Selection Studies of Zinc Finger - DNA Recognition

by

Edward John Rebar

B.S., Biochemistry
Rutgers University, 1989

Submitted to the Department of Biology in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

at the
Massachusetts Institute of Technology

May 1997

© 1997 Edward J. Rebar
All rights reserved

The author hereby grants to MIT permission to reproduce and to
distribute publicly paper and electronic copies of this thesis document in
whole or in part.

Signature of Author.. 5/30/97 ....
Department of Biology

Certified by .... May 30, 1997 ....
Carl O. Pabo, Professor of Biophysics and Structural Biology
Thesis Supervisor

Accepted by ....
Richard Young, Professor of Biology
Chairman, Biology Graduate Committee

Selection Studies of Zinc Finger - DNA Recognition

by

Edward John Rebar

Submitted to the Department of Biology in partial fulfillment
of the requirements for the degree of Doctor of Philosophy

Abstract

This thesis describes the development and application of methods for affinity-selecting zinc fingers with new DNA sequence specificities. These methods are used in studies of zinc finger-DNA recognition and in strategies for making proteins with desired specificities.

Chapter 1 discusses three related areas which provide the context and motivation for these studies. First, key features of DNA-binding proteins are described, with emphasis on the modular nature of these proteins and the possibility of using 'mix and match' approaches to design novel properties. Next, this chapter considers instances when such design efforts might require DNA-binding domains with novel specificities, and discusses the advantages of using selection methods for obtaining domains which bind desired DNA sequences. Finally, the zinc finger is introduced as one of nature's most adaptable DNA-binding motifs, and principles of zinc finger-DNA recognition are discussed and briefly reviewed.

Chapter 2 provides an overview of phage display methodology and describes the development of a phage display system for zinc finger proteins. A gene encoding the three fingers of Zif268 was cloned into the genome of filamentous phage as a fusion with a phage coat protein gene. Conditions were discovered which enabled the resultant 'Zif phage' to specifically interact with Zif target site on a solid support.

Chapter 3 describes experiments which demonstrated that phage display selection methods could be used to obtain fingers with new DNA sequence specificities. Combinatorial cassette mutagenesis was used to randomize key base - contacting residues of Zif finger 1. The resultant library of proteins was expressed on the surface of phage, and affinity selection methods were used to isolate proteins with new DNA-sequence specificities. These experiments were published as "Zinc Finger Phage: Affinity Selection of Fingers with New DNA-Binding Specificities" (Rebar, E.J. and Pabo, C.O. (1994) *Science*, **263**, 671).

The appendix to chapter 3 provides expanded protocols for chapter 3 and a more thorough discussion of several issues related to the construction and selection of zinc finger phage libraries. This appendix was published as part of the chapter: "Phage Display Methods for Selecting Zinc Finger Proteins with Novel DNA-Binding Specificities" (Rebar, E.J., Greisman, H.A., and Pabo, C.O. (1996) *Methods Enzymol.*, **267**, 129).

Chapter 4 describes a series of selections designed to yield pools of zinc fingers which bind to subsites from a key region of the HIV-1 promoter. Twenty selections were performed, of which sixteen yielded pools of fingers with desired specificities. Consensus sequences were determined for these successful pools. These consensus sequences, combined with the results of structural studies and other selection experiments, reveal the base contact preferences of the zinc finger motif. These preferences are described, and context effects which may influence these preferences are briefly discussed.

Chapter 5 describes an initial attempt to make multifinger proteins with completely novel specificities. Finger pools obtained in chapter 4 were linked together to form libraries of three-finger proteins, and these libraries were selected for binding to target sequences in the promoter region of HIV-1.

Thesis supervisor: Carl O. Pabo

To my parents, Edward and Leona Rebar,
and my sisters, Vicky, Amy and Sarah.

Acknowledgments

Numerous people have helped to make my years at MIT both enjoyable and intellectually enriching. In particular, Carl Pabo, my advisor, has played an especially positive role in my graduate career. Carl's intellectual intensity and sincere concern for his students makes him an outstanding mentor, and it has been my privilege and pleasure to work with Carl during these years. I wish to thank Carl for his guidance, patience, and candid advice throughout these studies, and also for providing a limitless supply of free lab coffee.

I wish also to thank members of my thesis and defense committees - Bob Sauer, Peter Kim, David Baltimore, Tania Baker and Ann Hochschild - for their time and insightful comments. I am especially grateful to Bob for allowing me to use his lab during the initial stages of this work.

Many members of the Pabo lab, both past and present, have helped to make my graduate studies more productive and fun. To Mark Rould, Amy Dunn, Tim Benson, Monicia Elrod-Erickson, Scot Wolfe, Ernest Fraenkel, Dan Chasman, Joel Pomerantz, WenQing Xu, Nikola Pavletich, Jin-Soo Kim, Sandra Fay-Richard, Bryan Wang, Juli Klemm, Kristen Chambers, and Harvey Greisman, I offer my gratitude for many helpful discussions and for many good times. In addition, I thank past members of the Sauer lab - Wendell Lim, Brigitte Raumann, Bronwen Brown, and Jim Hu - for helping me to learn molecular biology and for patiently answering many really dumb questions.

Outside of the lab, I have had the good fortune to experience many friendships and happy moments which are too numerous to mention here. However, I would like to thank Jun Horiuchi, Arie Berggrun and Mike Giffin for their friendship during these past few years and for providing many laughs and excellent excuses to grab some beers and shoot the breeze.

Finally, I would like to thank my family for encouragement and support throughout these studies, and for helping me to retain a perspective of the things that really matter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Sequence-Specific DNA-Binding Proteins:

General Themes:

Sequence-specific DNA-binding proteins play key roles in cell biology. By binding to particular DNA sites, these proteins define the precise chromosomal locations of important cellular events, including replication initiation[1], transcription initiation[2], V(D)J recombination[3, 4], and spindle attachment during mitosis[5]. These proteins also contribute to the properties of large chromosome segments, such as telomeres[6] and regions of transcriptional regulation. Estimates from large-scale sequencing efforts suggest that site-specific DNA-binding proteins represent from 1% to 10% of all eukaryotic genes[7]. The importance of these proteins is underscored by the consequences of dysfunction. Mutations which disrupt protein-DNA recognition have been shown to cause numerous heritable diseases[8-11] and to contribute to the development of cancer[12, 13].

Because of their central roles in cell biology, DNA-binding proteins have been the focus of intense study. Original efforts centered on the characterization of prokaryotic transcriptional regulators such as the lac and lambda repressor proteins[14-16]. Subsequent work has encompassed hundreds of DNA-binding proteins from numerous organisms. Several broad themes have emerged from this work. First, contacts between bases and proteins in the major groove play a key role in protein-DNA recognition. As pointed out by Rich and coworkers[17], the major groove edge of each base-pair features a distinctive pattern of hydrogen bonding and methyl groups (figure 1).

Proteins can recognize these patterns by displaying complementary arrays of hydrogen-bonding and hydrophobic residues on their surface, and this 'direct readout' of base composition plays a key role in sequence recognition by many DNA-binding proteins[18]. Other modes of protein-DNA recognition are also observed, such as minor groove binding[19, 20] and the 'sensing' of sequence-dependent DNA distortability[21], but to date considerably fewer examples of these recognition modes have been reported.

A second observation of these studies is that many DNA-binding proteins can be grouped into families which use structurally similar domains for DNA recognition. Thus, although thousands of DNA-binding proteins have been identified, most recognize DNA using one of just a dozen or so different types of DNA-binding domains, or 'motifs'. Figure 2 shows the structures and DNA docking arrangements of three well-characterized motifs. Most motifs characterized to date share several key features. First, most motifs insert an α–helix into the major groove upon binding to DNA, and this enables amino acids at key helix positions to make base contacts[18]. Other arrangements are also observed, including base recognition by loops[22, 23] and ß-sheets[24, 25], but these appear to be less common. Second, to varying degrees, most motifs may be adapted to bind new DNA sequences. This is usually accomplished by the replacement of amino acids which make base contacts. Third, despite this adaptability, a given motif typically exhibits a conserved structure and docking arrangement in different protein-DNA complexes. This last fact simplifies the study of protein-DNA recognition, since principles determined for one protein-DNA

complex are often applicable to other complexes involving the same motif.

A third theme to emerge from these studies is that many DNA-binding proteins possess a modular organization. Often, the DNA-binding domains work independently of, and are separable from, domains responsible for other protein functions. It appears that most eukaryotic transcription factors feature this organization[26, 27]. In addition, certain restriction enzymes[28] and recombinases[29] possess separable DNA-recognition and catalytic domains. This organization facilitates a reductionist approach to the study of DNA-binding proteins -- a given protein may be characterized one domain at a time. This organization also suggests that one may 'mix and match' DNA-binding domains with other protein functional regions to design DNA-binding proteins with new and useful properties.

'Mix and Match' Design Possibilities:

Recent studies have demonstrated some of the potential applications of a 'mix and match' approach to DNA-binding protein design. In most of these studies DNA-binding domains were swapped between proteins or grafted onto proteins which previously did not bind to DNA. For example, restriction enzymes with new specificities were constructed by grafting DNA-binding domains onto nonspecific nucleases[30] and by replacing the DNA-binding domains of modular restriction enzymes such as Fok 1[31, 32]. In other studies, drug-inducible eukaryotic transcription factors were created by splicing together novel combinations of DNA-binding domains and dimerization

elements[33] or activation regions[34]. In yet other studies, DNA-binding domains were linked to a variety of different peptides, including combinatorial peptide libraries, to create key components of *in vivo*[35, 36] and *in vitro*[37] selection systems. Finally, in several recent experiments, two or more DNA-binding domains were grafted together to create hybrid proteins with composite sequence specificities[38-41]. The success of these studies illustrates the functional modularity of many DNA-binding domains and other protein functional regions. Given this modularity, and given the large number and ongoing discovery of new types of functions (e.g., recombinases, nucleases, methylases, transcriptional activators and repressors etc.), 'mix and match' design strategies have the potential to yield diverse new DNA-binding proteins with useful properties and applications.

To date, most of these design efforts have simply 'borrowed' DNA-binding domains from other proteins in order to achieve appropriate sequence recognition. This approach has been suitable because many applications are compatible with proteins exhibiting a wide range of sequence specificities. For example, the selection applications described above required site-specific DNA-binding proteins, but virtually any specificity could have been used successfully in these studies[35-37]. In contrast, some of the most exciting potential applications for designed DNA-binding proteins require the recognition of particular, predetermined DNA sequences. For example, potential medical applications could require the recognition of particular gene alleles, characteristic viral DNA sequences, discrete chromosomal breakpoints, or chemical modifications of key base-pairs. Also, possible research uses (e.g. the

design of transcriptional regulators for novel gene combinations) might require the recognition of specific sequences common to arbitrary sets of promoters. To design DNA-binding proteins for such applications it will not be sufficient to simply 'borrow' DNA-bindng domains from other proteins. Rather, such applications will require the development of methods for making DNA-binding domains with novel, predetermined DNA sequence specificities.

Methods for Making DNA-Binding Domains With Desired Specificities:

The most obvious way to make a domain with a novel, predetermined DNA sequence specificity is to 'reprogram' a DNA-binding motif so that it binds to the desired target sequence. Given the importance of base contacts for DNA recognition, one would expect this to require the mutation of key base contacting positions to residues which could interact favorably with the chosen target site. How difficult might this be? It is clear that many amino acids prefer to contact particular bases. For example, glutamine and arginine may adopt especially favorable hydrogen-bonding geometries with adenine and guanine, respectively (figure 3A)[17], and the frequent occurrence of these interactions in cocrystal structures suggests that these contacts are especially favorable for sequence recognition. At first glance, it might appear that mutations suitable for 'reprogramming' a DNA-binding motif could be identified based solely on these considerations.

However, two key factors complicate this approach. First, position- and orientation- dependent context effects can swamp out

intrinsic preferences for certain contacts. Thus the general preferences described above (e.g. 'arginine recognizes guanine') may not apply at a particular base contact position. For example, in the zinc finger motif, two base contacting positions favor arginine-guanine contacts, but a third key position, which is closer to the bases, favors a *histidine*-guanine interaction (figure 3B). It appears that the side chain of arginine may be too long to interact effectively with guanine at this position. Second, it is clear that base contacting amino acids can interact with each other in ways that may modulate specificity. Figure 3C shows one such interaction, which is observed in many zinc finger-DNA complexes. Such interactions, which are quite diverse and difficult to predict, must be anticipated in any attempt to 'reprogram' a DNA-binding motif.

Given these considerations, which methods might successfully provide DNA-binding domains with desired specificities? Computational methods involving the explicit calculation of interaction energies may eventually provide the most powerful approaches for dealing with the complexities described above. However, given our inadequate understanding of the energetics of macromolecular recognition, such strategies are currently not feasible. Other computational approaches, which survey protein-DNA contact databases to identify interactions appropriate for a given positional- and orientational-context[42], offer greater promise in the short term but remain for the most part undeveloped. Simpler methods, including the various proposed protein-DNA recognition 'codes' may be useful in certain cases where limited changes are made to a well characterized

protein-DNA complex[43-45], but as yet no such 'code' has been used to make a DNA-binding domain with a novel, predetermined specificity.

