-2x10^{-9} M, 100mM KCl, pH 7.6; Figure 3)[81]; this is the the central sequence comprising base pairs 11-14. Because the ends of the duplex fragments stack end-to-end to form a pseudo-continuous duplex in the crystal, a second, weaker binding site has been created (base pairs 1-3 and 21); *en* binds to this site with a K_D of ~10^{-7} M.[81] Superimposing the two structures reveals that the protein conformations and DNA contacts are virtually identical. The complex that has been studied and presented is that at the central base pairs 11-14. This structure is consistent

with the vast body of genetic and biochemical data on homeodomain-DNA interactions.

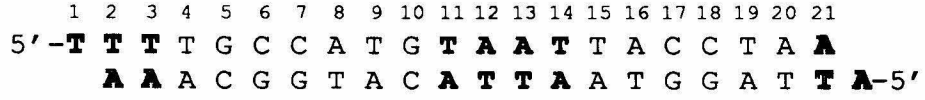


FIGURE 3. DNA sequence used for cocrystallization of the *engrailed* homeodomain. The overhanging 5′ ends pair with those of neighboring duplexes to form a pseudocontinuous double helix. Two monomers bind to this fragment. One makes its contacts at the 5′-TAAT site including base pairs 11-14 (shown in bold). The other homeodomain binds at the end of the DNA, where neighboring duplexes stack at a 5′- AAAT site on the lower strand (base pairs 1-3 and 21, shown in outline).

The *en* homeodomain contains three α helices and an extended amino-terminal arm (Figures 2 and 4). Helix 1 (residues 10-22) and helix 2 (residues 28-37) pack against each other and span the major groove; both of these helices, however, are too far from the DNA to make many contacts. Helix 3 (residues 42-58) is roughly perpendicular to the first two helices and lies in the major groove making extensive contacts with the DNA. The hydrophobic face of helix 3 packs against helices 1 and 2 to form the interior of the protein. No is seen evidence for the kink seen between the *Antp* recognition helices III and IV. The first few residues of the homeodomain appear to be disordered in the crystal, but residues 3-9 form an extended amino-terminal arm that crosses from the major groove into the minor groove and supplements contacts made by helix 3.

In helix 3, the recognition helix, it is interesting to note the critical roles played by Trp48, Phe49, Asn51, and Arg53. These residues are conserved in every one of the higher eukaryotic homeodomains compiled thus far. Trp48 and Phe49, part of the hydrophobic core, play a major role in stabilizing the structure and controlling the packing of helix 1 against helix 3; this is critical for DNA recognition, for it affects the spatial relationship of contacts made by

the amino-terminal arm and those made by helix 3. The invariant hydrophilic residues Asn51 and Arg52 make contacts in the major groove; Asn51 makes a pair of hydrogen bonds with adenine at base-pair 13, donating a hydrogen to the N7 position and accepting a hydrogen from N6. Arg53 makes hydrogen bonds with two phosphate groups.

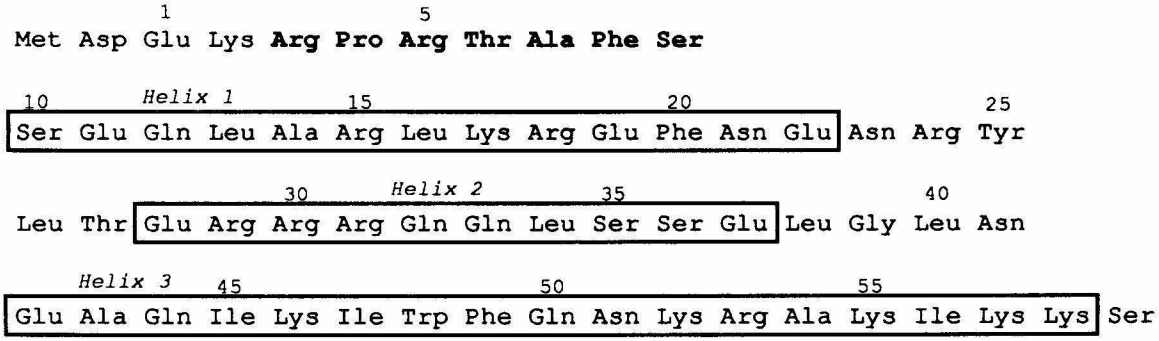


FIGURE 4. Sequence of the *engrailed* homeodomain. The protein fragment used for cocrystallization includes 60 amino acids from the Drosophila *engrailed* homeodomain; the cloning procedure adds a methionine at the amino terminus. The numbering scheme corresponds to that used in the NMR studies of the *Antennapedia* homeodomain. The three α helices are boxed and the amino-terminal arm is marked in bold.

Other residues in helix 3 make critical contacts with DNA. Ile47 makes a sequence-specific hydrophobic contact with the methyl group of thymine at base-pair 14; valine typically occurs at this position in other homeodomains, and it would also be able to make a similar contact. The side chain of Gln50 makes van der Waals contacts with the methyl group of thymine at base-pair 16. The homeodomain also makes an extensive set of contacts with the phosphodiester backbone. Arg31 contacts the phosphate between base-pairs 19 and 18. Arg53 contacts the next phosphate (between base-pairs 18 and 17). The next phosphate appears to be especially critical (between base-pairs 17 and 16); contacts are made from Arg52, Tyr25, and Lys57. Interestingly, helix 1

makes no contacts with DNA, whereas Tyr25 and Arg31 provide the only DNA contacts made from the loop or from helix 2. On the other strand of DNA, Lys55 contacts a phosphodiester oxygen between base-pairs 10 and 11. Trp48 may be able to make an electrostatic interaction with the phosphodiester oxygen between base-pairs 11 and 12. Thr6 makes side chain and main chain hydrogen bonds to the phosphodiester oxygen between base pairs 12 and 13 (a diagram summarizing these interactions is discussed later on in the text in comparison with the recently resolved MATα2 cocrystal structure).

The amino-terminal arm is a well conserved structure among homeodomains. Residues 3-5 form a well defined region of extended chain that recognizes the minor groove. The side chain of Arg5, the most highly conserved residue in this portion of the homeodomain, makes a hydrogen bond with O2 of thymine at base-pair 11. The electron density for Arg3 is not as well defined as it is for other side chains; it appears that the side chain of Arg3 makes a hydrogen bond with O2 of thymine at base pair 12 and/or with the sugar oxygen from adenosine at base pair 13. These minor groove contacts may explain the preference among homeodomains for A,T-rich sites, because of the appropriately situated hydrogen bond acceptors in the minor groove.[81] Whether or not these contacts can distinguish an AT from a TA base pair, as N3 of adenine and O2 of thymine occupy similar positions in the minor groove, is unclear.

The structure of the *en*-DNA complex suggests some general principles about homeodomain-DNA interactions.[81] It appears that highly conserved residues on helix 3 and the amino-terminal arm form a core recognition unit that is responsible for many of the contacts with the TAAT subsite. Recognition may be based on a set of contacts with a core consensus sequence,

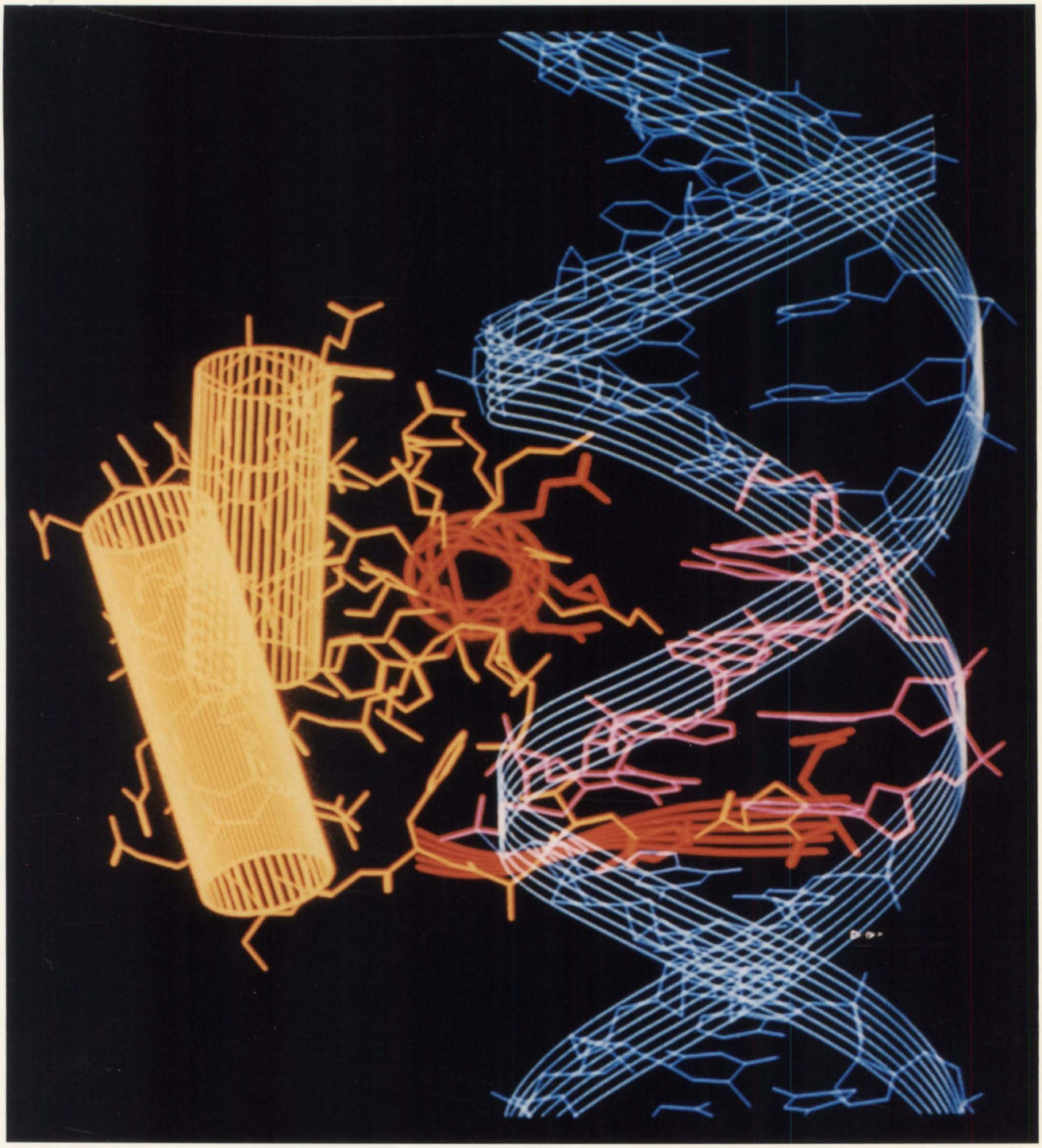
and modulating these interactions would generate new specificities and regulatory activities.

NMR studies on the *Antp* homeodomain and the crystal structure of *en* confirms that these homeodomains possess a helix-turn-helix unit very similar to that used by prokaryotic regulatory proteins; superposition of the backbone residues of the helix-turn-helix from λ repressor on that of *en* gives a room-mean-square value of 0.84Å, which is only slightly larger than distances obtained when superimposing HTH motifs of prokaryotic proteins.[81] However, the *en* cocrystal structure reveals that the homeodomain uses the HTH unit in a much different way than do prokaryotic HTH proteins. Residues near the amino terminus of λ repressor recognition helix make critical contacts to DNA, whereas residues in the center of the *en* extended recognition helix make critical contacts. In the λ repressor-operator complex,[25] helix 2 fits partially into the major groove, and its amino terminus contacts the phosphodiester backbone of DNA. In contrast, helix 2 of *en* lies above the major groove, and the DNA is rotated away from the amino terminus of the helix. Although the position of helix 2 is rather different in the two complexes, clearly it plays related roles; in both complexes, helix 2 packs against helix 3 and serves as a sort of "anchor."

Similarities between the Homeodomain and the Hin Recombinase DNA Binding Domain

A comparison of the homeodomain to the DNA binding domain of Hin recombinase from prokaryotic *Salmonella typhimurium* (see Chapter 2) shows that both proteins contain very similar amino-terminal structural elements that strongly contribute to sequence-specific binding. Although no three-dimensional structure exists of the Hin DNA binding domain,

FIGURE 5. Cocrystal structure of the *engrailed*-DNA complex. This is a photograph of the November 2, 1990, cover of *Cell* (*Cell*, **1990**, *63*) depicting the cocrystal structure presented by Carl Pabo and co-workers.[81] Presented is an end-on view of helix 3, shown as a red circle in the major groove; helices 1 and 2 are represented as yellow cylinders. The amino-terminal arm is shown as a the red ribbon projecting into the minor groove. The 5′-TAAT site is highlighted in pink.



Protein Sequence Alignment

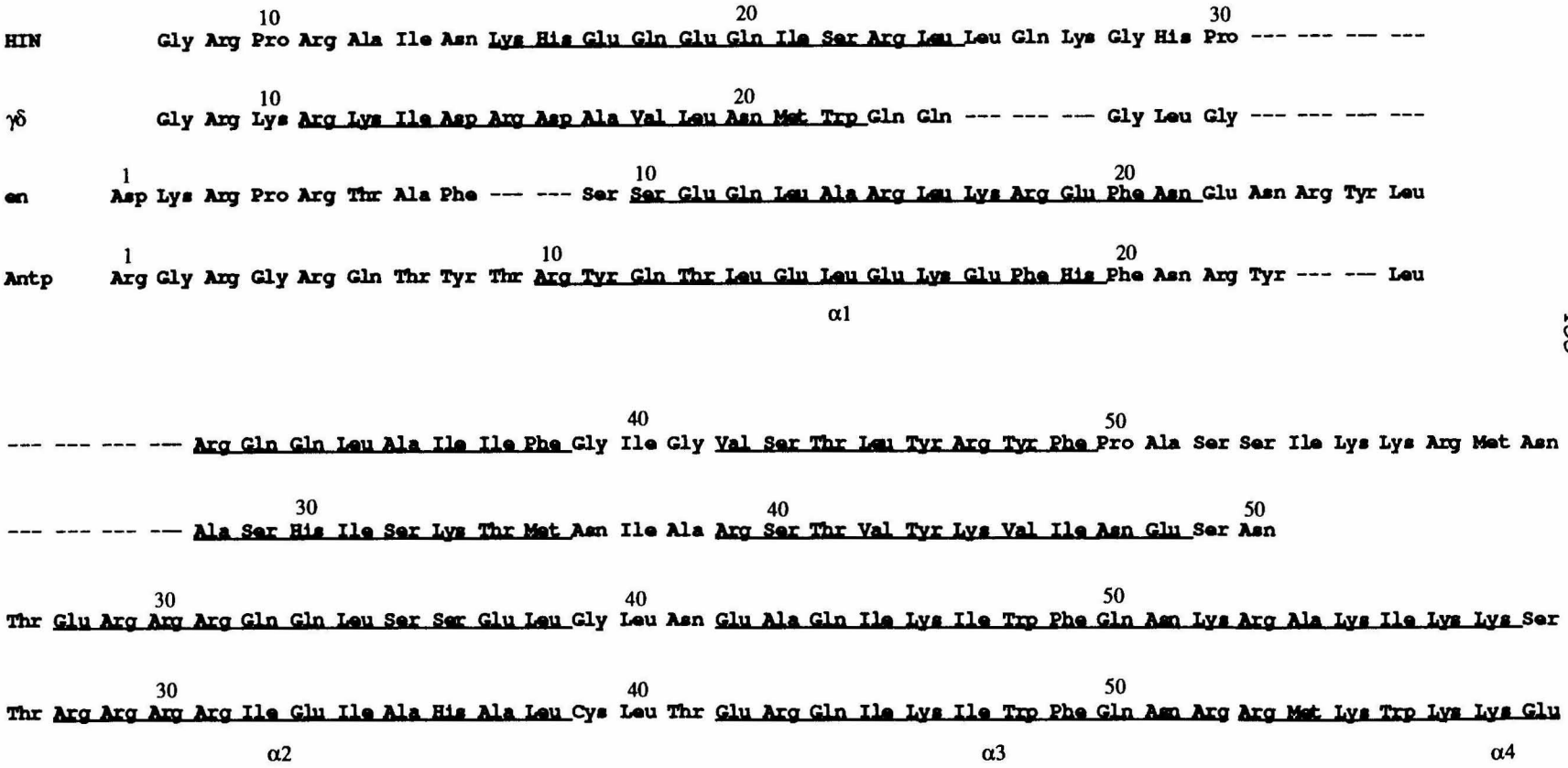


Figure 6. Comparison of the Hin and γδ resolvase DNA-binding domains with the *Antennapedia* and *engrailed* homeodomains. Putative α helices for Hin and γδ and the known α helices for *Antennapedia* and *engrailed* are underlined. Numbering scheme is taken from Affolter, *et al. Cell*, **64**, *1991*, 879.

statistical evaluations have shown that the Hin DNA binding domain most likely contains a helix-turn-helix motif, which appears to be more closely related to the helix-turn-helix of prokaryotic regulatory proteins rather than eukaryotic homeodomains; like the prokaryotes, the Hin recognition helix is believed to be approximately 10 amino acids in length, whereas homeodomains contain much longer recognition helices of ~17 residues. Alignment of the protein sequences of *en*, *Antp*, and Hin reveal that both hydrophobic and positively charged residues are highly conserved (Figure 6).[84] The conserved hydrophobic residues play an important role in the architecture of the hydrophobic core of the HTH unit, and this indicates that Hin possesses an HTH structure that folds similarly to the homeodomain HTH. For the alignment shown, the sequence identity between *en* and Hin is 27% (Figure 6).[84]

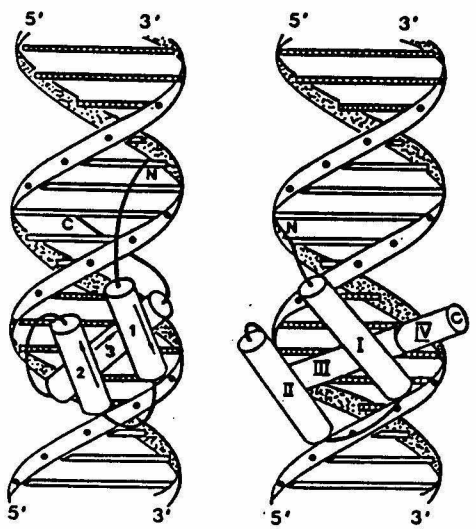


FIGURE 7. Left. Model of the DNA binding domain of Hin recombinase. α-Helices are indicated as cylinders and arrows point from the N→C termini. **Right.** Schematic model of the *Antennapedia* homeodomain-DNA complex studied by NMR. Note the kink between helices III and IV. This kink does not exist in the *engrailed* structure.

Homeodomains bind to operator sites as monomers with binding constants approximately four orders of magnitude higher than that of prokaryotic repressor DNA binding domains binding as monomers to half sites.[84] Like the homeodomain, the Hin DNA binding domain can also bind monomerically to half sites; normally, Hin recombinase binds as a dimer to a twofold, pseudosymmetric operator site. However, the Hin DNA binding domain retains the ability to bind monomerically, albeit with lower affinity, whereas this ability is generally not observed in prokaryotic repressor DNA binding domains.[84]

A remarkable feature of the homeodomain-DNA complex is that contacts are made between the extended amino-terminal arm of the protein and the minor groove of DNA (discussed below). In contrast, it is the HTH unit that provides all the proteins contacts to the operator in prokaryotic regulatory proteins; the only exception is the amino-terminal arm of λ repressor, which wraps around the operator in the major groove and plays a critical role in DNA binding[44]; removal of the arm results in a greater than 8000-fold reduction in binding affinity.[45] Deletion of the amino-terminal residues from the homeodomain greatly diminishes affinity for target DNA, suggesting that the amino-terminal arm contributes significantly toward the high affinity of homeodomains for their operators.

Minor groove interactions are also critical for the recognition of DNA by the Hin recombinase DNA binding domain (Chapters 2 and 3). Affinity cleaving[169] and dimethyl sulfate footprinting studies[214] show that there is a protein interaction in the minor groove. Additionally, a high degree of sequence homology exists among the amino-terminal residues of the Hin DNA binding domain and the homeodomains (Figure 6).[84] This homology is most pronounced between Hin and *en*, with residues 9-11 of Hin (Arg-Pro-

Arg) corresponding exactly to residues 3-5 of *en*. The direct involvement of both arginines with the minor groove in the *en*-DNA complex and the requirement of these three amino acids for binding of the Hin DNA binding domain to DNA argue that this sequence may interact similarly in both complexes. However, the well characterized DNA binding domains of prokaryotic HTH proteins make no contacts to the minor groove, except in the case of 434 repressor in which an internal loop makes water-mediated interactions with the minor groove (Chapter 1).

The observations discussed above provide evidence that the amino-terminal arm of the Hin DNA binding domain-DNA complex is similar to that of the homeodomain-DNA complex. The similarity between eukaryotic homeodomains and the prokaryotic Hin DNA binding domain may extend to other prokaryotic DNA binding proteins, such as γδ resolvase, which interacts with both the major and minor grooves.[141]

AFFINITY CLEAVING STUDIES ON [Fe•EDTA]*en* and Ni•GGH*en*

Using the numbering convention for homeodomains shown in Figure 3, the *engrailed* homeodomain, residues 2-59, was synthesized and Boc-GABA and tricyclohexyl EDTA (see Chapter 2) were covalently attached at the amino terminus to afford the EDTA-GABA-derivatized homeodomain, [EDTA]*en*. This chemically synthesized homeodomain contains [EDTA-GABA]-Lys-Arg-Pro-Arg- at the amino terminus, which is very similar to the Hin DNA binding domain, which contains [EDTA-GABA]-Gly-Arg-Pro-Arg-; thus [Fe•EDTA]*en* should provide affinity cleaving results directly comparable to that for [Fe•EDTA]Hin(139-190). The sequence of the synthesized homeodomain is shown in Figure 8. The DNA binding site used in the affinity cleaving studies is the same as that used in the *engrailed* cocrystal

complex (Figure 3). Two complementary oligonucleotides, a 26-mer and 18-mer, were synthesized and inserted into the multiple cloning site polylinker of plasmid pUC19[204]; for insertion and ligation purposes, an EcoR I restriction endonuclease site was placed at the 5′ end of the oligonucleotide insert (see Figure 3 as a reference), and a Pst I site was placed at the 3′ end. A 433 base-pair restriction fragment was used in these studies.

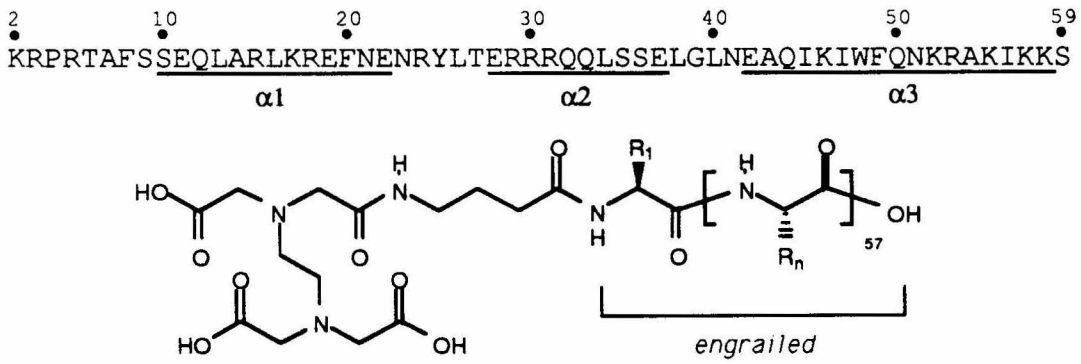
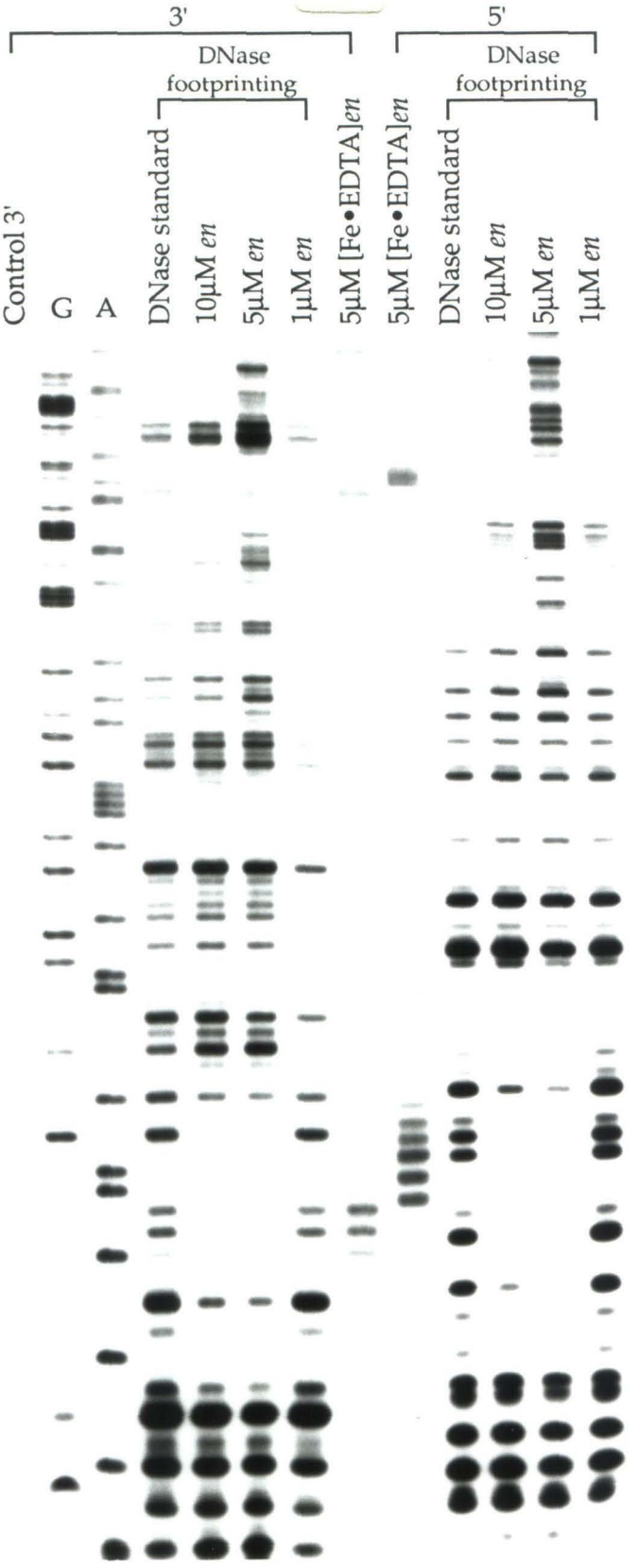


FIGURE 8. Top. Residues 2-59 (according to the numbering scheme in Figure 3) comprising the *engrailed* homeodomain; α-helices are underlined. **Bottom.** The EDTA metal chelator and GABA linker attached to *en*.

The autoradiogram in Figure 9 shows two extremely important results. First of all, the underivatized *en* (residues 2-59) gives very strong footprints at micromolar concentrations, a result similar to Hin(139-190). Because the binding constant for *engrailed* bound to the operator used for crystallization studies (discussed above) is on the order of 10^9 M^{-1}, strong footprints were expected; it is interesting to note that the *engrailed* monomer achieves such a strong DNA binding interaction without any cooperative enhancement of binding, whereas the full Hin recombinase, whose binding constant is also

FIGURE 9. Autoradiogram of a high-resolution, 8% polyacrylamide denaturing gel containing footprinting and affinity cleaving reactions of *engrailed* bound to the Hind III-EcoO 109 restriction fragment (433 base pairs) from pJSENGR. 3′ and 5′ end-labelled lanes are indicated in the figure. Lane 1, 3′ end-labelled DNA control; lane 2, G sequencing reaction; lane 3, A sequencing reaction. Lanes 4-7, 3′ end-labelled DNase footprinting; lane 4, DNase standard; lane 5, 10μM *en*; lane 6, 5μM *en*; lane 7, 1μM *en*. Lanes 8 and 9, affinity cleaving reactions (3′ and 5′ end-labels) with 5μM [Fe•EDTA]*en*. Lanes 10-13, 5′ end-labelled DNase footprinting; lane 10, DNase standard; lane 11, 10μM *en*; lane 12, 5μM *en*; lane 13, 1μM *en*. DNase footprinting reaction mixtures (10μl) contained ^{32}P end-labelled DNA fragment (10,000 cpm),*en*, TKMC buffer, pH 7.0, 0.1 mg/ml tRNA, and DNase solution. Reactions were initiated by the addition of 2μL DNase solution and allowed to proceed for 3 min at 22°C followed by addition of 1.5μL DNase footprinting stop and ethanol precipitation. Affinity cleaving reaction mixtures (10μl) contained ^{32}P end-labelled DNA fragment (10,000 cpm), [Fe•EDTA]*en*, 1 mM sodium ascorbate, 50 mM tris-acetate, pH 7.0, 20 mM NaCl, and 0.1 mg/ml tRNA. Reactions were initiated by the addition of sodium ascorbate (1 mM), and were allowed to proceed for 30 min at 22°C. Reactions were terminated by ethanol precipitation.


3'
5'
DNase footprinting
DNase footprinting
Control 3'
G
A
DNase standard
10μM *en*
5μM *en*
1μM *en*
5μM [Fe•EDTA]*en*
5μM [Fe•EDTA]*en*
DNase standard
10μM *en*
5μM *en*
1μM *en*

around 10^9 M^{-1} (discussed in Chapter 3), binds cooperatively to a dimeric binding site. Secondly, [Fe•EDTA]*en* is capable of binding and cleaving DNA with sequence specificity; a strong affinity cleaving pattern is seen at the *engrailed* binding site. This pattern is shifted to the 3′ side indicating that the affinity cleaving moiety lies in or near the minor groove, as was expected. The cleavage on the 5′ end-labelled DNA strand appears to be more intense than that for the 3′ strand; this reproducible result indicates that the affinity cleaving moiety is closer to the 5′ strand.

The affinity cleaving experiments also show that [Fe•EDTA]*en* binds to sites other than the *engrailed* operator (5′-TAAT); these other sites are all 4-6 base pair tracts of A's and T's–for example, one site contains 5′-TTAAT and another is 5′-TTTTA. In Figure 9, a weaker affinity cleaving pattern appears about ten base pairs above the cleaving pattern at the *engrailed* binding site. The *engrailed* cocrystal structure shows that the DNA undergoes very little distortion upon binding to the homeodomain, and that the DNA remains in B-form. Because ten base pairs of DNA are approximately equivalent to one turn of B-DNA,[41] it was speculated that the weaker cleavage was caused by the diffusible oxidant generated in the minor groove of the *engrailed* site spilling over into an adjacent minor groove.

In order to test this possibility, the metal-chelating tripeptide Gly-Gly-His was covalently attached to the amino terminus (Lys2) of *en*; from previous work, it is known the metal-GGH moiety does not produce a diffusible DNA cleaving agent.[213] It is also suspected that cleavage occurs only in the minor groove. Nickel(II) and copper(II) form square-planar complexes with GGH (Figure 10), and this complex can slide into the minor groove. Although the mechanism for cleavage of DNA by these complexes is unclear, the cleaving agent is suspected of being a non-diffusible metal-oxo or ligand-

oxo moiety.[213] When GGH-derivatized Hin(139-190) is incubated with Cu(II), weak cleavage at the Hin binding sites is exhibited[213]; however, Cu•GGH*en* shows no sequence-specific DNA cleavage. However, Ni•GGH*en* shows strong cleavage of DNA at the *engrailed* operator site as well as other sites containing A,T tracts, including some A,T tracts in which virtually no affinity cleavage is exhibited (Figures 11 and 12). Control reactions in Figures 11 and 12 demonstrate that GGH*en*, Ni(II), and an oxidant such as monoperoxyphthalic acid are all required for cleavage; the absence of any of these components results in no cleavage.

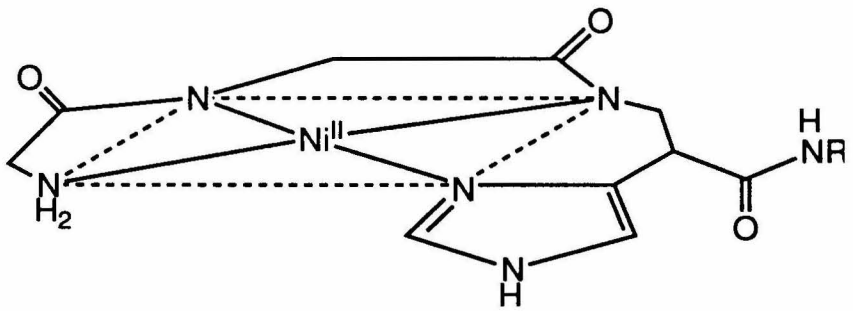


FIGURE 10. The square-planar Ni(II)•Gly-Gly-His complex.

Importantly, the weaker cleavage site, which is one turn away from the *engrailed* site, shows strong cleavage; thus, cleavage at this site is due to protein binding and not to DNA cleaving oxidant diffusing into an adjacent minor groove. The sequence at this weaker cleavage site is 5′-AATT, which is rather similar to the 5′-TAAT sequence of the *engrailed* operator; this weaker site is actually the EcoR I site (5′-GAATTC) used for insertion of the synthesized oligonucleotide duplex into pUC19. The weaker site was inadvertently designed into the DNA. In the *engrailed* cocrystal structure, a second homeodomain binding site was also inadvertently created when the DNA duplexes stack to form a pseudocontinuous helix (discussed above) with

FIGURE 11. Autoradiogram of a high-resolution, 8% polyacrylamide denaturing gel containing affinity cleaving reactions of *engrailed* bound to the Hind III-EcoO 109 restriction fragment (433 base pairs) from pJSENGR. All lanes contain 5′ end-labelled DNA. Lane 1, 5′ end-labelled DNA control; lane 2, A sequencing reaction; lane 3, G sequencing reaction. Lanes 4-7, affinity cleaving reactions; lane 4, 5μM [Fe•EDTA]*en*; lane 5, 5μM [Fe•EDTA]*en*; lane 6, 1μM [Fe•EDTA]*en*. Lanes 7-10, control reactions as indicated in the legend on the right of the figure. Lane 7 does not contain Ni(II) or monoperoxyphthalic acid; lane 8 does not contain monoperoxyphthalic acid; lane 9 does not contain GGH*en* or monoperoxyphthalic acid; lane 10 does not contain GGH*en*. Lanes 11-13 contain all components necessary for cleavage. Lane 11, 5μM Ni•GGH*en*; lane 12, 2.5μM Ni•GGH*en*; lane 13, 1μM Ni•GGH*en*. Affinity cleaving reaction mixtures (10μl) with [Fe•EDTA]*en* contained ^{32}P end-labelled DNA fragment (10,000 cpm), [Fe•EDTA]*en*, 1 mM sodium ascorbate, 50 mM tris-acetate, pH 7.0, 20 mM NaCl, and 0.1 mg/ml tRNA. Reactions were initiated by the addition of sodium ascorbate (1 mM), and were allowed to proceed for 30 min at 22°C. Reactions were terminated by ethanol precipitation. Affinity cleaving reaction mixtures (10μl) with Ni•GGH*en* contained ^{32}P end-labelled DNA fragment (10,000 cpm), Ni•GGH*en*, 20 mM phosphate buffer, pH 7.5, 20 mM NaCl, 0.1 mg/ml tRNA, and 5μM monoperoxphthalic acid. All components except monoperoxyphthalic acid were incubated for 20 min at 37°C. Reactions were initiated by the addition of monoperoxyphthalic acid and allowed to proceed for 30 min at 22°C followed by ethanol precipitation. Reactions were dried and 50μL of 0.1N butylamine was added; reactions proceeded for 30 min at 90°C.

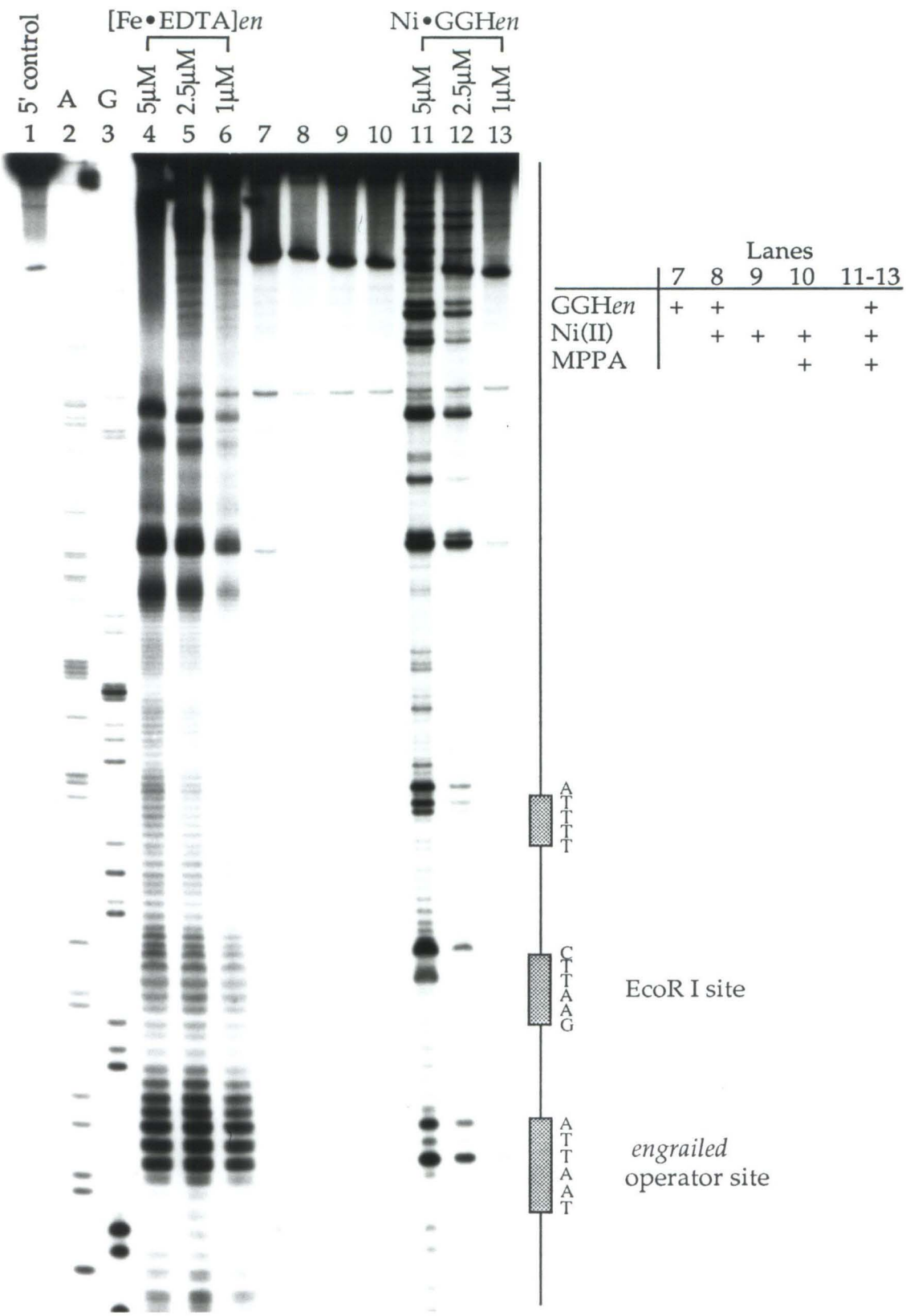
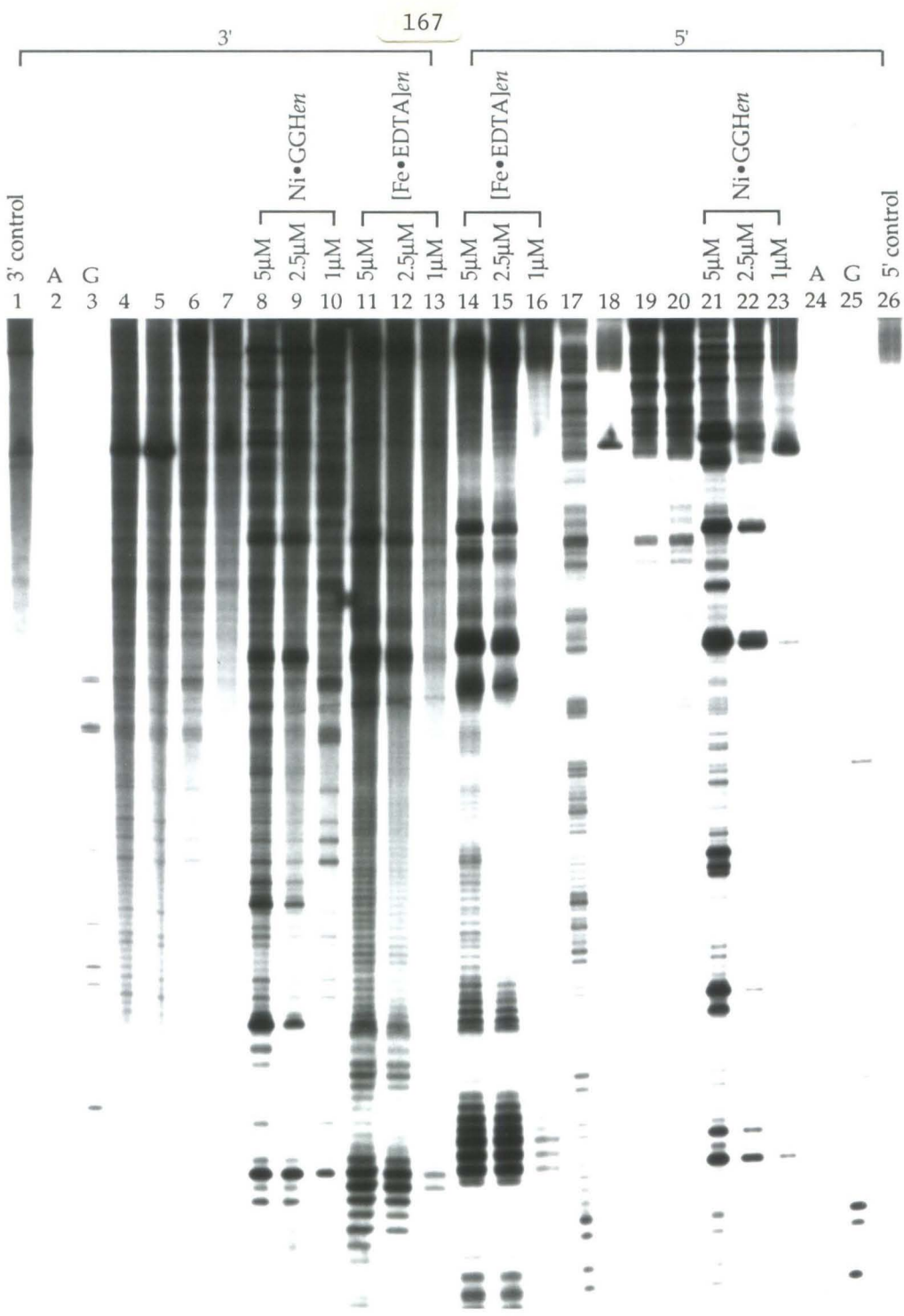

5' control
A
G
[Fe•EDTA]*en*
5μM
2.5μM
1μM
Ni•GGH*en*
5μM
2.5μM
1μM
1 2 3 4 5 6 7 8 9 10 11 12 13
Lanes
7 8 9 10 11-13
GGH*en* + + + +
Ni(II) + + + +
MPPA + +
ATTTT
CTTAAG
EcoR I site
ATTAAT
engrailed
operator site

FIGURE 12. Autoradiogram of a high-resolution, 8% polyacrylamide denaturing gel containing affinity cleaving reactions of *engrailed* bound to the Hind III-EcoO 109 restriction fragment (433 base pairs) from pJSENGR. 3′ and 5′ end-labelled DNA reactions are indicated in the figure. Lane 1, 3′ end-labelled DNA control; lane 2, A sequencing reaction; lane 3, G sequencing reaction. Lanes 4-7, control reactions as indicated in the legend at the bottom of the figure. Lane 4 does not contain Ni(II) or monoperoxyphthalic acid; lane 5 does not contain monoperoxyphthalic acid; lane 6 does not contain GGH*en* or monoperoxyphthalic acid; lane 7 does not contain GGH*en*. Lanes 8-10 contain all components necessary for cleavage. Lane 8, 5μM Ni•GGH*en*; lane 9, 2.5μM Ni•GGH*en*; lane 10, 1μM Ni•GGH*en*. Lanes 11-13, affinity cleaving reactions; lane 11, 5μM [Fe•EDTA]*en*; lane 12, 5μM [Fe•EDTA]*en*; lane 13, 1μM [Fe•EDTA]*en*. Lanes 14-16, affinity cleaving reactions; lane 14, 5μM [Fe•EDTA]*en*; lane 15, 5μM [Fe•EDTA]*en*; lane 16, 1μM [Fe•EDTA]*en*. Lanes 17-20, control reactions as indicated in the legend at the bottom of the figure. Lane 17 does not contain Ni(II) or monoperoxyphthalic acid; lane 18 does not contain monoperoxyphthalic acid; lane 19 does not contain GGH*en* or monoperoxyphthalic acid; lane 20 does not contain GGH*en*. Lanes 21-23 contain all components necessary for cleavage. Lane 21, 5μM Ni•GGH*en*; lane 22, 2.5μM Ni•GGH*en*; lane 23, 1μM Ni•GGH*en*. Lane 24, A sequencing reaction; lane 25, G sequencing reaction; lane 26, 5′ end-labelled DNA control. Affinity cleaving reaction mixtures (10μl) with [Fe•EDTA]*en* contained ^{32}P end-labelled DNA fragment (10,000 cpm), [Fe•EDTA]*en*, 1 mM sodium ascorbate, 50 mM tris-acetate, pH 7.0, 20 mM NaCl, and 0.1 mg/ml tRNA. Reactions were initiated by the addition of sodium ascorbate (1 mM), and were allowed to proceed for 30 min at 22°C. Reactions were terminated by ethanol precipitation. Affinity cleaving reaction mixtures (10μl) with Ni•GGH*en* contained ^{32}P end-labelled DNA fragment (10,000 cpm), Ni•GGH*en*, 20 mM phosphate buffer, pH 7.5, 20 mM NaCl, 0.1 mg/ml tRNA, and 5μM monoperoxphthalic acid. All components except monoperoxyphthalic acid were incubated for 20 min at 37°C. Reactions were initiated by the addition of monoperoxyphthalic acid and allowed to proceed for 30 min at 22°C followed by ethanol precipitation. Reactions were dried and 50μL of 0.1N butylamine was added; reactions proceeded for 30 min at 90°C.

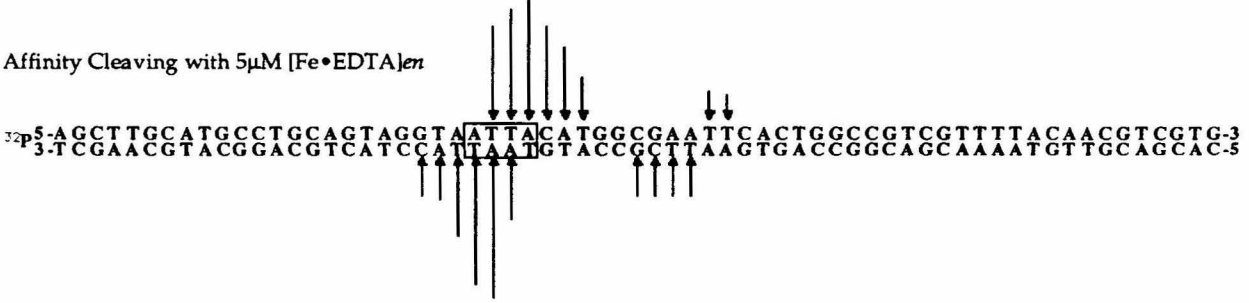


	Lanes				
	4	5	6	7	8-10
	17	18	19	20	21-23
GGH*en*	+	+			+
Ni(II)		+	+	+	+
MPPA				+	+

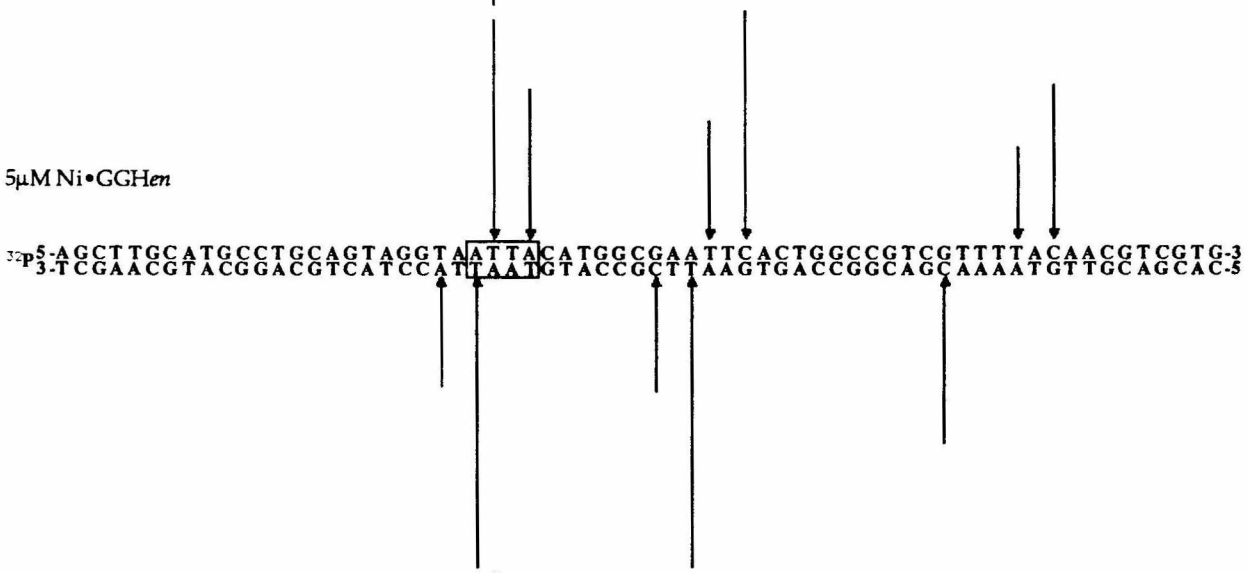
FIGURE 13. Histogram data from Figures 9, 11, and 12. Arrow heights indicate extent of cleavage.

DNase Footprinting with 5μM *en*

Affinity Cleaving with 5μM [Fe•EDTA]*en*



5μM Ni•GGH*en*



2.5μM Ni•GGH*en*

1μM Ni•GGH*en*

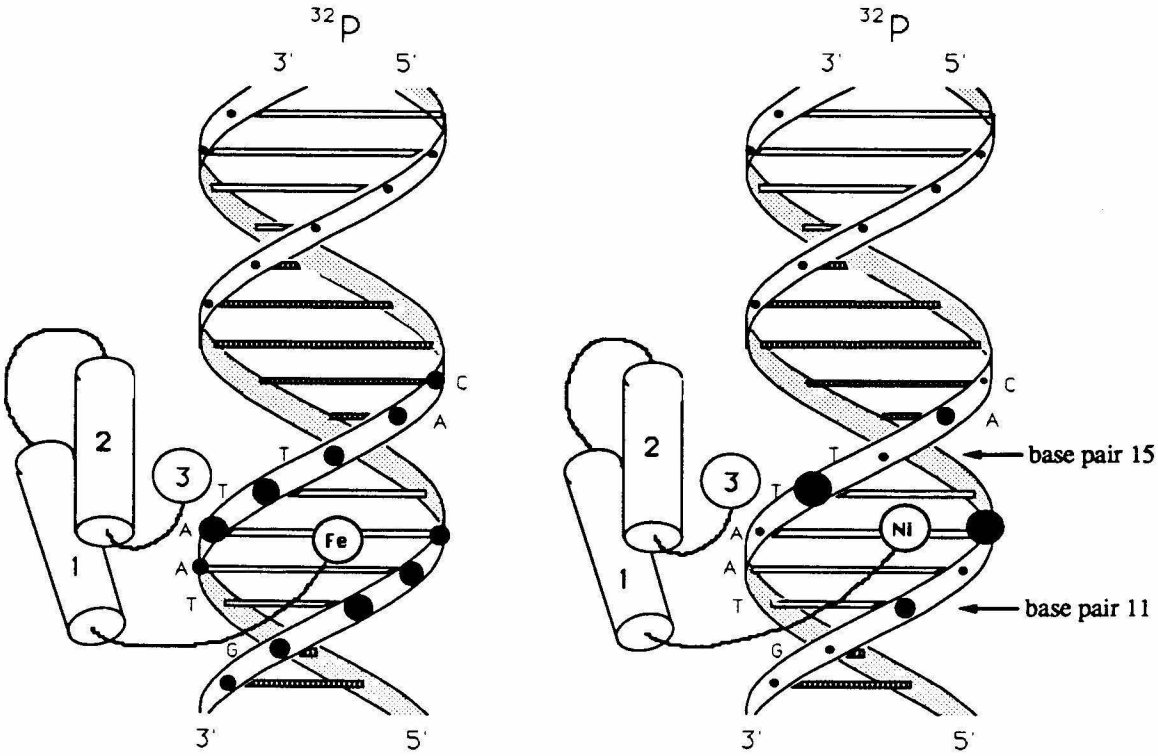


FIGURE 14 . Models of the affinity cleaving proteins [Fe•EDTA]*en* and Ni•GGH*en*. Affinity cleaving sites are indicated by black filled circles; size of circles represents extent of cleavage. **Left.** Affinity cleaving pattern from [Fe•EDTA]*en*. **Right.** Affinity cleaving pattern from Ni•GGH*en*.

sequence 5′-AAAT; in a conversation with Carl Pabo, he had stressed the importance of breaking up A,T tracts to avoid creating unintentional binding sites. It was not believed that the EcoR I site would provide a homeodomain binding site, especially such a strong site; it was also surprising that Ni•GGH*en* shows strong cleavage at the site 5′-TTTTA, at which no discernible affinity cleavage is exhibited (Figure 11).

The Ni•GGH*en* DNA cleaving pattern is striking. Intensity of cleavage is much stronger for Ni•GGH*en* than that for [Fe•EDTA]*en* at 5μM protein concentration (Figures 11, 12, and 13); this observation is corroborated by previous work with the corresponding Hin proteins. As the concentration of protein decreases, however, from 5μM to 2.5μM, and finally to 1μM, the intensity of cleavage at the *engrailed* binding site drops off drastically for Ni•GGH*en*, whereas the intensity of cleavage produced by [Fe•EDTA]*en* diminishes only slightly. An explanation for this result is that the Ni•GGH moiety may provide an unfavorable interaction with DNA causing binding affinity to decrease somewhat. Unlike [Fe•EDTA]*en*, which contains a four-carbon linker between the *engrailed* amino terminus (Lys2) and the EDTA•Fe moiety, Ni•GGH is directly attached to Lys2 and may interfere with the Arg3, and perhaps even the Arg5, interaction with the minor groove.

Similarly, at the EcoR I binding site, the intensity of cleavage produced by [Fe•EDTA]*en* hardly diminishes, whereas that for Ni•GGH*en* vanishes. Although at 5μM Ni•GGH*en*, the cleavage intensities at the EcoR I and *engrailed* binding sites are equivalently strong, as protein concentration decreases to 1μM, the cleavage pattern disappears at EcoR I, but still remains, albeit weakly, at the *engrailed* site. Both the affinity cleaving data from [Fe•EDTA]*en*, which shows much weaker cleavage at the EcoR I site than at

the *engrailed* site, and the data from Ni•GGH*en* show that binding is weaker at the EcoR I site than at the *engrailed* site.

It would appear that at higher protein concentrations, the Ni•GGH cleaving agent is much more efficient at cleaving DNA than the EDTA•Fe moiety, for cleavage produced by Ni•GGH*en* at 5μM concentration is much stronger than the cleavage produced by 5μM [Fe•EDTA]*en*. The Ni•GGH*en* cleaving pattern is very striking, for it shows strong cleavage at one base pair followed by weaker cleavage two base pairs away (Figures 11 and 12), and this pattern is exhibited at the other Ni•GGH*en* binding sites. This result is in direct contrast with work by Dave Mack on [Ni•GGH]Hin(139-190); this protein gives strong cleavage at a single base pair with only slight cleavage exhibited at an adjacent base pair.[213]

The possibility that Ni•GGH*en* is binding in two orientations on the DNA has been offered as an explanation as to why the Ni•GGH*en* cleavage pattern differs from that of [Ni•GGH]Hin(139-190); the Ni•GGH*en* cleavage pattern shows a strongly cleaved band, followed by a skipped band, then a weaker band (Figure 13). Because the *engrailed* binding site is palindromic (5′-TAAT), the question arises as to whether the homeodomain can bind in two orientations: i.e., the recognition helix (helix 3) in Figure 14 would be turned 180° in the plane perpendicular to the page, and thus one binding orientation is preferred over the other, giving stronger and weaker bands. This explanation for the Ni•GGH*en* cleavage pattern may be possible. Other Ni•GGH*en* binding sites show the same type of cleavage pattern, and it may be that Ni•GGH*en* can bind in two orientations at these sites also. However, the [Fe•EDTA]*en* cleaving pattern does not appear to contain two cleavage patterns shifted by two base pairs superimposed upon each other, and thus it seems that this explanation for the Ni•GGH*en* cleavage pattern is unlikely.

Computer modeling studies were performed on the *engrailed* homeodomain cocrystal structure. The Fe•EDTA-GABA moiety was attached to the amino terminus of *engrailed* (Figure 15). Modeling studies showed that the most intense affinity cleavage should be exhibited 1-2 base pairs from the amino terminus of the homeodomain; modeling exactly correlated with the affinity cleaving data that showed the most intense cleavage at the expected base pairs (Figure 16). In Figure 16, the base pairs exhibiting the most intense affinity cleaving bands are highlighted; on the left DNA strand in Figure 16, base-pairs 13 and 14 show the most intense cleavage, whereas on the right DNA strand, base-pairs 11-13 show the most intense cleavage. From the *engrailed* cocrystal structure, these affinity cleaving results were expected. Therefore, the structural data obtained from affinity cleaving can yield accurate information about protein-DNA structures, especially information about the positioning of protein elements with respect to DNA. Additionally, attachment of the Ni•GGH DNA cleaving moiety to a protein can yield structural information about protein elements that lie in the minor groove. Both the EDTA•Fe and Ni•GGH affinity cleaving moieties can provide valuable structural information about the local protein elements to which they are attached, and the affinity cleaving patterns from [Fe•EDTA]*en* and Ni•GGH*en* are in the same location on the DNA. The results from these two affinity cleaving proteins corroborate and complement each other, and prove that affinity cleaving is a powerful technique for gaining structural information about complexes between affinity cleaving molecules and DNA.

FIGURE 15. Modeling of the Fe•EDTA-GABA moiety attached to the amino terminus of the *engrailed* homeodomain. The coordinates of the cocrystal structure were received from Carl Pabo,[81] and the Fe•EDTA-GABA moiety was built using the Insight II computer modeling system from Biosym Technologies. *Engrailed* is shown as a green ribbon with protruding side chains; helix 3 is the green circle lying in the major groove. The Fe•EDTA-GABA moiety is shown in yellow and the DNA duplex in pink.

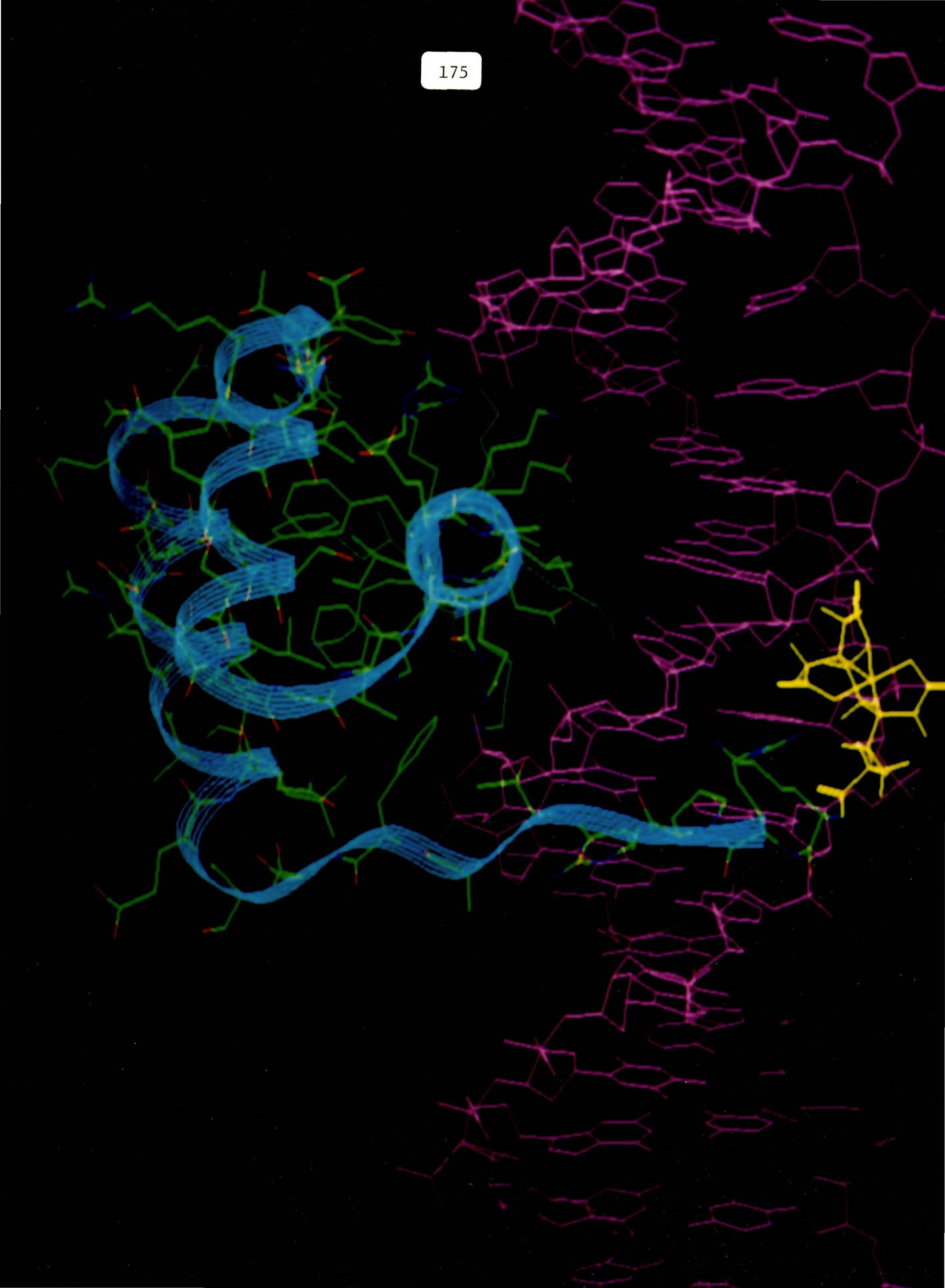
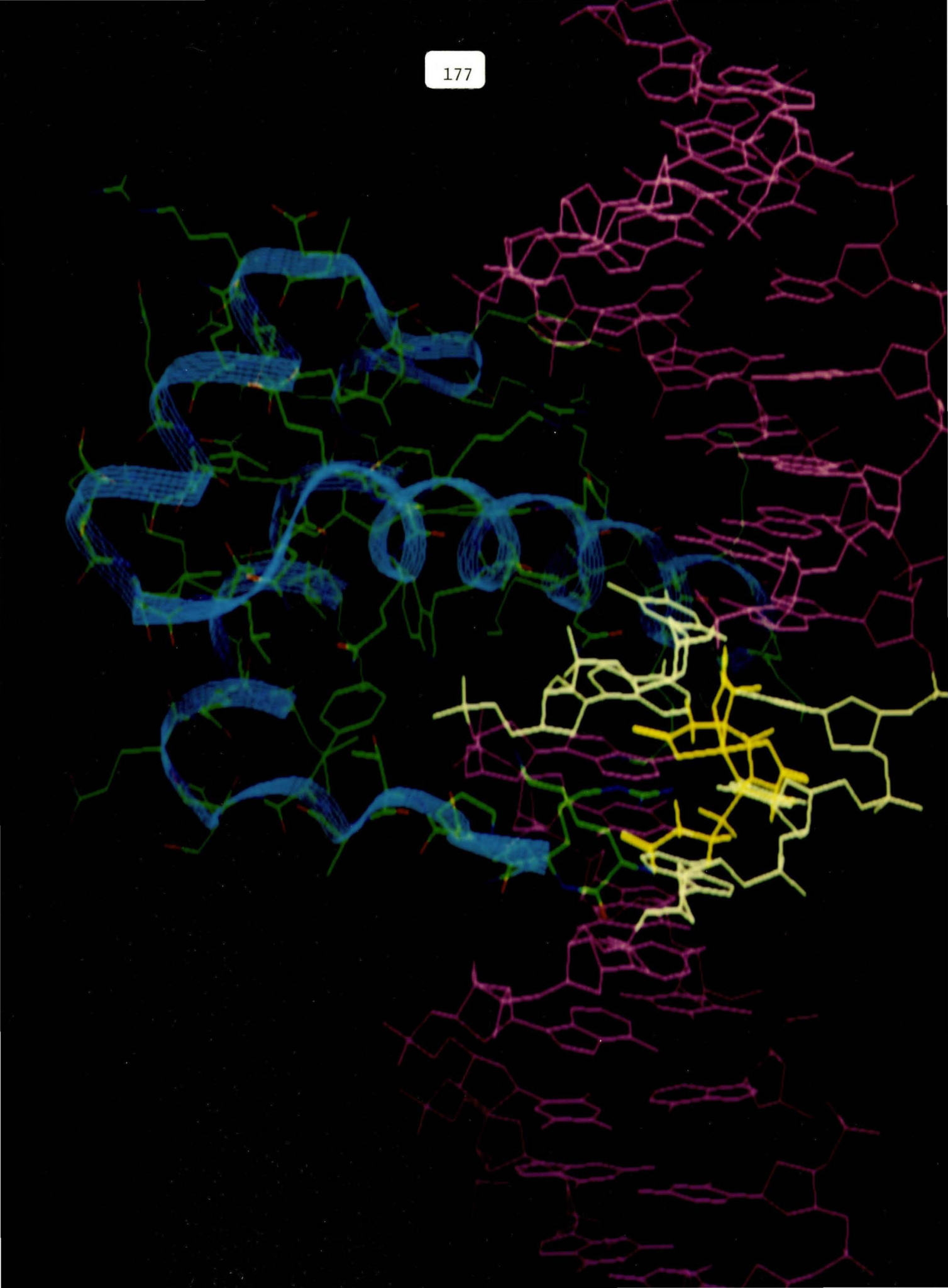


FIGURE 16. Closer view of the Fe•EDTA-GABA moiety attached to the amino terminus of the *engrailed* homeodomain. The coordinates of the cocrystal structure were received from Carl Pabo.[81] Bases at which strong cleavage bands are produced are highlighted in white.



A GENERAL MODEL FOR HOMEODOMAIN-DNA INTERACTIONS

A year after the *engrailed*-DNA cocrystal structure was determined, the cocrystal of the yeast MATα2 homeodomain complexed to a consensus operator was solved to 2.7Å resolution.[215] Comparison of this structure with the Drosophila *engrailed* homeodomain-DNA complex shows that the protein fold is highly conserved, despite a three-residue insertion in the loop between helices 1 and 2 in MATα2 and a mere 27% sequence identity between the two homeodomains; sequence identity between the MATα2 and *Antennapedia* homeodomain is only 28%.[215] The orientation of the MATα2 recognition helix on the DNA is also conserved; this arrangement is maintained by side-chain contacts with the DNA phosphodiester-sugar backbone that are identical between the two homeodomain complexes. These residues are conserved among all homeodomains, and this result suggests a general model for homeodomain-DNA interactions.[215]

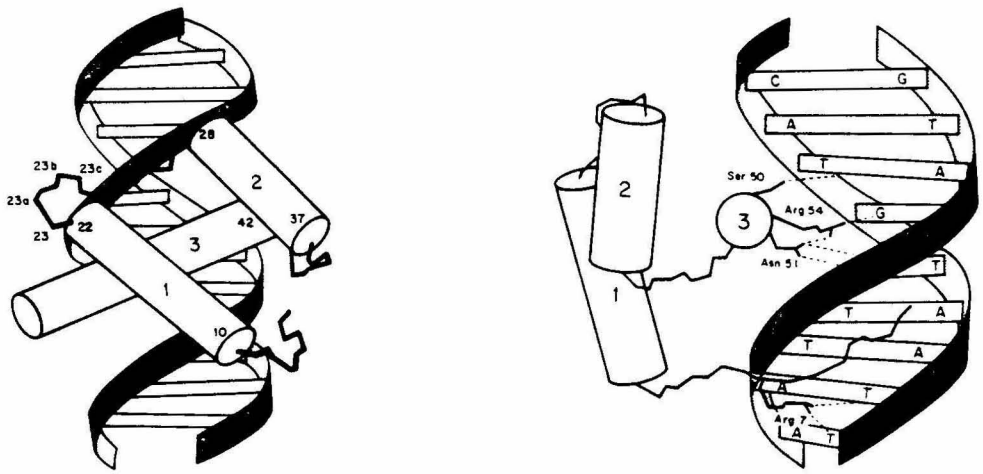


FIGURE 17. Sketch of the α2 homeodomain-DNA complex. Critical contacts are shown on the right.

The yeast α2 repressor, the product of the MATα2 gene, is a mating-type regulatory protein that specifies cell identity in yeast and is a particularly well characterized member of the homeodomain superfamily of DNA binding proteins.[73,211] The intact α2 repressor contains 210 residues; the homeodomain occurs near the carboxyl terminus.[215] The α2 homeodomain, like *engrailed* and *Antennapedia*, contains a helix-turn-helix unit and an amino-terminal arm that binds in the minor groove. As in the case of *engrailed*, recognition helix 3 of α2 is a single extended helix, and the DNA exhibits only minor distortions from B-form DNA upon binding of α2.

Side chains from α2 contact three base pairs in the major groove and two base pairs in the minor groove (Figure 17). The major groove contacts are made by side chains in helix 3. Asn51 and Arg54 form hydrogen bonds to each other and to base-pairs 10 and 11; Ser50 makes a van der Waals contact to base-pair 9. The extended amino-terminal arm fits into the minor groove, and Arg7 makes contacts with the O2 of thymine in both base pairs 14 and 15. Unlike *engrailed* and Hin recombinase, α2 does not contain an Arg-Pro-Arg recognition element in the minor groove. The α2 homeodomain makes an extensive set of contacts with the DNA backbone, and this presumably makes a large contribution to positioning of helix 3 in the major groove as well as to the binding energy. A total of 8 residues interact with the DNA backbone–7 contacts with phosphate groups and one contact with a sugar ring. These homeodomain contacts to the DNA backbone are highly conserved between the *engrailed* and α2 homeodomains, whereas more unique contacts are made between individual side chains and specific base pairs.

The DNA operator used in the α2 cocrystal complex is very similar to the *engrailed* operator. Both operators contain the sequence 5'-CATGTAATT, and the numbering scheme for the α2 binding site referred to below will be

the same as in Figure 3; alignment of various α2 binding sites revealed that this sequence is also the α2 consensus sequence. Within this conserved site, the 5′-TGTA sequence is invariant[215]; studies on mutant operators have shown that the strongest reduction in α2 binding occurs from changes in the 5′-TGT sequence (base pairs 9-11). The crystal structure shows that the α2 recognition helix is centered on this sequence and makes side-chain contacts with all three base pairs.[215]

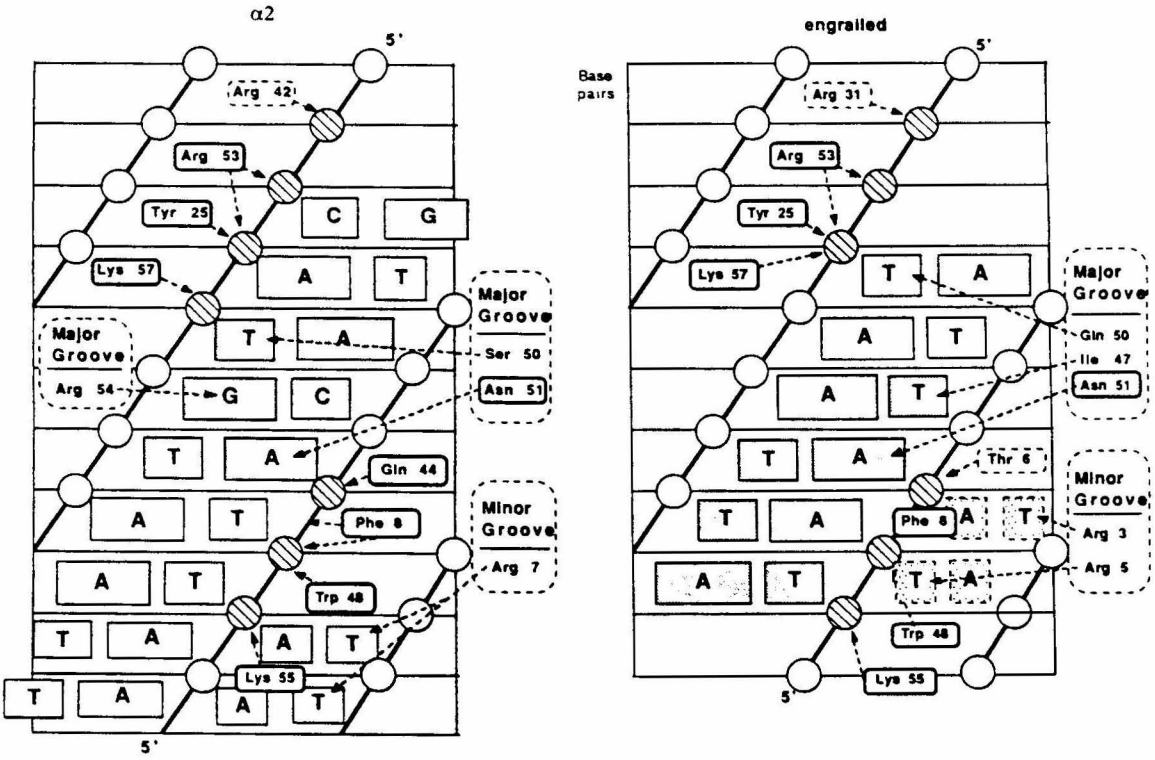


FIGURE 18. Sketch summarizing all the contacts made by the α2 and *engrailed* homeodomains. The DNA is represented as a cylindrical projection; phosphates are shown as circles. Residues that are identical in the two proteins are encircled in solid lines; those that are different are encircled with dashed lines. The TGT core of the α2 site and the TAAT core of *engrailed* are stippled.

Comparison of α2 binding operators also shows that the base pair at position 14 is highly conserved, and in fact, it is always either an A•T or T•A base pair.[215] Arg7 contacts both base pairs 14 and 15 in the minor groove. As

discussed in Chapter Two, both A•T or T•A base pairs have a similar distribution of hydrogen bond donors and acceptors in the minor groove, thereby making discrimination in the minor groove between the two base pairs difficult. Biochemical and genetic studies demonstrate that Arg7 confers on α2 this base preference; the invariance at base pair 15, however, is not entirely due to the Arg7 contact. Likely, other proteins that contact the intact α2 repressor upon binding to DNA contribute to this invariance.[215]

A detailed comparison between the *engrailed*-DNA and α2-DNA complexes shows that the two structures are very similar.[215] To compare the recognition helix-DNA interactions, the two homeodomain-DNA complexes were aligned by superimposing the protein backbones of the recognition helices and allowing the position of the DNA to be determined by the protein. Remarkably, this comparison results in superposition of the two DNA duplexes with the bases in precise register and nearly superimposing phosphates. Given the significant differences in sequence between the two homeodomains, this result was unanticipated, and it indicates that the recognition helix has a conserved docking mechanism that has been tightly conserved throughout evolution.

These results also suggest that homeodomains can be expected to bind to DNA in the same manner and adopt similar orientations for the recognition helix and amino-terminal arm relative to DNA. The question of how different homeodomains recognize different sites lies in the contacting residues in helix 3 and the amino-terminal arm. Asn51, an invariant residue among homeodomains whose side chain projects from helix 3, is expected to contact an adenine in every homeodomain-DNA complex. The side chains of Asn47, Ser50, and Arg54 of α2 also project from helix 3 into the major groove; these three amino acids, however, vary considerably among

homeodomains and should be crucial in dictating sequence specificity. Additionally, arginines on the amino-terminal arm of homeodomains make specific contacts to the minor groove; these include Arg3 and Arg5 of *engrailed* and Arg7 of α2, which contact conserved bases and assist in determining site selectivity.

CONCLUSION

Eukaryotic homeodomains appear to have a more highly conserved DNA binding mechanism than prokaryotic helix-turn-helix proteins. Each prokaryotic helix-turn-helix domain typically makes a different set of DNA backbone contacts, and thus the helix-turn-helix unit is oriented uniquely. Only when a subclass of prokaryotic proteins is examined does there emerge a conserved set of phosphate contacts, and hence, a conserved orientation of the recognition helix on the DNA; such a similarity has been studied on the 434 and λ repressors (Chapter 2).

Although Hin recombinase is a prokaryotic protein believed to contain the helix-turn-helix motif, the Hin DNA binding domain does not exhibit many of the characteristics of other prokaryotic regulatory proteins. Prokaryotic helix-turn-helix proteins do not generally use recognition elements that bind in the minor groove; only 434 repressor has been shown to have a small element that interacts with the minor groove. Additionally, all of the DNA contacts are made by the helix-turn-helix unit in prokaryotes; only λ repressor uses an amino-terminal arm that recognizes the major groove. Thus, the Hin DNA binding domain, which contains DNA recognition elements other than the helix-turn-helix structure and has an amino-terminal arm that recognizes the minor groove, is unlike other known prokaryotic proteins, in this regard, and appears to be similar to the

eukaryotic homeodomain. The lengths of the recognition helices for Hin and *engrailed* are quite different; homeodomains have much longer recognition helices (~17 residues) than do prokaryotic regulatory proteins (~10 residues). It will be interesting to compare the affinity cleaving results produced by Hin(139-190) with the EDTA•Fe moiety attached to the carboxyl terminus, which is near the end of the recognition helix, with that of *engrailed* similarly derivatized at the carboxyl terminus.

The affinity cleaving patterns produced by [Fe•EDTA]*en* and Ni•GGH*en* are exactly as predicted by the *engrailed* -DNA cocrystal structure. Therefore, affinity cleaving is a valuable tool for obtaining structural information about molecular interactions with DNA, especially those complexes for which no NMR or crystallographic data exist. The affinity cleaving study on the *engrailed* homeodomain shows that maximal cleavage occurs 1-2 base pairs from the amino terminus of *engrailed* when the cleaving agent Fe•EDTA-GABA is covalently attached, and these results will aid in the design and interpretation of future affinity cleaving experiments.

MATERIALS AND METHODS

Proteins were synthesized and prepared as described in Chapter 2.

DNA Substrate

Construction of Plasmid pJSENGR. Standard techniques were used in plasmid construction.[205] Oligodeoxyribonucleotides were designed to place EcoR I and Pst I restriction endonuclease sites at the 5' and 3' ends of the insert, respectively. Automated oligonucleotide synthesis was performed on an Applied Biosystems 380B DNA Synthesizer using β-cyanoethyl-phosphoramidite chemistry.[216] All chemicals were purchased from Applied Biosystems. Removal of the oligonucleotide from the support was accomplished by treatment with ammonium hydroxide. Oligonucleotides were purified on 20% acrylamide gels (1:20 cross-link, 50% urea). The complementary oligonucleotides were allowed to anneal, and a phosphate group was added to each 5' end with T4 polynucleotide kinase and ATP (New England Biolabs and Calbiochem, respectively). pUC19 vector[204] (Life Technology Research Laboratories) was cut with EcoR I and Pst I, and the synthesized insert was ligated with T4 DNA ligase into the pUC19 multiple cloning site polylinker region (New England Biolabs). The transformation was conducted on competent cells purchased from Bethesda Research Laboratories; through α-complementation, these transformed DH5α competent cells were capable of producing blue/white colonies suitable for screening on agar plates containing Bluo-gal and IPTG. Colonies were chosen and grown in overnight cell cultures (5 ml); the cells were collected, lysed and the DNA was sequenced (Pharmacia T7 Sequencing Kit). The overnight cell culture yielding plasmid of the desired sequence was inoculated into large overnight cultures (500 ml). Cells were harvested, lysed, and the recovered plasmid was purified by cesium chloride density gradient.

Radioactive Labelling of Restriction Fragment. Plasmid pJSENGR was linearized by digestion with Hind III. Linearized plasmid pJSENGR was 3'-end-labelled with [α^{32}P]-dATP and DNA polymerase I Klenow fragment.[177] Linearized plasmid pJSENGR was 5'-end-labeled by dephosphorylation using calf intestinal alkaline phosphatase, followed by phosphorylation, using [γ^{32}P]-ATP and T4 polynucleotide kinase.[177] Labelled linearized plasmid pJSENGR was digested with EcoO 109, and the resulting labeled 433 base-pair DNA fragment was isolated using non-denaturing polyacrylamide gel electrophoresis.