At present, perhaps the most promising approach for making DNA-binding proteins with desired specificities is via selection methods. In these experiments, random mutations are introduced at one or several base contacting amino acids of a DNA-binding domain and the resulting library is selected for binding to a target DNA sequence. A powerful feature of selections is that very few assumptions are required to implement these methods. Often, it is sufficient merely to know which amino acids may make sequence-specific DNA contacts in order to accomplish an effective, efficient selection for new sequence specificities. An additional attraction of selection methods is that they provide information which can further our understanding of protein-DNA recognition. A variety of general, powerful, systems for selecting novel DNA-binding proteins have been described[1, 46-48] In chapter two, I outline the development of a phage-based selection system for this purpose.

Zinc Fingers:

Irrespective of the approach used for making proteins with new DNA specificities, it is necessary to choose a motif for such studies. A key concern in this choice should be DNA-binding adaptability. Ideally it should be possible to adapt the chosen motif to bind to a diverse range of sequences. This is an important consideration, because not all motifs are equal in this regard. For example, most natural binding sites for the helix-loop-helix motif are variations of the sequence

CAnnTG[49], and most binding sites for the C6 zinc cluster motif contain multiple 'CGG' sequences in various spacings and orientations[50-52]. This limited diversity in natural target sequences suggests that such motifs may not be very adaptable. Other motifs exhibit more flexibility. For example, homeodomains may be adapted to bind target sequences of the general form nAAnnn[43]. Binding sites the various ribbon-helix-helix proteins are also quite diverse[53], although relatively few of these proteins have been characterized.

Perhaps the most attractive motif for such studies is the zinc finger. The zinc finger motif, which was first recognized in transcription factor IIIA (TFIIIA)[54], appears to offer an especially adaptable framework for constructing proteins with novel DNA sequence specificities. The binding sites of naturally-occurring zinc finger proteins are very diverse (table 1), and so it is clear that this motif may be adapted to bind to a wide variety of DNA sequences. The zinc finger is also one of the most common eukaryotic DNA-binding motifs[55, 56], and it has been estimated that the human genome encodes several hundred zinc finger proteins[55, 57]. The function of the majority of these proteins is unknown, but it is clear that zinc finger proteins play many varied roles in cell biology. Demonstrated functions include, ATP-dependent nucleosome disruption[58] and transcription activation[59], repression[60], and initiation[61]. The use of zinc fingers by a large number and wide variety of DNA-binding proteins further underscores the functional versatility of this motif.

Structurally, the zinc finger is quite simple. The typical finger contains about 30 amino acids, of which only seven are key consensus residues (figure 4). Numerous NMR[62-65] and crystallographic[66-69]

studies show that this sequence folds into a compact structure consisting of an α–helix and two strands of ß-sheet (figure 4). In the folded finger, the consensus cysteine and histidine residues coordinate a zinc ion, which provides much of the folding stability for the finger. Complementary studies demonstrate the critical importance of zinc binding for finger folding and function. For example, removal of zinc from a zinc finger protein causes unfolding of fingers and loss of DNA-binding activity. On the other hand, 'zinc finger' peptides consisting of little more than the key consensus residues and alanines adopt the proper fold upon binding of zinc[70].

The typical zinc finger protein contains multiple fingers, although some proteins contain just a single finger[65]. To date, the maximum number of fingers found in a single protein is 37[71]. It is clear that many multifinger proteins use at least some of their fingers for purposes other than DNA recognition[68, 72]. Other functions which have been reported for zinc fingers include roles in mediating protein-protein interactions[73-76], RNA recognition[77-79] and possibly binding to RNA-DNA hybrid duplexes[69, 80].

Zinc Finger-DNA recognition:

To date, four x-ray cocrystal structures and one NMR solution structure have been determined for natural zinc finger proteins complexed with their target sites[65-69, 81]. These structures illustrate several key features of zinc finger - DNA recognition. First, when a multifinger protein binds to DNA, its tandem fingers typically 'line up' in the major groove with the amino-termini of the α–helices in close

proximity to the bases (figure 5). In this binding arrangement, the fingers interact with a series of adjacent DNA subsites and typically space themsleves along the DNA with a periodicity of one finger every three base pairs. This general organization of multifinger protein - DNA complexes is illustrated by the sketch of the Zif268 - DNA cocrystal structure provided in figure 5.

These structures also show that four positions within the zinc finger motif make most of the base contacts. These 'base contact positions' include the residue immediately preceding the α–helix, and include the second, third and sixth residues of this helix (positions '-1', '2', '3' and '6', figure 6). A variety of functional studies support the importance of these positions for base recognition. For instance, several studies have shown that it is possible to impart the specificity of one finger onto another by exchanging this region of the α–helix[82-84]. Also, alanine scanning mutagenesis of the ADR1 protein implicated amino acids in this region in DNA-binding specificity[85]. Also, a mutational reversion study of NGF1-A zinc fingers was consistent with the important role of these fingers in base contacts[86]. Furthermore, a statistical survey of residue conservation in the zinc finger motif was consistent with three of these positions -- '-1', '3' and '6' -- serving as key determinants of DNA-binding specificity in the zinc finger motif[56].

These structures also show that the potential length of the binding site for an individual zinc finger is about four base pairs. Within this site, contacts to a given base pair typically originate from a single base contact position of the zinc finger motif (figure 7). Since zinc fingers of multifinger proteins exhibit a three - base pair periodicity but may contact four base pairs of DNA, the binding sites of

adjacent fingers in a multifinger protein will typically overlap by one base-pair. However, because positions 2 and 6 typically contact bases on opposite DNA strands (figure 7), this arrangement does not necessarily lead to steric conflicts. Thus, a multifinger protein bound to DNA can be viewed as a series of tandem fingers bound to adjacent and partially overlapping subsites.

Given this organization, and given the existence of zinc finger proteins containing up to 37 fingers, it seems natural to wonder about the maximum number of tandem fingers which may complex with DNA. Several studies and obsevations suggest possible limitations in this respect. For example, binding studies of proteins containing three, four or five fingers consistently show that adding a fourth and/or fifth finger to a three finger protein yields relatively modest affinity gains[87-89]. Also, cocrystal structures of natural zinc finger proteins, which include two-, three-, four-, and five-finger proteins, exhibit a maximum of four tandem fingers bound to DNA[66-69, 81]. [In the five finger structure (GLI), one finger does not touch the DNA at all, and one of the remaining four fingers makes no base contacts[68].] Finally, to date, no zinc finger protein has been proven to use more than about four tandem fingers to bind to DNA. Together, these considerations suggest that restraints exist which limit the maximum number of fingers that may productively contact DNA, and that these restraints may exert effects in complexes containing as few as about five tandem fingers. The mechanism for this limitation is unclear. However, zinc finger proteins clearly distort DNA upon binding[90, 91] and it has been suggested that the DNA may operate as a 'molecular spring' in zinc finger-DNA recognition. According to this proposal, the binding of each

additional finger of a multifinger protein requires increasing amounts of energy in order to further distort the DNA[87].

The tandem, periodic binding of fingers in protein-DNA complexes has also suggested the exciting possibility that zinc finger-DNA recognition might be functionally modular. If this were true, then it could be possible to 'mix and match' zinc fingers to obtain proteins with novel DNA specificities, provided that constraints imposed by overlapping subsites were satisfied. Certain design studies and biological systems provide some evidence for this proposal. For example, studies by Berg and coworkers have demonstrated that it is possible in at least some cases to rearrange the fingers of a designed zinc finger protein to obtain a new protein with the expected DNA-binding specificity[92]. In addition, mRNAs encoding certain natural zinc finger proteins undergo alternate splicing which brings together different sets of zinc fingers and creates proteins with alternate DNA-binding specificities[93, 94]. However, instances of rearranged zinc finger proteins exhibiting very poor and/or nonspecific DNA binding have also been reported[40, 92], and so it is clear that zinc fingers are not perfectly modular and that not all combinations of zinc fingers work well together.

The DNA-binding arrangement of zinc fingers has also suggested the possibility that a molecular 'code' might govern zinc finger-DNA recognition. Since each of the four base contact positions typically contacts a different base position, it is tempting to speculate that one could create fingers with desired DNA specificities by simply programming in the proper amino acid at each base contact position. However, numerous biochemical studies have shown that zinc finger-

DNA recognition is not so simple. For example, mutational studies have demonstrated that the substitution of amino acids at a single base contact position can change DNA-binding specificity at multiple base-pairs[95, 96]. Other studies have shown that changing multiple base contact positions simultaneously can often yield fingers which bind DNA poorly and/or exhibit unexpected specificities[97]. Thus, although there are clearly position-dependent recognition preferences (e.g., asparagine at base contact position 3 usually recognizes an adenine at the appropriate base pair in the zinc finger subsite (figure 7)), it is clear that these preferences do not constitute simple, general code, and cannot be freely combined to obtain fingers with expected specificities.

Structural and biochemical studies of zinc fingers and zinc finger-DNA complexes also show that peptides adjacent to the amino-terminal zinc finger can contribute energy to DNA binding. In the tramtrack-DNA structure, for instance, the amino terminal finger has seven extra residues which form an additional strand of ß-sheet. These residues make no DNA-contacts, but rather assist binding by stabilizing the folded form of this finger[66]. The amino terminal finger of SWI5 also requires an extra 22 amino acids for proper folding and optimal binding to DNA. The NMR structure of this finger shows that the extra residues form a strand of ß-sheet and a short α–helix which may contact the DNA[98]. Adr1 has not been structurally characterized, but biochemical studies show that approximately 20 residues adjacent to the amino terminal finger are required for optimal binding. This region probably makes direct DNA contacts, since it contains many basic residues and since mutations which enhance binding can be

selected in this region[99]. Finally, a fragment of the GAGA protein consisting of just a single zinc finger and a 33 residue amino terminal peptide has been shown to exhibit the DNA binding specificity and high affinity of the full length protein[100]. A recent NMR structure of this fragment in complex with DNA suggests that the amino terminal peptide makes DNA contacts with the bases and phosphate backbone[65].

Approaches for Making Fingers With New Specificities:

In recent years numerous studies have described the construction of zinc fingers with new DNA sequence specificities. These studies have tested different methods for modifying zinc finger specificity and have also revealed insights into zinc finger - DNA recognition. These studies are summarized here, and are grouped into categories according to how the new fingers were obtained.

Swaps:

Several early studies sought to discover those zinc finger positions responsible for determining DNA-binding specificity. The basic strategy of these studies was to swap sets of residues between different fingers, and then to test the resulting fingers for DNA-binding specificity. Using this approach, Charnay and coworkers[82] substituted the base contacting positions of finger 2 of Krox-20 with sets of amino acids from the base contacting positions of other zinc fingers, and demonstrated the importance of these positions for determining binding preference[82, 101]. In related studies, Thiesen and

coworkers replaced the amino-terminal half of the α–helix of finger 2 in SP1 with the same region from several different fingers, and showed that the resulting fingers exhibited new DNA-sequence specificities[83, 84]. These experiments were important because they identified the specificity-determining regions of the zinc finger motif. They also demonstrated for the first time that it was possible to modify the DNA-binding specificity of individual fingers.

Screens:

In an attempt to isolate mutants of Adr1 with novel DNA-binding specificities, Young and coworkers used combinatorial cassette mutagenesis to individually randomize key base contact positions of Adr1 finger one. Fingers with new specificities were then identified by screening the resulting proteins for binding to variations of the Adr1 target sequence. Using this approach, three fingers were discovered which contained single amino acid substitutions and which preferred to bind DNA sites differing from the Adr1 target by single base changes. Young and coworkers then constructed three double mutant proteins (corresponding to all possible pairs of single mutations) and tested these proteins for binding specificity. Of the three double mutants, only one displayed the binding specificity expected from the properties of its constituent single mutations. This work showed that it was possible to obtain fingers with new DNA-binding specificities by changing amino acids at individual base contacting positions (as opposed to changing whole sets of base contacting residues). However, this work also showed that zinc finger-

DNA recognition could not be described as a simple position-dependent code[97].

Rule- and Database-Guided Studies:

In an extensive series of experiments, Berg and coworkers used a novel approach to make large numbers of fingers with new DNA specificities. In these experiments, the design of new fingers was guided by two factors: i) the favorable hydrogen-bonding geometries of certain amino acid-base contacts (e.g. such as those in figure 3) and ii) the frequencies of different types of amino acids at the base contact positions of naturally occurring zinc fingers. An example of this strategy is provided by their initial study, in which they desired to change the DNA-binding specificity of finger 2 of SP1 from GCGG to GCAG. Their approach was as follows: First, the arginine at position -1 of this finger, which normally contacts the underlined guanine, was changed to a glutamine. This substitution was chosen because the geometry of a glutamine - adenine interaction suggests that this contact should be both specific and energetically favorable (figure 3). Next, positions 2 and 3 were changed to serine and aspartate because, when position -1 is a glutamine, serine and aspartate are the most common residues at positions 2 and 3. The DNA-binding properties of the finger were then characterized, and it was demonstrated the finger possessed a new specificity[102]. In subsequent studies, a series of SP1 finger 2 variants were created using similar methods and the specificities of these fingers were characterized using rapid binding assays. These studies generated a considerable number of fingers with

new sequence specificities. However, it was often difficult to predict the specificities of the new fingers[92, 103].