DNA Cleaving Experiments

Reaction Conditions for Affinity Cleaving with [Fe•EDTA]en. [EDTA]*en* was allowed to equilibrate with ferrous ammonium sulfate (1:1) for 10 minutes, 22°C. Reaction mixtures (10μl) contained ^{32}P end-labelled DNA fragment (10,000 cpm), [Fe•EDTA]*en*, 1 mM sodium ascorbate, 50 mM tris-acetate, pH 7.0, 20 mM NaCl, and 0.1 mg/ml tRNA (Sigma Chemical, Type XX). All components except sodium ascorbate were incubated for 20 min at 37°C. Reactions were initiated by the addition of sodium ascorbate (1 mM), and were allowed to proceed for 30 min at 22°C. Reactions were terminated by ethanol precipitation. Reaction products were analyzed by denaturing gel electrophoresis on 8% polyacrylamide gels (1:20 crosslink, 7M urea). After electrophoresis, gels were dried and autoradiographed. Autoradiograms were analyzed by laser densitometry.

Reaction Conditions for Affinity Cleaving with Ni•GGHen. GGH*en* was allowed to equilibrate with nickel(II) acetate (1:1) for 10 minutes, 22°C. Reaction mixtures (10μl) contained ^{32}P end-labelled DNA fragment (10,000 cpm), Ni•GGH*en*, 20 mM phosphate buffer, pH 7.5, 20 mM NaCl, 0.1 mg/ml tRNA (Sigma Chemical, Type XX), and 5μM monoperoxyphthalic acid. All

components except monoperoxyphthalic acid were incubated for 20 min at 37°C. Reactions were initiated by the addition of monoperoxyphthalic acid and allowed to proceed for 30 min at 22°C followed by ethanol precipitation. Following lyophilization, 50μL of 0.1N butylamine were added to each reaction tube; the tubes were heated for 30 min at 90°C. The tubes were then lyophilized and analyzed by gel electrophoresis as above.

Reaction Conditions for DNase Footprinting. Reaction mixtures (10μl) contained ^{32}P end-labelled DNA fragment (10,000 cpm),*en*, TKMC buffer (10mM tris, 10mM KCl, 100mM $MgCl_2$, 50mM $CaCl_2$, pH 7.0), 0.1 mg/ml tRNA (Sigma Chemical, Type XX), and DNase solution. All components except DNase were incubated for 20 min at 37°C. Reactions were initiated by the addition of 2μL DNase solution (1μL 0.33mg/ml DNase stock, 20μL 50mM DTT, 979μL water for total volume of 1ml) and allowed to proceed for 3 min at 22°C followed by addition of 1.5μL DNase footprinting stop (3M ammonium acetate and 250mM EDTA). Reactions were ethanol precipitated, dried and analyzed by gel electrophoresis as above.

REFERENCES

1. Jacob, F.; Monod, J. *Journal of Molecular Biology* **1961**, *3*, 318-356.

2. Lewin, B. *Genes IV*; Oxford University Press: New York, 1990.

3. Ptashne, M. *A Genetic Switch: Gene Control and Phage lambda*; Cell Press and Blackwell Scientific Publications: Cambridge, MA, 1986.

4. Ptashne, M.; Johnson, A. D.; Pabo, C. O. *Scientific American* **1982**, *247*, 128-140.

5. Johnson, A. D.; Poteete, A. R.; Lauer, G.; Sauer, R. T.; Ackers, G. K.; Ptashne, M. *Nature* **1981**, *294*, 217-223.

6. Johnson, A. D.; Pabo, C. O.; Sauer, R. T. *Methods in Enzymology* **1980**, *65*, 839-856.

7. Johnson, A.; Meyer, B. J.; Ptashne, M. *Proceedings of the National Academy of Sciences of the United States of America* **1978**, *75*, 1783-1787.

8. Maniatis, T.; Ptashne, M.; Maurer, R. *Cold Spring Harbor Symposium in Quantitative Biology* **1973**, *38*, 857-868.

9. Ohlendorf, D. H.; Matthews, B. W. *Annual Review of Biophysics and Bioengineering* **1983**, *12*, 259-284.

10. Pabo, C. O.; Sauer, R. T. *Annual Review of Biochemistry* **1984**, *53*, 293-321.

11. von Hippel, P. H.; Berg, O. G. *Proceedings of the National Academy of Sciences of the United States of America* **1986**, *83*, 1608-1612.

12. Ollis, D. L.; White, S. W. *Chemical Reviews* **1987**, *87*, 981-995.

13. Schlief, R. *Science* **1988**, *241*, 1182-1187.

14. Brennan, R. G.; Matthews, B. W. *Trends in Biochemical Sciences* **1989**, *14*, 286-290.

15. Steitz, T. A. *Quarterly Reviews of Biophysics* **1990**, *23*, 205-280.

16. Harrison, S. C.; Aggarwal, A. K. *Annual Review of Biochemistry* **1990**, *59*, 933-969.

17. Brennan, R. G. *Current Opinion in Structural Biology* **1991**, *1*, 80-88.

18. Harrison, S. C. *Nature* **1991**, *353*, 715-719.

19. Matthews, B. W. *Nature* **1988**, *335*, 294-295.

20. Nelson, H. C. M.; Finch, J. T.; Luisi, B. F.; Klug, A. *Nature* **1987**, *330*, 221-226.

21. Hogan, M. E.; Austin, R. H. *Nature* **1987**, *329*, 263-266.

22. Wharton, R. P.; Ptashne, M. *Trends in Biochemical Sciences* **1986**, *11*, 71-73.

23. Wharton, R. P.; Ptashne, M. *Nature* **1985**, *316*, 601-605.

24. Sauer, R. T.; Yocum, R. R.; Doolittle, R. F.; Lewis, M.; Pabo, C. O. *Nature* **1982**, *298*, 447-451.

25. Jordan, S. R.; Pabo, C. O. *Science* **1988**, *242*, 893-899.

26. Aggarwal, A. K.; Rodgers, D. W.; Drottar, M.; Ptashne, M.; Harrison, S. C. *Science* **1988**, *242*, 899-907.

27. Pabo, C. O.; Aggarwal, A. K.; Jordan, S. R.; Beamer, L. J.; Obeysekare, U. R.; Harrison, S. C. *Science* **1990**, *247*, 1210-1213.

28. Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987**, *326*, 563-567.

29. Creighton, T. E. *Nature* **1987**, *326*, 547-548.

30. Sali, D.; Bycroft, M.; Fersht, A. R. *Nature* **1988**, *335*, 740-743.

31. Anderson, W. F.; Ohlendorf, D. H.; Takeda, Y.; Matthews, B. W. *Nature* **1981**, *290*, 754-758.

32. Pabo, C. O.; Krovatin, W.; Jeffrey, A.; Sauer, R. T. *Nature* **1982**, *298*, 441-446.

33. Lewis, M.; Jeffrey, A.; Wang, J.; Ladner, R.; Ptashne, M.; Pabo, C. O. *Cold Spring Harbor Symposium in Quantitative Biology* **1983**, *47*, 435-440.

34. McKay, D. B.; Weber, I. T.; Steitz, T. A. *Journal of Biological Chemistry* **1982**, *257*, 9518-9524.

35. Steitz, T. A.; Weber, I. T.; Ollis, D.; Brick, P. *Journal of Biomolecular Structure and Dynamics* **1983**, *1*, 1023-1037.

36. Ohlendorf, D. H.; Anderson, W. F.; Fisher, R. G.; Takeda, Y.; Matthews, B. W. *Nature* **1982**, *298*, 718-723.

37. Bushman, F. D.; Anderson, J. E.; Harrison, S. C.; Ptashne, M. *Nature* **1985**, *316*, 651-653.

38. Weber, I. T.; Steitz, T. A. *Proceedings of the National Academy of Sciences of the United States of America* **1984**, *81*, 3973-3977.

39. Anderson, J. E.; Ptashne, M.; Harrison, S. C. *Nature* **1985**, *316*, 596-601.

40. Anderson, J. E.; Ptashne, M.; Harrison, S. C. *Nature* **1987**, *326*, 846-852.

41. Dickerson, R. E. *Scientific American* **1983**, *249*, 94-111.

42. Wolberger, C.; Dong, Y.; Ptashne, M.; Harrison, S. C. *Nature* **1988**, *335*, 789-795.

43. Mondragon, A.; Harrison, S. *Journal of Molecular Biology* **1990**, *219*, 321-334.

44. Clarke, N. D.; Beamer, L. J.; Goldberg, H. R.; Berkower, C.; Pabo, C. O. *Science* **1991**, *254*, 267-270.

45. Eliason, J. L.; Weiss, M. A.; Ptashne, M. *Proceedings of the National Academy of Sciences of the United States of America* **1985**, *82*, 2339-2343.

46. Matthews, B. W.; Ohlendorf, D. H.; Anderson, W. F.; Fisher, R. G.; Takeda, Y. *Cold Spring Harbor Symposium in Quantitative Biology* **1982**, *47*, 427-433.

47. Hochschild, A.; Ptashne, M. *Cell* **1986**, *44*, 925-933.

48. Mossing, M. C.; Sauer, R. T. *Science* **1990**, *250*, 1712-1715.

49. Hubbard, A. J.; Bracco, L. P.; Eisenbeis, S. J.; Gayle, R. B.; Beaton, G.; Caruthers, M. H. *Biochemistry* **1990**, *29*, 9241-9249.

50. Luisi, B. F.; Sigler, P. B. *Biochimica et Biophysica Acta* **1990**, *1048*, 113-126.

51. Joachimiak, A.; Kelley, R. L.; Gunsalus, R. P.; Yanofsky, C.; Sigler, P. B. *Proceedings of the National Academy of Sciences of the United States of America* **1983**, *80*, 668-672.

52. Schevitz, R. W.; Otwinowski, Z.; Joachimiak, A.; Lawson, C. L.; Sigler, P. B. *Nature* **1985**, *317*, 782-786.

53. Lawson, C. L.; Sigler, P. B. *Nature* **1988**, *333*, 869-871.

54. Lawson, C. L.; Zhang, R.-G.; Schevitz, R. W.; Otwinowski, Z.; Joachimiak, A.; Sigler, P. B. *Proteins: Structure, Function, and Genetics* **1988**, *3*, 18-31.

55. Otwinowski, Z.; Schevitz, R. W.; Zhang, R.-G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. *Nature* **1988**, *335*, 321-329.

56. Staacke, D.; Walter, B.; Kisters-Woike, B.; von Wilcken-Bergmann, B.; Müller-Hill, B. *The EMBO Journal* **1990**, *9*, 1963-1967.

57. Schultz, S. C.; Shields, G. C.; Steitz, T. A. *Science* **1991**, *253*, 1001-1007.

58. Wüthrich, K. *Science* **1989**, *243*, 45-50.

59. Kaptein, R.; Boelens, R.; Scheek, R. M.; van Gunsteren, W. F. *Biochemistry* **1988**, *27*, 5389-5395.

60. Geisler, N.; Weber, K. *Biochemistry* **1977**, *16*, 938-943.

61. Geisler, N.; Weber, K. *FEBS Letters* **1978**, *87*, 215-218.

62. Ogata, R. T.; Gilbert, W. *Proceedings of the National Academy of Sciences of the United States of America* **1978**, *75*, 5851-5854.

63. Nick, H.; Arndt, K.; Boschelli, F.; Jarema, M. A. C.; Lillis, M.; Sadler, J.; Caruthers, M.; Lu, P. *Proceedings of the National Academy of Sciences of the United States of America* **1982**, *79*, 218-222.

64. Scheek, R. M.; Zuiderweg, E. R. P.; Klappe, K. J. M.; van Boom, J. H.; Kaptein, R.; Rüterjans, H.; Beyreuther, K. *Biochemistry* **1983**, *22*, 228-235.

65. Kaptein, R.; Zuiderweg, E. R. P.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. F. *Journal of Molecular Biology* **1985**, *182*, 179-182.

66. Boelens, R.; Scheek, R. M.; van Boom, J. H.; Kaptein, R. *Journal of Molecular Biology* **1987**, *193*, 213-216.

67. Lehming, N.; Sartorius, J.; Niemoller, M.; Genenger, G.; von Wilcken-Bergmann, B.; Müller-Hill, B. *The EMBO Journal* **1987**, *6*, 3145-3153.

68. Lehming, H.; Sartorius, J.; Oehler, S.; von Wilcken-Bergmann, B.; Müller-Hill, B. *Proceedings of the National Academy of Sciences of the United States of America* **1988**, *85*, 7947-7951.

69. Lamerichs, R. M. J. N.; Boelens, R.; van der Marel, G. A.; van Boom, J. H.; Kaptein, R.; Buck, F.; Fera, B.; Rüterjans, H. *Biochemistry* **1989**, *28*, 2985-2991.

70. Karslake, C.; Schroeder, S.; Wang, P. L.; Gorenstein, D. G. *Biochemistry* **1990**, *29*, 6578-6584.

71. Shin, J. A.; Ebright, R. H.; Dervan, P. B. *Nucl. Acid. Res.* **1991**, *19*, 5233-5236.

72. Matthews, B. W.; Ohlendorf, D. H.; Anderson, W. F.; Takeda, Y. *Proceedings of the National Academy of Sciences of the United States of America* **1982**, *79*, 1428-1432.

73. Gehring, W. J. *Science* **1987**, *236*, 1245-1252.

74. Ghysen, A.; Dambly-Chaudiere, C. *Genes & Development* **1988**, *2*, 495-501.

75. Holland, P. W. H.; Hogan, B. L. M. *Genes & Development* **1988**, *2*, 773-782.

76. Levine, M.; Hoey, T. *Cell* **1988**, *55*, 537-540.

77. Robertson, M. *Nature* **1988**, *336*, 522-524.

78. Akam, M. *Cell* **1989**, *57*, 347-349.

79. Gehring, W. J.; Müller, M.; Affolter, M.; Percival-Smith, A.; Billeter, M.; Qian, Y. Q.; Otting, G.; Wüthrich, K. *Trends in Genetics* **1990**, *6*, 323-329.

80. Hanes, S. D.; Brent, R. *Science* **1991**, *251*, 426-430.

81. Kissinger, C. R.; Liu, B.; Martin-Blanco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* **1990**, *63*, 579-590.

82. Mihara, H.; Kaiser, E. T. *Science* **1988**, *242*, 925-927.

83. Qian, Y. Q.; Billeter, M.; Otting, G.; Müller, M.; Gehring, W. J.; Wüthrich, K. *Cell* **1989**, *59*, 573-580.

84. Affolter, M.; Percival-Smith, A.; Müller, M.; Billeter, M.; Qian, Y. Q.; Otting, G.; Wüthrich, K.; Gehring, W. J. *Cell* **1991**, *64*, 879-880.

85. Otting, G.; Qian, Y. Q.; Billeter, M.; Müller, M.; Affolter, M.; Gehring, W. J.; Wüthrich, K. *The EMBO Journal* **1990**, *9*, 3085-3092.

86. Percival-Smith, A.; Müller, M.; Affolter, M.; Gehring, W. J. *The EMBO Journal* **1990**, *9*, 3967-3974.

87. Damante, G.; Di Laura, R. *Proceedings of the National Academy of Sciences of the United States of America* **1991**, *88*, 5388-5392.

88. Ekker, S. C.; Young, K. E.; von Kessler, D. P.; Beachy, P. A. *The EMBO Journal* **1991**, *10*, 1179-1186.

89. Yang, C.-C.; Nash, H. A. *Cell* **1989**, *57*, 869-880.

90. Landschulz, W. H.; Johnson, P. F.; McKnight, S. L. *Science* **1988**, *240*, 1759-1764.

91. Marx, J. L. *Science* **1988**, *242*, 1377-1378.

92. Agre, P.; Johnson, P. F.; McKnight, S. L. *Science* **1989**, *246*, 922-926.

93. Landschulz, W. H.; Johnson, P. F.; McKnight, S. L. *Science* **1989**, *243*, 1681-1688.

94. O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Science* **1989**, *243*, 538-542.

95. O'Shea, E. K.; Rutkowski, R.; Stafford, W. F.; Kim, P. S. *Science* **1989**, *245*, 646-648.

96. Turner, R.; Tjian, R. *Science* **1989**, *243*, 1689-1694.

97. Hu, J. C.; O'Shea, E. K.; Kim, P. S.; Sauer, R. T. *Science* **1990**, *250*, 1400-1403.

98. Shuman, J. D.; Vinson, C. R.; McKnight, S. L. *Science* **1990**, *249*, 771-774.

99. Talanian, R. V.; McKnight, C. J.; Kim, P. S. *Science* **1990**, *249*, 769-771.

100. Oakley, M. G.; Dervan, P. B. *Science* **1990**, *248*, 847-850.

101. Vinson, C. R.; Sigler, P. B.; McKnight, S. L. *Science* **1989**, *246*, 911-916.

102. Murre, C.; McCaw, P. S.; Baltimore, D. *Cell* **1989**, *56*, 777-783.

103. Murre, C.; McCaw, P. s.; Vaessin, H.; Caudy, M.; Jan, L. Y.; Jan, Y. N.; Cabrera, C. V.; Buskin, J. N.; Hauschka, S. D.; Lassar, A. B.; Weintraub, H.; Baltimore, D. *Cell* **1989**, *58*, 537-544.

104. Peterson, C. A.; Gordon, H.; Hall, Z. W.; Paterson, B. M.; Blau, H. M. *Cell* **1990**, *62*, 493-502.

105. Benezra, R.; Davis, R. L.; Lockshon, D.; Turner, D. L.; Weintraub, H. *Cell* **1990**, *61*, 49-59.

106. Blackwood, E. M.; Eisenman, R. N. *Science* **1991**, *251*, 1211-1217.

107. Klug, A.; Rhodes, D. *Trends in Biochemical Sciences* **1987**, *12*, 464-469.

108. Berg, J. M. *Journal of Biological Chemistry* **1990**, *265*, 6513-6516.

109. Neuhaus, D.; Rhodes, D. *Current Opinion in Structural Biology* **1991**, *1*, 268-270.

110. Parraga, G.; Horvath, S. J.; Eisen, A.; Taylor, W. E.; Hood, L.; Young, E. T.; Klevit, R. E. *Science* **1988**, *241*, 1489-1492.

111. Lee, M. S.; Gippert, G. P.; Soman, K. V.; Case, D. A.; Wright, P. E. *Science* **1989**, *245*, 635-637.

112. Omichinski, J. G.; Clore, G. M.; Appella, E.; Sakaguchi, K.; Gronenborn, A. M. *Biochemistry* **1990**, *29*, 9324-9334.

113. Berg, J. M. *Proceedings of the National Academy of Sciences of the United States of America* **1988**, *85*, 99-102.

114. Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809-817.

115. South, T. L.; Blake, P. R.; Sowder, R. C.; Arthur, L. O.; Henderson, L. E.; Summers, M. F. *Biochemistry* **1990**, *29*, 7786-7789.

116. Trainor, C. D.; Evans, T.; Felsenfeld, G.; Boguski, M. S. *Nature* **1990**, *343*, 92-96.

117. Krizek, B. A.; Amann, B. T.; Kilfoil, V. J.; Merkle, D. L.; Berg, J. M. *Journal of the American Chemical Society* **1991**, *113*, 4518-4523.

118. Merkle, D. L.; Schmidt, M. H.; Berg, J. M. *Journal of the American Chemical Society* **1991**, *113*, 5450-5451.

119. Berg, J. M. *Cell* **1989**, *57*, 1065-1068.

120. Rhodes, D.; Schwabe, J. W. R. *Nature* **1991**, *352*, 478-469.

121. Luisi, B. F.; Xu, W. X.; Otwinowski, Z.; Freedman, L. P.; Yamamoto, K. R.; Sigler, P. B. *Nature* **1991**, *352*, 497-505.

122. Frederick, C. A.; Grable, J.; Melia, M.; Samudzi, C.; Jen-Jacobson, L.; Wang, B.-C.; Greene, P.; Boyer, H. W.; Rosenberg, J. M. *Nature* **1984**, *309*, 327-331.

123. McClarin, J. A.; Frederick, C. A.; Wang, B.-C.; Greene, P.; Boyer, H. W.; Grable, J.; Rosenberg, J. M. *Science* **1986**, *234*, 1526-1541.

124. Phillips, S. E. V.; Manfield, I.; Parsons, I.; Davidson, B. E.; Rafferty, J. B.; Somers, W. S.; Margarita, D.; Cohen, G. N.; Saint-Girons, I.; Stockley, P. G. *Nature* **1989**, *341*, 711-715.

125. Rafferty, J. B.; Somers, W. S.; Saint-Girons, I.; Phillips, S. E. V. *Nature* **1989**, *341*, 705-710.

126. Knight, K. L.; Sauer, R. T. *Proceedings of the National Academy of Sciences of the United States of America* **1989**, *86*, 797-801.

127. Knight, K. L.; Sauer, R. T. *Journal of Biological Chemistry* **1989**, *264*, 13706-13710.

128. Vershon, A. K.; Kelley, R. D.; Sauer, R. T. *Journal of Biological Chemistry* **1989**, *264*, 3267-3273.

129. Zagorski, M. G.; Bowie, J. U.; Vershon, A. K.; Sauer, R. T.; Patel, D. J. *Biochemistry* **1989**, *28*, 9813-9825.

130. Breg, J. N.; van Opheusden, J. H. J.; Burgering, M. J. M.; Boelens, R.; Kaptein, R. *Nature* **1990**, *346*, 586-589.

131. Frampton, J.; Leutz, A.; Gibson, T. J.; Graf, T. *Nature* **1989**, *342*, 134.

132. Gabrielsen, O. S.; Sentenac, A.; Fromageot, P. *Science* **1991**, *253*, 1140-1143.

133. Weber, G. L.; Westin, E. H.; Clarke, M. F. *Science* **1990**, *249*, 1291-1293.

134. Furukawa, Y.; Piwnica-Worms, H.; Ernst, T. J.; Kanakura, Y.; Griffin, J. D. *Science* **1990**, *250*, 805-808.

135. Luscher, B.; Christenson, E.; Litchfield, D. W.; Krebs, E. G.; Eisenman, R. N. *Nature* **1990**, *344*, 517-522.

136. DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D. *Science* **1989**, *243*, 622-628.

137. Hill, C. P.; Anderson, D. H.; Wesson, L.; DeGrado, W. F.; Eisenberg, D. *Science* **1990**, *249*, 543-546.

138. O'Neil, K. T.; Hoess, R. H.; DeGrado, W. F. *Science* **1990**, *249*, 774-778.

139. Anderson, J.; Ptashne, M.; Harrison, S. C. *Proceedings of the National Academy of Sciences of the United States of America* **1984**, *81*, 1307-1311.

140. Abdel-Meguid, S. S.; Grindley, N. D. F.; Templeton, N. S.; Steitz, T. A. *Proceedings of the National Academy of Sciences of the United States of America* **1984**, *81*, 2001-2005.

141. Graham, K.; Dervan, P. *Journal of Biological Chemistry* **1990**, *27*, 16534-16540.

142. Ogata, R.; Gilbert, W. *Journal of Molecular Biology* **1979**, *132*, 709-728.

143. Kaptein, R.; Boelens, R.; Scheek, R. M.; van Gunsteren, W. F. *Biochemistry* **1988**, *27*, 5389-5395.

144. Kaptein, R.; Lamerichs, R.; Boelens, R.; Rullmann, J. A. *Biochem. Pharm.* **1990**, *40*, 89-96.

145. Zieg, J.; Silverman, M.; Hilmen, M. I.; Simon, M. *Science* **1977**, *196*, 170-172.

146. Johnson, R. C.; Bruist, M. F.; Glaccum, M. B.; Simon, M. I. *Cold Spring Harb. Symp. Quant. Biol.* **1984**, *49*, 751-760.

147. Johnson, R. C.; Simon, M. I. *Cell* **1985**, *41*, 781-791.

148. Simon, M. I.; Zieg, J.; Silverman, M.; Mandel, G.; Doolittle, R. *Science* **1980**, *209*, 1370-1374.

149. Johnson, R. C.; Bruist, M. F.; Simon, M. I. *Cell* **1986**, *46*, 531-539.

150. Bruist, M. F.; Glasgow, A. C.; Johnson, R. C.; Simon, M. I. *Gene. Dev.* **1987**, *1*, 762-772.

151. Johnson, R. C.; Glasgow, A. C.; Simon, M. I. *Nature* **1987**, *329*, 462-465.

152. Johnson, R. C.; Simon, M. I. *Trend. Genet.* **1987**, *3*, 262-267.

153. Sluka, J. P. Ph. D. Thesis, California Institute of Technology, 1988.

154. Galas, D. J.; Schmitz, A. *Nucl. Acid. Res.* **1978**, *5*, 3157.

155. Tullius, T. D.; Dombroski, B. A.; Churchill, M. E. A.; Kam, L. *Meth. Enzym.* **1987**, *155*, 537.

156. Hertzberg, R. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 313.

157. Hertzberg, R. P.; Dervan, P. B. *Biochemistry* **1984**, *23*, 3934.

158. Dervan, P. B. *Science* **1986**, *232*, 464-471.

159. Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861.

160. Taylor, J.; Schultz, P.; Dervan, P. *Tetrahedron* **1984**, *40*, 457-465.

161. Tullius, T. D.; Dombroski, B. A. *Science* **1985**, *230*, 679-681.

162. Youngquist, R. S.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 2565.

163. Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574.

164. Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 6840.

165. Moser, H.; Dervan, P. B. *Science* **1987**, *238*, 645.

166. Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 976.

167. Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059.

168. Strobel, S. A.; Moser, H. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1988**, *110*, 7927.

169. Sluka, J.; Horvath, s.; Bruist, M.; Simon, M.; Dervan, P. *Science* **1987**, *238*, 1129-1132.

170. Mack, D.; Sluka, J.; Shin, J.; Griffin, J.; Simon, M.; Dervan, P. *Biochemistry* **1990**, *29*, 6561-6567.

171. Chen, C. B.; Sigman, D. S. *Science* **1987**, *237*, 1197.

172. Sluka, J.; Griffin, J.; Mack, D.; Dervan, P. *Journal of the American Chemical Society* **1990**, *112*, 6369-6374.

173. Merrifield, B. *Science* **1986**, *232*, 341-347.

174. Stewart, J. M.; Young, J. D. *Solid Phase Peptide Synthesis, 2nd Edition*; Pierce Chemical Co.: Rockford, IL, 1984.

175. Kent, S. *Annual Review of Biochemistry* **1988**, *57*, 957.

176. Clark-Lewis, I.; Aebersold, R.; Ziltener, H.; Schrader, J.; Hood, L.; Kent, S. *Science* **1986**, *213*, 134-139.

177. Maxam, A.; Gilbert, W. *Methods in Enzymology* **1980**, *65*, 499-560.

178. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Meth. Enzym.* **1986**, *130*, 132-181.

179. Goodisman, J.; Dabrowiak, J. C. *J. Biomol. Struct. Dyn.* **1985**, *2*, 967-979.

180. Ackers, G. K.; Johnson, A. D.; Shea, M. A. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 1129-1133.

181. Ackers, G. K.; Shea, M. A.; Smith, F. R. *J. Mol. Biol.* **1983**, *170*, 223-242.

182. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. F. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462-8466.

183. Wade, W. S.; Dervan, P. B. unpublished results.

184. Singleton, S. F.; Dervan, P. B. unpublished results.

185. Johnston, R. F.; Pickett, S. C.; Barker, D. L. *Electrophoresis* **1990**, *11*, 355-360.

186. Marquardt, D. W. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431-441.

187. Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: New York, 1969.

188. Fried, M.; Crothers, D. M. *Nucleic Acids Research* **1981**, *9*, 6505-6524.

189. Bruist, M. F.; Horvath, S. J.; Hood, L. E.; Steitz, T. A.; Simon, M. I. *Science* **1987**, *235*, 777.

190. Matthews, B. W.; Ohlendorf, D. H.; Anderson, W. F.; Fisher, R. G.; Takeda, Y. *Trends in Biochemical Sciences* **1983**, 25-29.

191. Zuiderweg, E. R. P.; Kaptein, R.; Wüthrich, K. *European Journal of Biochemistry* **1983**, *137*, 279-292.

192. Stob, S.; Scheek, R. M.; Boelens, R.; Kaptein, R. *FEBS Letters* **1988**, *239*, 99-104.

193. Sadler, J. R.; Sasmor, H.; Betz, J. L. *Proceedings of the National Academy of Sciences of the United States of America* **1983**, *80*, 6785-6789.

194. Simons, A.; Tils, D.; von Wilcken-Bergmann, B.; Müller-Hill, B. *Proceedings of the National Academy of Sciences of the United States of America* **1984**, *81*, 1624-1628.

195. Zuiderweg, E.; Billeter, M.; Boelens, R.; Scheek, R.; Wüthrich, K.; Kaptein, R. *FEBS Letters* **1984**, *174*, 243-247.

196. Takeda, Y.; Ohlendorf, D.; Anderson, W.; Matthews, B. *Science* **1983**, *221*, 1020-1026.

197. Ebright, R. H. *Proceedings of the National Academy of Sciences of the United States of America* **1986**, *83*, 303-307.

198. Lamerichs, R.; Boelens, R.; van der Marel, G.; van Boom, J.; Kaptein, R. *European Journal of Biochemistry* **1990**, *194*, 629-637.

199. Allen, T.; Wick, K.; Matthews, K. *Journal of Biological Chemistry* **1991**, *266*, 6113-6119.

200. Ebright, R. H. *Journal of Biomolecular Structure and Dynamics* **1985**, *3*, 281-297.

201. Ebright, R. *Methods in Enzymology* **1991**, in press.

202. Dervan, P. B. *Methods in Enzymology* **1991**, in press.

203. Betz, J. L.; Sasmor, H. M.; Buck, F.; Insley, M. Y.; Caruthers, M. H. *Gene* **1986**, *50*, 123-132.

204. Yanisch-Perron, C.; Viera, J.; Messing, J. *Gene* **1985**, *33*, 103-119.

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

206. Walldorf, U.; Fleig, R.; Gehring, W. J. *Proceedings of the National Academy of Sciences of the United States of America* **1989**, *86*, 9971-9975.

207. Scott, M. P.; Weiner, A. J. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 4115-4119.

208. McGinnis, W.; Levine, M. S.; Hafen, E.; Kuroiwa, A.; Gehring, W. J. *Nature* **1984**, *308*, 428-433.

209. Laughon, A.; Scott, M. P. *Nature* **1984**, *310*, 25-31.

210. Carrasco, A. E.; McGinnis, W.; Gehring, W. J.; De Robertis, E. M. *Cell* **1984**, *37*, 409-414.

211. Shepherd, J. C. W.; McGinnis, W.; Carrasco, A. E.; De Robertis, E. M.; Gehring, W. J. *Nature* **1984**, *310*, 70-71.

212. Affolter, M.; Percival-Smith, A.; Müller, M.; Leupin, W.; Gehring, W. J. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87*, 4093-4097.

213. Mack, D. P. Ph. D. Thesis, California Institute of Technology, 1991.

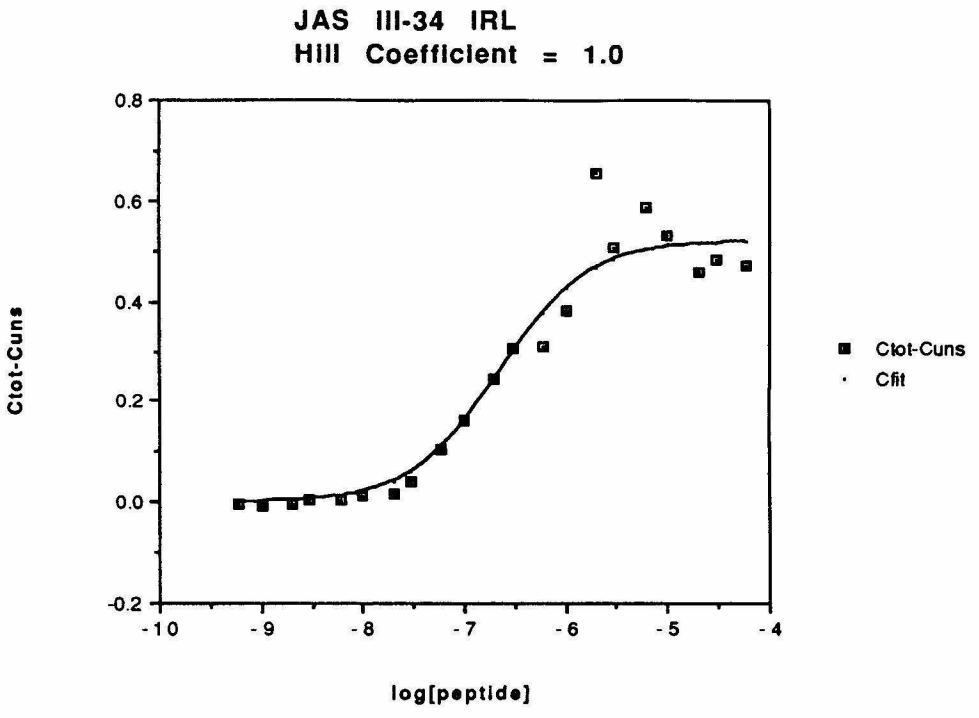
214. Glasgow, A. C.; Bruist, M. F.; Simon, M. I. *J. Biol. Chem.* **1989**, *264*, 10072-10082.

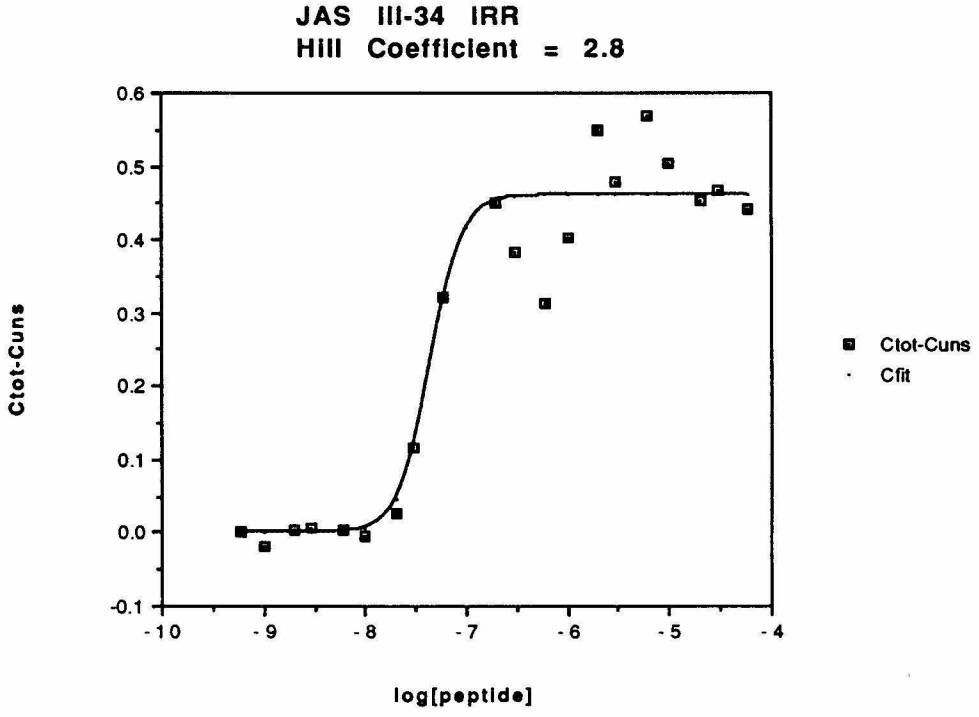
215. Wolberger, C.; Vershon, A. K.; Liu, B.; Johnson, A. D.; Pabo, C. O. *Cell* **1991**, *67*, 517-528.

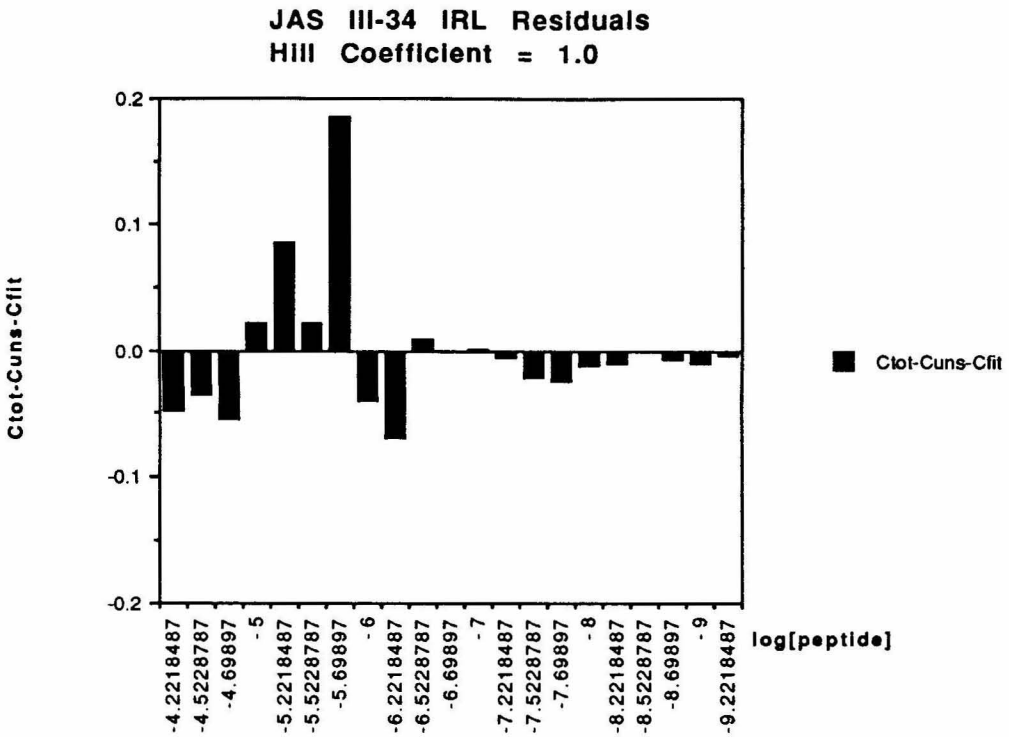
216. Gait, M. J. *Oligonucleotide Synthesis*; IRC Press: Oxford, 1984.

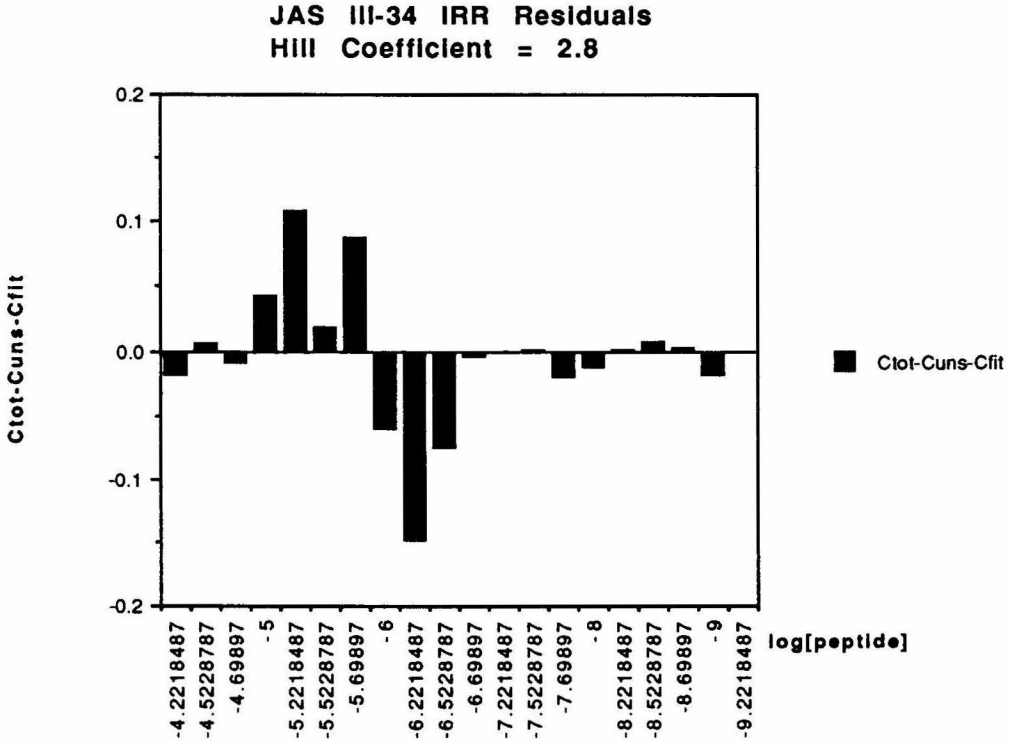
APPENDIX

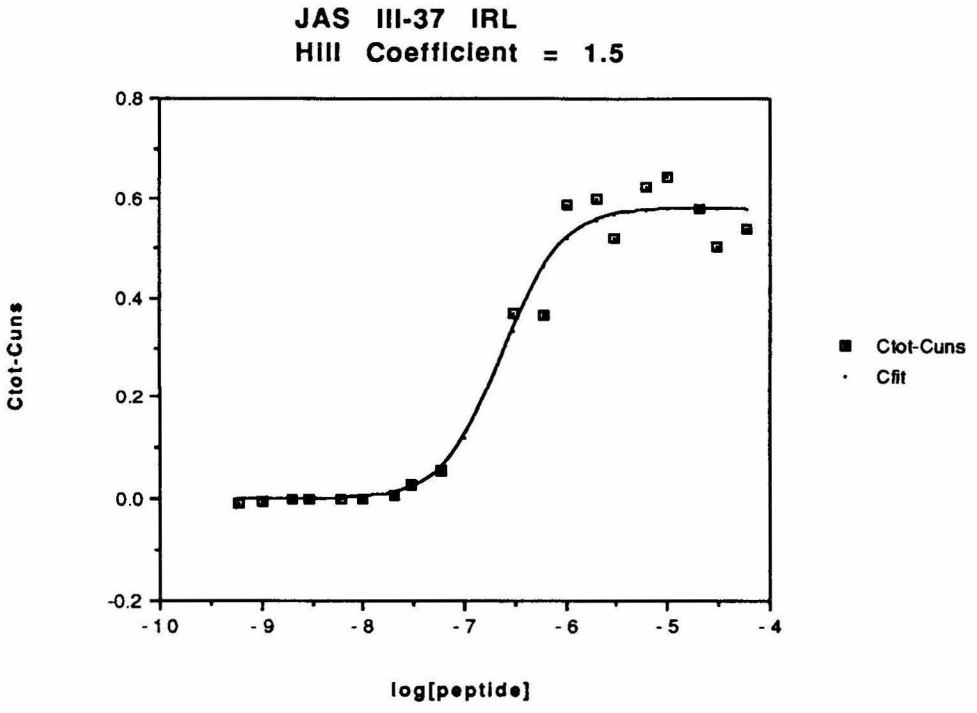
[Fe•EDTA]Hin(139-190). On the following pages are shown the binding isotherms and residuals for gels JAS III-34 and 37. On top are data for the *hix*L IRL binding site, and on the bottom are data for the *hix*L IRR binding site. Residuals measure the difference between the obtained data points and the fit curve.

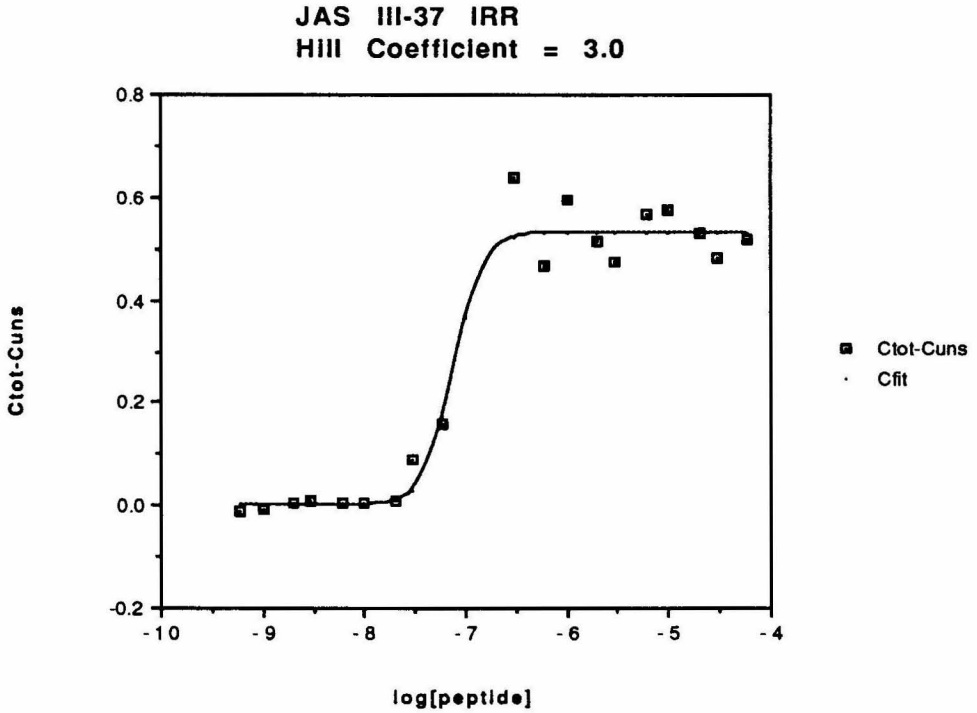

JAS III-34 IRL
Hill Coefficient = 1.0
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-10
-9
-8
-7
-6
-5
-4
log[peptide]
Ctot-Cuns
Cfit

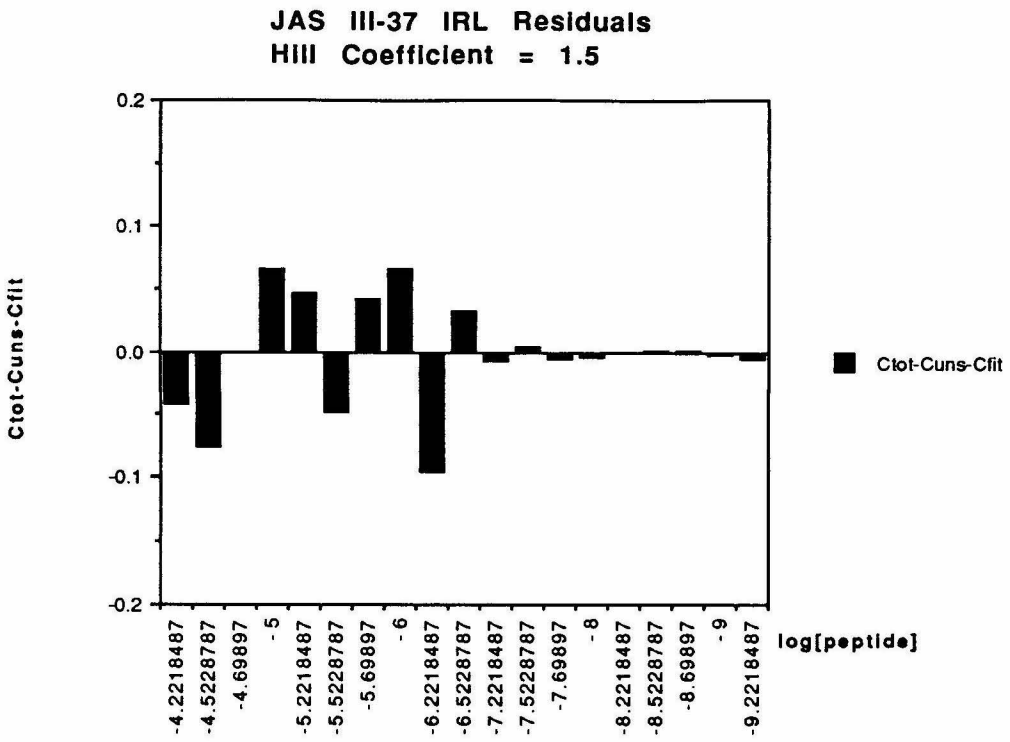

JAS III-34 IRR
Hill Coefficient = 2.8
Ctot-Cuns
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-10
-9
-8
-7
-6
-5
-4
log[peptide]
Ctot-Cuns
Cfit

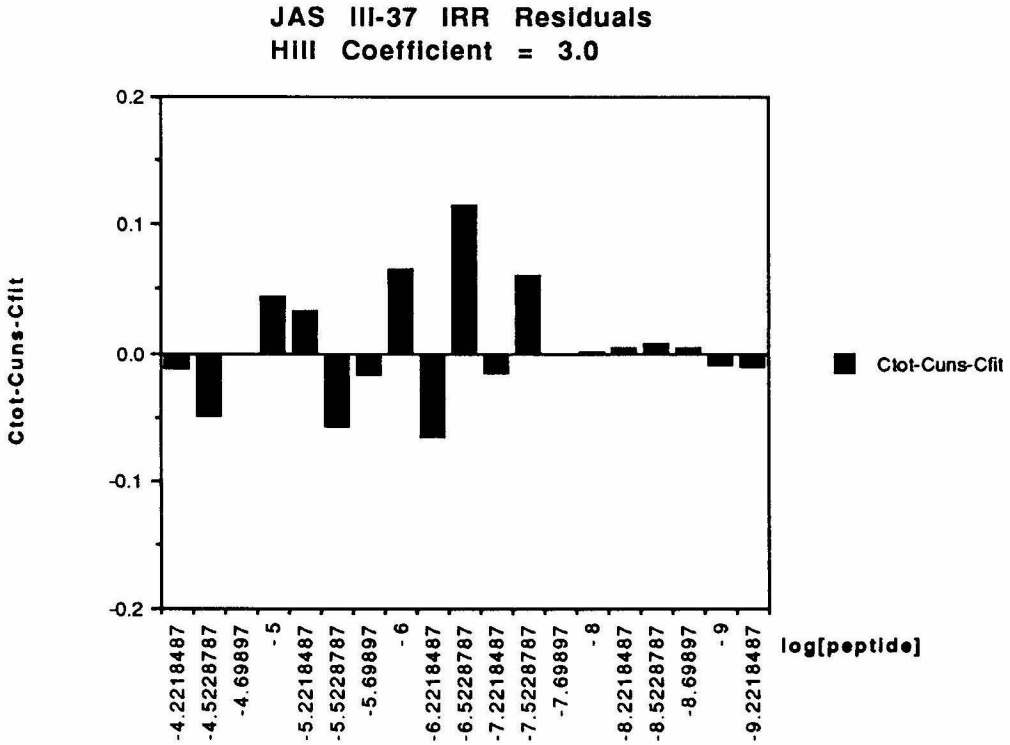

JAS III-34 IRL Residuals
Hill Coefficient = 1.0
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
-9
-9.2218487
log[peptide]
Ctot-Cuns-Cfit


JAS III-34 IRR Residuals
Hill Coefficient = 2.8
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
-9
-9.2218487
log[peptide]
Ctot-Cuns-Cfit


JAS III-37 IRL
Hill Coefficient = 1.5
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-10
-9
-8
-7
-6
-5
-4
log[peptide]
Ctot-Cuns
Cfit


JAS III-37 IRR
Hill Coefficient = 3.0
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-10
-9
-8
-7
-6
-5
-4
log[peptide]
Ctot-Cuns
Cfit

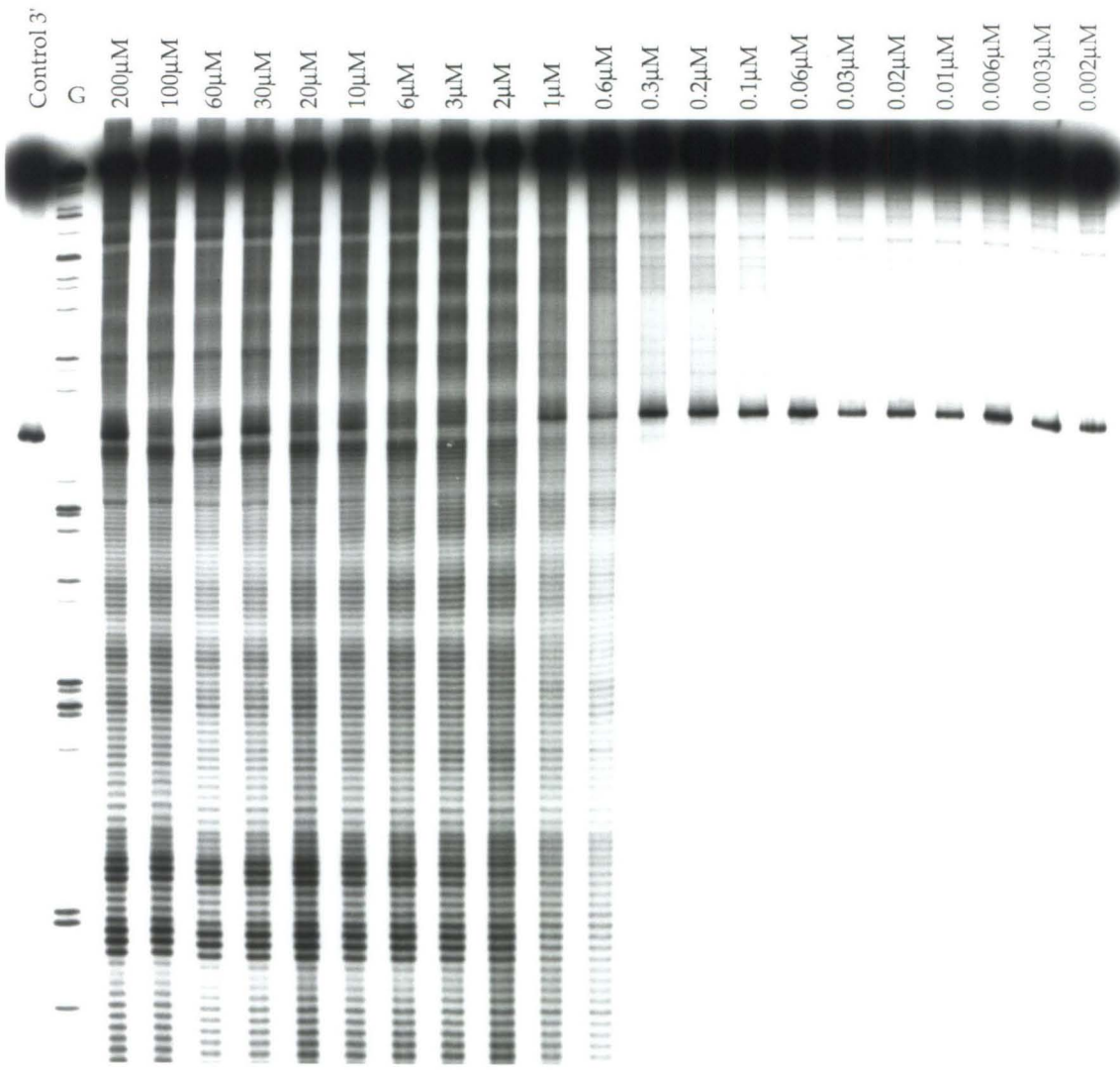

JAS III-37 IRL Residuals
Hill Coefficient = 1.5
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
-9
-9.2218487
log[peptide]
Ctot-Cuns-Cfit

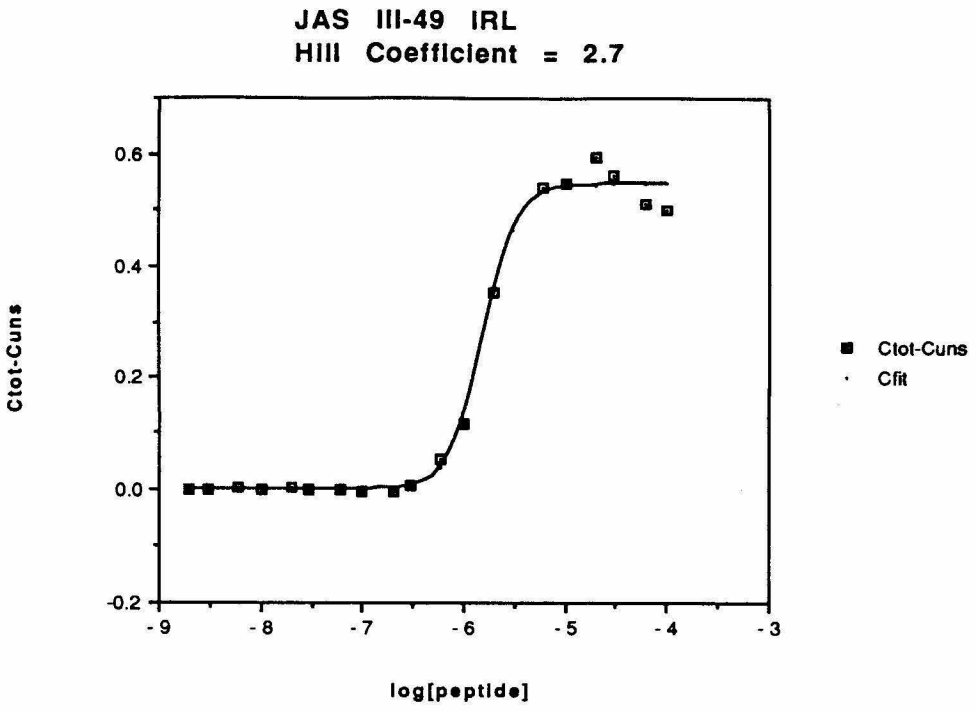

JAS III-37 IRR Residuals
Hill Coefficient = 3.0
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
-9
-9.2218487
log[peptide]
Ctot-Cuns-Cfit

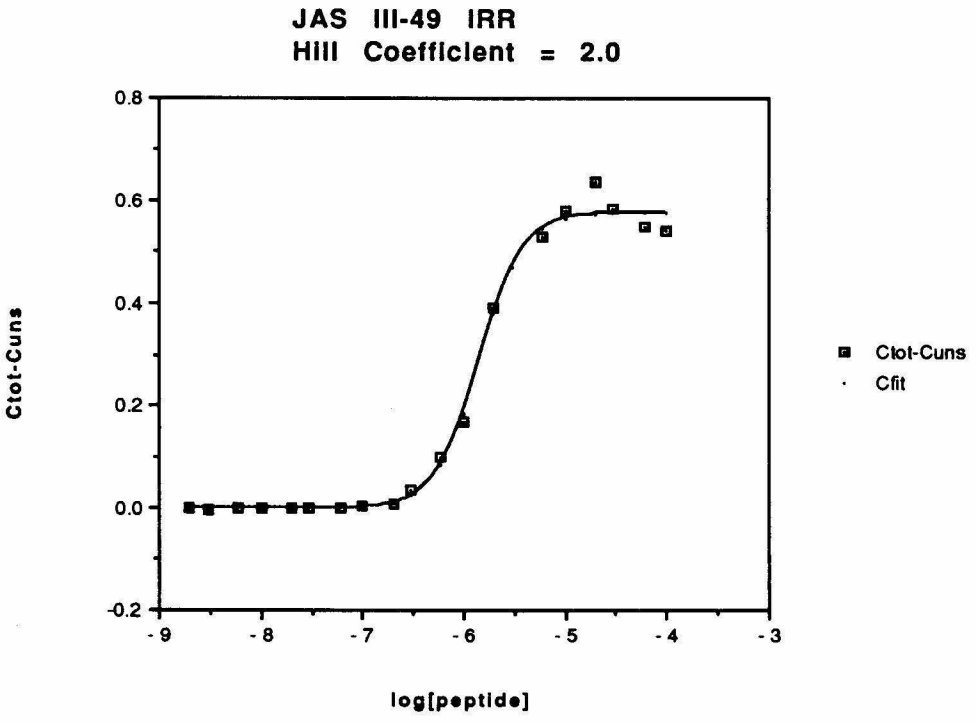
[Fe•EDTA]Hin(139-184). Autoradiogram of a quantitative affinity cleaving gel of [Fe•EDTA]Hin(139-184) bound to the 3′ end-labelled Xba I-Spe I fragment (218 base pairs) from pMFB36. This autoradiogram is from the gel labelled JAS III-51 (see JAS notebook III). Lane 1, control; lane 2, G sequencing reaction. Lanes 3-23 contain decreasing concentrations of protein as indicated. Each reaction contains 20mM phosphate buffer, pH 7.5, 20mM NaCl, and 1mM dithiothreitol. Each reaction proceeds at room temperature for 12 minutes. Reactions are stopped and extracted with 200μL of a 2:1 solution of phenol:chloroform; followed by butanol extraction and ethanol precipitation. Reactions are taken up in formamide loading buffer and loaded onto an 8% polyacrylamide denaturing gel.

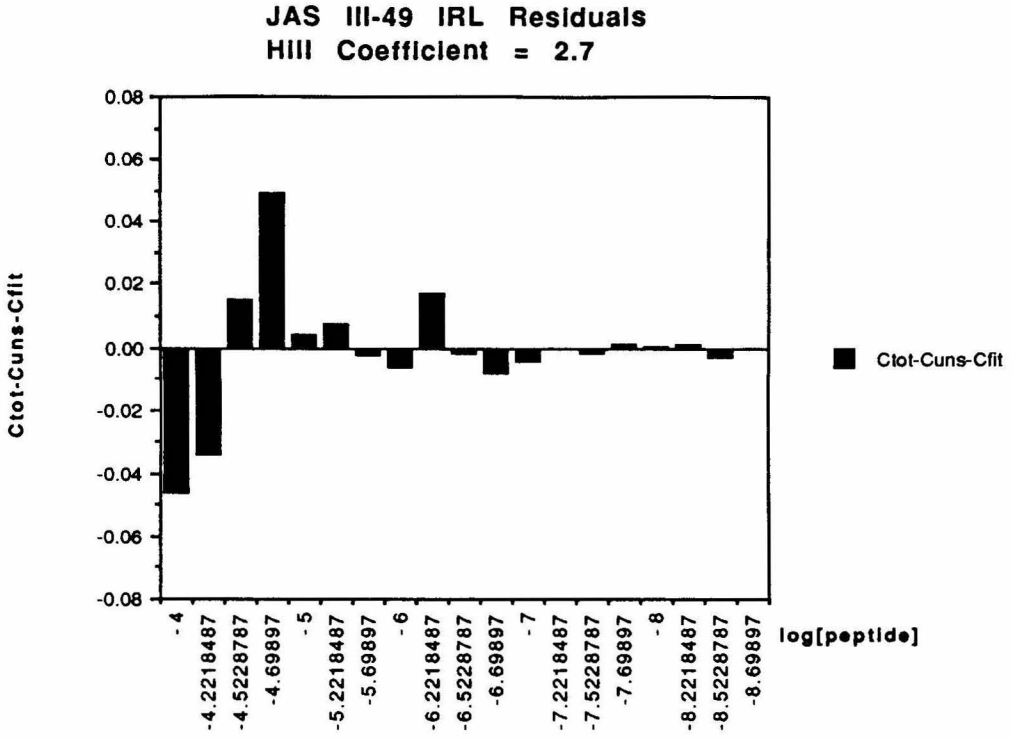
On the following pages are shown the binding isotherms and residuals for gels JAS III-49, 50, 51, and 52. On top are data for the *hix*L IRL binding site, and on the bottom are data for the *hix*L IRR binding site. Residuals measure the difference between the obtained data points and the fit curve.

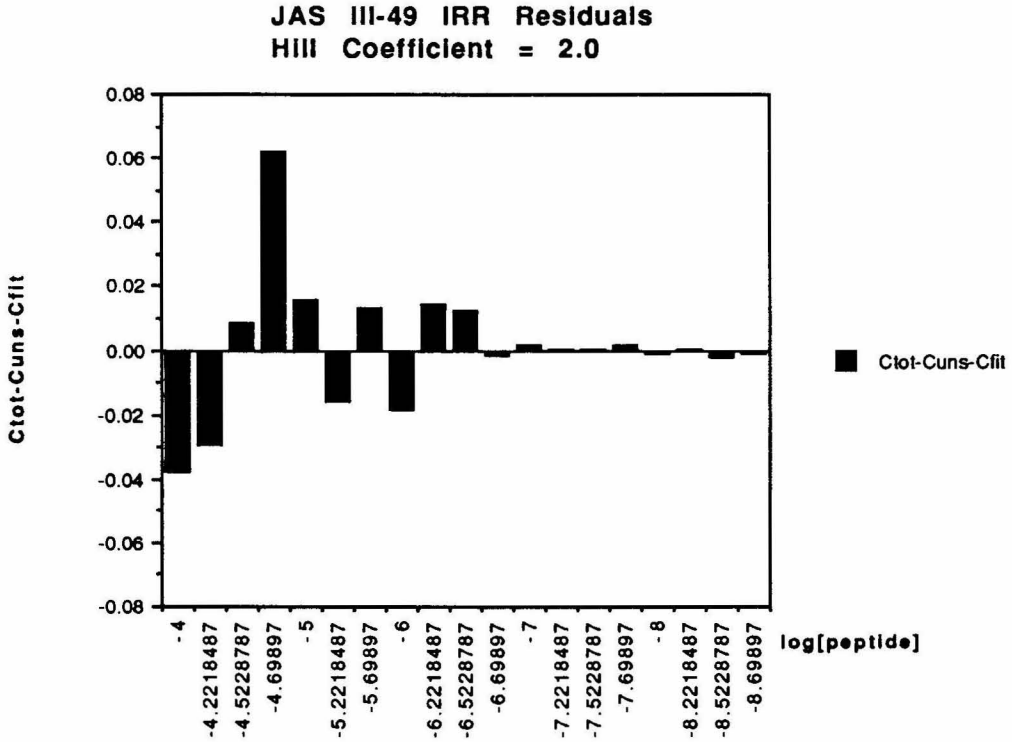
Quantitative Affinity Cleaving
[Fe•EDTA] Hin(139-184)

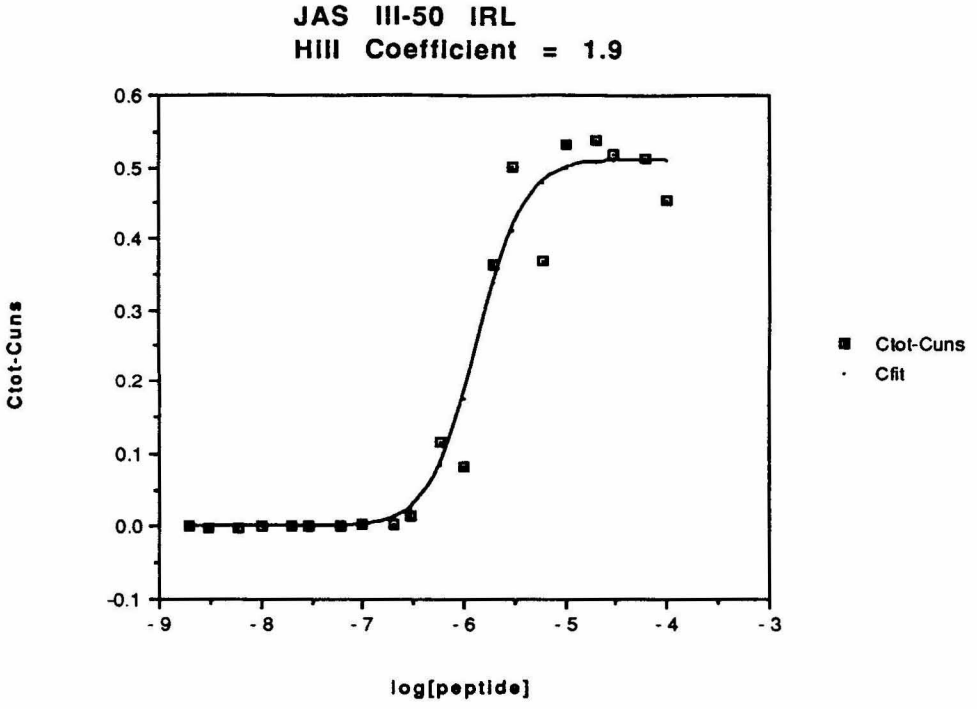


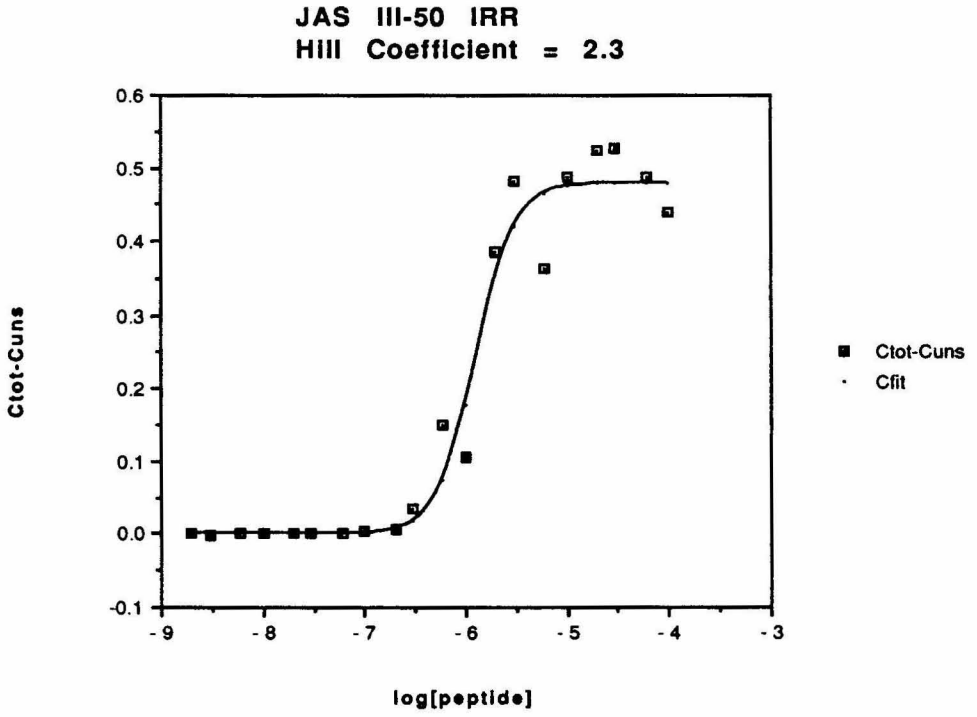

JAS III-49 IRL
HIII Coefficient = 2.7
Ctot-Cuns
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

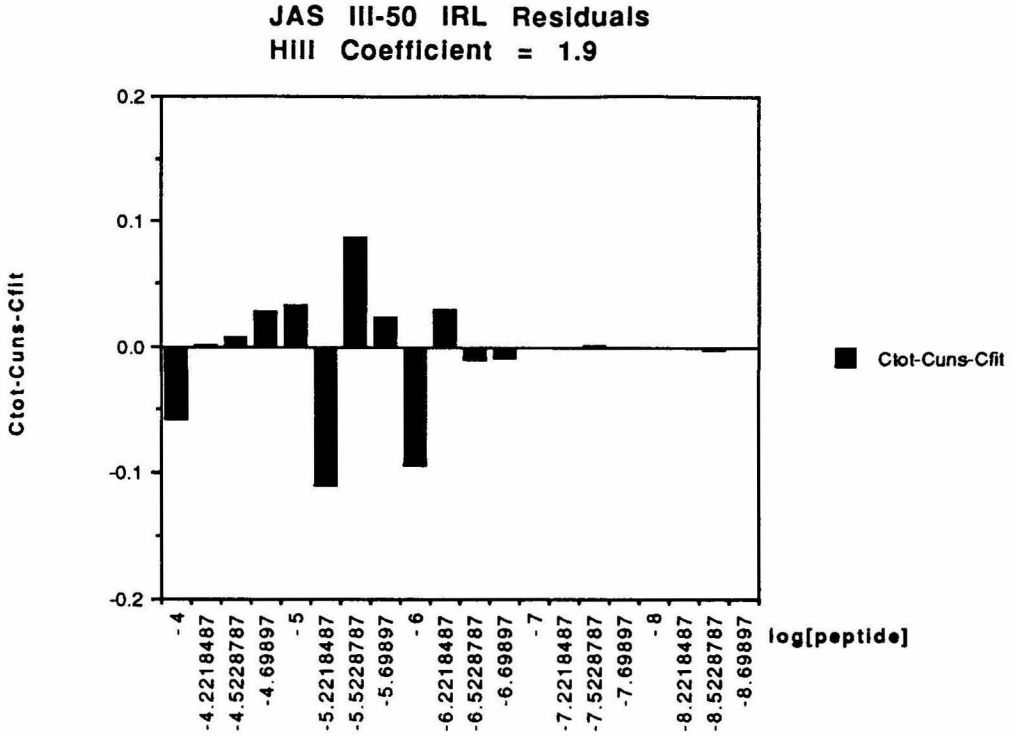

JAS III-49 IRR
HIII Coefficient = 2.0
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

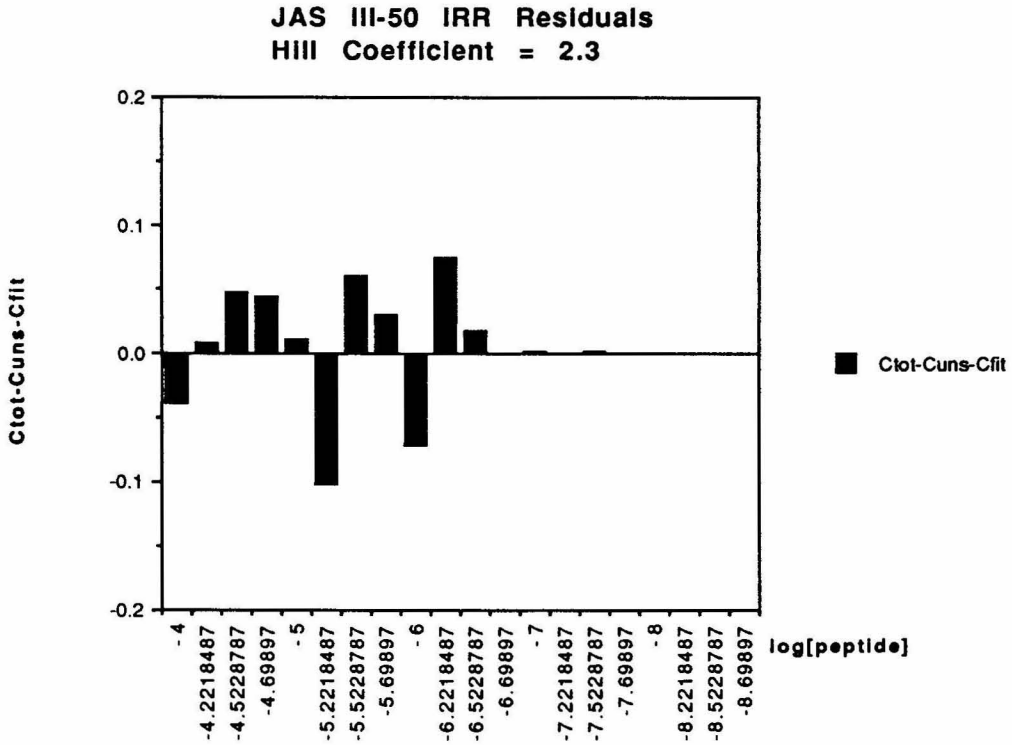

JAS III-49 IRL Residuals
HIII Coefficient = 2.7
Ctot-Cuns-Cfit
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
-0.08
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

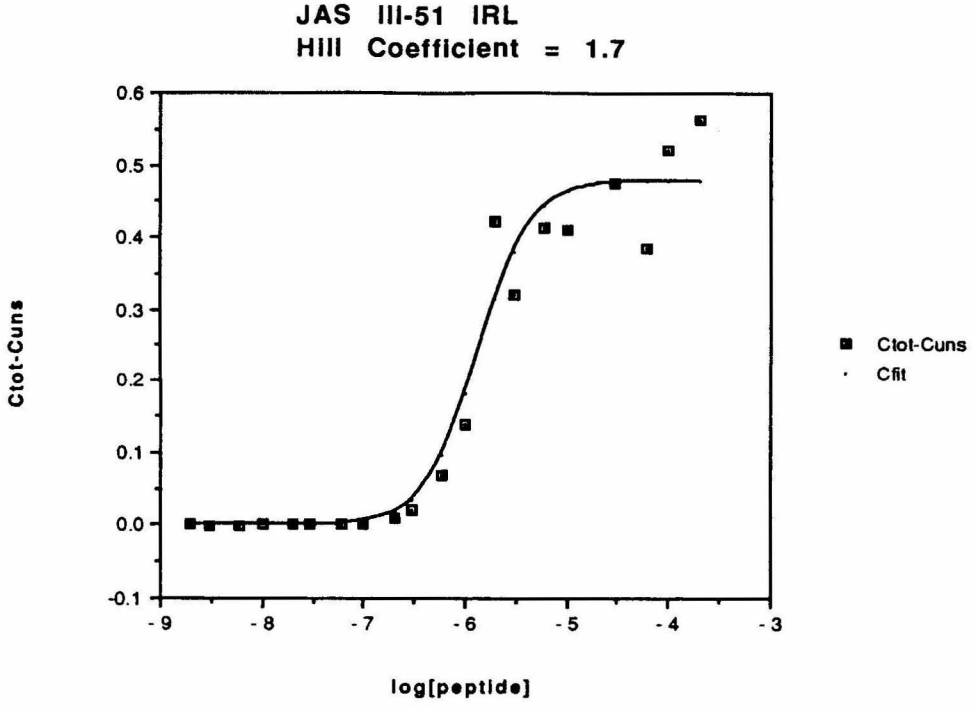

JAS III-49 IRR Residuals
HIII Coefficient = 2.0
Ctot-Cuns-Cfit
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
-0.08
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

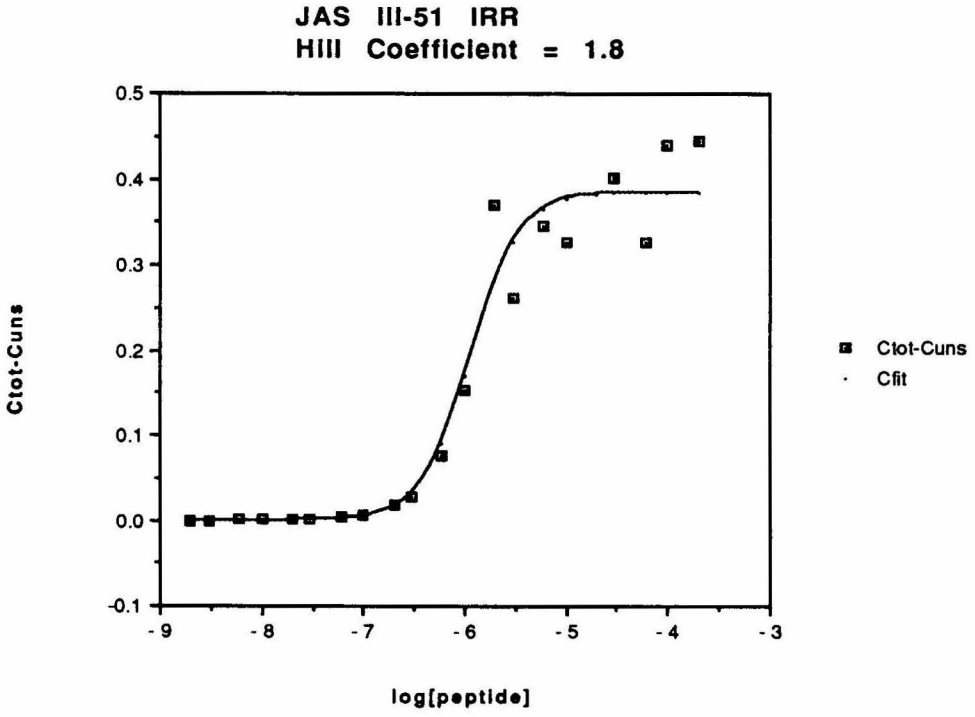

JAS III-50 IRL
Hill Coefficient = 1.9
Ctot-Cuns
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

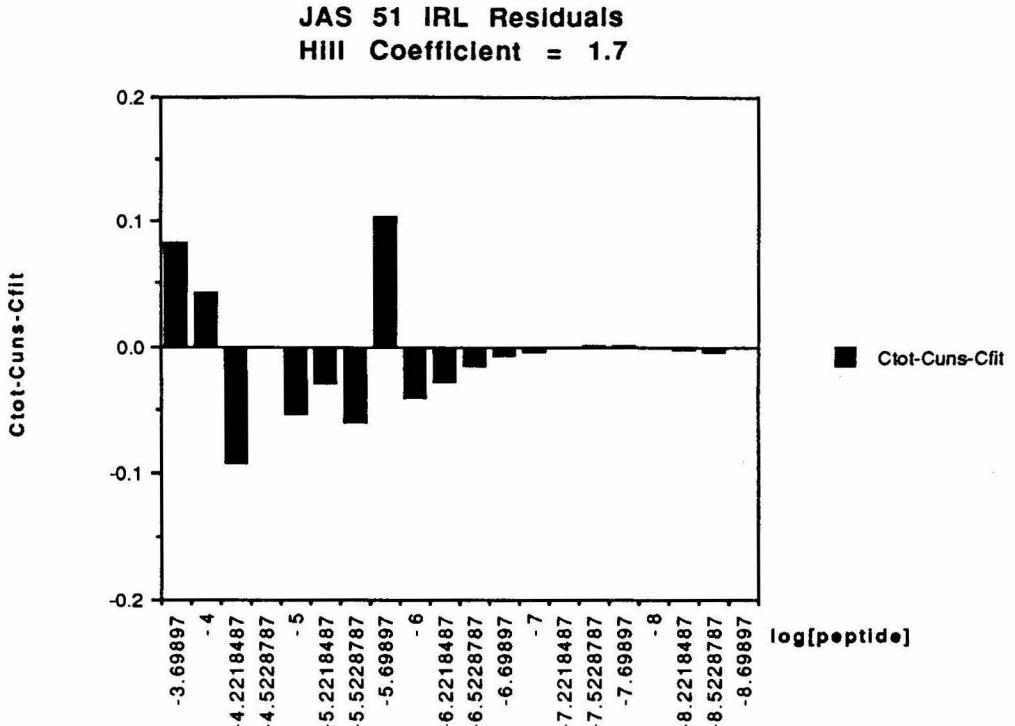

JAS III-50 IRR
Hill Coefficient = 2.3
Ctot-Cuns
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