In a second series of experiments, Berg and coworkers demonstrated the usefulness of a 'consensus finger' motif for constructing fingers with novel DNA specificities. In the first such study, a database of all known zinc fingers was used to derive a sequence for a 'consensus finger'. This finger was shown to fold properly and to bind zinc tightly[104]. Next, it was demonstrated that peptides containing three consensus fingers could bind to DNA with high affinity, provided that they contained the proper amino acids at the base contact positions[92]. Experiments similar to those described in the previous paragraph were used to make and characterize numerous consensus fingers exhibiting new DNA-binding specificities[105].

Selections:

To date all selections of zinc fingers for new specificities have used systems based on phage display. In initial studies, a phage display system was developed for the zinc fingers of Zif268. Base contact positions of finger 1 were then randomized using combinatorial cassette mutagenesis and the resulting library was selected for binding to three different four - base pair DNA sequences. Two of these selections yielded proteins which bound to the new target sites specifically and with high affinity[106]. Chapter 2 describes the development of this phage display system, and chapter 3 describes the use of this system to select fingers with novel specificities. The appendix to chapter 3 provides detailed protocols for chapters 2 and 3

and discusses key considerations for constructing zinc finger phage libraries and selecting fingers with new sequence specificities.

In several subsequent studies by others, each finger of Zif268 has been randomized and selected for binding to new DNA sequences. In one such study, Wells and coworkers randomized the four base contacting residues of Zif finger 1 and selected the resulting library for binding to a total of eight different target sequences. They achieved three successful selections which yielded new fingers that bound specifically to their selection targets. Intriguingly, they also observed that the overall charge of their selected residue sets tended to be zero, although many sets featured three or four charged amino acids[107]. In other studies, Choo and Klug randomized Zif finger 2 at its base contacting positions, and also at three additional positions ('1', '5' and '8'). They performed 28 selections and obtained 16 fingers with new sequence specificities[108]. In more recent studies, Greisman and Pabo performed a series of selections in which they randomized each finger of Zif at its base contacting positions and at two additional positions ('1' and '5'), and selected these fingers for binding to nine new sequences. Every selection was successful. In addition, these selections were performed as part of a wider effort to make three finger proteins which could bind to completely new DNA target sites, and these studies marked the first clear demonstration that the zinc finger motif could be used to make proteins with high affinity and specificty for entirely novel sequences[109]. Together these studies and others[110] have confirmed the importance of the base-contacting positions for determining finger specificity and the effectiveness of phage display methods for affinity - selecting new fingers.

In more recent studies, I have used phage display methods to select a library of zinc fingers for binding to an panel of twenty different four - base pair target sites from a region of the HIV-1 promoter surrounding the TATA site. In sixteen of these selections, pools of fingers were obtained which exhibit specificity and affinity for the selection targets. This selection experiment differs from previous studies in several key respects. First, its design provides a basis for assessing the DNA-binding adaptibility of the zinc finger motif. Second, it tests the influence of a variety of phosphate contacts and other variables on zinc finger recognition which were not considered in earlier studies. Also, the results of this study, when combined with structural information and the results of other selections, provides an updated picture of the preferred contacts of the zinc finger motif. These studies are discussed in chapter 4.

Finally, this study has yielded pools of fingers which may prove useful as subunits for the design of multifinger proteins which can bind sequences in the promoter of HIV-1. In chapter 5, I describe initial studies of a combinatorial 'mix and match' approach for using these pools to make multifinger proteins with novel, desired sequence specificities.

References:

1.   Li, J.L. and I. Herskowitz. Isolation of *ORC6*, a Component of the Yeast Origin Recognition Complex by a One-Hybrid System. Science, 1993. **262**: p. 1870-1974.

2.   Burley, S.K. The TATA box binding protein. Curr. Opin. Struct. Biol., 1996. **6**: p. 69-75.

3.   Difilippantonio, M.J., *et al.* RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. Cell, 1996. **87**: p. 253-262.

4.   vanGent, D.C., D.A. Ramsden, and M. Gellert. The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell, 1996. **85**: p. 107-113.

5.   Pluta, A.F., *et al.* The Centromere: Hub of Chromosomal Activities. Science, 1995. **270**: p. 1591-1594.

6.   Marcand, S., E. Gilson, and D. Shore. A Protein-Counting Mechanism for Telomere Length Regulation in Yeast. Science, 1997. **275**: p. 986-990.

7.   Adams, M.D., *et al.* 3,400 new expressed sequence tags identify diversity of transcripts in human brain. Nature Genet., 1993. **4**: p. 256-267.

8.   Sakai, T., *et al.* Oncogenic germ-line mutations in Sp1 and ATF sites in the human retinoblastoma gene. Nature, 1991. **353**: p. 83-86.

9.   Pelletier, J., *et al.* Germline Mutations in the Wilm's Tumor Suppressor Gene Are Associated with Abnormal Urogenital Development in Denys-Drash Syndrome. Cell, 1991. **67**: p. 437-447.

10.   Little, M., *et al.* DNA binding capacity of the WT1 protein is abolished by Denys-Drash syndrome WT1 point mutations. Hum. Mol. Genet., 1995. **4**: p. 351-358.

11.   Koivisto, U., *et al.* A single-base substitution in the proximal Sp1 site of the human low density lipoprotein receptor promoter as a cause of heterozygous familial hypercholesterolemia. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 10526-10530.

12.   Cho, Y., *et al.* Crystal Structure of a p53 Tumor Suppressor-DNA Complex: Understanding Tumorigenic Mutations. Science, 1994. **265**: p. 346-355.

13.   Little, M.H., *et al.* Zinc finger point mutations within the WT1 gene in Wilms tumor patients. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 4791-4795.

14.   Kercher, M.A., P. Lu, and M. Lewis. *Lac* repressor-operator complex. Curr Opin Struct Biol, 1997. **7**: p. 76-85.

15.   Lewis, M., *et al.* Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer. Science, 1996. **271**: p. 1247-1254.

16.   Sauer, R.T., S.R. Jordan, and C.O. Pabo. λ Repressor: A Model System For Understanding Protein-DNA Interactions and Protein Stability. Adv. Prot. Chem., 1990. **40**: p. 1-61.

17.   Seeman, N.C., J.M. Rosenberg, and A. Rich. Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl. Acad. Sci. U.S.A., 1976. **73**: p. 804-808.

18.   Ades, S.E. The Engrailed Homeodomain: Determinants of DNA-Binding Affinity and Sepecificity. Ph. D. thesis., 1995.

19.   Kissinger, C.R., *et al.* Crystal Structure of an engrailed Homeodomain-DNA Complex at 2.8 Å Resolution: A Framework for Understanding Homeodomain-DNA Interactions. Cell, 1990. **63**: p. 579-590.

20.   Schumacher, M.A., *et al.* Crystal Structure of LacI Member, PurR, Bound to DNA: Minor Groove Binding by α-helices. Science, 1994. **266**: p. 763-770.

21.   Koudelka, G.B., *et al.* DNA twisting and the affinity of bacteriophage 434 operator for bacteriophage 434 repressor. Proc. Natl. Acad. Sci. U.S.A., 1988. **85**: p. 4633-4637.

22.   Ghosh, G., *et al.* Structure of NF-κB homodimer bound to a κB site. Nature, 1995. **373**: p. 303-310.

23.   Muller, C.W., *et al.* Structure of the NF-κB p50 homodimer bound to DNA. Nature, 1995. **373**: p. 311-317.

24.   Raumann, B.E., *et al.* DNA recognition by β-sheets in the Arc repressor-operator crystal structure. Nature, 1994. **367**: p. 754-757.

25.   Somers, W.S. and S.E. Phillips. Crystal structure of the *met* repressor-operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. Nature, 1992. **359`**: p. 387-393.

26.   Pabo, C.O. and R.T. Sauer. Transcription Factors: Structural Families and Principles of Recognition. Annu. Rev. Biochem., 1992. **61**: p. 1053-1095.

27.   Frankel, A.D. and P.S. Kim. Modular Structure of Transctiption Factors: Implication for Gene Regulation. Cell, 1991. **65**: p. 717-719.

28.   Li, L., L.P. Wu, and S. Chandrasegaran. Functional domains in *Fok* I restriction endonuclease. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 4275-4279.

29.   Abdel-Meguid, S.S., *et al.* Cleavage of the site-specific recombination protein γδ resolvase: The smaller of two fragments binds DNA specifically. Proc. Natl. Acad. Sci. U.S.A., 1984. **81**: p. 2001-2005.

30.   Nagaoka, M., *et al.* A Novel Zinc Finger-Based DNA Cutter: Biosynthetic Design and Highly Selective DNA Cleavage. J. Am. Chem. Soc., 1994. **116**: p. 4085-4086.

31.   Kim, Y., J. Cha, and S. Chandrasegaran. Hybrid restriction enzymes: Zinc finger fusions to *Fok* I cleavage domain. Proc. Natl. Acad. Sci. U.S.A., 1996. **93**: p. 1156-1160.

32.   Kim, Y. and S. Chandrasegaran. Chimeric restriction endonuclease. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 883-887.

33.   Licitra, E.J. and J.O. Liu. A three-hybrid system for detecting small ligand-protein receptor interactions. Proc. Natl. Acad. Sci. U.S.A., 1996. **93**: p. 12817-12821.

34.   Gossen, M. and H. Bujard. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 5547-5551.

35.   Gyuris, J., *et al.* Cdi1, a Human G1 and S Phase Protein Phosphatase That Associates with Cdk2. Cell, 1993. **75**: p. 791-803.

36.   Colas, P., *et al.* Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. Nature, 1996. **380**: p. 584-550.

37.   Cull, M.G., J.F. Miller, and P.J. Schatz. Screening for receptor ligands using large libraries of peptides linked to the C terminus of the *lac* repressor. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 1865-1869.

38.   Park, C., J.L. Campbell, and W.A. Goddard. Protein stitchery: design of a protein for selective binding to a specific DNA sequence. Proc. Natl. Acad. Sci. U.S.A, 1992. **89**: p. 9094-9096.

39.   Pomerantz, J.L., P.A. Sharp, and C.O. Pabo. Structure-Based Design of Transcription Factors. Science, 1995. **267**: p. 93-96.

40.   Choo, Y., I. Sánchez-García, and A. Klug. *In vivo* repression by a site-specific DNA-binding protein designed against an oncogenic sequence. Nature, 1994. **372**: p. 642-645.

41.   Chatterjee, S. and K. Struhl. Connecting a promoter-bound protein to TBP bypasses the need for a transcriptional activation domain. Nature, 1995. **374**: p. 820-822.

42.   Pabo, C.O. and L. Nekludova. unpublished results. 1996.

43.   Damante, G., *et al.* A molecular code dictates sequence-specific DNA recognition by homeodomains. EMBO J., 1996. **15**: p. 4992-5000.

44.   Suzuki, M. and N. Yagi. DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 12357-12361.

45.   Lehming, N., *et al.* Mutant lac repressors with new specificities hint at rules for protein-DNA recognition. EMBO J., 1990. **9**: p. 615-621.

46.   Mossing, M.C., J.U. Bowie, and R.T. Sauer. A Streptomycin Selection for DNA-Binding Activity. Methods Enzymol., 1991. **208**: p. 604-619.

47.   Wang, M.M. and R.R. Reed. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. Nature, 1993. **364**: p. 121-126.

48.   Elledge, S.J., *et al.* Genetic selection for genes encoding sequence-specific DNA-binding proteins. Proc. Natl. Acad. Sci. U.S.A., 1989. **86**: p. 3689-3693.

49.   Ellenberger, T. Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. Curr. Opin. Struct. Biol., 1994. **4**: p. 12-21.

50.   Hellauer, K., M. Rochon, and B. Turcotte. A Novel DNA Binding Motif for Yeast Zinc Cluster Proteins: the Leu3p and Pdr3p Transcriptional Activators Recognize Everted Repeats. Mol. Cell. Biol., 1996. **16**: p. 6096-6102.

51.   Liang, S.D., *et al.* DNA Sequence Preferences of GAL4 and PPR1: How a Subset of Zn2Cys6 Binuclear Cluster Cluster Proteins Recognizes DNA. Mol. Cell. Biol., 1996. **16**: p. 3773-3780.

52.   Zhang, L. and L. Guarente. The C6 zinc cluster dictates asymmetric binding by HAP1. EMBO J., 1996. **15**: p. 4676-4681.

53.   Raumann, B.E., K.L. Knight, and R.T. Sauer. Dramatic changes in DNA-binding specificity caused by single residue substitutions in an Arc/Mnt hybrid repressor. Nature Struct. Biol., 1995. **2**: p. 1115-1122.

54.   Miller, J., A.D. McLachlan, and A. Klug. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J., 1985. **4**: p. 1609-1614.

55.   Pellegrino, G.R. and J.M. Berg. Identification and characterization of "zinc finger" domains by the polymerase chain reaction. Proc. Natl. Acad. Sci. U.S.A., 1991. **88**: p. 671-675.

56.   Jacobs, G.H. Determination of the base recognition positions of zinc fingers from sequence analysis. EMBO J., 1992. **11**: p. 4507-4517.