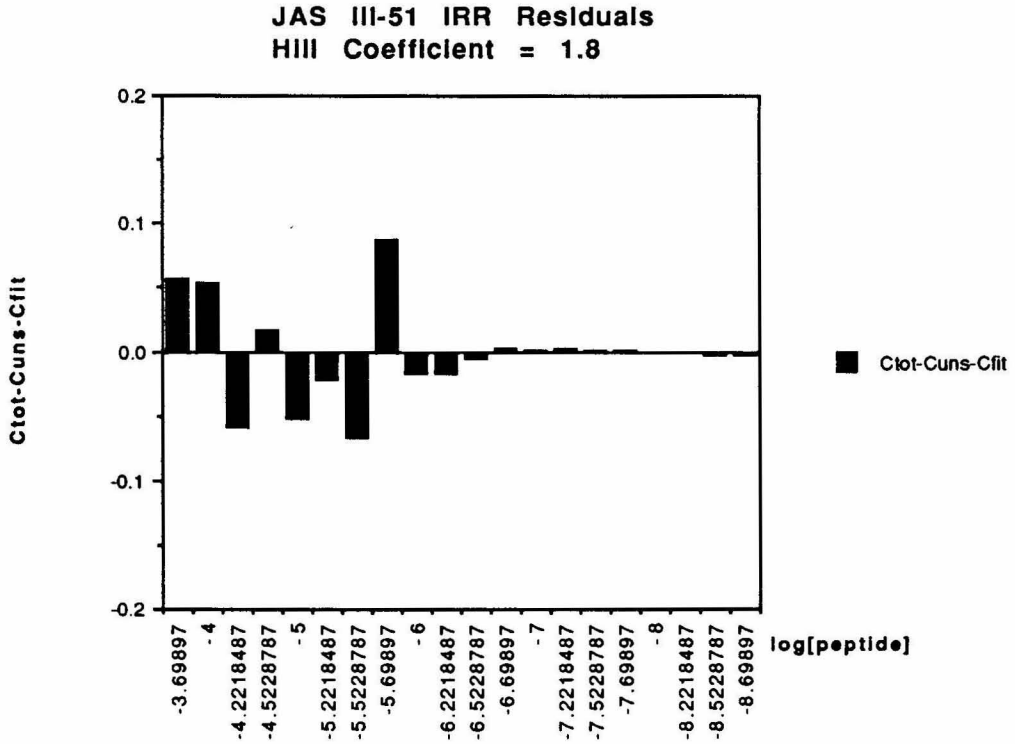

JAS III-50 IRL Residuals
HIII Coefficient = 1.9
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

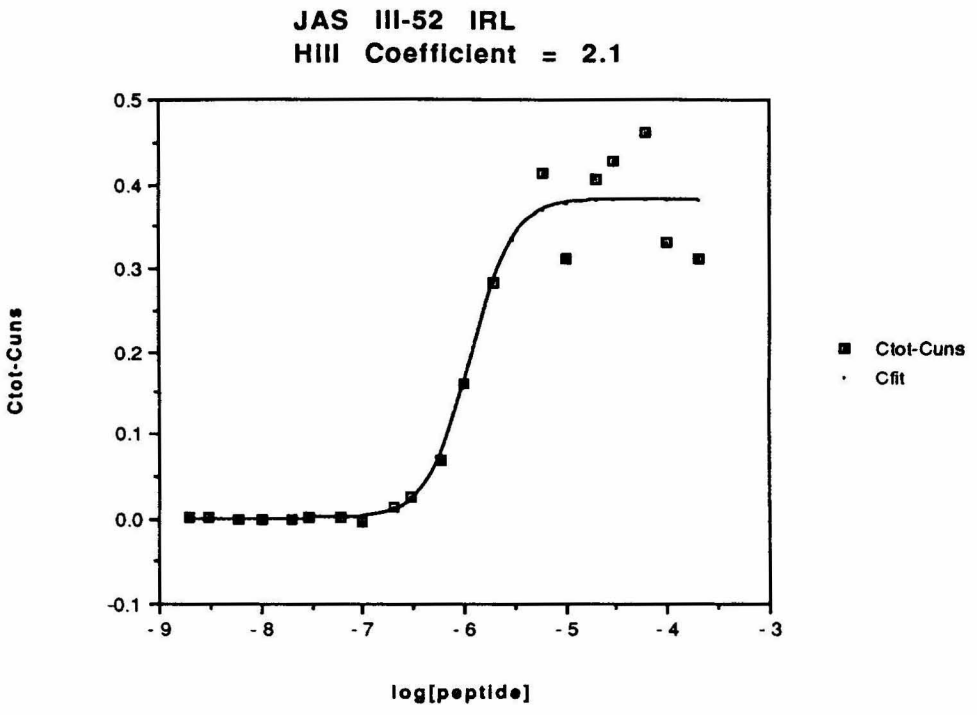

JAS III-50 IRR Residuals
HIII Coefficient = 2.3
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

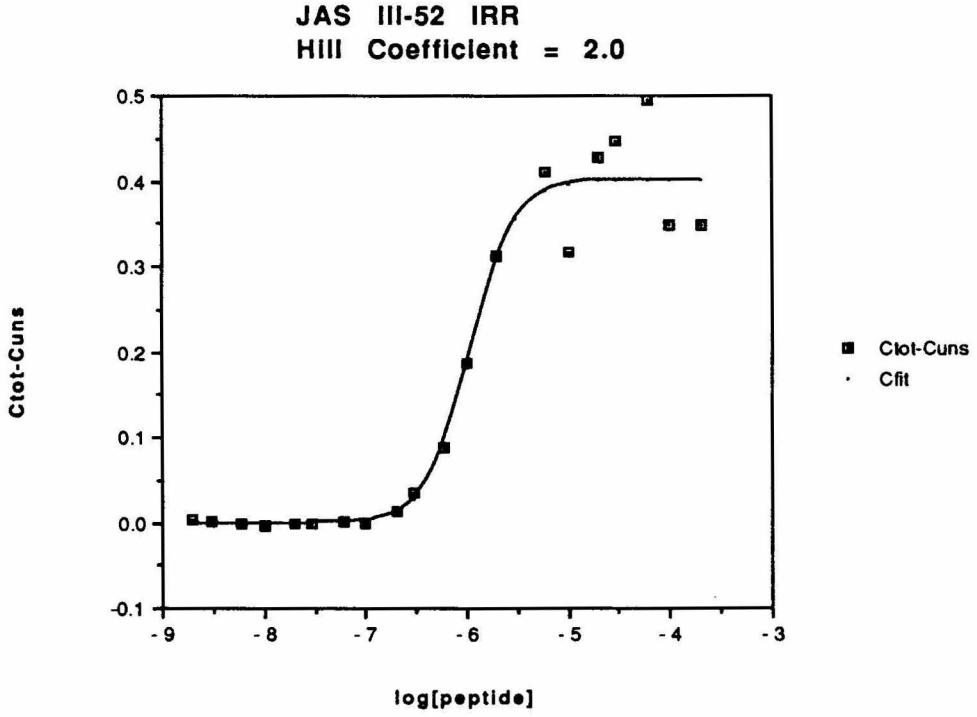

JAS III-51 IRL
HIII Coefficient = 1.7
Ctot-Cuns
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

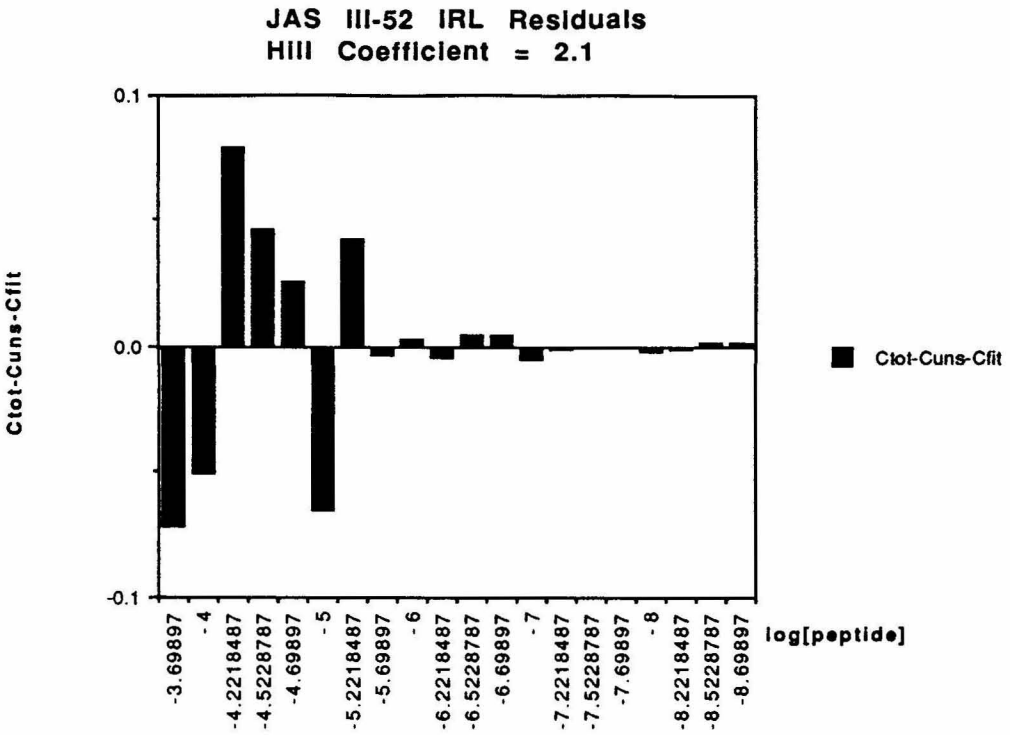

JAS III-51 IRR
HIII Coefficient = 1.8
Ctot-Cuns
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

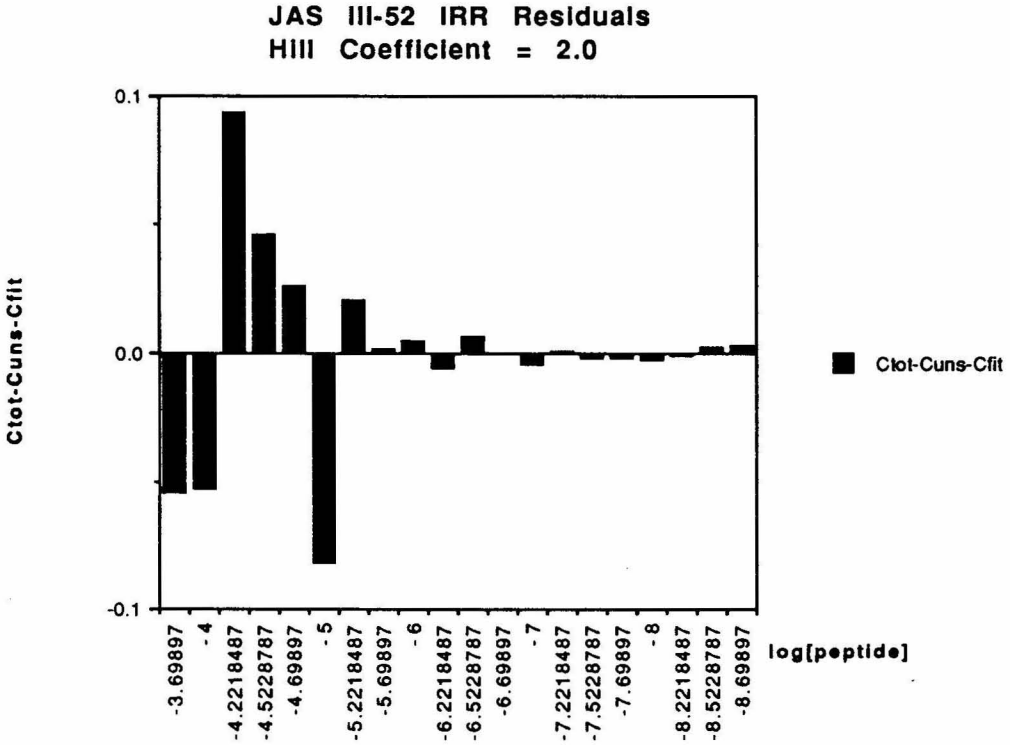

JAS 51 IRL Residuals
Hill Coefficient = 1.7
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-3.69897
-4
-4.2218487
-4.5228787
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit


JAS III-51 IRR Residuals
Hill Coefficient = 1.8
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-3.69897
-4
-4.2218487
-4.5228787
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit


JAS III-52 IRL
HIII Coefficient = 2.1
Ctot-Cuns
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit


JAS III-52 IRR
HIII Coefficient = 2.0
Ctot-Cuns
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

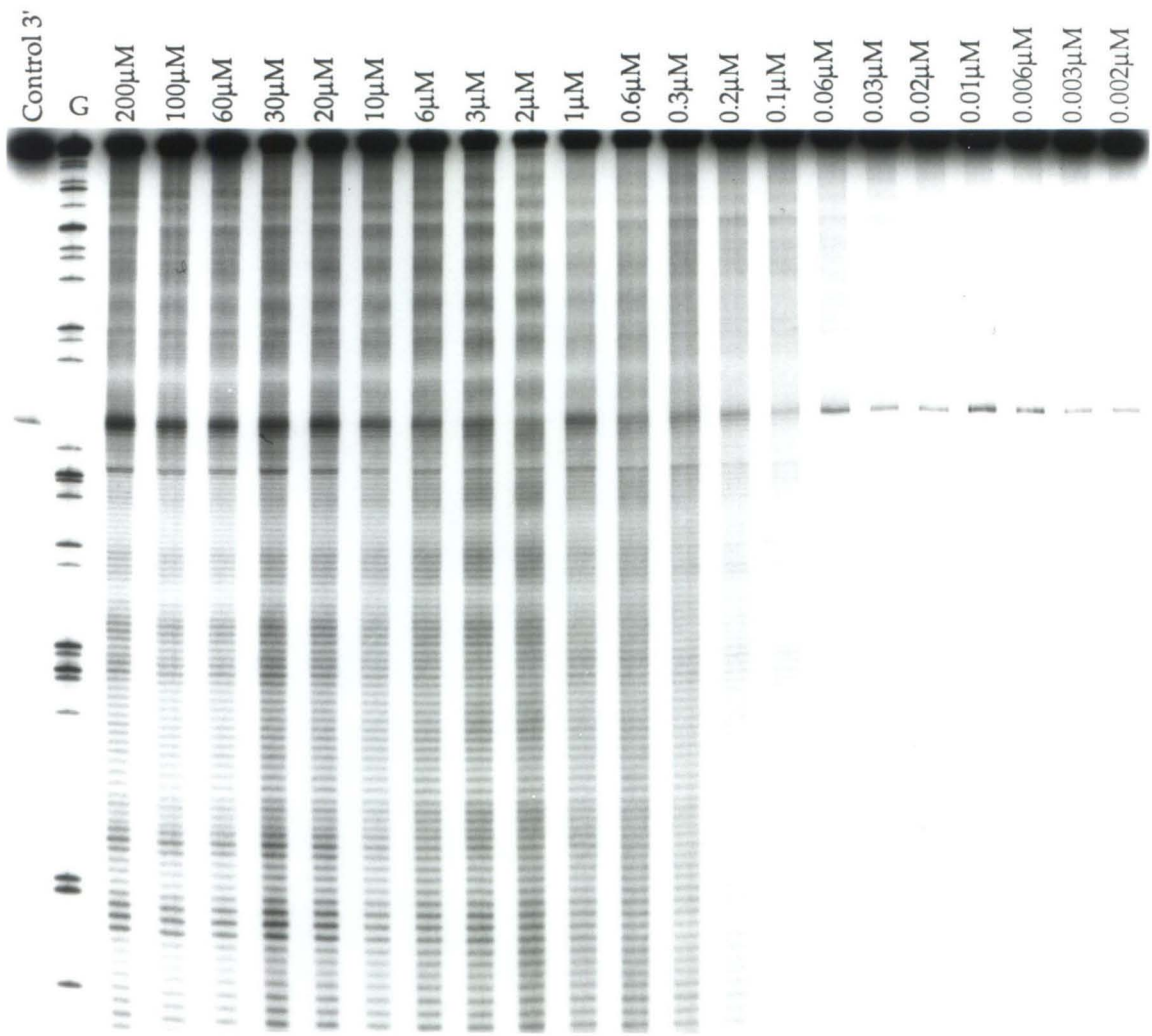

JAS III-52 IRL Residuals
Hill Coefficient = 2.1
Ctot-Cuns-Cfit
0.1
0.0
-0.1
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

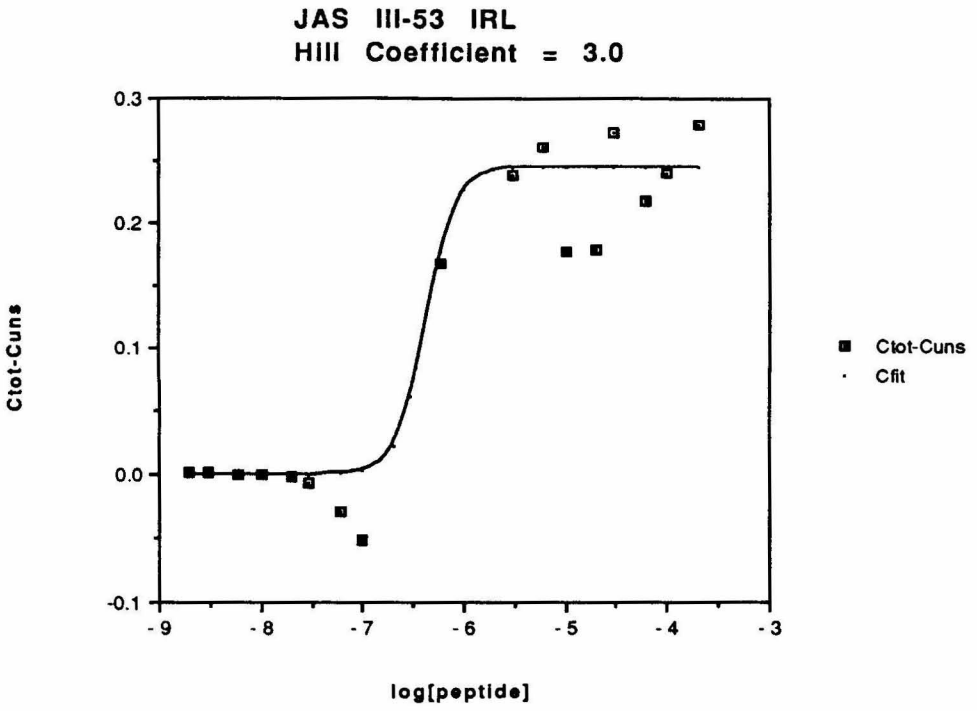

JAS III-52 IRR Residuals
Hill Coefficient = 2.0
Ctot-Cuns-Cfit
0.1
0.0
-0.1
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

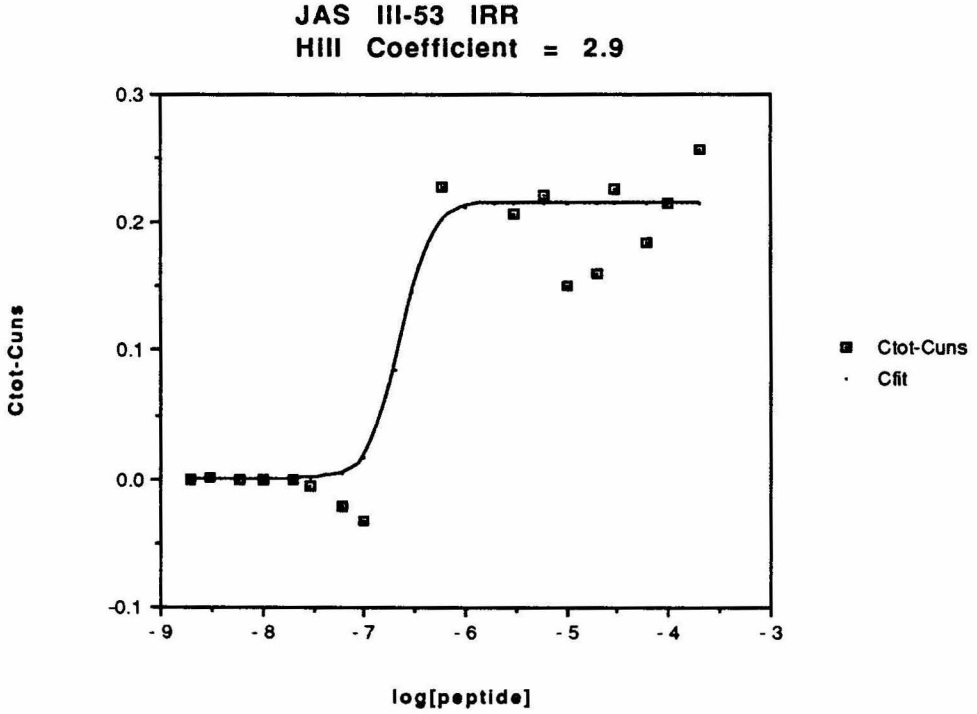
[Fe•EDTA]Hin(139-190)R140→K. Autoradiogram of a quantitative affinity cleaving gel of [Fe•EDTA]Hin(139-190)R140→K bound to the 3′ end-labelled Xba I-Spe I fragment (218 base pairs) from pMFB36. This autoradiogram is from the gel labelled JAS III-56 (see JAS notebook III). Lane 1, control; lane 2, G sequencing reaction. Lanes 3-23 contain decreasing concentrations of protein as indicated. Each reaction contains 20mM phosphate buffer, pH 7.5, 20mM NaCl, and 1mM dithiothreitol. Each reaction proceeds at room temperature for 30 minutes. Reactions are stopped and extracted with 200μL of a 2:1 solution of phenol:chloroform; followed by butanol extraction and ethanol precipitation. Reactions are taken up in formamide loading buffer and loaded onto an 8% polyacrylamide denaturing gel.

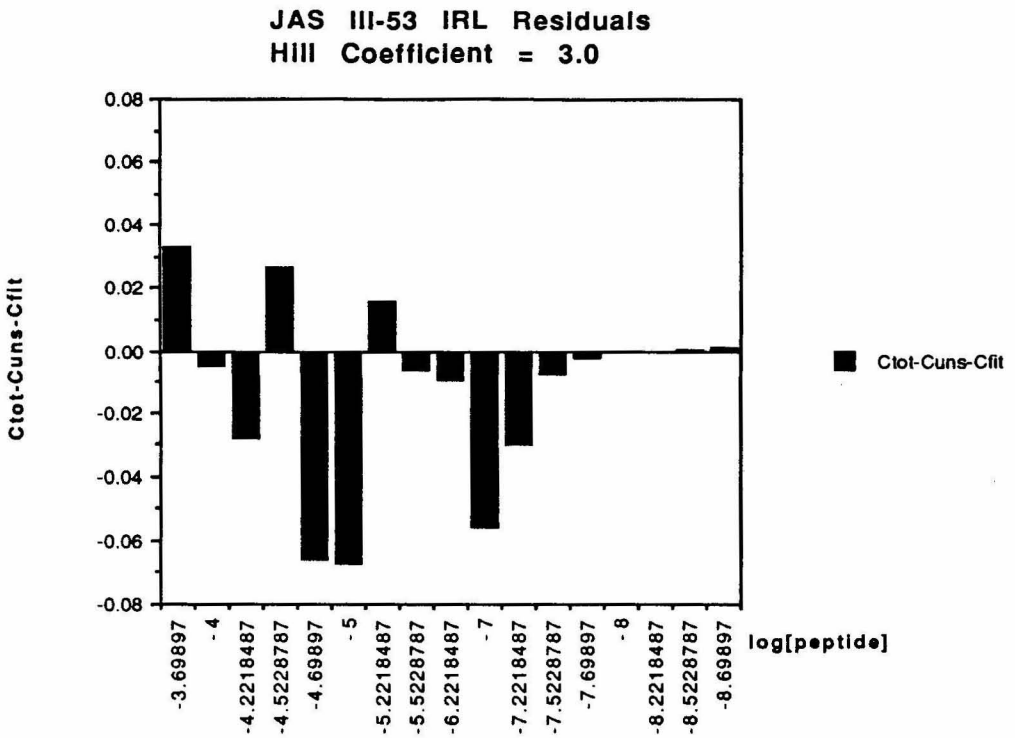
On the following pages are shown the binding isotherms and residuals for gels JAS III-53, 55, and 56. On top are data for the *hix*L IRL binding site, and on the bottom are data for the *hix*L IRR binding site. Residuals measure the difference between the obtained data points and the fit curve.

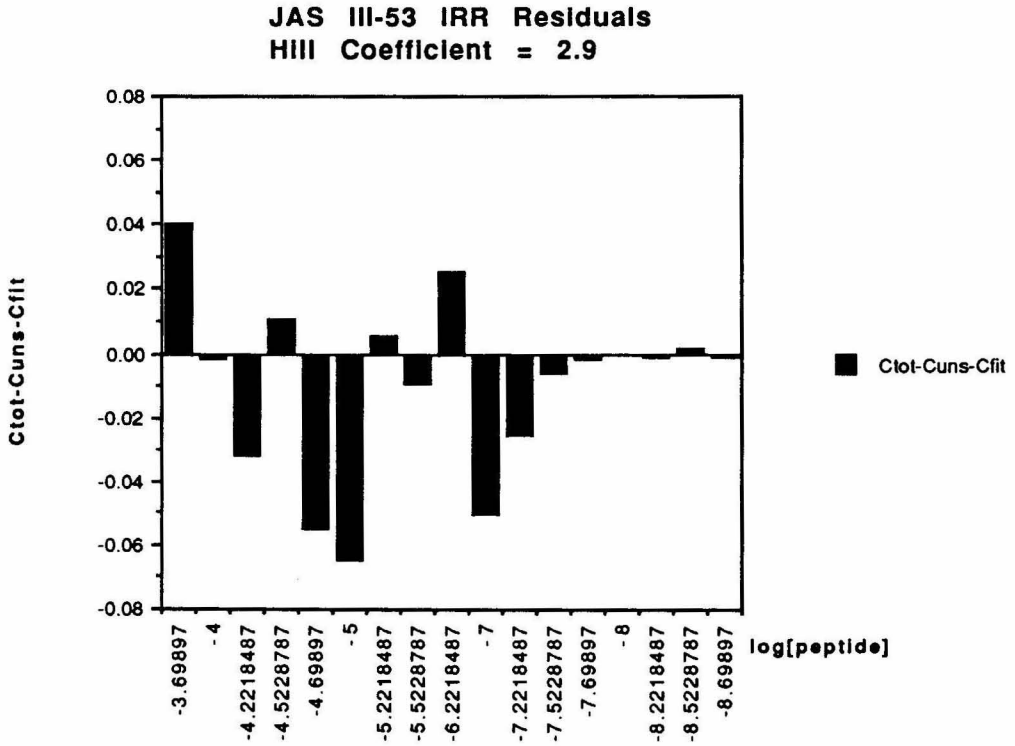
Quantitative Affinity Cleaving
[Fe•EDTA] Hin(139-190)R140→K

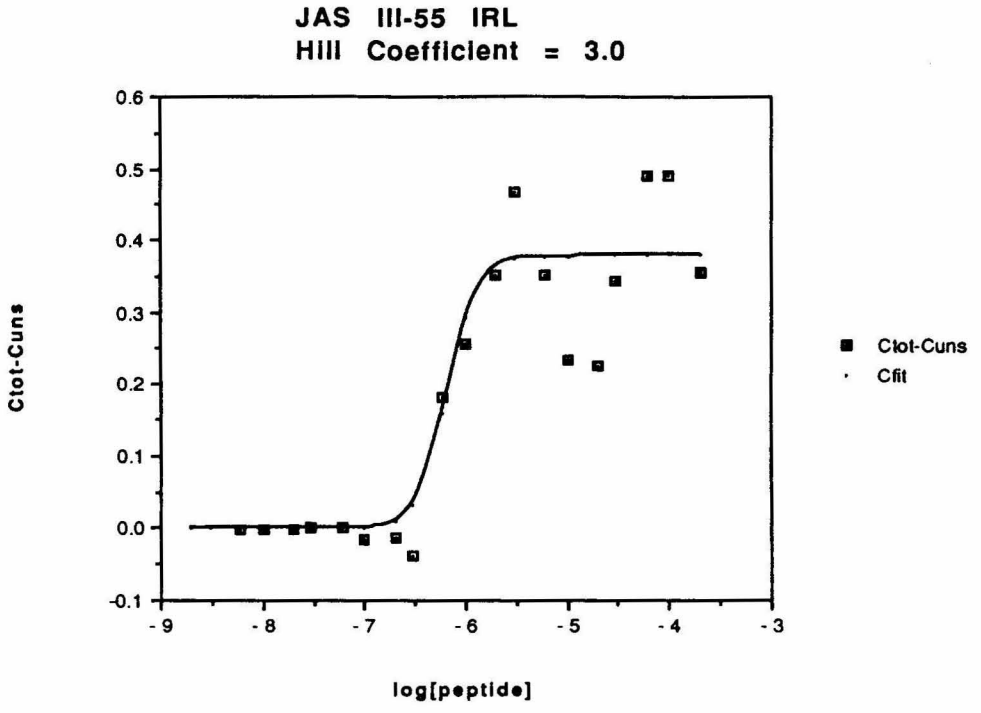


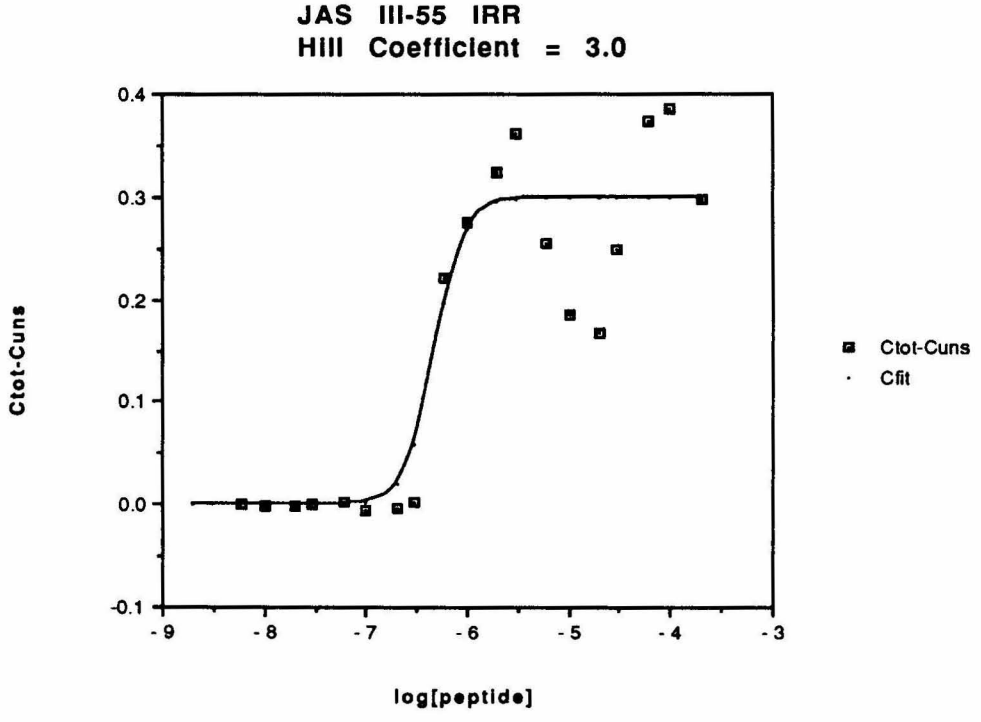

JAS III-53 IRL
HIII Coefficient = 3.0
Ctot-Cuns
log[peptide]
Ctot-Cuns
Cfit

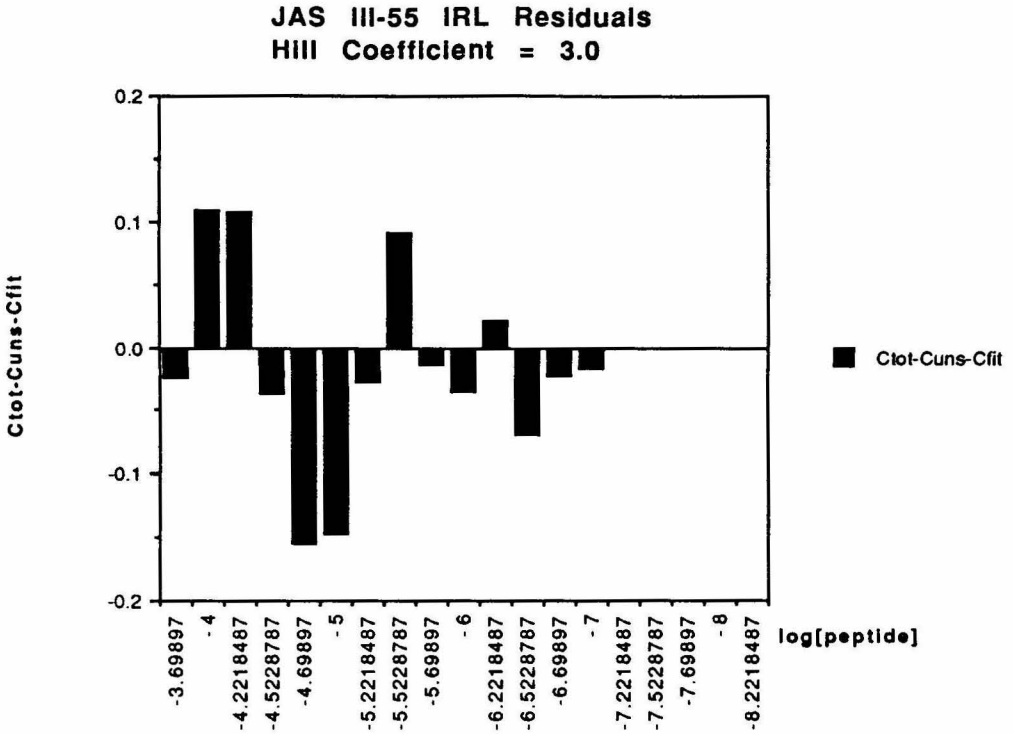

JAS III-53 IRR
HIII Coefficient = 2.9
Ctot-Cuns
log[peptide]
Ctot-Cuns
Cfit

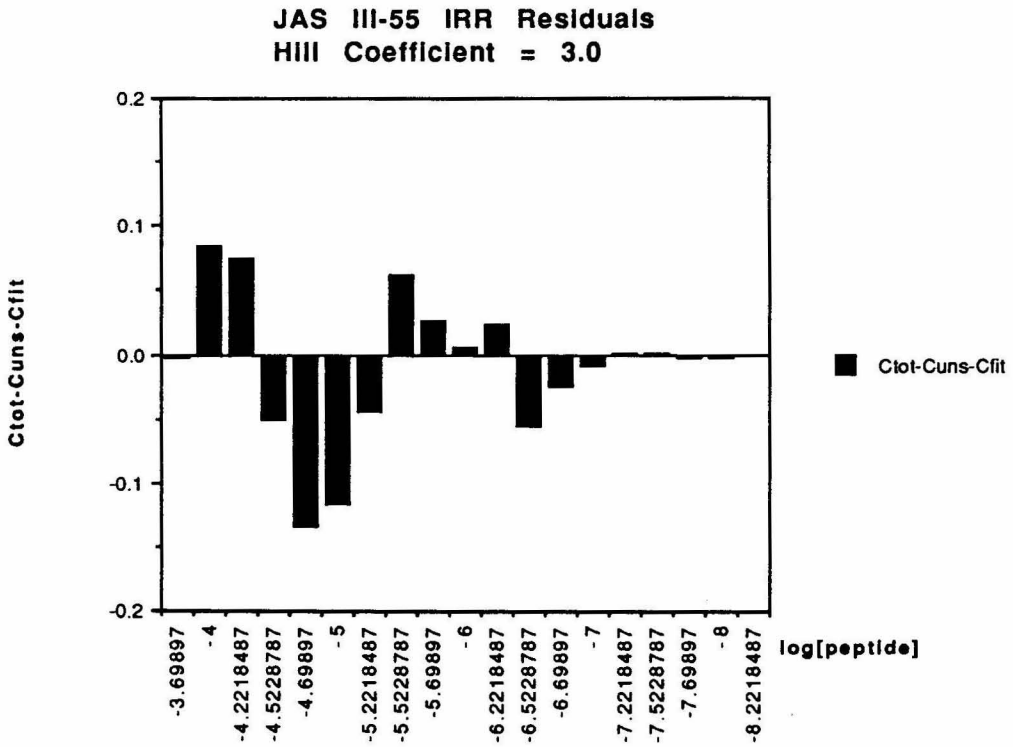

JAS III-53 IRL Residuals
Hill Coefficient = 3.0
Ctot-Cuns-Cfit
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
-0.08
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-6.2218487
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

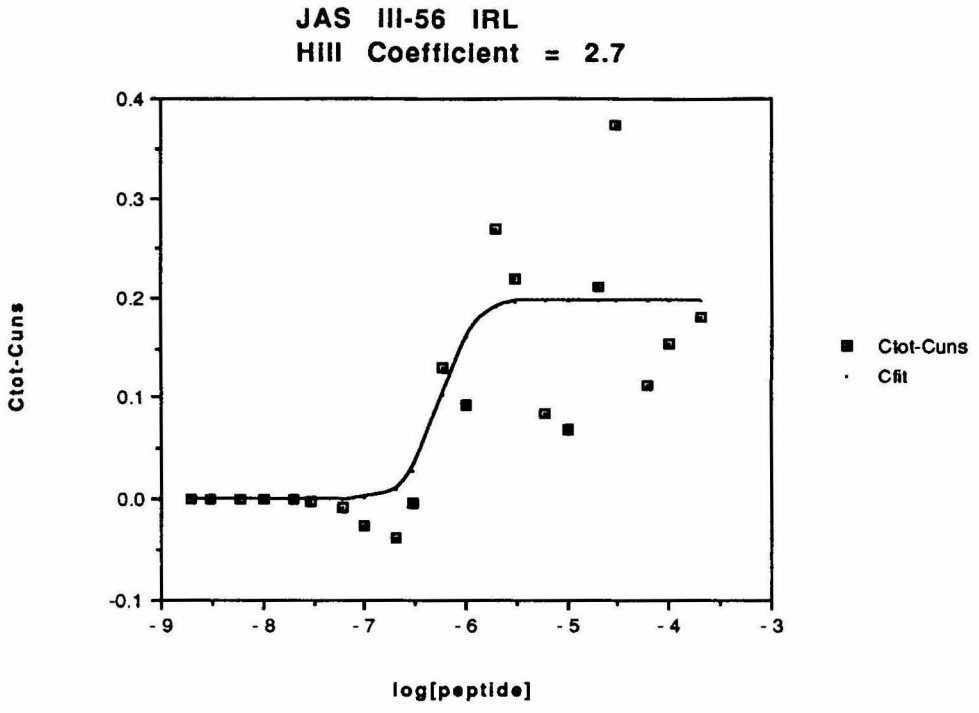

JAS III-53 IRR Residuals
Hill Coefficient = 2.9
Ctot-Cuns-Cfit
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
-0.08
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-6.2218487
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