57.   Bray, P., *et al.* Characterization and mapping of human genes cncoding zinc finger proteins. Proc. Natl. Acad. Sci. U.S.A., 1991. **88**: p. 9563-9567.

58.   Tsukiyama, T., P.B. Becker, and C. Wu. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature, 1994. **367**: p. 525-532.

59.   Schmidt, M.C., Q. Zhou, and A.J. Berk. Sp1 Activates Transcription without Enhancing DNA-Binding Activity of the TATA Box Factor. Mol. Cell. Biol. 1989. **9**: p. 3299-3307.

60.   Witzgall, R., *et al.* The Kruppel-associate box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 4514-4518.

61.   Usheva, A. and T. Shenk. TATA-Binding Protein-Independent Initiation: YY1, TFIIB, and RNA Polymerase II Direct Basal Trnascription on Supercoiled Template DNA. Cell, 1994. **76**: p. 1115-1121.

62.   Lee, M.S., *et al.* Three-Dimensional solution Structure of a Single Zinc Finger DNA-Binding Domain. Science, 1989. **245**: p. 635-637.

63.   Omichinski, J.G., *et al.* High-Resolution Solution Structure of the Double Cys2His2 Zinc Finger from the Human Enhancer Binding Protein MBP-1. Biochemistry, 1992. **31**: p. 3907-3917.

64.   Klevit, R.E., J.R. Herriott, and S.J. Horvath. Solution Structure of a Zinc Finger Domain of Yeast ADR1. Prot. Struct. Funct. Genet., 1990. **7**: p. 215-226.

65.   Omichinski, J.G., *et al.* The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. Nat. Struct. Biol., 1997. **4**: p. 122-132.

66.   Fairall, L., *et al.* The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature, 1993. **366**: p. 483-487.

67.   Pavletich, N.P. and C.O. Pabo. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. Science, 1991. **252**: p. 809-817.

68.   Pavletich, N.P. and C.O. Pabo. Crystal Structure of a Five-Finger GLI-DNA Complex: New Perspectives on Zinc Fingers. Science, 1993. **261**: p. 1701-1707.

69.   Houbaviy, H.B., *et al.* Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. Proc. Natl. Acad. Sci. U.S.A., 1996. **93**: p. 13577-13582.

70.   Michael, S.F., *et al.* Metal binding and folding properties of a minimalist Cys2His2 zinc finger peptide. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 4796-4800.

71.   Klug, A. and J.W. Schwabe. Zinc fingers. FASEB J., 1995. **9**: p. 597-604.

72.   Kim, J., *et al.* The DNA-Binding and Enhancer-Blocking Domains of the *Drosophila suppressor of Hairy-wing* Protein. Mol. Cell. Biol., 1996. **16**: p. 3381-3392.

73.   Perkins, N.D., *et al.* An Interaction between the DNA-Binding Domains of RelA(p65) and Sp1 Mediates Human Immunodeficiency Virus Gene Activation. Mol. Cell. Biol., 1994. **14**: p. 6570-6583.

74.   Sun, L., A. Liu, and K. Georgopoulos. Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. EMBO J., 1996. **15**: p. 5358-5369.

75.   Maheswaran, S., *et al.* The WT1 gene product stabilizes p53 and inhibits p53-mediated apoptosis. Genes & Dev., 1995. **9**: p. 2143-2156.

76.   Lee, J., K.M. Galvin, and Y. Shi. Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. Proc. Natl. Acad. Sci. U.S.A., 1993. **90**: p. 6145-6149.

77.   Zang, W. and P.J. Romaniuk. Characterization of the 5 S RNA Binding Activity of *Xenopus* Zinc Finger Protein p43. J. Mol. Biol., 1995. **245**: p. 549-558.

78.   Theunissen, O., *et al.* RNA and DNA Binding Zinc Fingers in Xenopus TFIIIA. Cell, 1992. **71**: p. 679-690.

79.   Clemens, K.R., *et al.* Molecular Basis for Specific Recognition of Both RNA and DNA by a Zinc Finger Protein. Science, 1993. **260**: p. 530-533.

80.   Shi, Y. and J.M. Berg. Specific DNA-RNA Hybrid Binding by Zinc Finger Proteins. Science, 1995. **268**: p. 282-284.

81.   Elrod-Erickson, M., *et al.* Zif268 protein-DNA complex refined at 1.6A: a model system for understanding zinc finger-DNA interactions. Structure, 1996. **15**: p. 1171-1180.

82.   Nardelli, J., *et al.* Base sequence discrimination by zinc-finger DNA-binding domains. Nature, 1991. **349**: p. 175-178.

83.   Thiesen, H. and B. Schroder. Amino Acid Substitutions In The SP1 Zinc Finger Domain Alter The DNA Binding Affinity To Cognate SP1 Target Site. Biochem. Biophys. Res. Comm., 1991. **175**: p. 333-338.

84.   Thiesen, H. and C. Bach. Determination of DNA binding specificities of mutated zinc finger domains. FEBS, 1991. **283**: p. 23-26.

85.   Thukral, S.K., M.L. Morrison, and E.T. Young. Alanine scanning site-directed mutagenesis of the zinc fingers of transcription factor ADR1: Residues that contact DNA and that transactivate. Proc. Natl. Acad. Sci. U.S.A., 1991. **88**: p. 9188-9192.

86.   Wilson, T.E., *et al.* In Vivo Mutational Analysis of the NGF1-A Zinc Fingers. J. Biol. Chem., 1922. **267**: p. 3718-3724.

87.   Shi, Y. Ph. D. thesis. 1995.

88.   Limb, C., E.J. Rebar, and C.O. Pabo. unpublished. 1994.

89.   White, S.A. and C.O. Pabo. unpublished. 1997.

90.   Nekludova, L. and C.O. Pabo. Distinctive DNA conformation with enlarged major groove is found in Zn-finger-DNA and other protein-DNA complexes. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 6948-6952.

91.   Shi, Y. and J.M. Berg. DNA Unwinding Induced by Zinc Finger Protein Binding. Biochemistry, 1996. **35**: p. 3845-3848.

92.   Desjarlais, J.R. and J.M. Berg. Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. Proc. Natl. Acad. Sci. U.S.A., 1993. **90**: p. 2256-2260.

93.   Gogos, J.A., *et al.* Sequence Discrimination by Alternatively Spliced Isoforms of a DNA Binding Zinc Finger Domain. Science, 1992. **257**: p. 1951-1955.

94.   Hsu, T., *et al.* Multipl Zinc Finger Forms Resulting from Developmentally Regulated Alternative Splicing of a Transcription Factor Gene. Science, 1992. **257**: p. 1946-1950.

95.   Cheng, C. and E.T. Young. A Single Amino Acid Substitution in Zinc Finger 2 of Adr1p Changes its Binding Specificity at two Positions in UAS1. J. Mol. Biol., 1995. **251**: p. 1-8.

96.   Kim, C.A. and J.M. Berg. Serine at Position 2 in the DNA Recognition Helix of a Cys2-His2 Zinc Finger Peptide is Not, in General, Responsible for Base Recognition. J. Mol. Biol., 1995. **252**: p. 1-5.

97.   Thukral, S.K., M.L. Morrison, and E.T. Young. Mutations in the Zinc Fingers of ADR1 That Change the Specificity of DNA Binding and Transactivation. Mol. Cell. Biol., 1992. **12**: p. 2784-2792.

98.   Dutnall, R.N., D. Neuhaus, and D. Rhodes. The solution structure of the first zinc finger domain of SWI5: a novel structural extension to a common fold. Structure, 1996. **4**: p. 599-611.

99.   Camier, S., N. Kacherovsky, and E.T. Young. A Mutation outside theTwo zinc Fingers of ADR1 Can Suppress Defects in Either Finger. Mol. Cell. Biol., 1992. **12**: p. 5758-5767.

100.   Pedone, P.V., *et al.* The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding. Proc. Natl. Acad. U.S.A., 1996. **93**: p. 2822-2826.

101.   Nardelli, J., T. Gibson, and P. Charnay. Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis. Nucl. Acids Res., 1992. **20**: p. 4137-4144.

102.   Desjarlais, J.R. and J.M. Berg. Redesigning the DNA-Binding Specificity of a Zinc Finger Protein: A Dat Base-Guided Approach. Prot. Struct. Funct. Genet., 1992. **12**: p. 101-104.

103. Desjarlais, J.R. and J.M. Berg. Toward rules relating zinc finger protein sequences and DNA binding site preferences. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 7345-7349.

104. Krizek, B.A., et al. A Consensus Zinc Finger Peptide: Design, High-Affinity Metal Binding, a pH-Dependent Structure, and a His to Cys Sequence Variant. J. Am. Chem. Soc., 1991. **113**: p. 4518-4523.

105. Desjarlais, J.R. and J.M. Berg. Length-encoded multiplex binding site determination: Application to zinc fingers. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 11099-11103.

106. Rebar, E.J. and C.O. Pabo. Zinc Finger Phage: Affinity Selection of Fingers with New DNA-Bindig Specificities. Science, 1994. **263**: p. 671-673.

107. Jamieson, A.C., S. Kim, and J.A. Wells. In Vitro Selection of Zinc Fingers with Altered DNA-Binding Specificity. Biochemistry, 1994. **33**: p. 5689-5695.

108. Choo, Y. and A. Klug. Toward a code for the interactions of zinc fingers with DNA: Selection of randomized fingers displayed on phage. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 11163-11167.

109. Greisman, H.A. and C.O. Pabo. A General Strategy for Selecting High-Affinity Zinc Finger Proteins for Diverse DNA Target Sites. Science, 1997. **275**: p. 657-661.

110. Wu, H., W. Yang, and C.F. Barbas. Building zinc fingers by selection: Toward a therapeutic application. Proc. Natl. Acad. Sci. U.S.A., 1995. **92**: p. 344-348.

111. Cheng, C., et al. Identification of Potential Target Genes for Adr1p through Characterization of Essential Nucleotides in UAS1. Mol. Cell. Biol., 1994. **14**: p. 3842-3852.

112. Funabiki, T., B.L. Kreider, and J.N. Ihle. The carboxyl domain of zinc fingers of the Evi-1 myeloid transforming gene binds a consensus sequence of GAAGATGAG. Oncogene, 1994. **9**: p. 1575-1581.

113. Delwel, R., et al. Four of the Seven Zinc Fingers of the Evi-1 Myeloid-Transforming Gene Are Required for Sequence-Specific Binding

to GA(C/T)AAGA(T/C)AAGATAA. Mol. Cell. Biol., 1993. **13**: p. 4291-4300.

114. Ikeda, K. and K. Kawakami. DNA binding through distinct domains of zinc-finger-homeodomain protein AREB6 has different effects on gene transcription. Eur. J. Biochem., 1995. **233**: p. 73-82.

115. Sekido, R., *et al.* The d-Crystallin Enhancer-Binding Protein dEF1 Is a Repressor of E2-Box-Mediated Gene Activation. Mol. Cell. Biol., 1994. **14**: p. 5692-5700.

116. Thiesen, H. and C. Bach. Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. Nucleic Acids Res., 1990. **18**: p. 3203-3209.

117. Swirnoff, A.H. and J. Milbrandt. DNA-Binding Specificity of NGF1-A and Related Zinc Finger Transcription Factors. Mol. Cell. Biol., 1995. **15**: p. 2275-2287.

118. Ma, P.C., *et al.* Crystal Structure of MyoD bHLH Domain-DNA Complex: Perspectives on DNA Recognition and Implications for Transcriptional Activation. Cell, 1994. **77**: p. 451-459.

Table 1:   DNA sequence specificities of zinc finger proteins
            determined from site selection experiments

This survey focused on specificities determined via site selections because such studies provide the most unbiased assessment of DNA binding site. Sources are as follows: ADR1: Cheng *et al.*[111]; CF2-1: Gogos *et al.*[93]; Evi-1 (COOH fingers): Funabiki *et al.*[112]; Evi-1 ('domain 1'): Delwel *et al.*[113]; AREB6 (both NH2 and COOH fingers): Ikeda and Kawakami[114]; dEF1: Sekido *et al.*[115]; SP1: Thiesen and Bach[116]; Zif268: Swirnoff and Milbrandt[117];

Table 1:DNA sequence specificities of zinc finger proteins determined from site selection experiments

Protein	Consensus sequence from site selections
ADR1	GG(G/A)G
CF2-I	GTATATTATATA
Evi-1 (COOH fingers)	GAAGATGAG
Evi-1 ('domain 1')	GA(C/T)AAGA(T/C)AAGATAA
AREB6 (NH2 fingers)	(GT/TG)TCACCTGT
AREB6 (COOH fingers)	(C/T)ACCT(G/T)T
dEF1	CACCT
SP1	GGGGCGGGT
Zif268	TGCGT(A/G)GGCG(T/G)

Figure 1:   The four base pairs

The four 'watson - crick' base pairs of normal duplex DNA are shown.   Black circles denote the attachment points of deoxyribose sugars.   Beneath each base pair is shown its characteristic pattern of major groove functional groups.

Figure 1: The four base pairs

guanine - cytosine   cytosine - guanine   adenine - thymine   thymine - adenine

minor groove

major groove

pattern of major groove functional groups
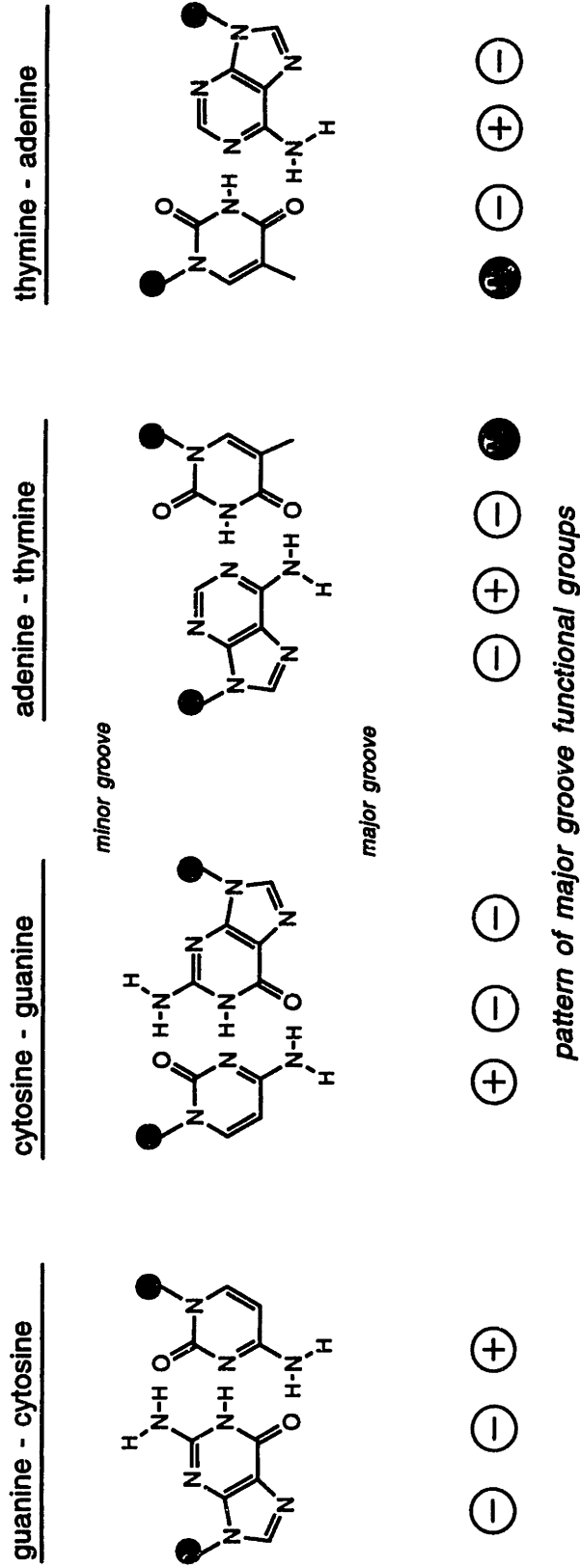
symbols: (+) hydrogen bond donor;   (|) hydrogen bond acceptor;   ● methyl;

Figure 2:   Representative motifs complexed with DNA

Sketches are shown which illustrate the general docking arrangement of three well-characterized DNA - binding motifs. Sketches were reproduced from the following sources: homoeodomain: Kissinger et al.[19]; zinc finger: Pavletich and Pabo[67]; helix-loop-helix: Ma et al.[118]. References for target site entries are: homeodomain: Damante et al.[43]; helix-loop-helix: Ma et al.[118].

Figure 2: Representative motifs complexed with DNA:

motif:	helix-loop-helix	homeodomain	zinc finger
target sites:	CAnnTG (most)	nAAnnn (most)	no pattern
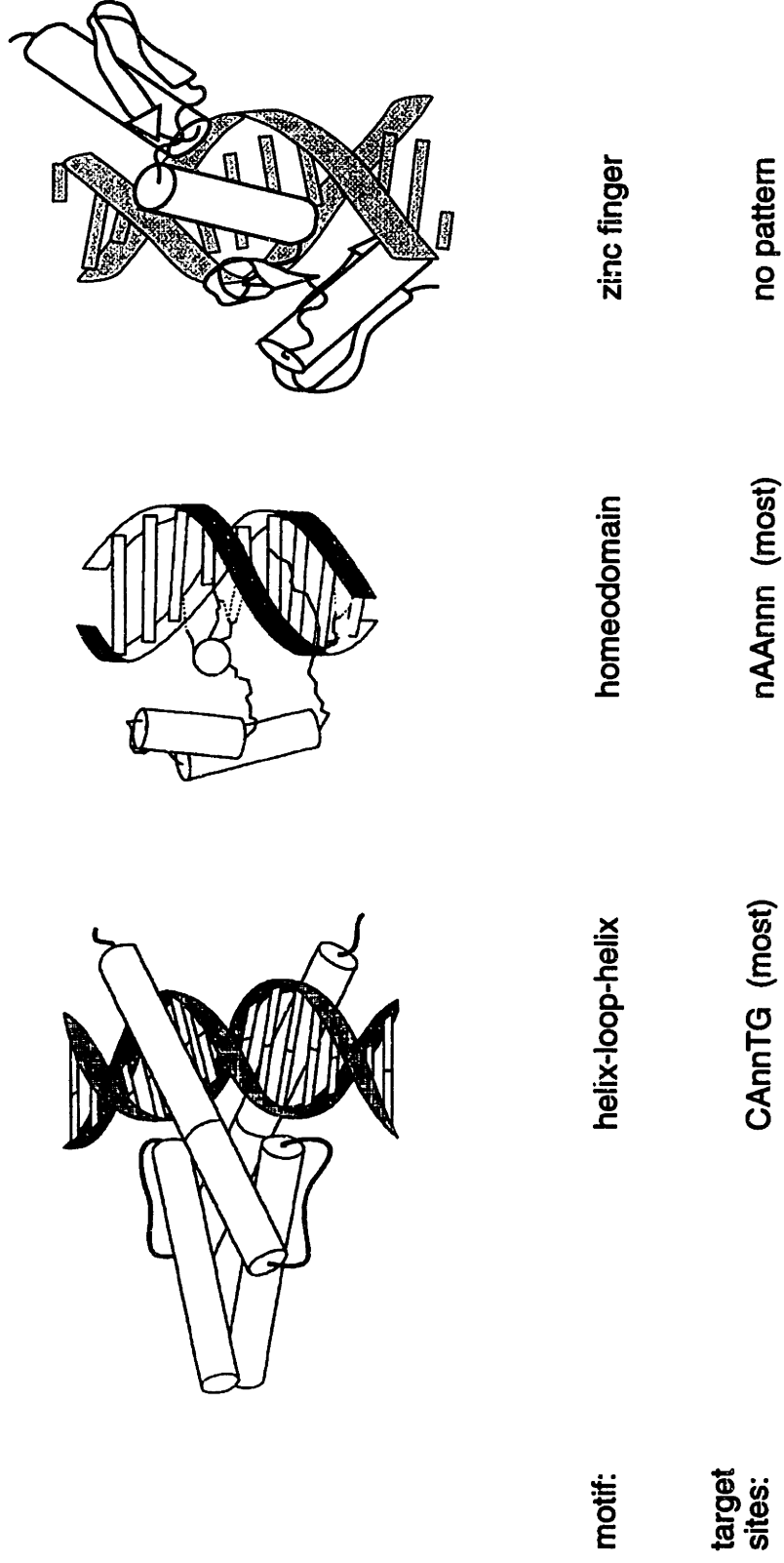
Figure 3A: Examples of favorable amino acid - base pair hydrogen bonding geometries

Base pairs are depicted at top [A: adenine; T: thymine; G: guanine; C: cytosine]. Black circles denote the attachment points of deoxyribose sugars. Dashed lines indicate hydrogen bonds between bases and amino acids.

Figure 3B: Histidine-guanine contact

Interaction between histidine and guanine observed in the Zif268 protein-DNA complex[67].

Figure 3C: Arginine-aspartate-guanine interaction

Cooperation between an arginine and an aspartate to recognize a guanine base. Reproduced from Pavletich and Pabo[67].

Figure 3A: Examples of favorable amino acid
- base pair hydrogen bonding geometries:
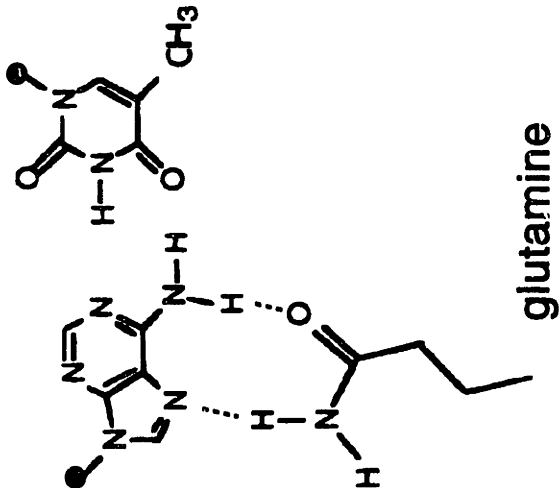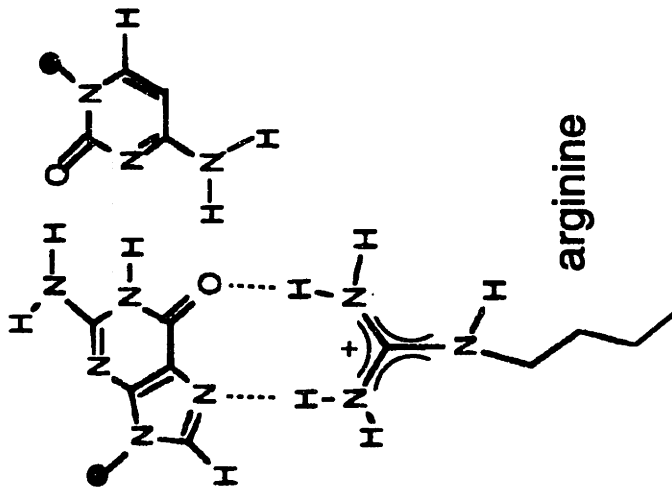
Figure 3B:
histidine-guanine contact

G C

histidine

Figure 3C:
arginine-aspartate-guanine interaction

G C
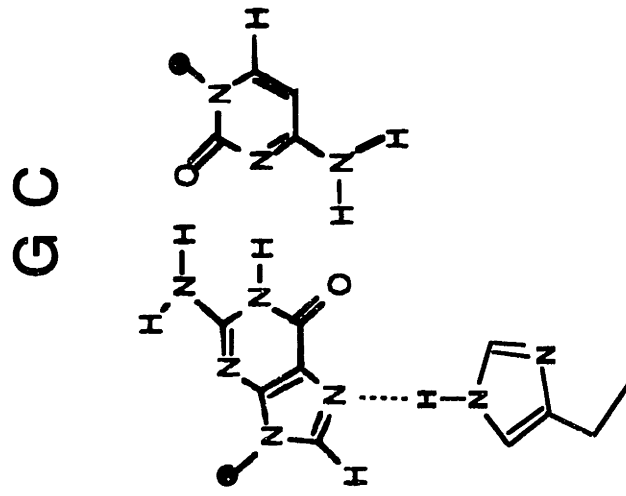
aspartate

arginine
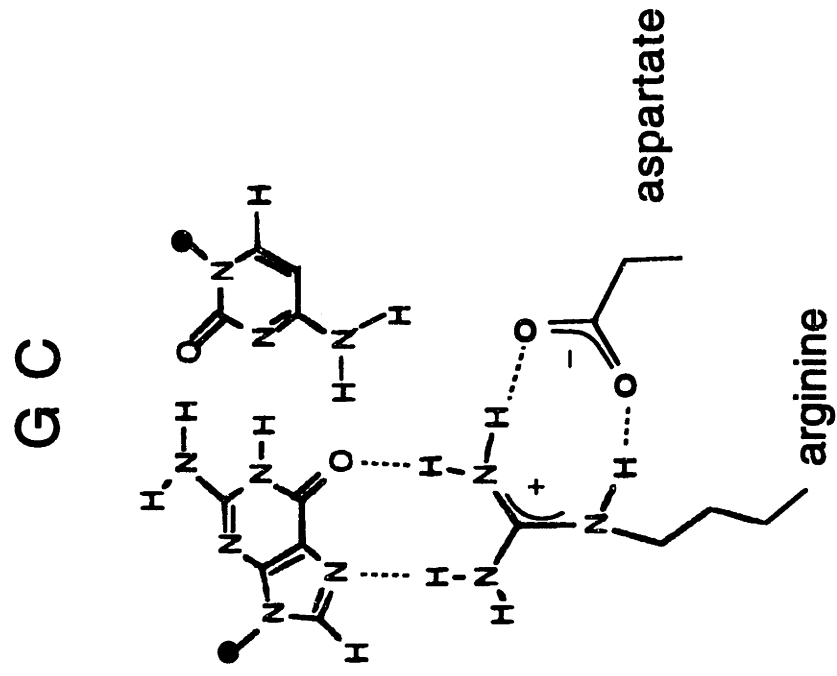
Figure 4:   Zinc finger consensus sequence and structure

Dashes in the consensus sequence indicate positions that exhibit no clear amino acid preference.   Brackets highlight optional insertions between the two zinc-coordinating histidines.   Elements of secondary structure are indicated as follows: arrows: strands of ß-sheet; cylinders: α-helices; thin lines: coils and turns.   The white circle in the folded finger indicates the approximate location of the coordinated zinc ion.