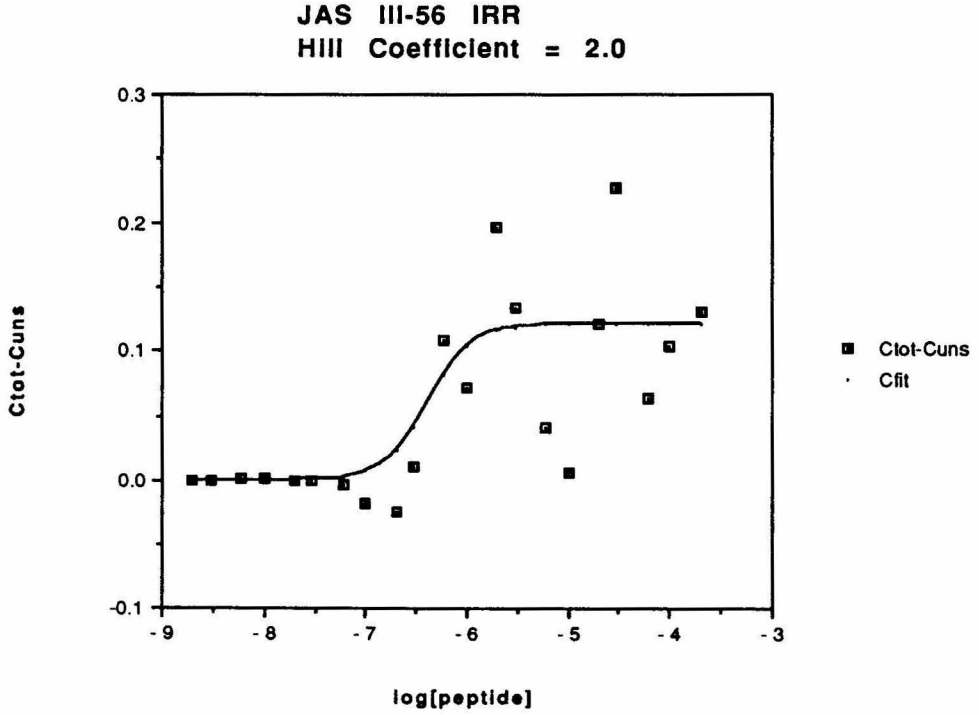

JAS III-55 IRL
Hill Coefficient = 3.0
Ctot-Cuns
log[peptide]
Ctot-Cuns
Cfit

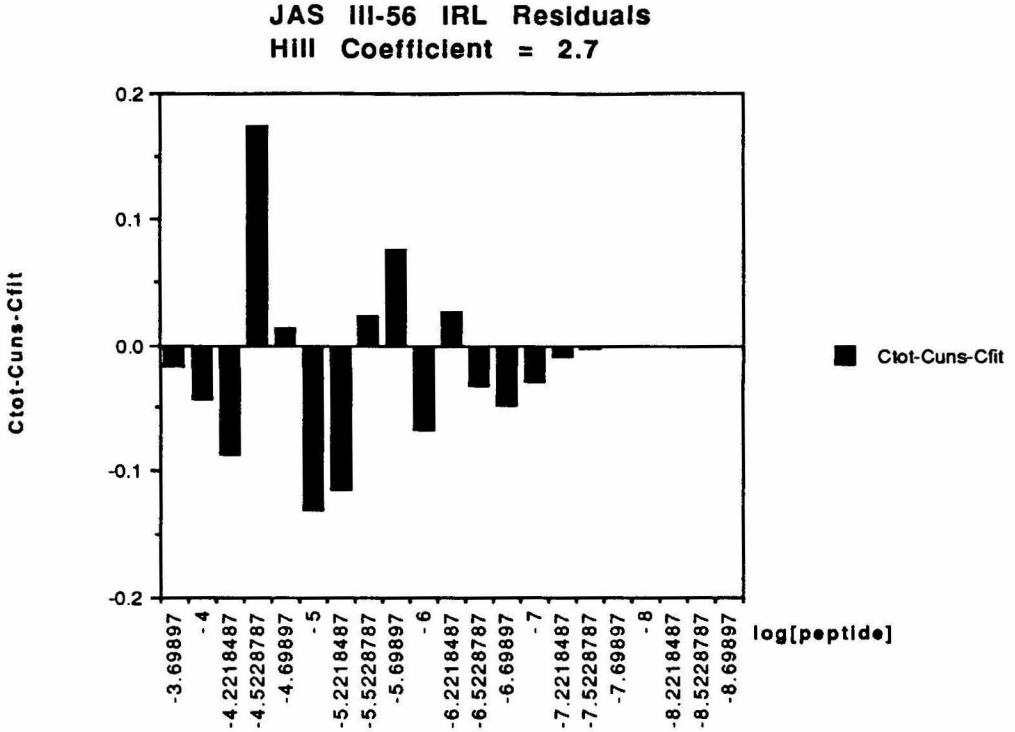

JAS III-55 IRR
Hill Coefficient = 3.0
Ctot-Cuns
log[peptide]
Ctot-Cuns
Cfit

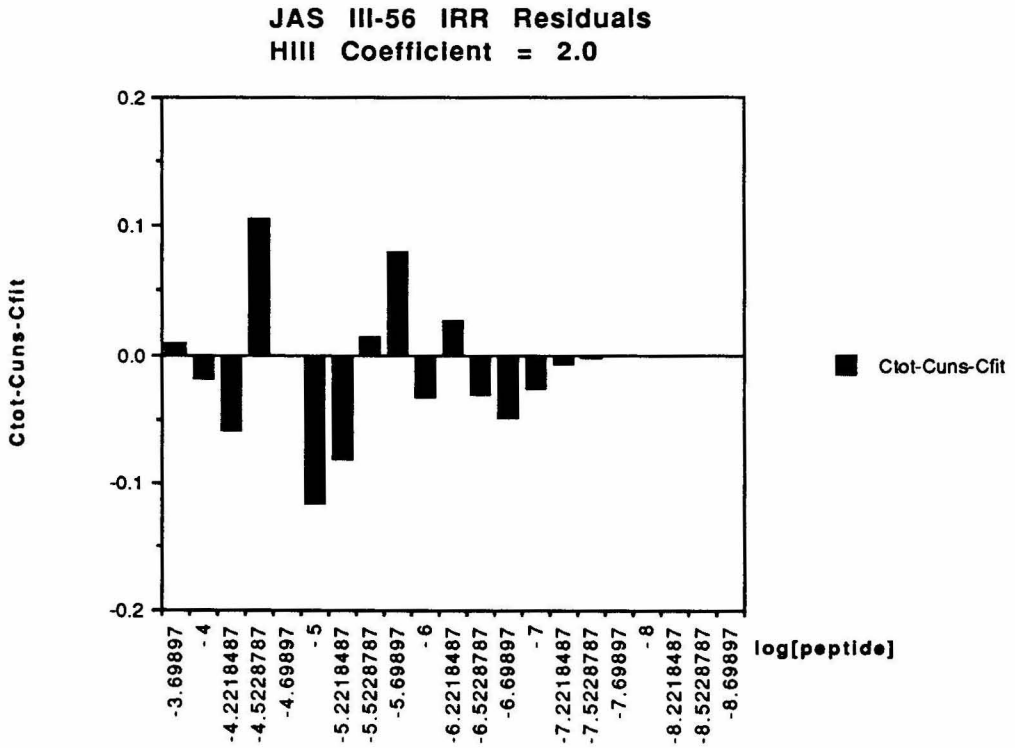

JAS III-55 IRL Residuals
HIII Coefficient = 3.0
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
log[peptide]
Ctot-Cuns-Cfit


JAS III-55 IRR Residuals
HIII Coefficient = 3.0
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
log[peptide]
Ctot-Cuns-Cfit


JAS III-56 IRL
Hill Coefficient = 2.7
Ctot-Cuns
0.4
0.3
0.2
0.1
0.0
-0.1
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit


JAS III-56 IRR
Hill Coefficient = 2.0
Ctot-Cuns
0.3
0.2
0.1
0.0
-0.1
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

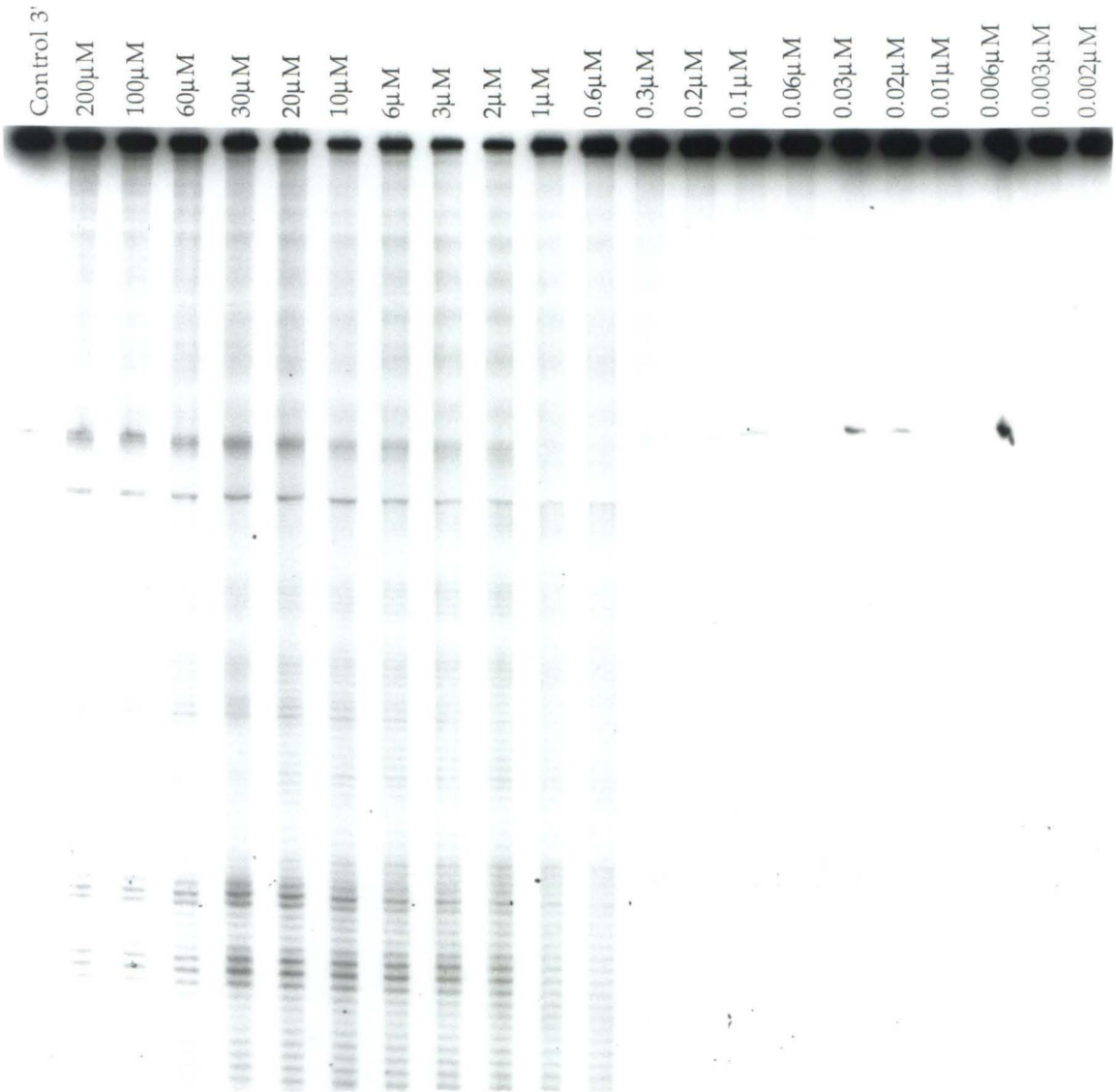

JAS III-56 IRL Residuals
Hill Coefficient = 2.7
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

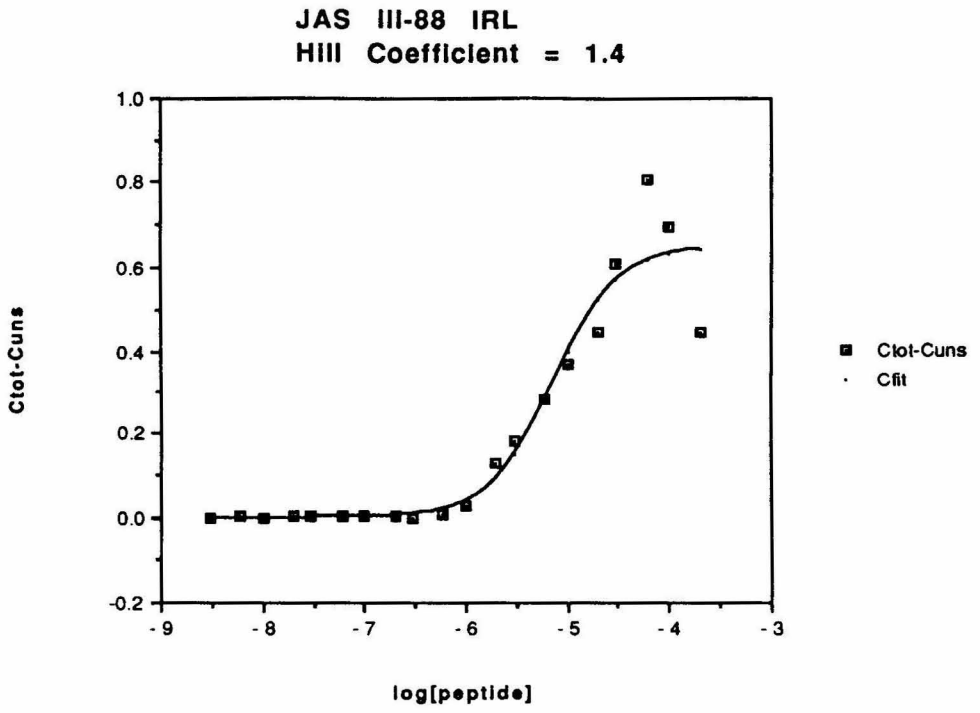

JAS III-56 IRR Residuals
Hill Coefficient = 2.0
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

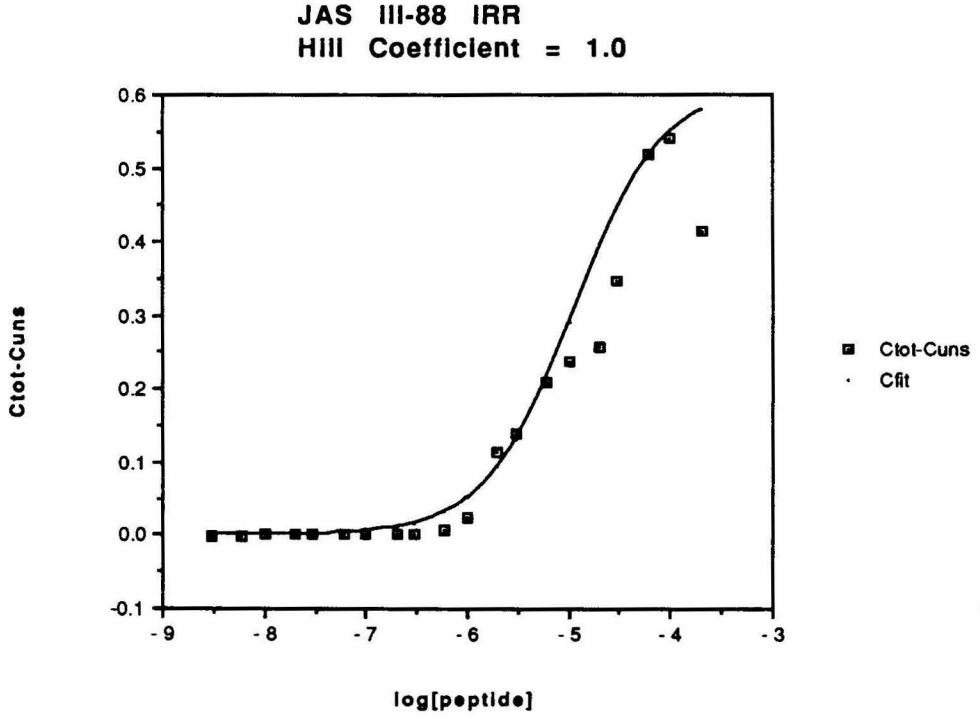
[Fe•EDTA]Hin(139-190)R140→A. Autoradiogram of a quantitative affinity cleaving gel of [Fe•EDTA]Hin(139-190)R140→A bound to the 3′ end-labelled Xba I-Spe I fragment (218 base pairs) from pMFB36. This autoradiogram is from the gel labelled JAS III-91 (see JAS notebook III). Lane 1, control. Lanes 2-22 contain decreasing concentrations of protein as indicated. Each reaction contains 20mM phosphate buffer, pH 7.5, 20mM NaCl, and 1mM dithiothreitol. Each reaction proceeds at room temperature for 35 minutes. Reactions are stopped and extracted with 200μL of a 2:1 solution of phenol:chloroform; followed by butanol extraction and ethanol precipitation. Reactions are taken up in formamide loading buffer and loaded onto an 8% polyacrylamide denaturing gel.

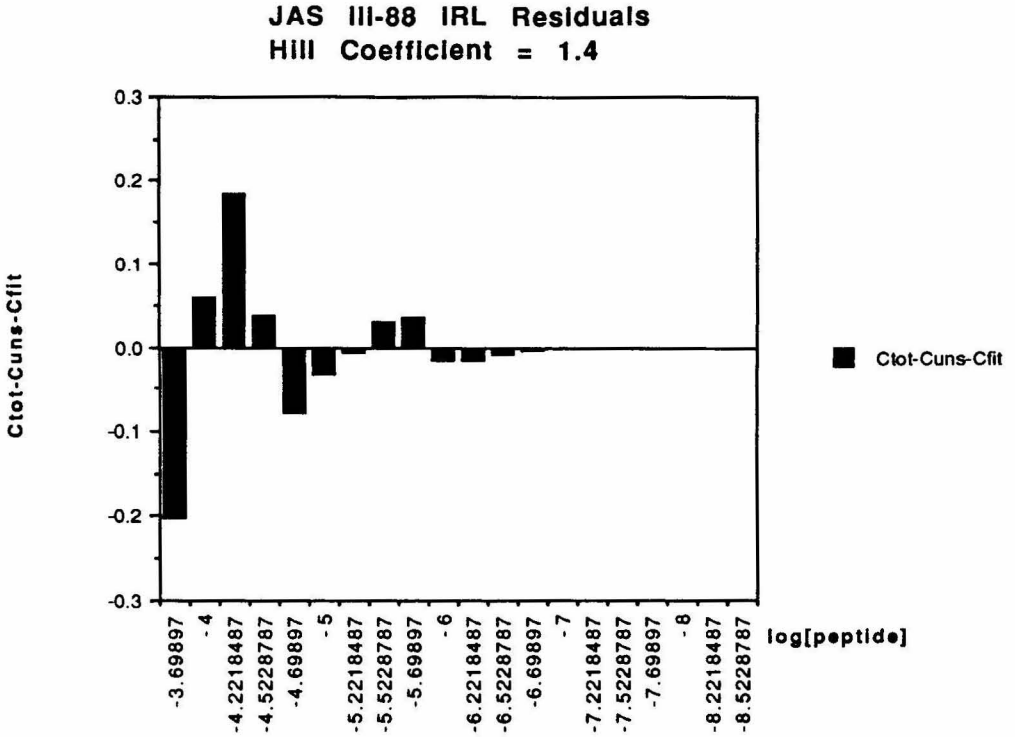
On the following pages are shown the binding isotherms and residuals for gels JAS III-88, 90, 91, and 92. On top are data for the *hix*L IRL binding site, and on the bottom are data for the *hix*L IRR binding site. Residuals measure the difference between the obtained data points and the fit curve.

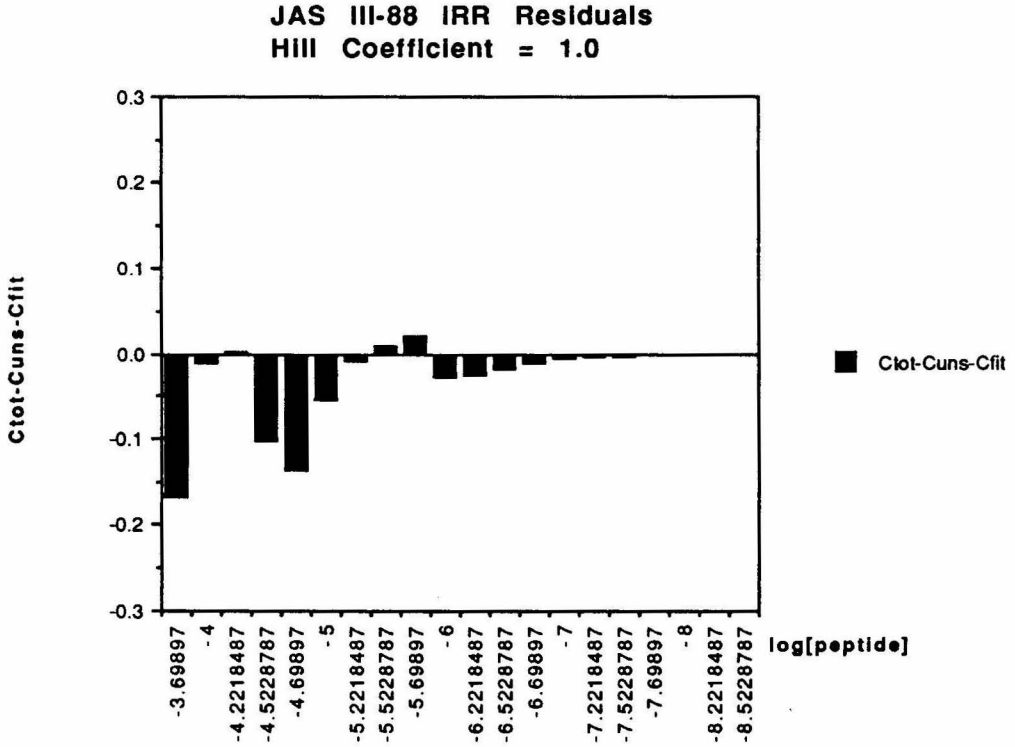
Quantitative Affinity Cleaving
[Fe•EDTA] Hin(139-190)R140→A

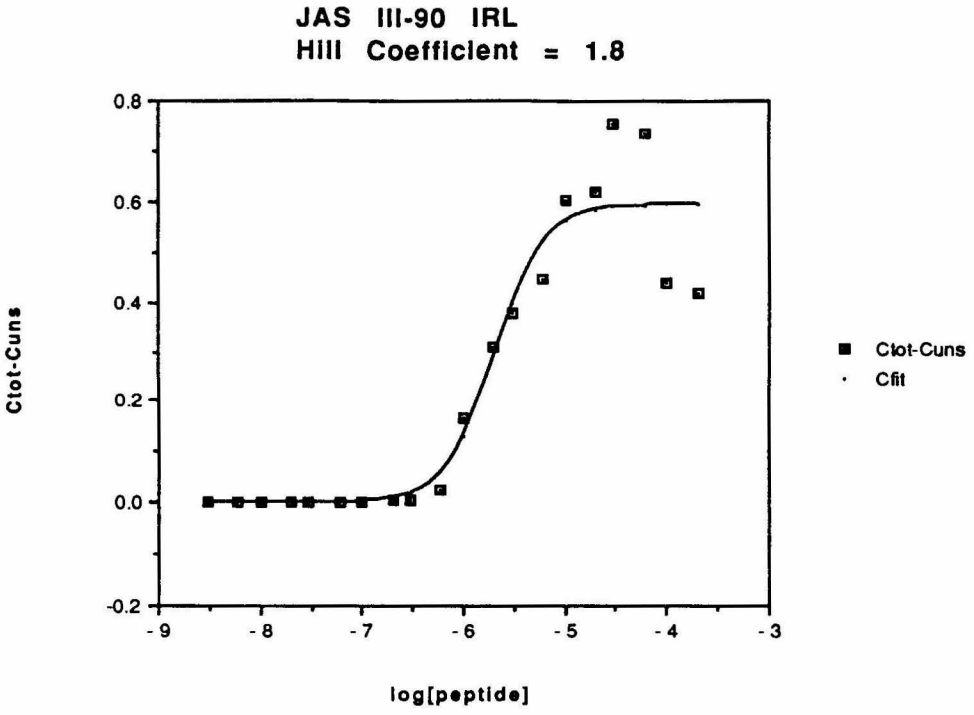


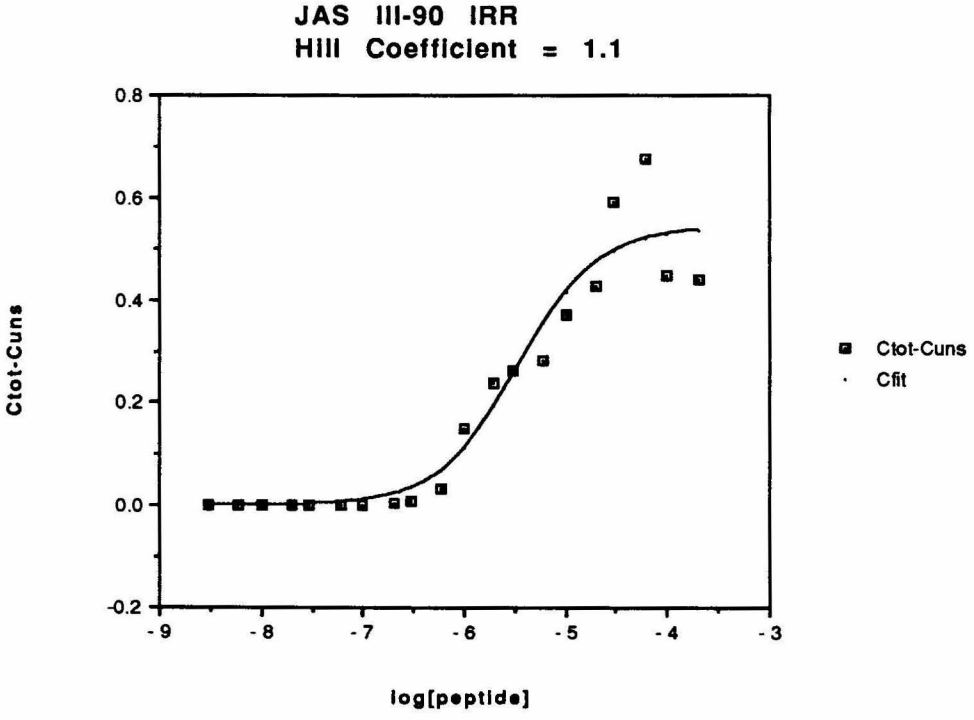

JAS III-88 IRL
Hill Coefficient = 1.4
Ctot-Cuns
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

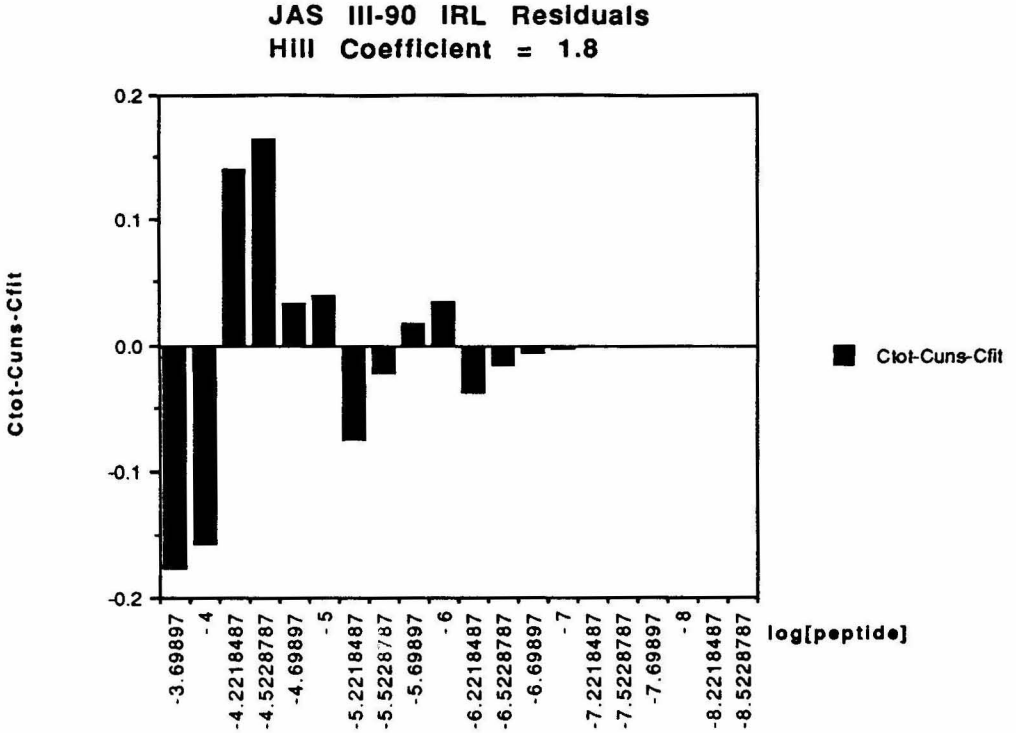

JAS III-88 IRR
Hill Coefficient = 1.0
Ctot-Cuns
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

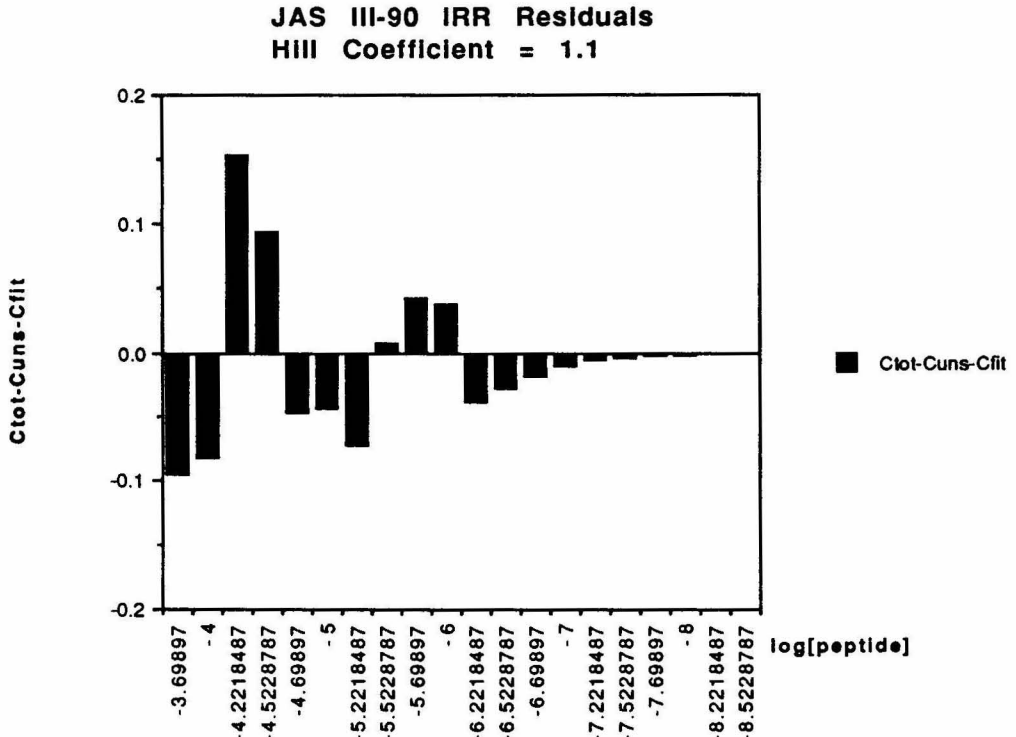

JAS III-88 IRL Residuals
Hill Coefficient = 1.4
Ctot-Cuns-Cfit
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
log[peptide]
Ctot-Cuns-Cfit

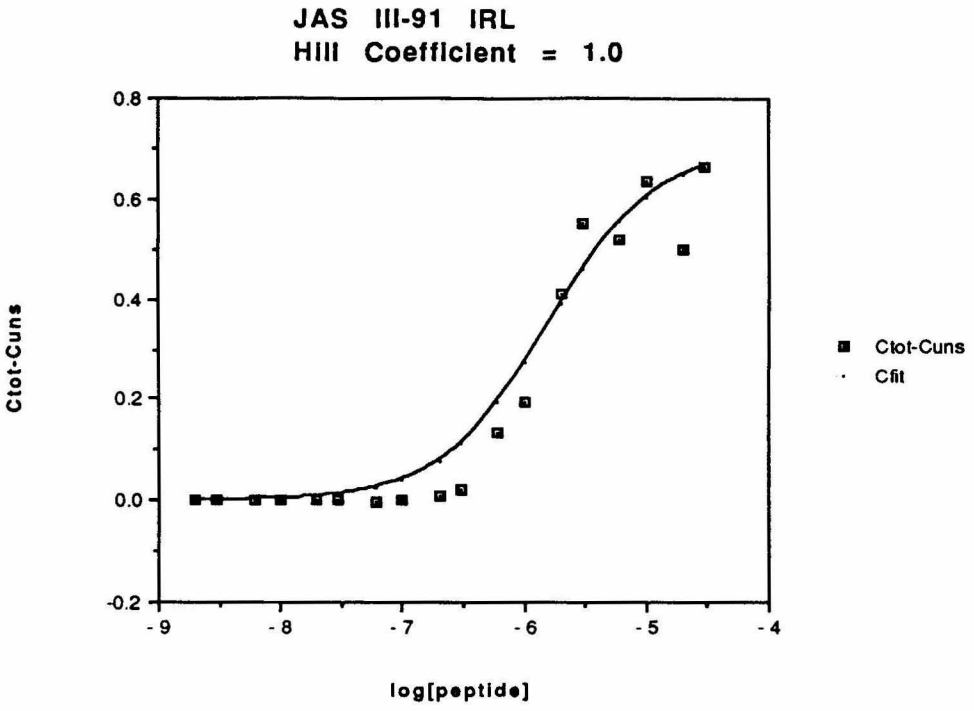

JAS III-88 IRR Residuals
Hill Coefficient = 1.0
Ctot-Cuns-Cfit
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
log[peptide]
Ctot-Cuns-Cfit

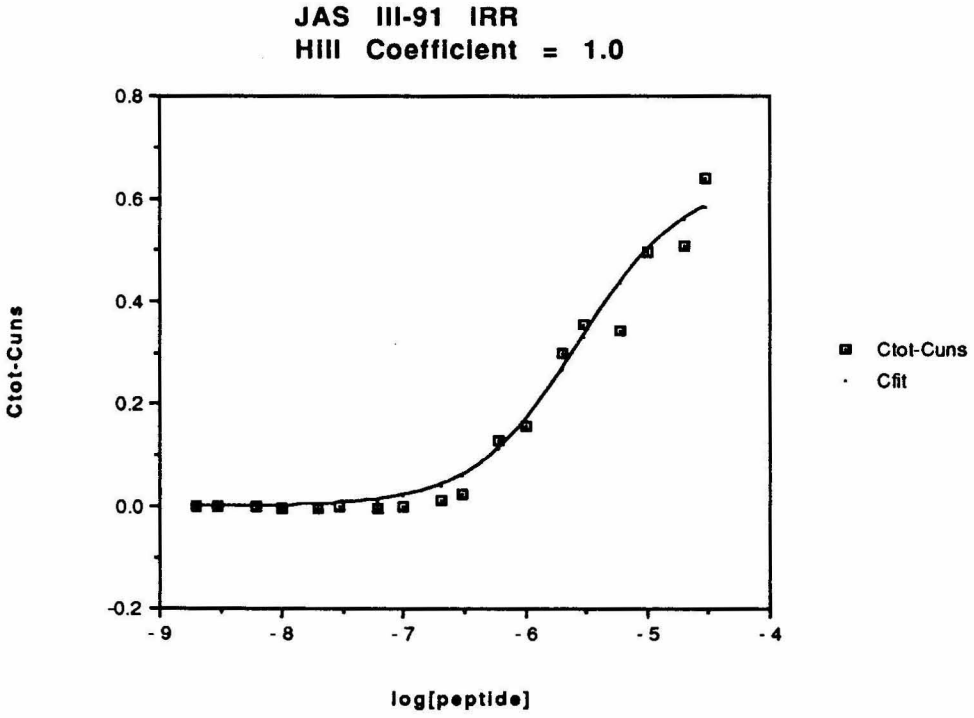

JAS III-90 IRL
Hill Coefficient = 1.8
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

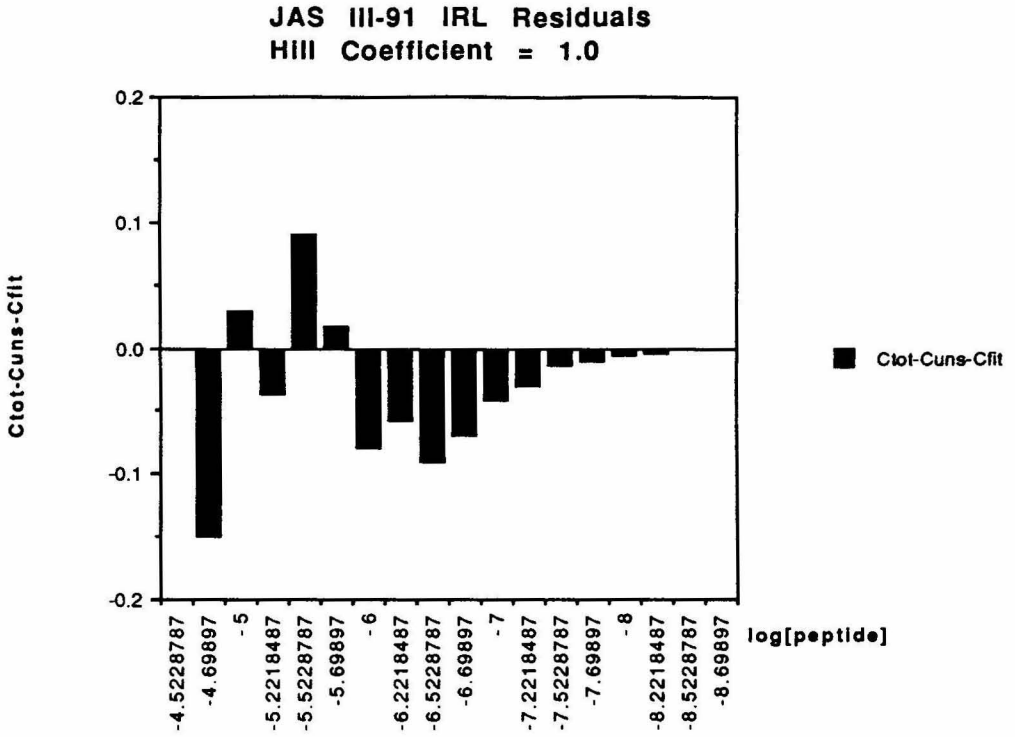

JAS III-90 IRR
Hill Coefficient = 1.1
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

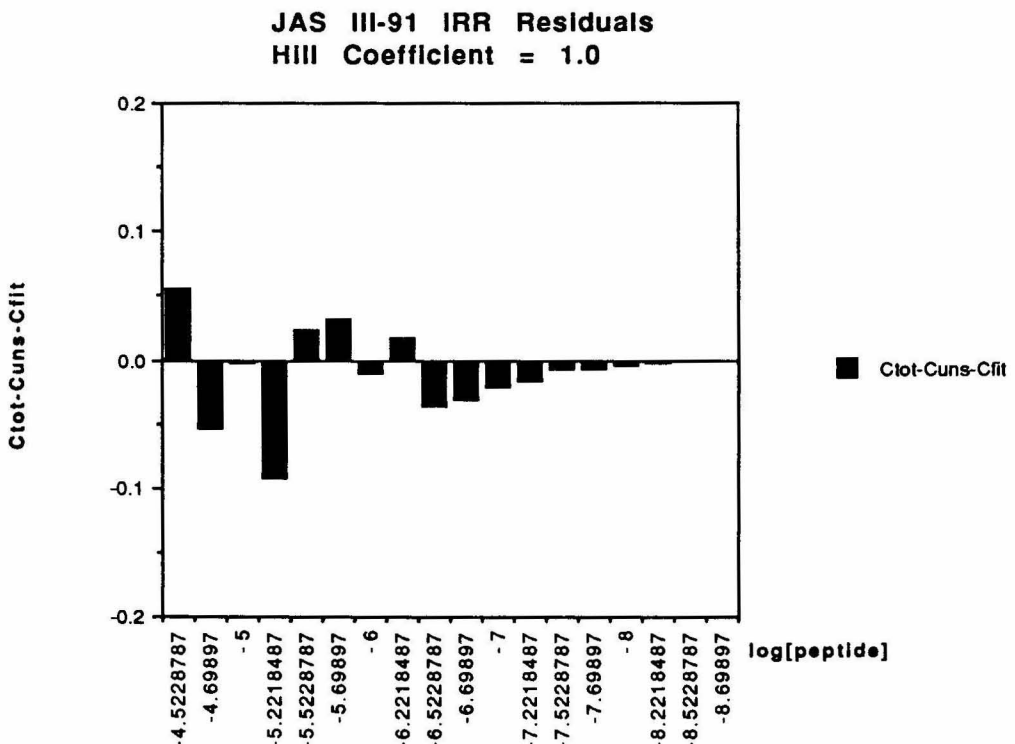

JAS III-90 IRL Residuals
Hill Coefficient = 1.8
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
log[peptide]
Ctot-Cuns-Cfit