Figure 4: Zinc finger consensus sequence and structure:

consensus:

- - - - - F/Y . C . [-.] . C . . . F/Y . . . . L . . H . . [-.] . H . .

secondary
structure:

folded
finger:
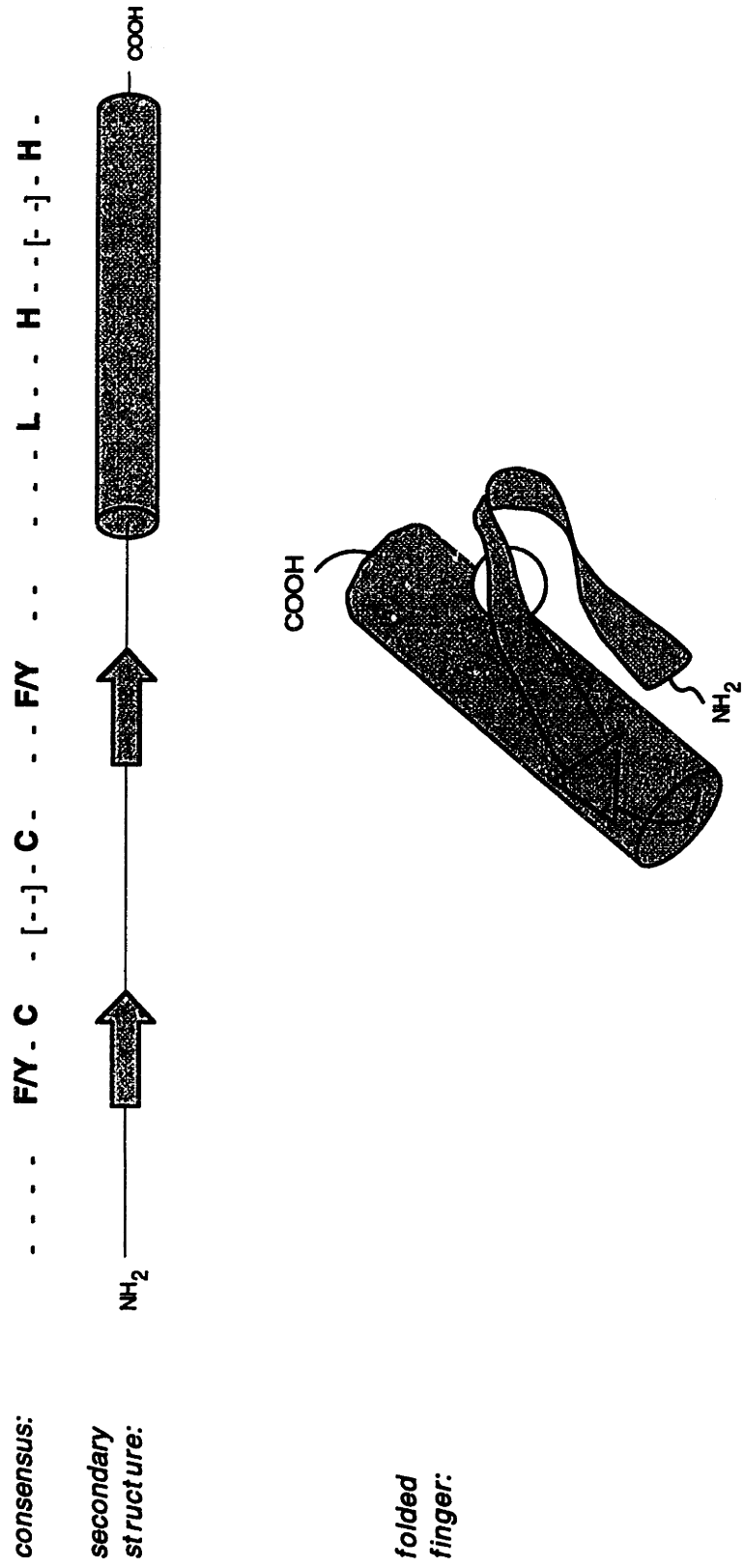
Figure 5: Sketch of Zif268 bound to its target site

Sketch of Zif268 bound to its target site, showing base contacts from finger 1[67]. Note, for clarity, coordinated zinc ions are not shown.

Figure 5: Sketch of Zif268 bound to its DNA target site

COOH

Finger 3

5'

3'

Finger 2

NH₂

R18   D20
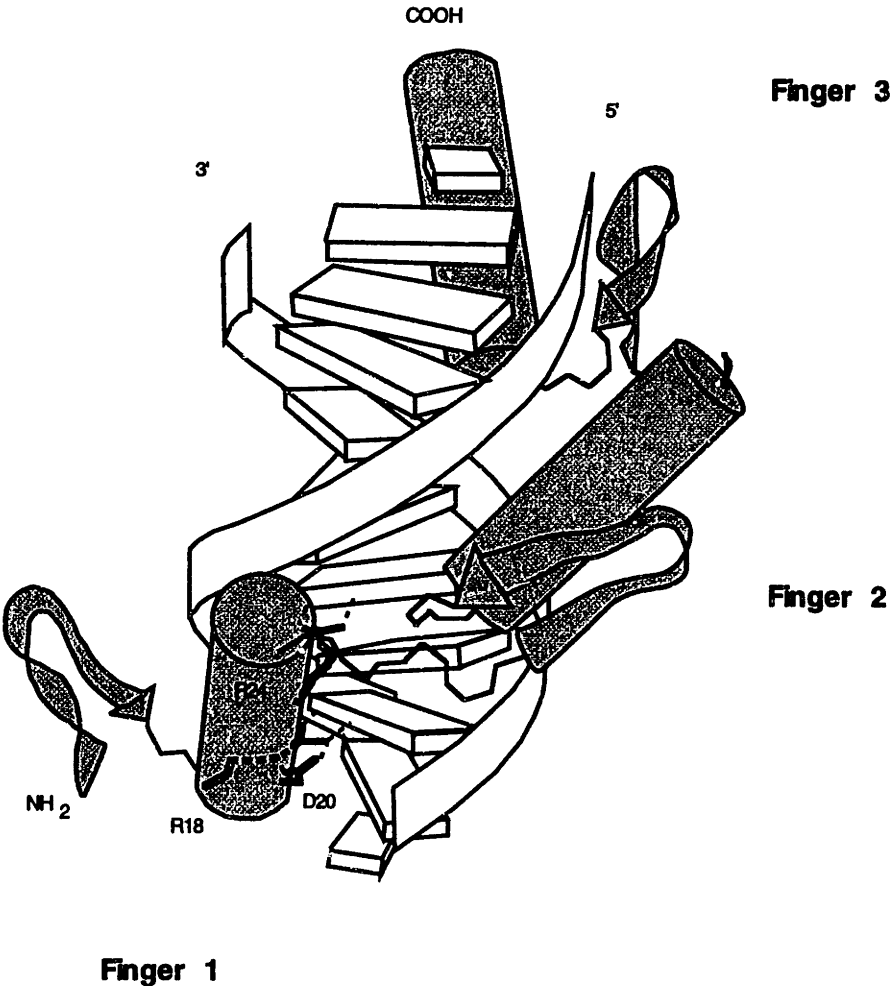
Finger 1

Figure 6:   Base-contacting fingers observed in structural studies

To date, structural studies have visualized base contacts from thirteen natural zinc fingers in five different protein-DNA complexes. These fingers include Zif268 fingers 1, 2 and 3[67, 81], tramtrack fingers 1 and 2[66], YY1 fingers 1, 2, 3 and 4[69], the GAGA finger[65], and GLI fingers 2, 4 and 5[68]. Amino acid sequences are shown for these thirteen fingers, and each finger is identified at left (Zif 1 denotes finger 1 of Zif268, etc.). Note that GLI fingers 1 and 3 are excluded, because GLI finger 1 does not contact DNA and GLI finger 3 makes only phosphate contacts.

Squares highlight residues that make base contacts in the complexes and circles highlight residues that make phosphate contacts. The four 'base contact positions' are shaded in gray. The numbers marking the 'base contact positions' indicate the location of each residue relative to the start of the α-helix. The secondary structure of the zinc finger domain is sketched at the bottom of the figure. The arrows indicate strands of the ß-sheet and the cylinder shows the location of the α-helix.

Figure 6: Base-contacting fingers observed in structural studies

base contact positions

				1		2 3		6		
Zif 1	E R P	Y A C P V E S C D R F		S R		S D E		R H	I R – I H T
Zif 2	G Q K P	F Q C R – – I C M R N F		S R		S D H		T H	I R – T H T
Zif 3	G E K P	F A C D – – I C G R K F		A R		S D E R K		R H	T K – I H L
GLI 2	G ERK E	F V C H W G G C S R ELRPF		K A		Q Y M L V		H	M R – R H T
GLI 4	G E K P	Y M C E H E G C S K A F		S N		A S D R A K		H	Q N R T H S
GLI 5	N E K P	Y V C K L P G C T K R Y		T D		P S S L R K H			H V K T V H G
TTK 1	T K E G E H T	Y R C K – – V C S R V Y		T A		I S N F		C R H	Y V T S H K
TTK 2	R N V K V	Y P C P – – F C F K E F		T R		K B N M T A H			H V K I I H K
YY1 1	M E P R T	I A C P H K G C T K M F		R D		N S A M R K H			L H – T H G
YY1 2	– P R V	H V C A – – E C G K A F		V E		S S L K G H			Q L – V H T
YY1 3	G E K P	F Q C T F E G C G R F		S E		D E N L R T H			V R – I H T
YY1 4	G D R P	Y V C P F D G C N K K F		A Q		S T N L K S H I L			– T H AKAKNNQ
GAGA	E Q P	A T C P – – I C Y A V I		R C		S F N L R R H			L E L R H FAKPGV

○ phosphate contacts
☐ base contacts

Figure 7: Zinc finger - base contacts observed in structural studies.

The base contacts from figure 6 are summarized, highlighting the identity and location of each contacted base. Contacts are organized to illustrate the tendency of each base contact position to interact with a preferred base. References for the contacts shown are the same as those mentioned in the legend to figure 6.

Figure 7: Zinc finger - base contacts observed in structural studies
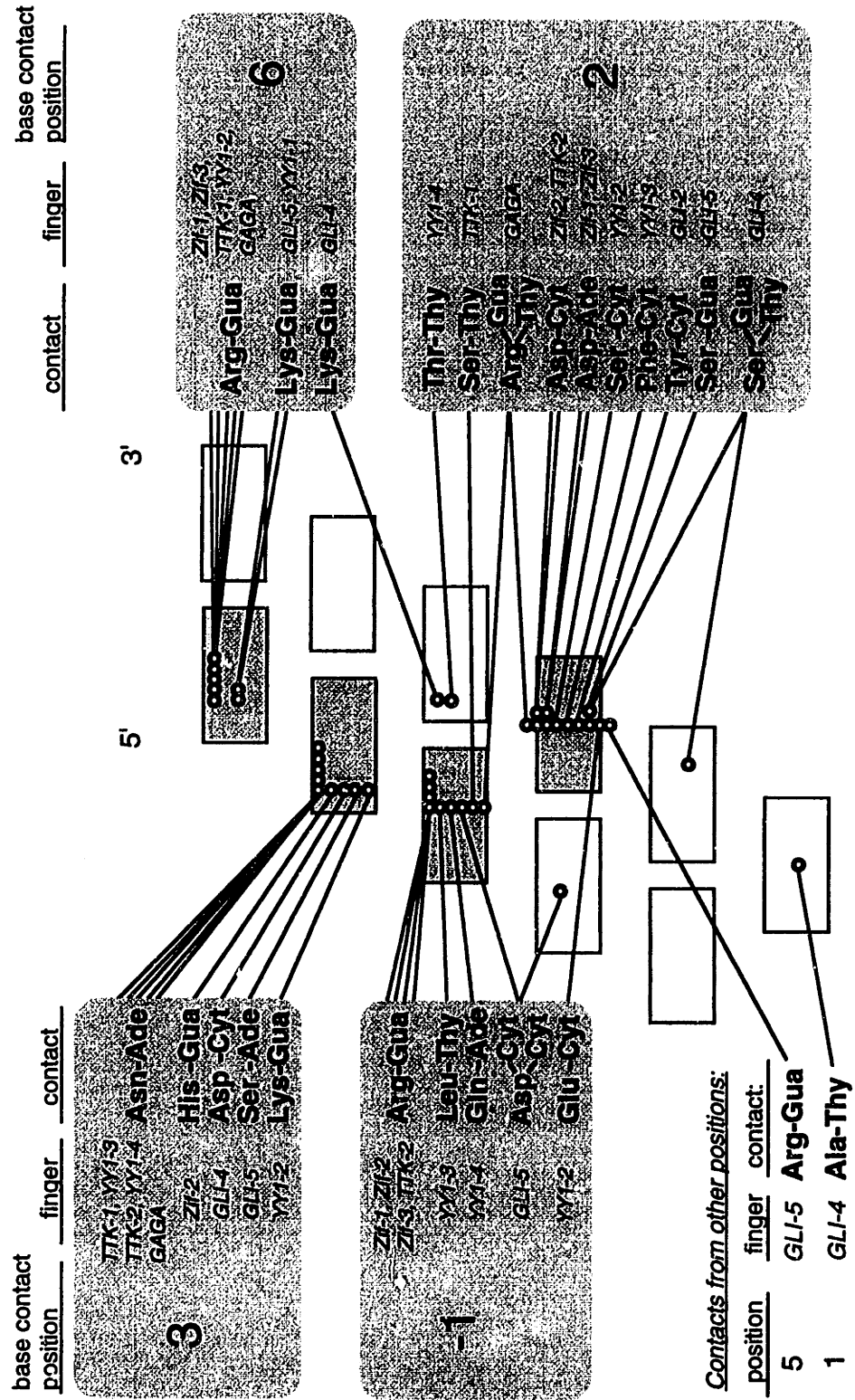
CHAPTER 2

DEVELOPMENT OF A PHAGE DISPLAY SYSTEM FOR ZINC FINGER PROTEINS

Introduction:

DNA-binding proteins play critical roles in cell biology, and the design of proteins with novel sequence specificities or functions may have important applications in research, biotechnology and medicine. Recent reports have highlighted some of the potential of these design efforts. For example, proteins with designed DNA-binding specificities have been used to regulate the transcription of specific genes[1, 2]. DNA-binding domains also have been attached to other proteins - such as nucleases[3] and general transcription factors [4] to create hybrid proteins with interesting properties. These designed DNA-binding proteins offer great promise as research tools and may eventually be used in gene therapy. However, many potential applications will require binding to novel target sites, and so the general usefulness of these proteins may depend on our ability to design or select DNA-binding domains with desired DNA sequence specificities.

The zinc finger motif, which was first recognized in transcription factor IIIA (TFIIIA)[5], appears to offer an especially attractive framework for constructing proteins with novel DNA sequence specificities. The zinc finger is one of the most common eukaryotic DNA-binding motifs[6, 7], and this family of proteins can recognize a diverse set of DNA sequences (see examples cited in Pavletich and Pabo[8]). Crystallographic studies of zinc finger proteins bound to DNA show that amino acids at six positions can make base contacts[8-11], and it appears that the identity of four of these residues plays an especially important role in determining the sequence

preference of a zinc finger[12, 13]. Zinc finger proteins also exhibit a modular organization, which may allow one to "mix and match" fingers to obtain proteins with novel DNA-binding specificities[14]. By changing a few key residues within zinc fingers, and by changing the order of fingers in multifinger proteins, it may be possible to make a zinc finger protein that would recognize any desired target site on double stranded DNA.

To explore the potential of TFIIIA-like zinc fingers for designing DNA-binding proteins with new sequence specificities, we developed a selection system based on filamentous phage display[15]. In this system[16] DNA encoding the zinc fingers of Zif268[17] is fused to the gene for the bacteriophage pIII protein, and the hybrid protein is expressed on the surface of phage. This chapter provides an overview of phage display methodology and describes key experiments used to develop and test zinc finger phage.

<u>**Phage display principles and background**</u>:

At its most fundamental level, phage display is a strategy for expressing a protein in a form which remains linked to its gene. Phage display takes advantage of the fact that bacteriophage assembly can tolerate insertions into, and extensions of, certain coat proteins. The method can be summarized briefly: i) first, a vector is prepared which contains the 'target' gene of interest cloned as a fusion with a phage coat protein gene; ii) the vector is transfected into bacteria; iii) the vector expresses the target gene as a fusion with the phage coat

protein, and this hybrid protein is assembled into new phage particles; and iv) the vector is packaged into these phage. The power of phage display stems from the fact that the resulting phage contain the target gene and exhibit biological properties of the target protein. Therefore, they may be used for *in vitro* genetic selections.

To date, the vast majority of phage display experiments have used fd bacteriophage (also called f1 or M13). [A few experiments have used other phage, including T7 (Novagen, 1996) and lambda[18, 19].] Fd is a relatively simple bacteriophage which encodes 10 genes and which contains just five proteins in the mature phage particle [20]. The phage is shaped like a long, flexible cylinder, with several copies of one particular coat protein - 'pIII' - protruding from one end (fig. 1A). Fd is an unusual bacteriophage because it does not lyse its host. Rather, progeny phage escape the infected cell using an extrusion process which leaves the cell membrane intact[20]. It appears that fd phage assemble during extrusion, since no phage particles are visible in infected *E. coli*[20] and since all coat proteins are anchored to the bacterial inner membrane prior to phage assembly[21]. According to current models, the fd phage extrusion/assembly process is directional, with the 'pIII' end of the phage being extruded and assembled last.

To date, most fd phage display experiments have used fusions with the pIII protein. Several copies of pIII are clustered at one end of the fd phage particle (fig. 1A)[22]. PIII contains at least four functionally distinct regions (fig. 1B): i) an 18-residue signal sequence at the amino terminus which is clipped off in the mature pIII; an

apparently globular, protease-resistant domain encompassing the remaining amino-terminal half of the protein[23-25]; a carboxy-terminal segment which inserts into the viral coat[23, 26]; and an intervening glycine-rich spacer region[23, 27]. The globular domain of pIII is important for phage infectivity. This domain mediates the docking of phage to f-pili prior to infection[24, 28] and plays a crucial role in the transport of phage DNA into the target cell[29]. Deletion of this domain abolishes phage infectivity without affecting phage assembly[30]. In contrast, the carboxy-terminal segment is vital for proper phage assembly. Deletion of this segment results in the production of 'polyphage': very long phage which contain multiple copies of genome[30]. The intervening region seems to function largely as a flexible spacer.

PIII is a secreted protein. Synthesis of the amino terminal 'signal sequence' directs the protein to the bacterial export apparatus. The remainder of the protein is then translocated across the bacterial inner membrane during translation except for the carboxy-terminal segment, which remains in the membrane and anchors pIII to it[26]. PIII must be secreted for proper function, as blockage secretion results in the production of nonviable phage.

The first successful phage display experiments fused protein fragments and peptide libraries to pIII. In initial studies, Smith inserted fragments of EcoR1 endonuclease into the spacer region of pIII, and showed that these phage bound specifically to anti-EcoR1 antibodies immobilized on a solid support[31]. These first 'fusion phage', however, exhibited very low infectivities. In subsequent studies the insertion point for displayed peptides was moved to the junction of the

signal sequence and the globular domain of pIII, and this yielded improved phage function[32]. Shortly thereafter, several groups expressed combinatorial peptide libraries at this location and showed that affinity methods could be used to isolate individual peptides from these libraries[15, 33, 34]. In subsequent experiments, phage selections have been used to discover peptide ligands for a diversity of targets, including GPIIb/IIIa[35], streptavidin[34, 36], concanavalin A[34, 36-38], and SH3 domains[39, 40]. These methods have also been used to define the epitopes of numerous antibodies (for examples, see Stephan and Lane[41] and Scott and Smith[15]).

A variety of studies have also expressed intact proteins and protein libraries as fusions with pIII. Successfully displayed proteins have included protease inhibitors[42-48], cytokines[49, 50], hormones[51], RNA-binding domains[52], zinc fingers[16, 53-55], and enzymes such as glutathione transferase[56], alkaline phosphatase[57], and serine proteases[58]. To date, approximately 20 distinct protein folds have been successfully displayed on phage, and randomized libraries of these proteins have been selected for a wide variety of binding and catalytic properties. One particularly interesting application has involved the display and selection of antibodies. In an extensive series of studies, several groups have examined the potential of these methods for creating antibodies for novel haptens (for examples, see Garrad and Henner[59], Gram et al.[60] and Hoogenboom and Winter[61]), and it appears that these methods may provide an alternative to conventional hybridoma technology.

Although pIII has proven adequate for a variety of peptide and protein selections, several features could constrain its use for certain studies. First, and perhaps most importantly, pIII is a secreted protein. Except for the carboxy-terminal region, the entire protein must pass through the bacterial inner membrane during synthesis. Consequently, it might be difficult to use this system to display any protein which contains a membrane anchor sequence (typically a stretch of hydrophobic amino acids greater than about 12 to 15 residues long[62, 63]). Also, this system exposes displayed proteins to the oxidizing environments of the *E. coli.* periplasm and culture medium, and so special procedures may be required for proteins which contain reduced thiols. These features of pIII expression could explain why relatively few intracellular proteins have been phage-displayed (approximately 4).

An additional feature of pIII which may complicate selection studies is its valency on phage. A variety of experiments suggest that three to five copies of pIII are present at the end of each phage particle[25, 64-66]. When improper selection conditions are used, this multivalency may cause problems by enabling enough weak binders to survive affinity selection to swamp out tighter binding peptides. Concern for this problem has prompted the development of 'monovalent phage display' systems[51]. In these systems, phagemid vectors express the pIII-target protein fusion gene, and helper phage provide other protein products required for assembly of virus particles and packaging of phagemid DNA. The helper phage produce an excess of normal pIII so that the resulting phage particles contain at most one copy of

phagemid-encoded pIII fusion protein. This effectively reduces the displayed protein to monovalency. An alternate approach to this problem is to raise the stringency of a selection by using more rigorous binding and wash conditions[16, 48].

Development of a Zinc Finger Phage System:

To develop a zinc finger phage system, it was necessary to: i) construct a zinc finger phage vector; ii) produce zinc finger phage; and iii) discover conditions which enable phage to bind specifically to a DNA-coated solid support. Steps (i) and (ii) were achieved with relative ease. To construct a zinc finger phage vector, a DNA fragment encoding the three fingers of Zif268 was inserted into the pIII gene of the fd vector fUSE3 between regions encoding the signal sequence and globular domain (fig. 2). The resulting vector, fd-tet.Zif, produced usable titers of 'Zif phage' (0.5 -14 x 10^9 phage/ml, detected as tetracycline transducing units (TU)[67]), which suggested that the Zif-pIII fusion was efficiently transported across the bacterial membrane and incorporated into phage particles.

Discovering conditions which permitted phage to bind specifically to immobilized Zif binding site proved to be more difficult. In a series of initial experiments, we failed to demonstrate binding of phage to a solid support coated with Zif268 binding site, and the reasons for this failure were unclear. One possibility was that proteolysis and/or oxidation inactivated the phage-displayed zinc finger peptides. We were particularly concerned about the latter possibility, given that

zinc fingers require free thiols for zinc coordination and proper folding. Another possibility was that the phage-displayed zinc fingers were properly folded, but could not interact with the support-bound DNA sites. Our experiments used a variation of the 'biopanning' protocol, which had been developed by Smith and Scott to study binding of epitope-expressing phage to immobilized antibodies[67]. [Details of our protocol are provided in the Materials and Methods section.] Since our variation of the biopanning protocol contained several novel features (e.g., biotinylated DNA instead of antibodies, different rinse procedures and elution conditions) it was possible that our protocol was in some way unsuitable for studying the binding of Zif phage to DNA.

In order to counteract potential problems with oxidation and/or proteolysis, we took the following measures: First, to eliminate the possibility of oxidation, anaerobic conditions were used to produce and purify phage. Growth media was degassed prior to inoculation with Zif phage-expressing *E. coli.*, and cultures were grown in tightly capped bottles. All critical phage purification steps were performed in degassed buffers within an anaerobic chamber ($\leq$ 1 part per million (ppm) O_2). Also, to reduce proteolysis, cultures were grown at ambient temperature (~22C) and care was taken to harvest phage promptly after culture saturation. Furthermore, growth media was supplemented with zinc to ensure its availability for zinc finger folding. [Our phage production protocol is described in the materials and methods section of this chapter, and a more detailed discussion of Zif phage production considerations is provided in the appendix to chapter 3.]

In order to address the possibility that our biopanning protocol interfered with the formation of a phage-DNA complex, we sought an alternate means for detecting this interaction. In particular, a gel shift assay was developed to enable the rapid assessment of phage-DNA binding. Using this assay, we could demonstrate binding of Zif phage to a synthetic DNA fragment containing multiple Zif binding sites (fig. 3). This interaction was competed by free Zif protein and by cold binding site but not calf thymus DNA (fig. 3). Phage produced using the parent vector did not bind to Zif site (data not shown).

The gel shift assay enabled us to characterize features of phage-DNA binding relevant to our biopanning protocol. For example, using this assay, the half-life of the phage-DNA complex was determined to be about 30'. Since this is longer then the time scale of the biopanning washes (~10'), this suggested that rapid complex dissociation was not causing our inability to observe specific phage binding to the solid support. Also, we verified that our elution protocol (4M NaCl, ≥10') was sufficiently stringent to disrupt the Zif phage/DNA complex. These gel shift studies, and others, enabled us to test and exclude possible reasons for our failure to observe Zif phage binding to in our biopanning assay.

During these initial studies, Zif phage often exhibited behavior which suggested a tendency to aggregate and/or stick nonspecifically to surfaces. For example, in the gel shift experiments described above, the phage-bound DNA typically smeared back to the wells (fig. 3). Additionally, mock platings showed that, even in the absence of Zif target site, retention efficiencies for Zif phage were from 100 to 1000

times higher than retention efficiencies of phage which did not express Zif fingers (data not shown). To counteract this problem, we tested a panel of detergents and chaotropic agents for the ability to improve phage behavior in the gel shift assay, and two detergents (triton x-100 and NP-40) were discovered which could reduce smearing when used at high concentrations (.5 %) (fig. 4).