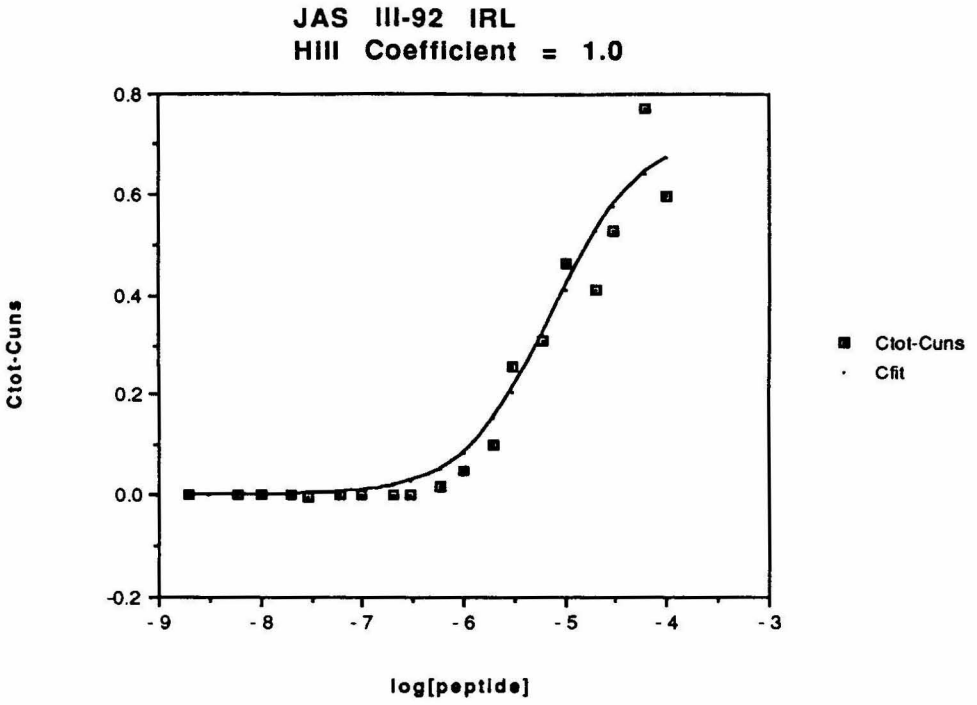

JAS III-90 IRR Residuals
Hill Coefficient = 1.1
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
log[peptide]
Ctot-Cuns-Cfit

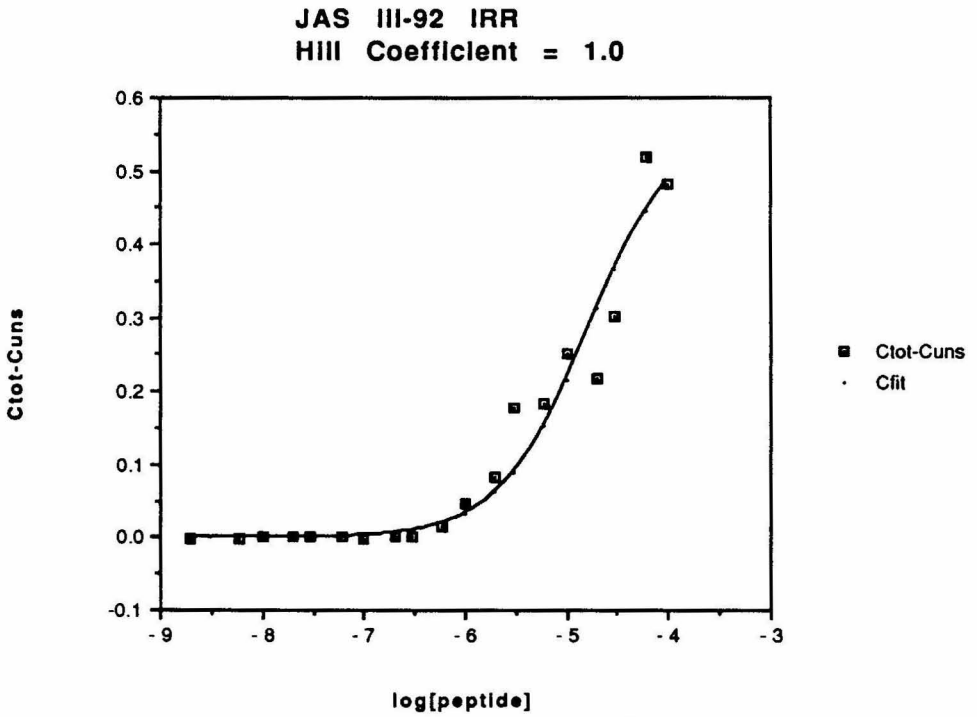

JAS III-91 IRL
Hill Coefficient = 1.0
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
log[peptide]
Ctot-Cuns
Cfit

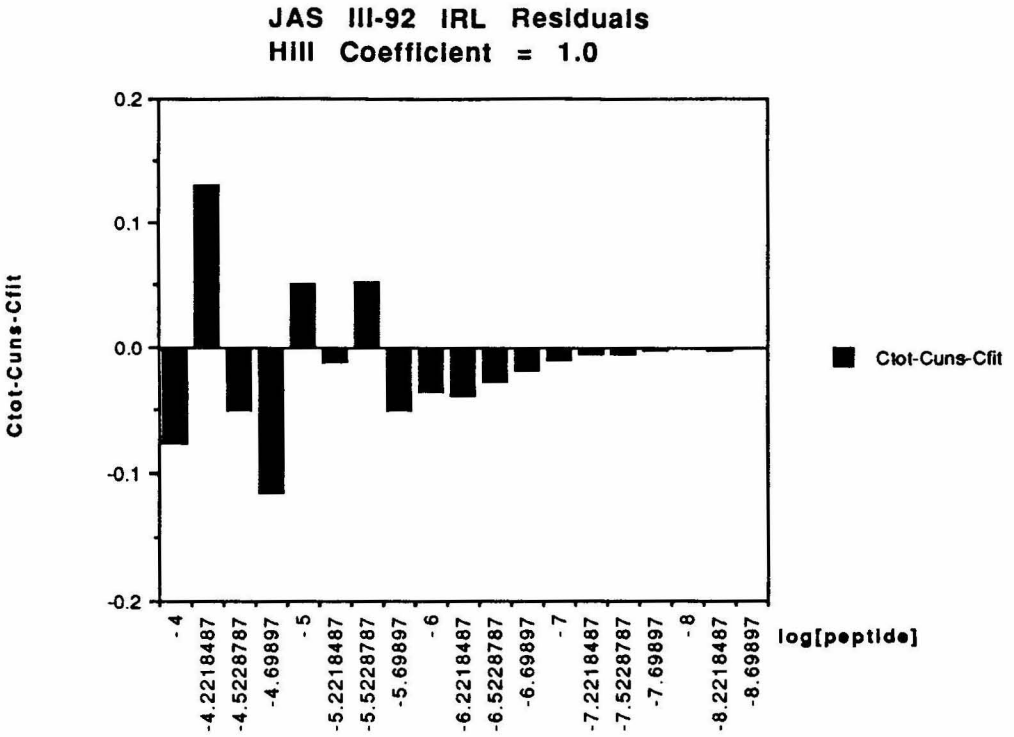

JAS III-91 IRR
Hill Coefficient = 1.0
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
log[peptide]
Ctot-Cuns
Cfit

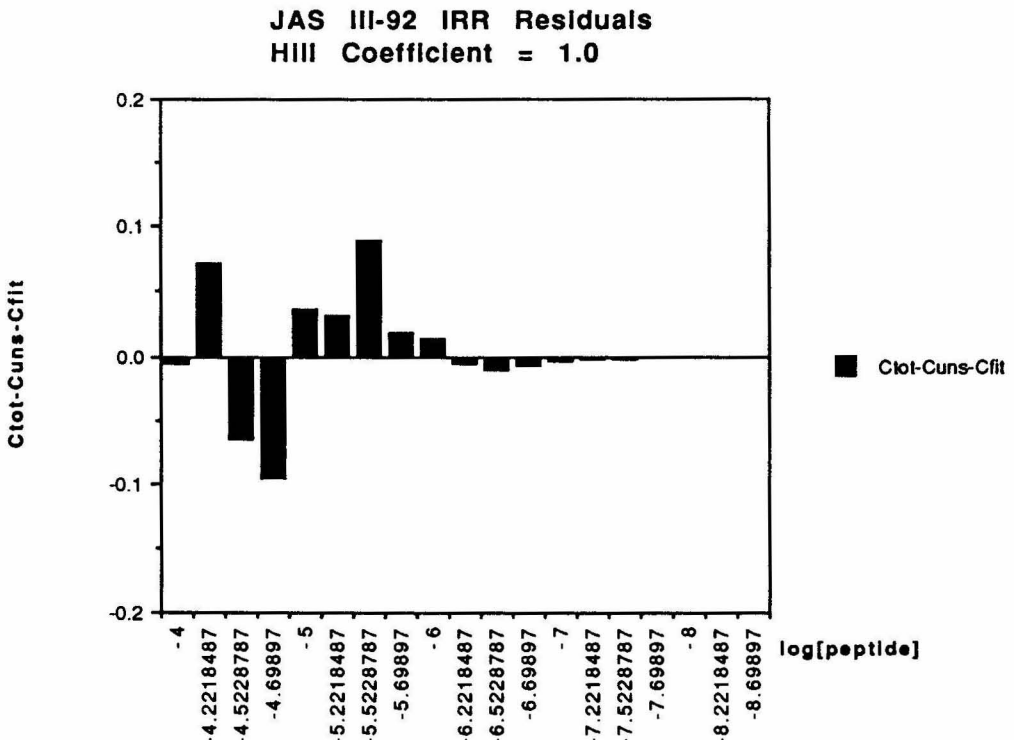

JAS III-91 IRL Residuals
Hill Coefficient = 1.0
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit


JAS III-91 IRR Residuals
Hill Coefficient = 1.0
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit


JAS III-92 IRL
HIII Coefficient = 1.0
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit


JAS III-92 IRR
HIII Coefficient = 1.0
Ctot-Cuns
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

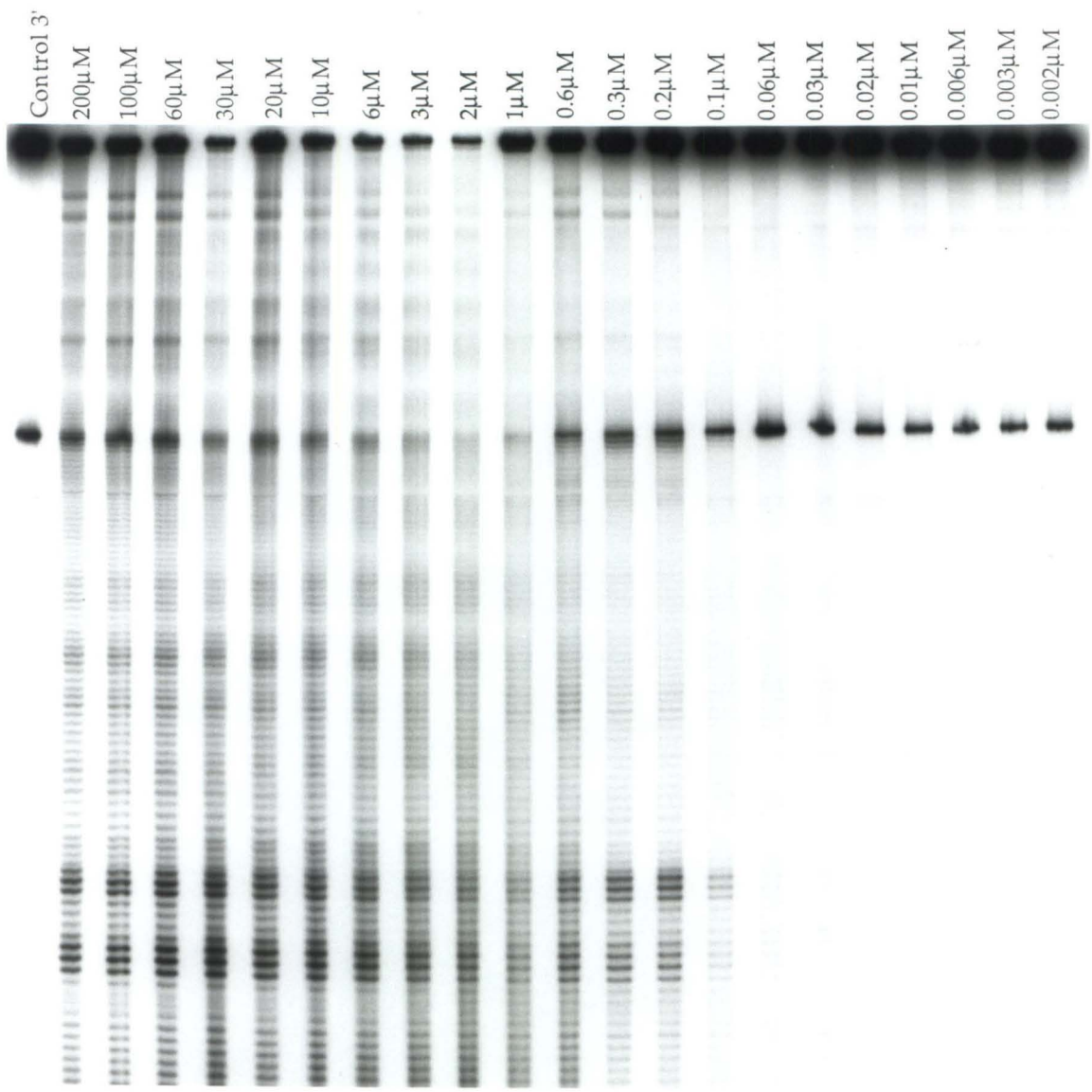

JAS III-92 IRL Residuals
Hill Coefficient = 1.0
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.69897
log[peptide]
Ctot-Cuns-Cfit

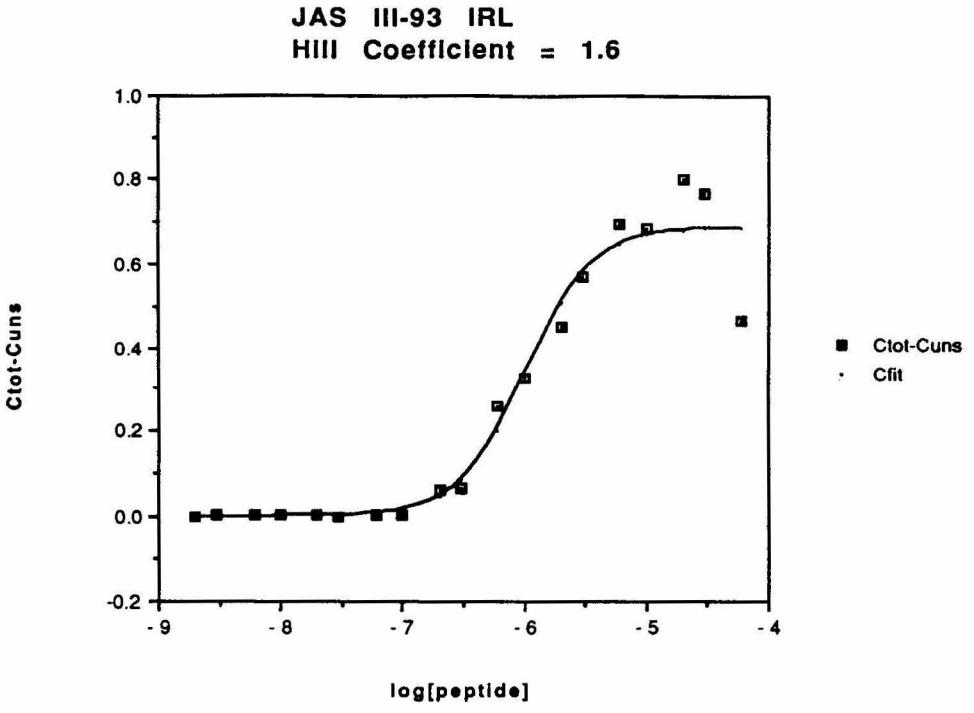

JAS III-92 IRR Residuals
Hill Coefficient = 1.0
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.69897
log[peptide]
Ctot-Cuns-Cfit

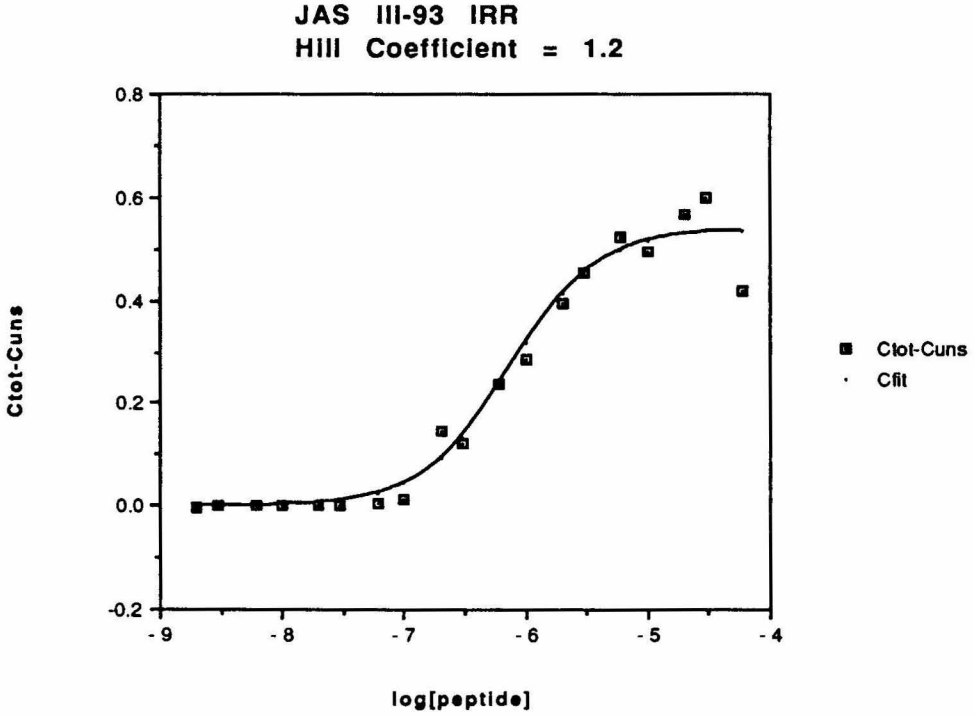
[Fe•EDTA]Hin(139-190)R142→K. Autoradiogram of a quantitative affinity cleaving gel of [Fe•EDTA]Hin(139-190)R142→K bound to the 3′ end-labelled Xba I-Spe I fragment (218 base pairs) from pMFB36. This autoradiogram is from the gel labelled JAS III-97 (see JAS notebook III). Lane 1, control. Lanes 2-22 contain decreasing concentrations of protein as indicated. Each reaction contains 20mM phosphate buffer, pH 7.5, 20mM NaCl, and 1mM dithiothreitol. Each reaction proceeds at room temperature for 30 minutes. Reactions are stopped and extracted with 200μL of a 2:1 solution of phenol:chloroform; followed by butanol extraction and ethanol precipitation. Reactions are taken up in formamide loading buffer and loaded onto an 8% polyacrylamide denaturing gel.

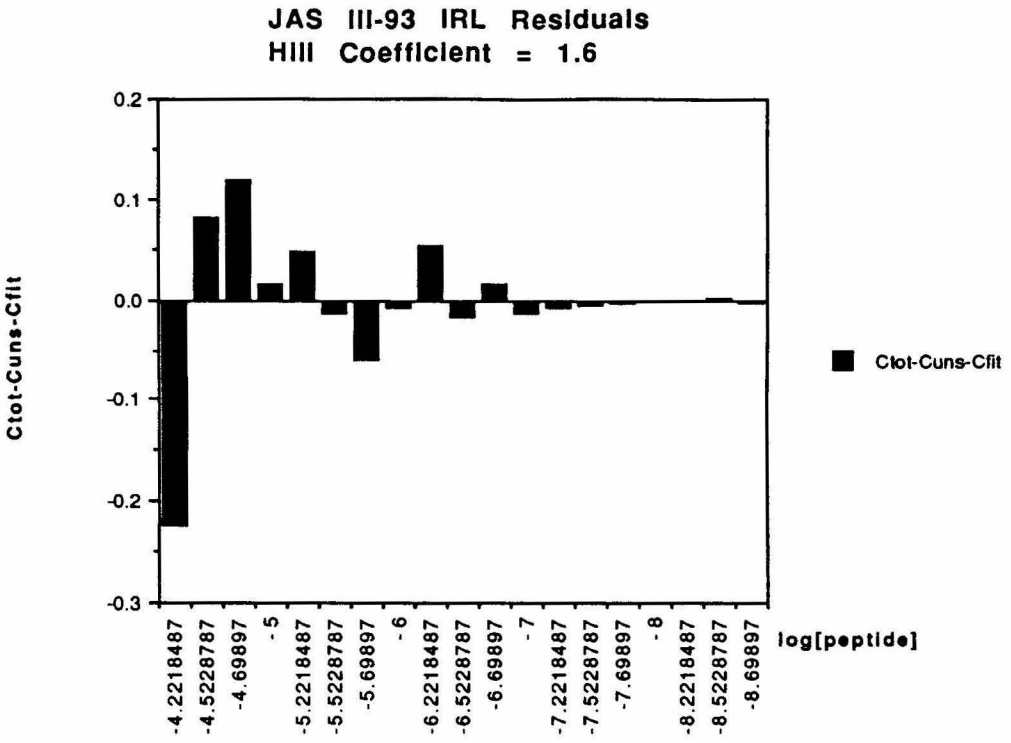
On the following pages are shown the binding isotherms and residuals for gels JAS III-93, 95, 96, and 97. On top are data for the *hix*L IRL binding site, and on the bottom are data for the *hix*L IRR binding site. Residuals measure the difference between the obtained data points and the fit curve.

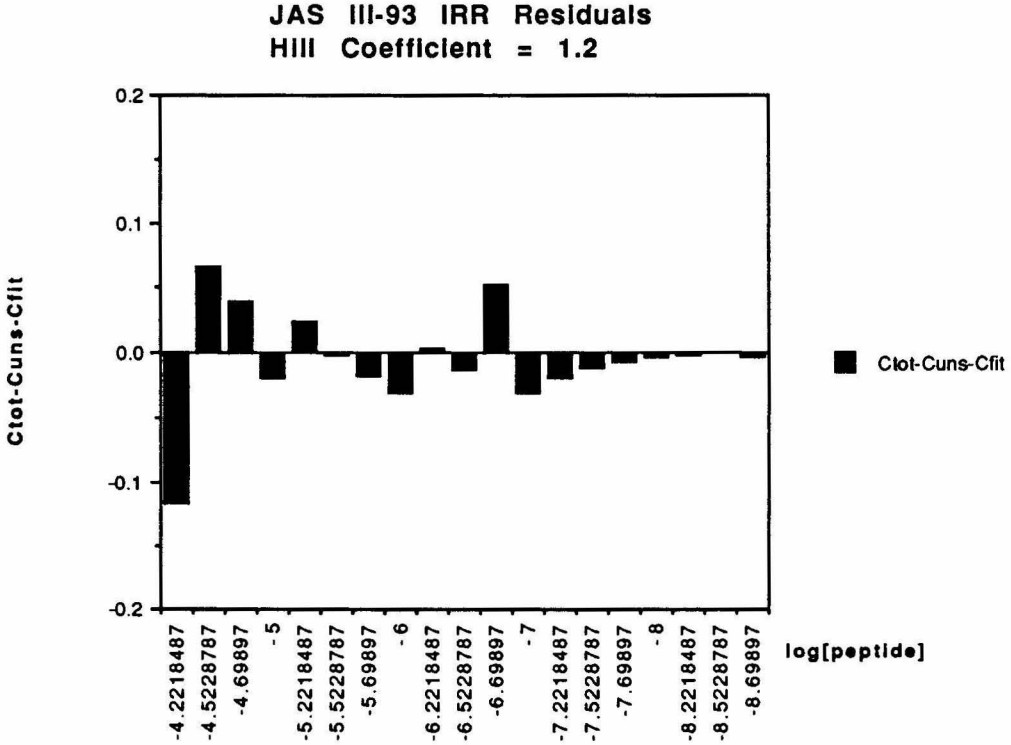
Quantitative Affinity Cleaving
[Fe•EDTA] Hin(139-190)R142→K

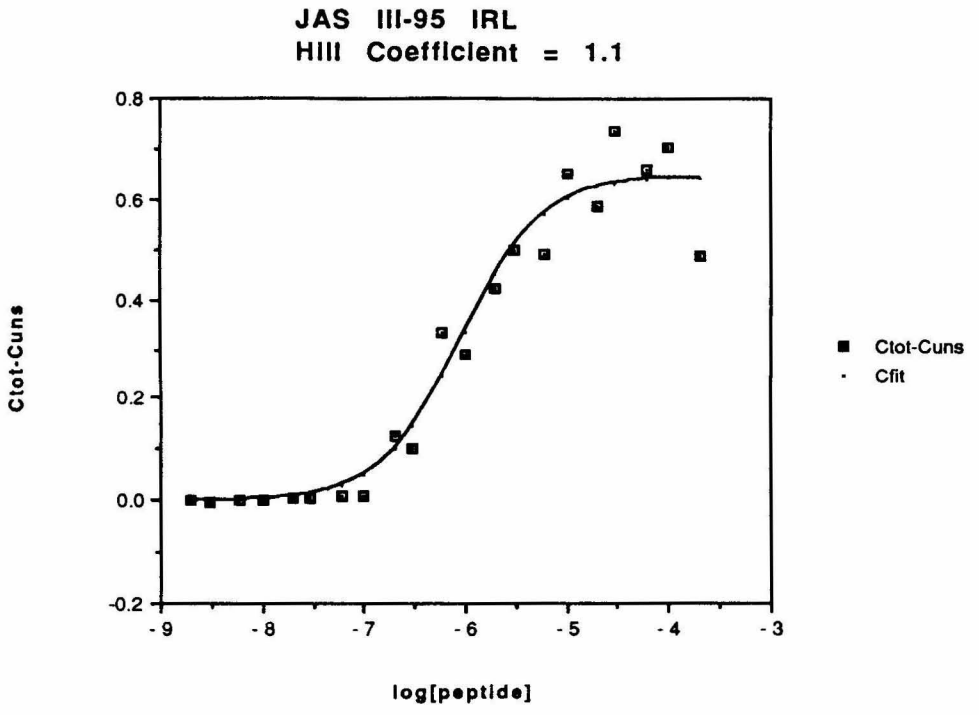


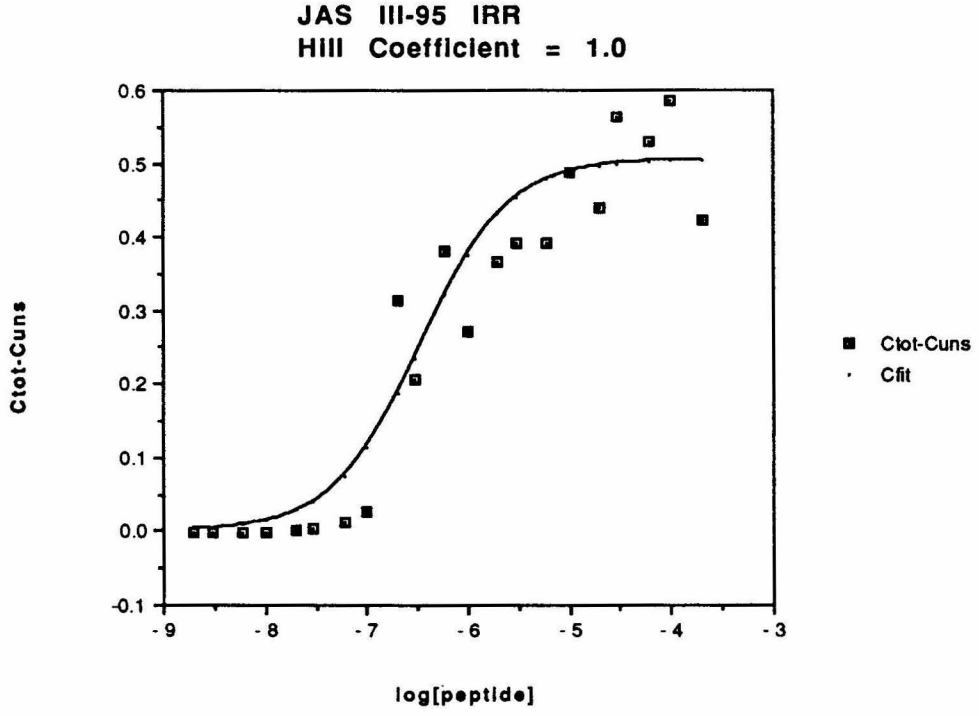

JAS III-93 IRL
HIII Coefficient = 1.6
Ctot-Cuns
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
log[peptide]
Ctot-Cuns
Cfit

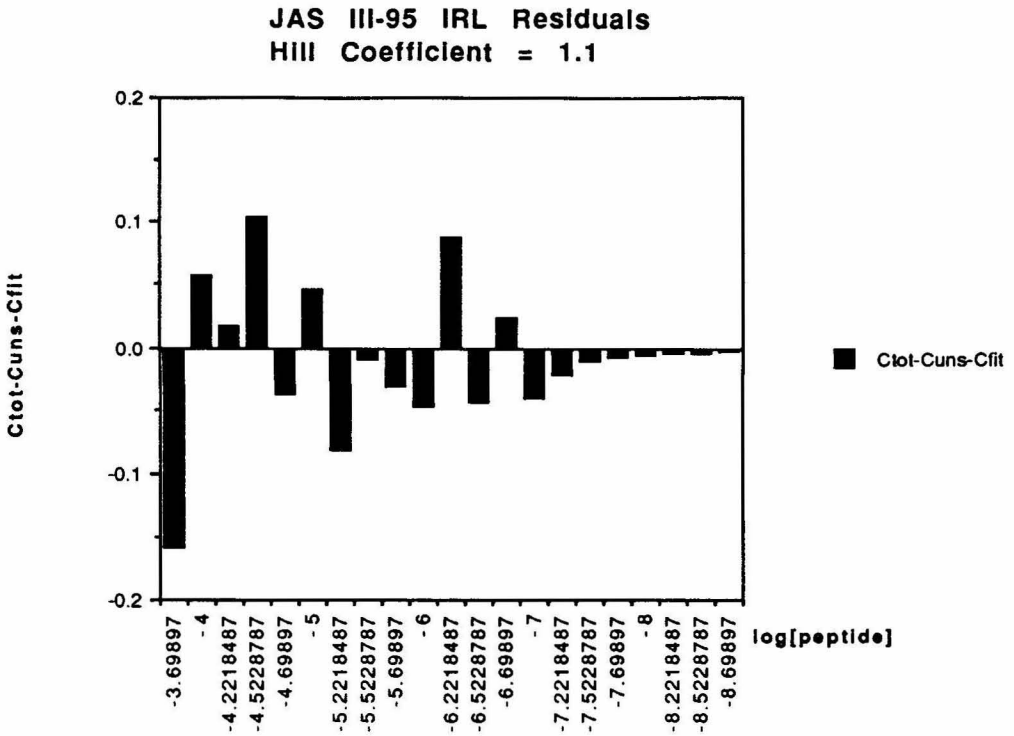

JAS III-93 IRR
HIII Coefficient = 1.2
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
log[peptide]
Ctot-Cuns
Cfit

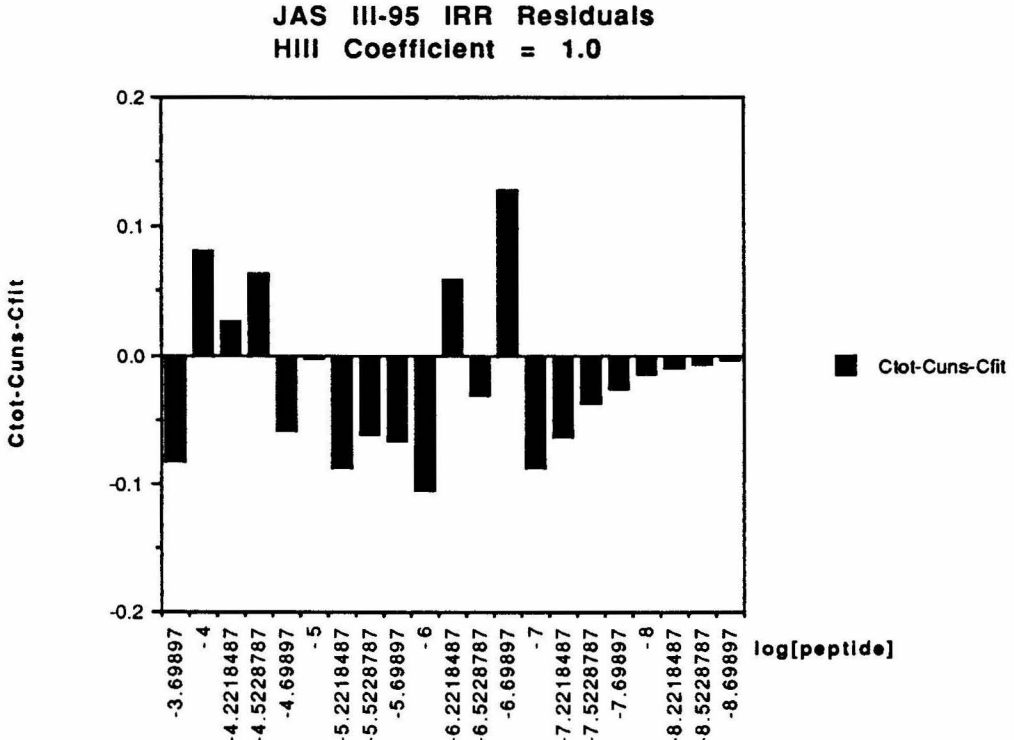

JAS III-93 IRL Residuals
Hill Coefficient = 1.6
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-0.3
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

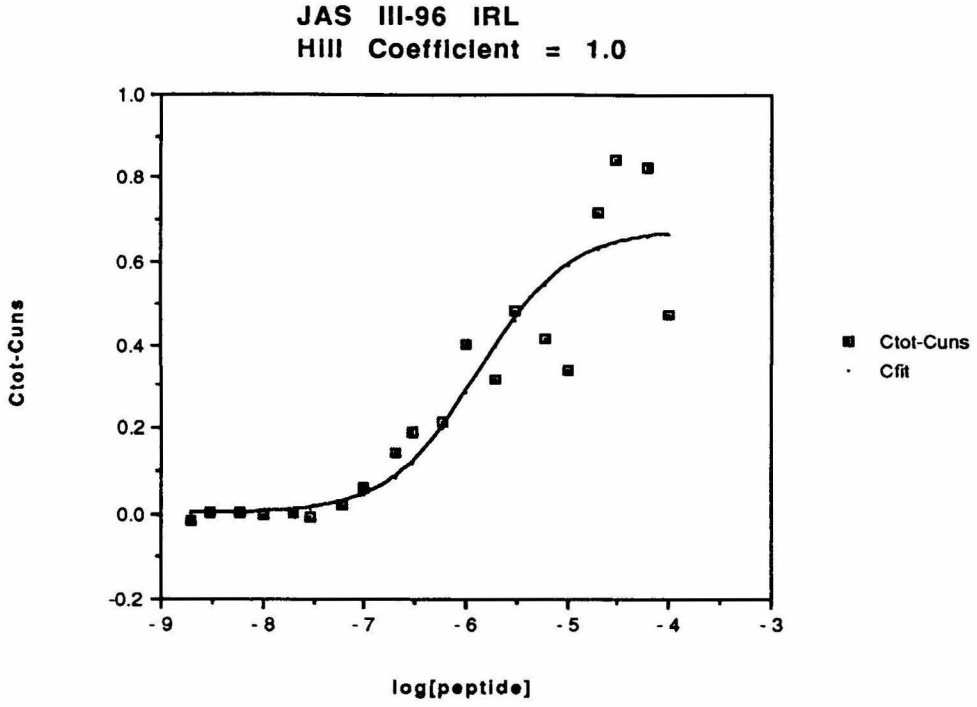

JAS III-93 IRR Residuals
Hill Coefficient = 1.2
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

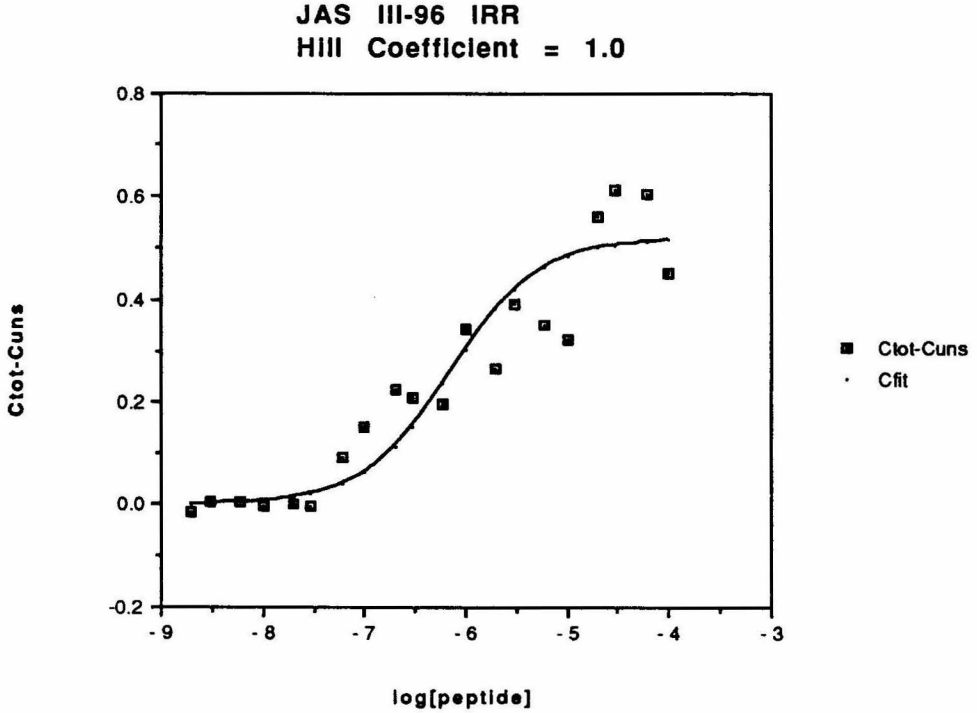

JAS III-95 IRL
HIII Coefficient = 1.1
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

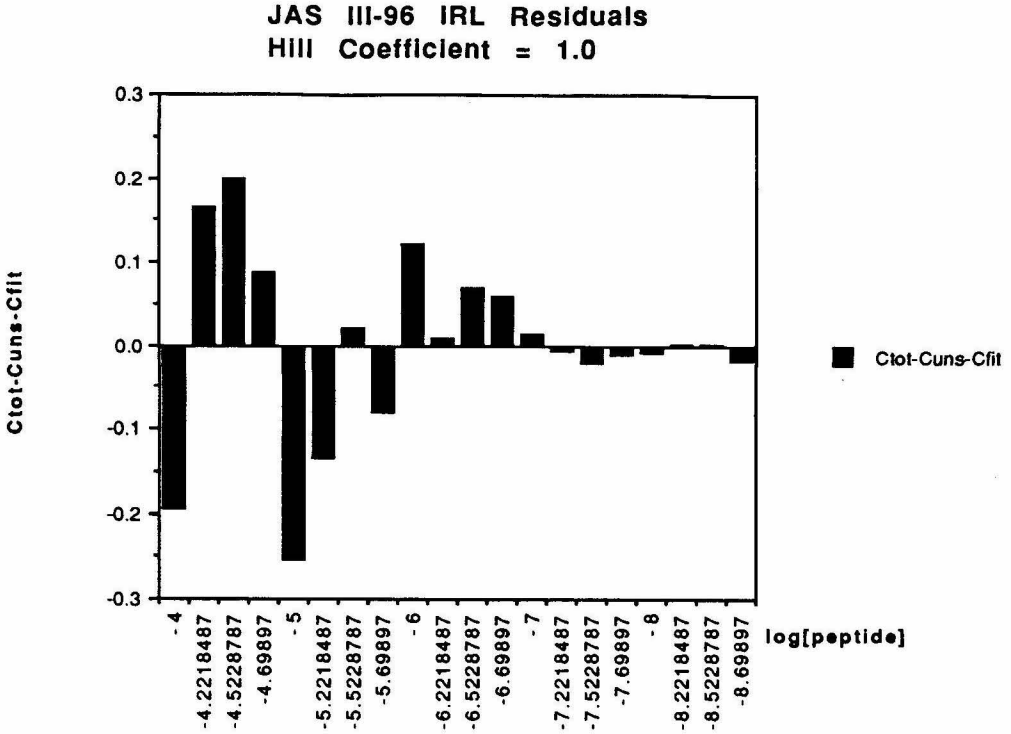

JAS III-95 IRR
HIII Coefficient = 1.0
Ctot-Cuns
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

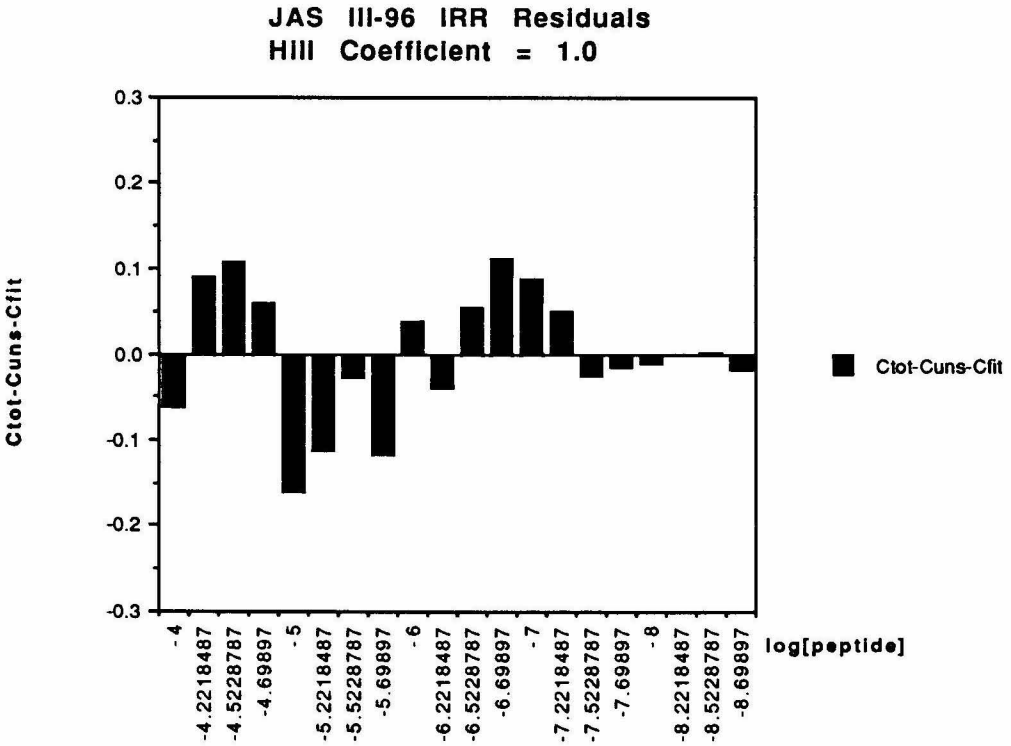

JAS III-95 IRL Residuals
Hill Coefficient = 1.1
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

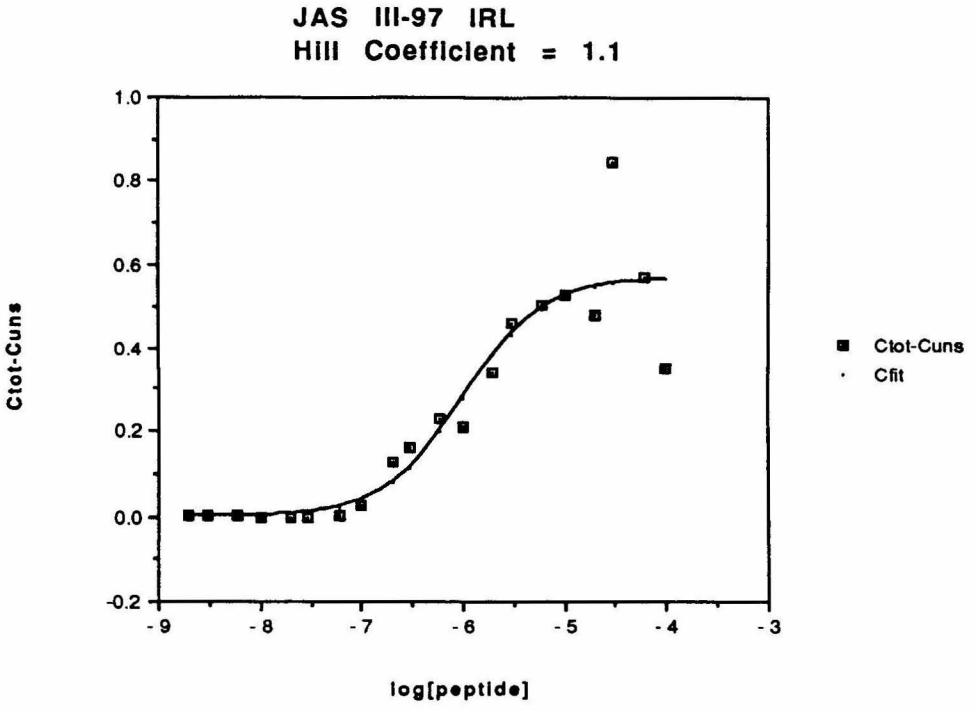

JAS III-95 IRR Residuals
Hill Coefficient = 1.0
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
-3.69897
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

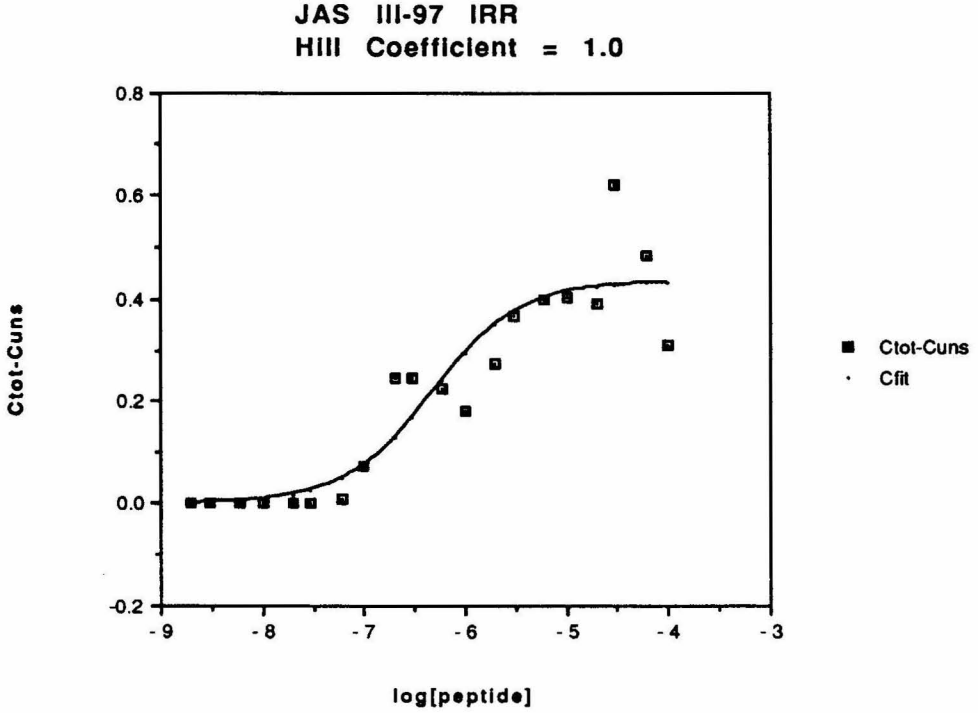

JAS III-96 IRL
Hill Coefficient = 1.0
Ctot-Cuns
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

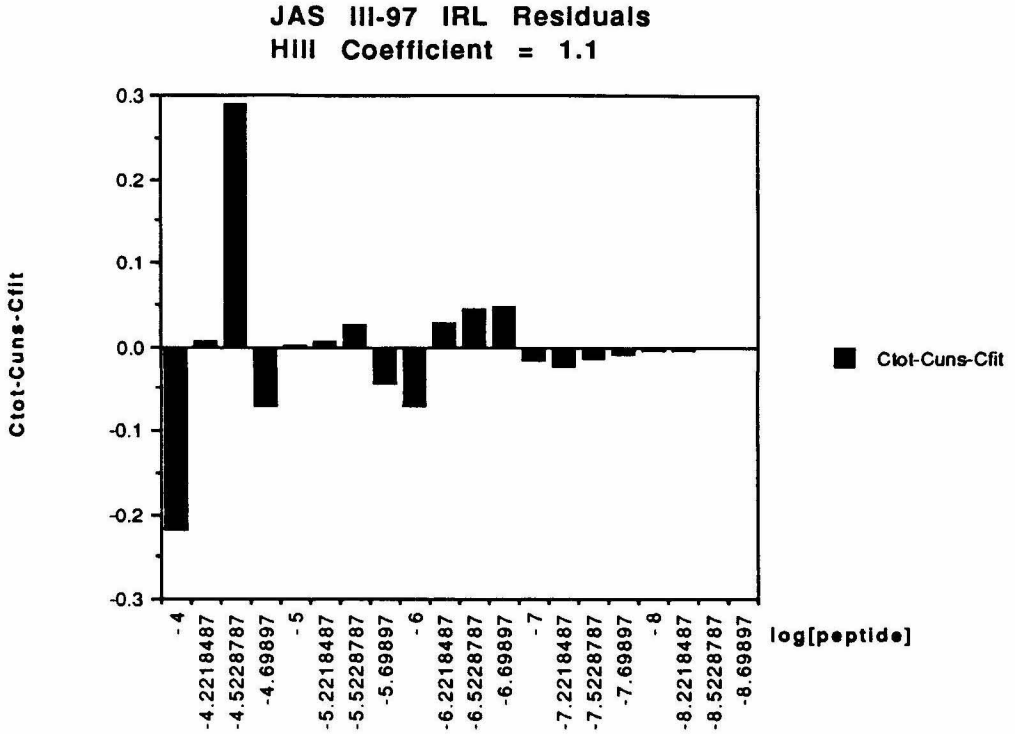

JAS III-96 IRR
Hill Coefficient = 1.0
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

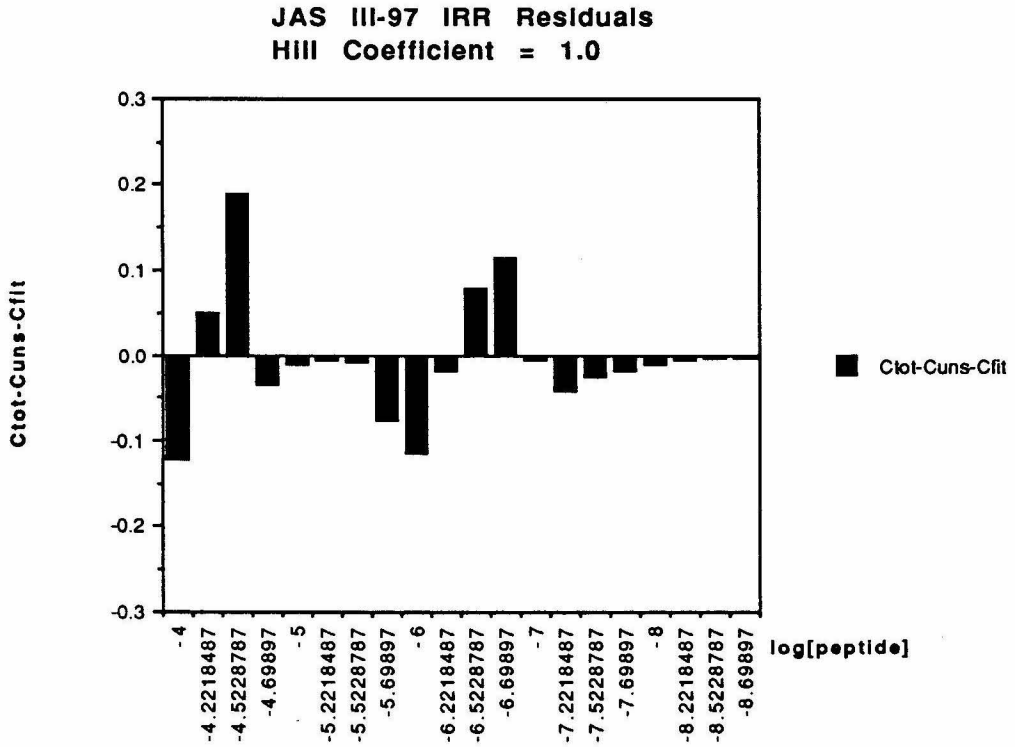

JAS III-96 IRL Residuals
Hill Coefficient = 1.0
Ctot-Cuns-Cfit
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit


JAS III-96 IRR Residuals
Hill Coefficient = 1.0
Ctot-Cuns-Cfit
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

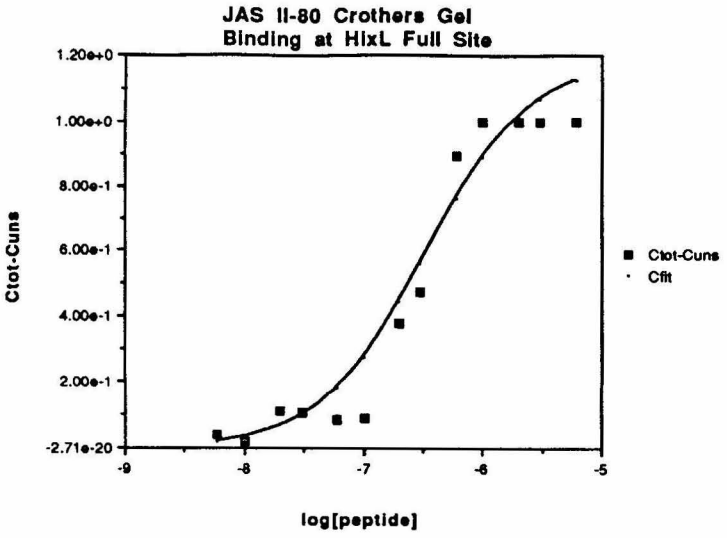

JAS III-97 IRL
Hill Coefficient = 1.1
Ctot-Cuns
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

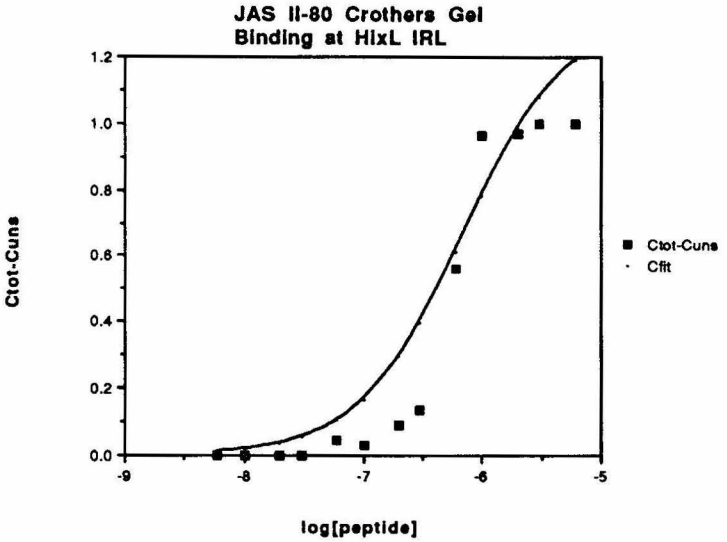

JAS III-97 IRR
Hill Coefficient = 1.0
Ctot-Cuns
0.8
0.6
0.4
0.2
0.0
-0.2
-9
-8
-7
-6
-5
-4
-3
log[peptide]
Ctot-Cuns
Cfit

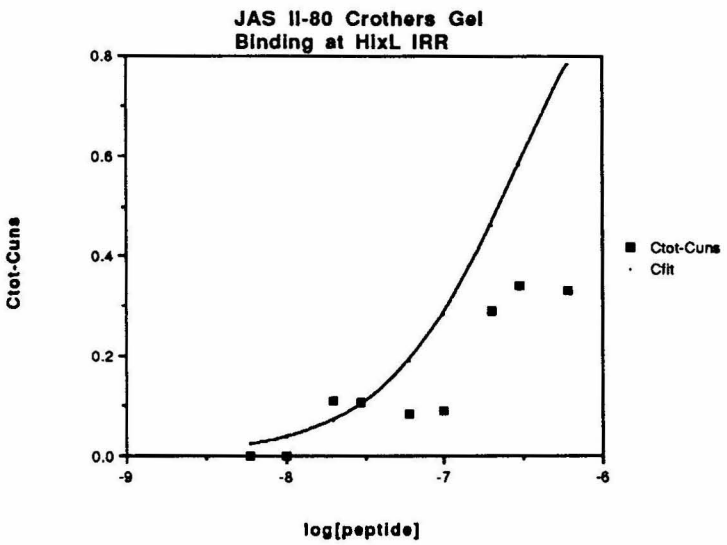

JAS III-97 IRL Residuals
Hill Coefficient = 1.1
Ctot-Cuns-Cfit
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

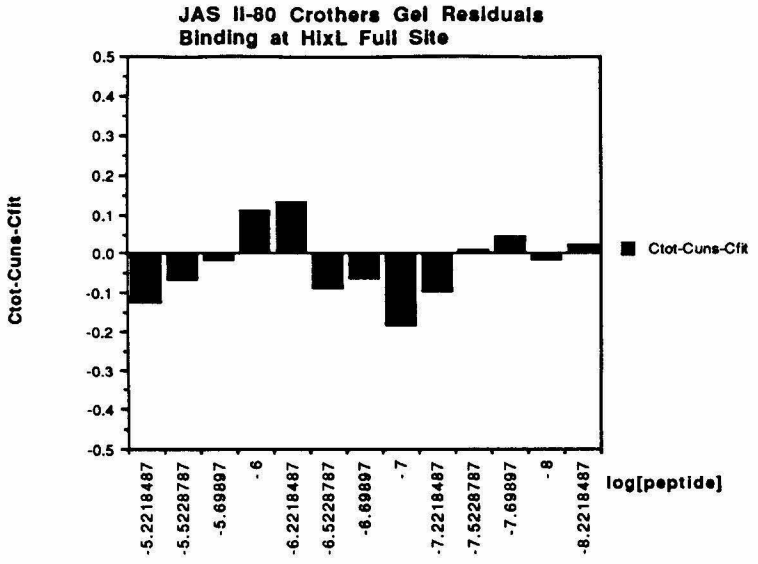

JAS III-97 IRR Residuals
Hill Coefficient = 1.0
Ctot-Cuns-Cfit
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-4
-4.2218487
-4.5228787
-4.69897
-5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
log[peptide]
Ctot-Cuns-Cfit

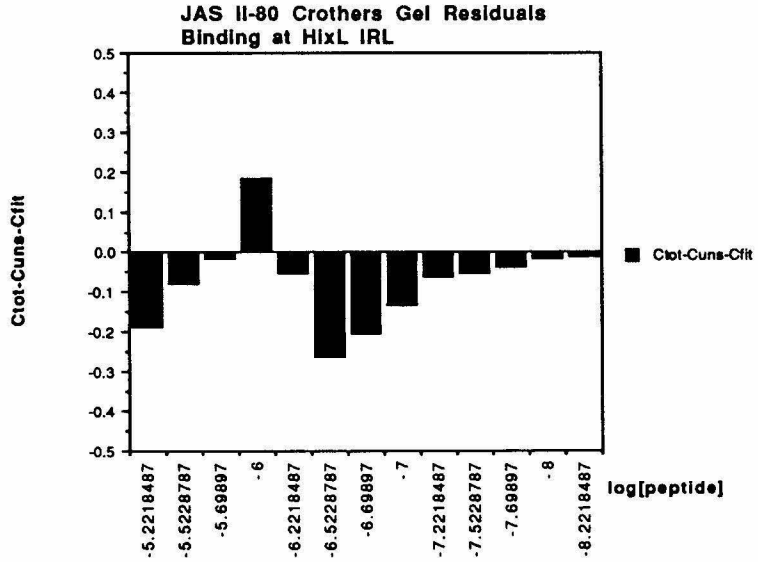
[Fe•EDTA]Hin(139-190). On the following pages are shown the binding isotherms and residuals for Crothers' gels JAS II-80, 81 and 95. On top are data for the *hix*L full site, followed by the*hix*L IRL binding site, and on the bottom are data for the *hix*L IRR binding site. Residuals measure the difference between the obtained data points and the fit curve.

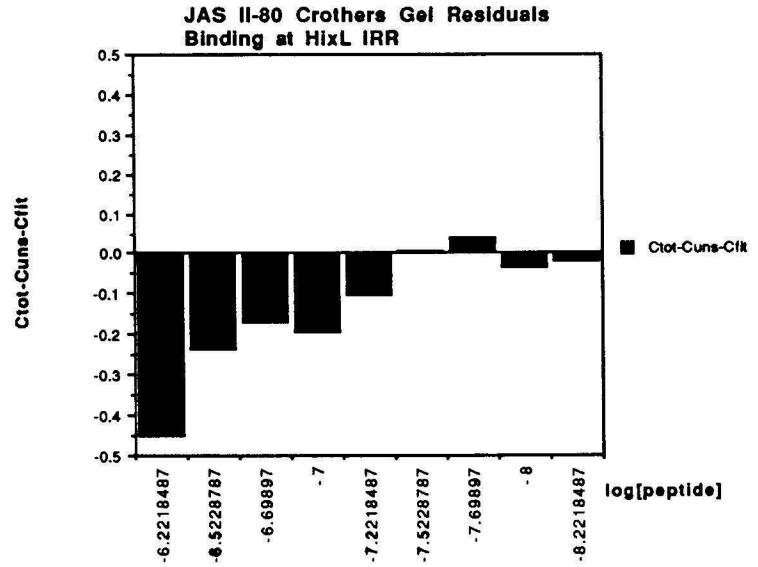

JAS II-80 Crothers Gel
Binding at HixL Full Site
1.20e+0
1.00e+0
8.00e-1
6.00e-1
4.00e-1
2.00e-1
-2.71e-20
Ctot-Cuns
-9
-8
-7
-6
-5
log[peptide]
Ctot-Cuns
Cfit

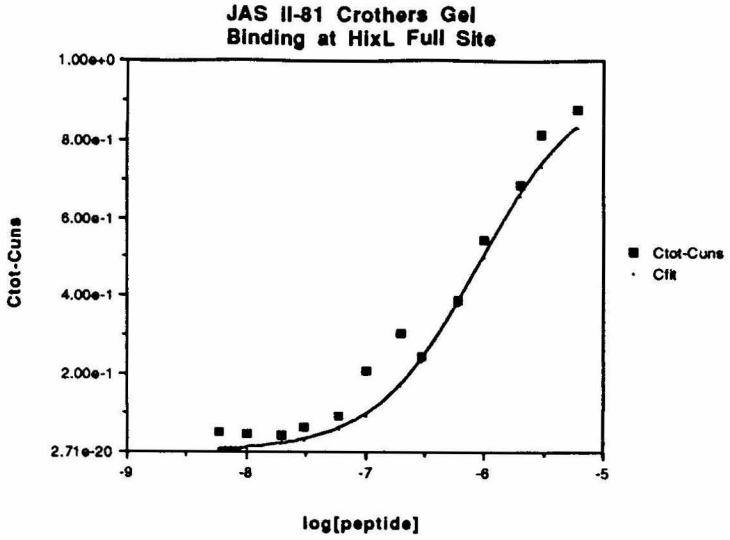

JAS II-80 Crothers Gel
Binding at HixL IRL
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Ctot-Cuns
-9
-8
-7
-6
-5
log[peptide]
Ctot-Cuns
Cfit

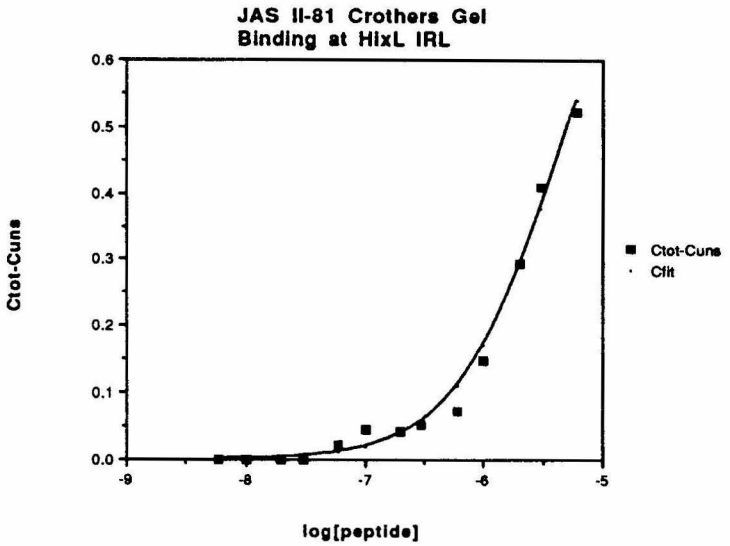

JAS II-80 Crothers Gel
Binding at HixL IRR
0.8
0.6
0.4
0.2
0.0
Ctot-Cuns
-9
-8
-7
-6
log[peptide]
Ctot-Cuns
Cfit

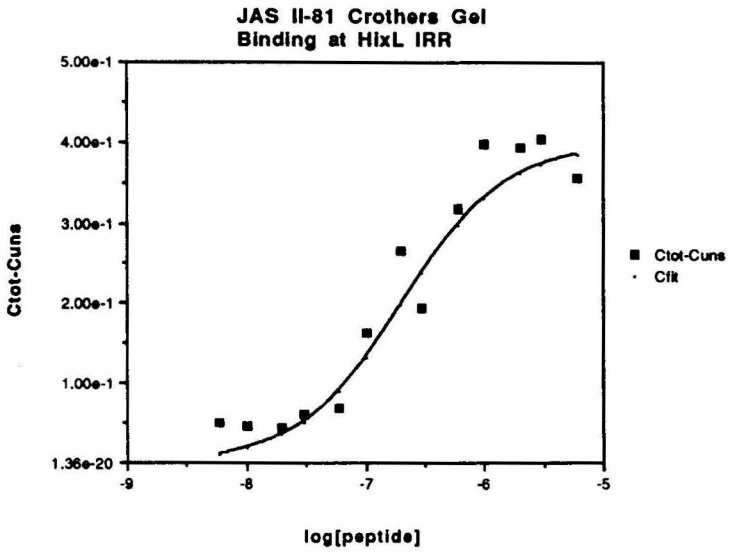

JAS II-80 Crothers Gel Residuals
Binding at HixL Full Site
Ctot-Cuns-Cfit
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
log[peptide]
Ctot-Cuns-Cfit

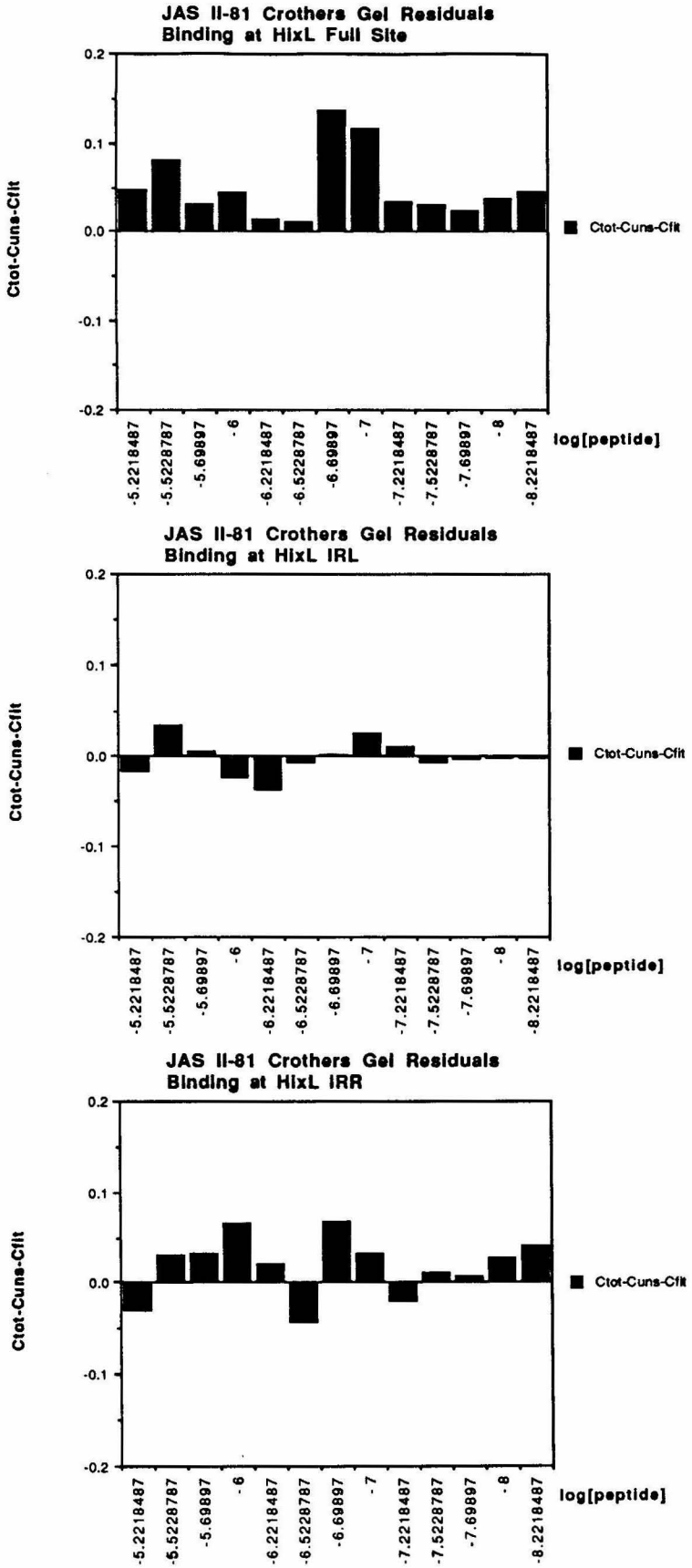

JAS II-80 Crothers Gel Residuals
Binding at HixL IRL
Ctot-Cuns-Cfit
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
log[peptide]
Ctot-Cuns-Cfit

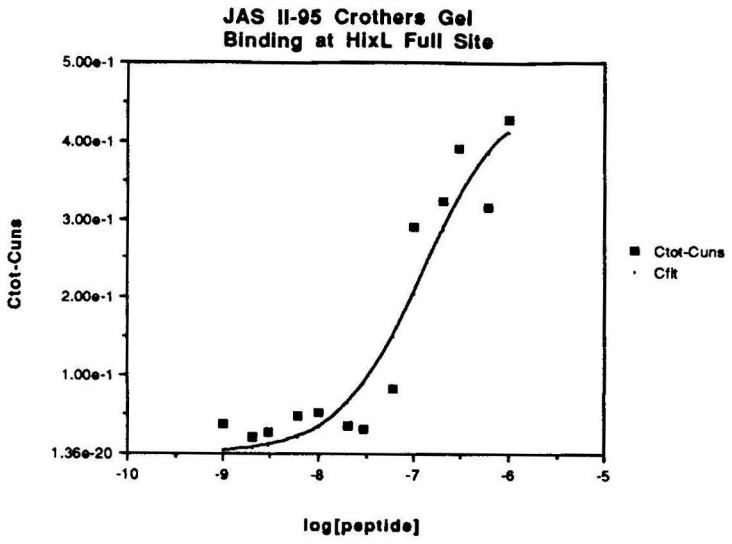

JAS II-80 Crothers Gel Residuals
Binding at HixL IRR
Ctot-Cuns-Cfit
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
log[peptide]
Ctot-Cuns-Cfit

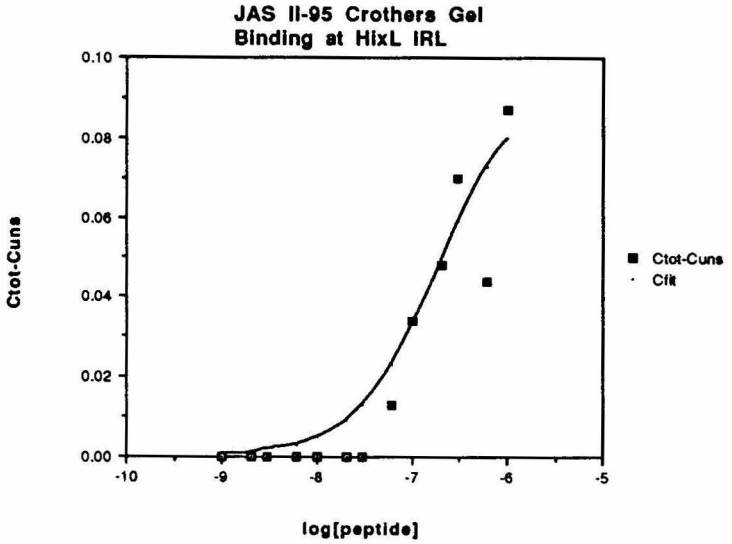

JAS II-81 Crothers Gel
Binding at HixL Full Site
1.00e+0
8.00e-1
6.00e-1
4.00e-1
2.00e-1
2.71e-20
Ctot-Cuns
-9
-8
-7
-6
-5
log[peptide]
Ctot-Cuns
Cfit

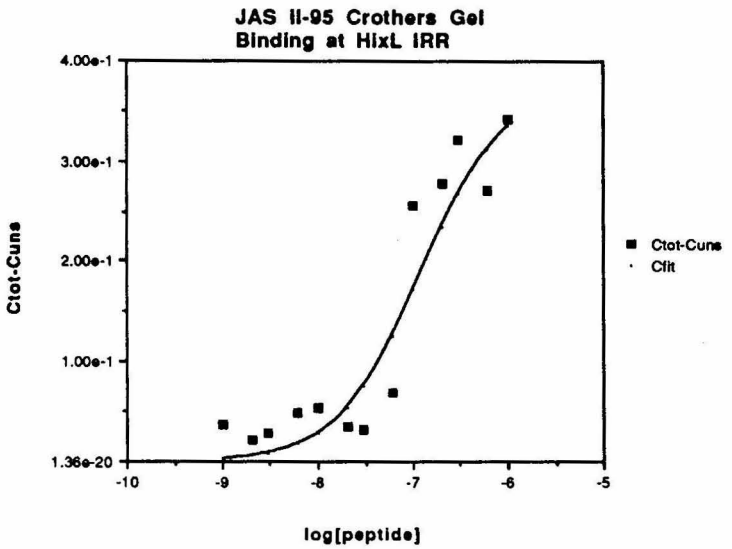

JAS II-81 Crothers Gel
Binding at HixL IRL
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Ctot-Cuns
-9
-8
-7
-6
-5
log[peptide]
Ctot-Cuns
Cfit

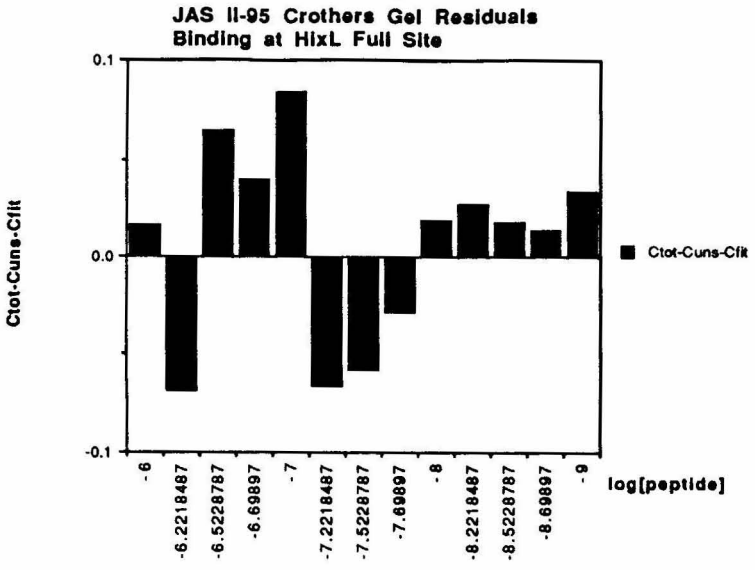

JAS II-81 Crothers Gel
Binding at HixL IRR
5.00e-1
4.00e-1
3.00e-1
2.00e-1
1.00e-1
1.36e-20
Ctot-Cuns
-9
-8
-7
-6
-5
log[peptide]
Ctot-Cuns
Cfit

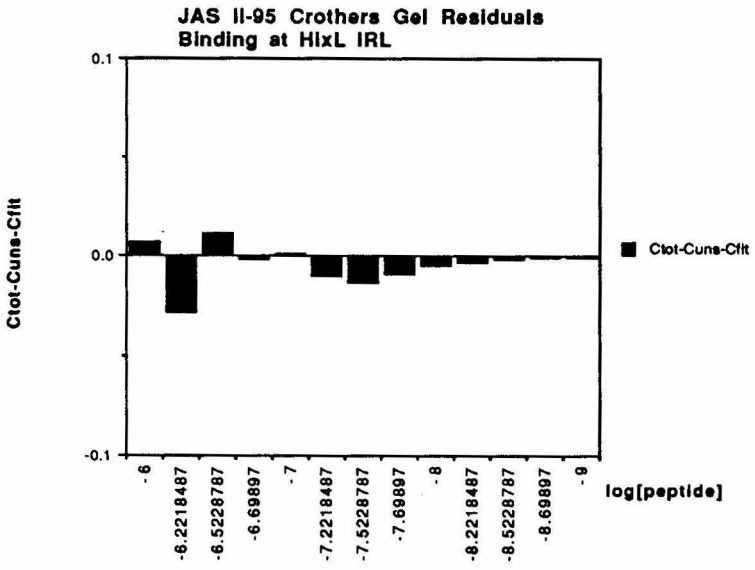

JAS II-81 Crothers Gel Residuals
Binding at HixL Full Site
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
Ctot-Cuns-Cfit
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
log[peptide]
JAS II-81 Crothers Gel Residuals
Binding at HixL IRL
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
Ctot-Cuns-Cfit
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
log[peptide]
JAS II-81 Crothers Gel Residuals
Binding at HixL IRR
Ctot-Cuns-Cfit
0.2
0.1
0.0
-0.1
-0.2
Ctot-Cuns-Cfit
-5.2218487
-5.5228787
-5.69897
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
log[peptide]

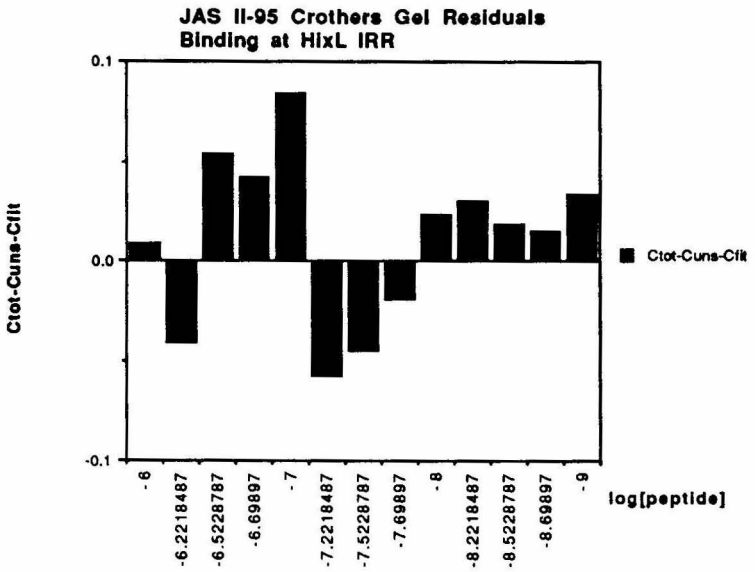

JAS II-95 Crothers Gel
Binding at HixL Full Site
5.00e-1
4.00e-1
3.00e-1
2.00e-1
1.00e-1
1.36e-20
Ctot-Cuns
-10
-9
-8
-7
-6
-5
log[peptide]
Ctot-Cuns
Cfit


JAS II-95 Crothers Gel
Binding at HixL IRL
0.10
0.08
0.06
0.04
0.02
0.00
Ctot-Cuns
-10
-9
-8
-7
-6
-5
log[peptide]
Ctot-Cuns
Cfit


JAS II-95 Crothers Gel
Binding at HixL IRR
4.00e-1
3.00e-1
2.00e-1
1.00e-1
1.36e-20
Ctot-Cuns
-10
-9
-8
-7
-6
-5
log[peptide]
Ctot-Cuns
Cfit


JAS II-95 Crothers Gel Residuals
Binding at HixL Full Site
Ctot-Cuns-Cfit
0.1
0.0
-0.1
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
-9
log[peptide]
Ctot-Cuns-Cfit


JAS II-95 Crothers Gel Residuals
Binding at HixL IRL
Ctot-Cuns-Cfit
0.1
0.0
-0.1
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
-9
log[peptide]
Ctot-Cuns-Cfit


JAS II-95 Crothers Gel Residuals
Binding at HixL IRR
Ctot-Cuns-Cfit
0.1
0.0
-0.1
-6
-6.2218487
-6.5228787
-6.69897
-7
-7.2218487
-7.5228787
-7.69897
-8
-8.2218487
-8.5228787
-8.69897
-9
log[peptide]
Ctot-Cuns-Cfit