We then tested whether triton x-100 or NP-40 could improve Zif phage plating behavior. As table 1 shows, the addition of either detergent to the wash buffers substantially improved the 'signal to noise' of phage retention efficiencies plated in the presence and absence of DNA. In particular, the use of triton x-100 resulted in a >100-fold difference in the retention efficiencies of Zif phage plated in the presence and absence of 23 nM target DNA. [Without detergent, virtually no difference was seen.] To demonstrate the specificity of this effect, we performed an additional plating, which is summarized in table 2. This experiment showed that high retention efficiencies required pre-coating of the solid support with streptavidin and the use of biotinylated Zif target site in the binding reactions. Furthermore, binding was disrupted when excess streptavidin, biotin or non-biotinylated Zif target site was included in the binding reactions, but was insensitive to excess nonspecific DNA (table 2).

In subsequent plating experiments we characterized and optimized this system. The first priority was to choose a standard solid support for these studies. Retention of phage was tested using six different types of microtiter plates and optimal behavior was observed with Falcon 3915 'Probind^{TM}' plates (table 3). These plates

were used in all subsequent plating experiments. In a second plating experiment, we tested whether a target DNA fragment bearing a single Zif site could mediate efficient phage retention. Our goal was to minimize the size of the target DNA fragment, since selections which used smaller targets would be less likely to yield proteins which could bind undesired DNA sequences. Phage were biopanned using a ten-fold dilution series of both the single- and multi-site DNA fragments, and both sites yielded similar phage retention efficiencies at 23 nM (the highest concentration tested for both sites) (fig. 5). At lower concentrations the multi-site DNA fragment was more effective, and we believe that this difference reflects the multivalency of pIII-zinc finger fusion proteins on each phage. A final plating tested the DNA-sequence specificity of Zif phage. Zif phage were biopanned using a wild-type and mutant Zif site, and phage recoveries obtained with the wild-type site were > 13 fold higher than those obtained with the mutant Zif site (table 4).

Throughout these tests, our highest phage retention efficiencies were approximately 1%, and it unclear why values much higher than this are not observed. [In subsequent studies, several potential reasons for this behavior have been excluded, including possibility that the amount of applied phage exceeds plate capacity, and the possibility that bound phage could not efficiently elute from the plate surface (data not shown)]. We note, however, that in affinity platings with phage expressing a variety of proteins, ~1% recovery of phage is typically the maximum value obtained[16, 35, 43, 45, 48, 57, 68]. Throughout

these studies, we consider 1% recovery of phage to represent full binding.

Conclusions:

The studies described in this chapter established the first phage display system for any DNA-binding protein, and the first affinity selection system for zinc fingers. Together, these studies i) demonstrated that DNA-binding domains, in particular zinc fingers, may be functionally expressed on the surface of filamentous phage; ii) demonstrated that DNA-binding phage may be affinity-selected using suitable binding sites attached to insoluble supports; and iii) defined protocols, reagents, and other experimental parameters which were crucial for later selection experiments. These experiments provided the foundation for subsequent studies in this thesis.

Materials and Methods:

Note: Procedures for individual experiments are provided with the table and figure legends.

Zif phage production

To produce Zif phage, MC1061 cells[67] which contained fd-tet.Zif were grown to saturation (room temperature, ~40 hours, no agitation) under anaerobic conditions in .05 l of Zif phage broth (see the chapter 3 appendix for a description of Zif phage broth). In an attempt to increase titer, cultures were grown in dialysis tubes (50 kd cut off) suspended in 1 liter of broth.

Phage was purified as follows: i) cells were cleared from the cultures via spinning at 7,800x G (4°C, 80 min); ii) phage particles were then pelleted by ultracentrifugation (171,000x G, 4°C, 6 hours); iii) phage pellets were resuspended in ~1/100 volume of binding buffer; and iv) residual insoluble matter was cleared via a brief spin in a microfuge, and the phage - containing supernatants were place in fresh microfuge tubes. These phage preps were stored anaerobically (<1 ppm O_2) on ice. Due to concerns about oxidation, all phage manipulations were performed so as to minimize exposure to O_2. In particular, all manipulations, except for the centrifugation and the ultracentrifugation, were performed in the anaerobic chamber.

Zif phage quantitation

Zif phage were titered as tetracycline transducing units (TTU) essentially as described[67] except that starved K91 cells had been stored at -80°C in a buffer containing: 67 mM NaCl, 42 mM $NH_4H_2PO_4$, and 14% glycerol. Titers of Zif phage cultures were 0.5 - 14 x 10^9 TTU/ml.

VCSM13 phage

VCSM13 phage was prepared as described in the chapter 3 appendix.

DNA: gel shift probes, biotinylated fragments, and competitor

DNA probes for gel shift studies were prepared and labeled essentially as described in footnote 15 of chapter 3. Competitor calf thymus DNA was prepared essentially as described in the chapter 3 appendix. Biotinylated DNA fragments for platings were prepared essentially as described in the chapter 3 appendix.

Zif protein

Zif protein was provided by Nikola P. Pavletich, and had been prepared as described in Pavletich and Pabo[9].

Binding buffer:

Binding buffer contained: 15 mM Hepes, 50 mM NaCl, 5 mM MgCl$_2$, 0.01 mM ZnCl$_2$, 0.1 mg/ml BSA, 5% glycerol (w/v), pH 7.8.

Elution buffer:

Elution buffer was identical to binding buffer except it contained 4 M NaCl;

0.05 M NaCl wash buffer:

0.05 M NaCl wash buffer contained: 15 mM Hepes, 50 mM NaCl, 5 mM MgCl$_2$, 0.01 mM ZnCl$_2$, 5% glycerol (w/v), and 0.5 % triton x-100 (w/v) pH 7.8.

Preparation of streptavidin - coated plates:

We prepared our streptavidin coated plates as described in the chapter 3 appendix.

References:

1.   Pomerantz, J.L., P.A. Sharp, and C.O. Pabo. Structure-Based Design of Transcription Factors. Science, 1995. **267**: p. 93-96.

2.   Choo, Y., I. Sánchez-García, and A. Klug. *In vivo* repression by a site-specific DNA-binding protein designed against an oncogenic sequence. Nature, 1994. **372**: p. 642-645.

3.   Kim, Y. and S. Chandrasegaran. Chimeric restriction endonuclease. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 883-887.

4.   Chatterjee, S. and K. Struhl. Connecting a promoter-bound protein to TBP bypasses the need for a transcriptional activation domain. Nature, 1995. **374**: p. 820-822.

5.   Miller, J., A.D. McLachlan, and A. Klug. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J., 1985. **4**: p. 1609-1614.

6.   Pellegrino, G.R. and J.M. Berg. Identification and characterization of "zinc finger" domains by the polymerase chain reaction. Proc. Natl. Acad. Sci. U.S.A., 1991. **88**: p. 671-675.

7.   Jacobs, G.H. Determination of the base recognition positions of zinc fingers from sequence analysis. EMBO J., 1992. **11**: p. 4507-4517.

8.   Pavletich, N.P. and C.O. Pabo. Crystal Structure of a Five-Finger GLI-DNA Complex: New Perspectives on Zinc Fingers. Science, 1993. **261**: p. 1701-1707.

9.   Pavletich, N.P. and C.O. Pabo. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. Science, 1991. **252**: p. 809-817.

10.   Fairall, L., *et al.* The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature, 1993. **366**: p. 483-487.

11.   Houbaviy, H.B., *et al.* Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. Proc. Natl. Acad. Sci. U.S.A., 1996. **93**: p. 13577-13582.

12.   Thukral, S.K., M.L. Morrison, and E.T. Young. Mutations in the Zinc Fingers of ADR1 That Change the Specificity of DNA Binding and Transactivation. Mol. Cell. Biol., 1992. **12**: p. 2784-2792.

13.   Desjarlais, J.R. and J.M. Berg. Toward rules relating zinc finger protein sequences and DNA binding site preferences. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 7345-7349.

14.   Desjarlais, J.R. and J.M. Berg. Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. Proc. Natl. Acad. Sci. U.S.A., 1993. **90**: p. 2256-2260.

15.   Scott, J.K. and G.P. Smith. Searching for Peptide Ligands with an Epitope Library. Science, 1990. **249**: p. 386-390.

16.   Rebar, E.J. and C.O. Pabo. Zinc Finger Phage: Affinity Selection of Fingers with New DNA-Bindig Specificities. Science, 1994. **263**: p. 671-673.

17.   Christy, B.A., L.F. Lau, and D. Nathans. A gene activated in mouse 3T3 cells by serum growth factors encodes a protein with "zinc finger" sequences. Proc. Natl. Acad. Sci. U.S.A., 1988. **85**: p. 7857-7861.

18.   Maruyama, I.N., H.I. Maruyama, and S. Brenner. Lambda-foo: a lambda phage vector for the expression of foreign proteins. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 8273-8277.

19.   Sternberg, N. and R. Hoess. Display of peptides and proteins on the surface of bacteriophage lambda. Proc. Natl. Acad. Sci. U.S.A., 1995. **92**: p. 1609-1613.

20.   Model, P. and M. Russel, *Filamentous Bacteriophage*, in *The Bacteriophages*, R. Calendar, Editor. 1988, Plenum Publishing: New York. p. 75-101.

21.   Endemann, H. and P. Model. Location of Filamentous Phage Minor Coat Proteins in Phage and in Infected Cells. J. Mol. Biol., 1995. **250**: p. 496-506.

22.   Woolford, J.L., H.M. Steinman, and R.E. Webster. Adsorption Protein of Bacteriophage f1: Solubilization in Deoxycholate and Localization in the f1 Virion. Biochemistry, 1977. **16**: p. 2694-2700.

23.   Kremser, A. and I. Rasched. The Adsorption Protein of Filamentous Phage fd: Assignment of Its Disulfide Bridges and Identification of the Domain Incorporated in the Coat. Biochemistry, 1994. **33**: p. 13954-13958.

24.   Armstrong, J., R.N. Perham, and J.E. Walker. Domain Structure of Bacteriophage fd Adsorption Protein. FEBS Let., 1981. **135**: p. 167-172.

25.   Gray, C.W., R.S. Brown, and D.A. Marvin. Adsorption Complex of Filamentous Virus. J. Mol. Biol., 1981. **146**: p. 621-627.

26.   Boeke, J.D. and P. Model. A prokaryotic membrane anchor sequence: Carboxyl terminus of bacteriophage f1 gene III protein retains it in the membrane. Proc. Natl. Acad. Sci. U.S.A., 1982. **79**: p. 5200-5204.

27.   Beck, E., *et al.* Nucleotide sequence of bacteriophage fd DNA. Nucleic Acids Res., 1978. **5**: p. 4495-4503.

28.   Jakes, K.S., N.G. Davis, and N.D. Zinder. A Hybrid Toxin from Bacteriophage f1 Attachment Protein and Colicin E3 Has Altered Cell Receptor Specificity. J. Bacteriol., 1988. **170**: p. 4231-4238.

29.   Stengele, I., *et al.* Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. J. Mol. Biol., 1990. **212**: p. 143-149.

30.   Crissman, J.W. and G.P. Smith. Gene-III Protein of Filamentous Phages: Evidence for a Carboxyl-Terminal Domain with a Role in Morphogenesis. Virology, 1984. **132**: p. 445-455.

31.   Smith, G.P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. Science, 1985. **228**: p. 1315-1317.

32.   Parmley, S.F. and G.P. Smith. Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. Gene, 1988. **73**: p. 305-318.

33.   Cwirla, S.E., *et al.* Peptides on phage: A vast library of peptides for identifying ligands. Proc. Natl. Acad. Sci. U.S.A., 1990. **87**: p. 6378-6382.

34.   Devlin, J.J., L.C. Panganiban, and P.E. Devlin. Random Peptide Libraries: A Source of Specific Protein Binding Molecules. Science, 1990. **249**: p. 404-406.

35.   O'neil, K.T., *et al.* Identification of Novel Peptide Antagonists for GPIIb/IIIa From a Conformationally Constrained Phage Peptide Library. Prot. Struct. Funct. Genet., 1992. **14**: p. 509-515.

36.   Kay, B.K., *et al.* An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. Gene, 1993. **128**: p. 59-65.

37.   Oldenburg, K.R., *et al.* Peptide ligands for a sugar-binding protein isolated from a random peptide library. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 5393-5397.

38.   Scott, S.K., *et al.* A family of concanavalin A-binding peptides from a hexapeptide epitope library. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 5398-5402.

39.   Schumacher, T.N., *et al.* Identification of D-Peptide Ligands Through Mirror-Image Phage Display. Science, 1996. **271**: p. 1854-1857.

40.   Rickles, R.J., *et al.* Phage display selection of ligand residues important for Src homology 3 domain binding specificity. Proc. Natl. Acad. Sci. U.S.A., 1995. **92**: p. 10909-10913.

41.   Stephen, C.W. and D.P. Lane. Mutant Conformation of p53: Precise Epitope Mapping Using a Filamentous Phage Epitope Library. J. Mol. Biol., 1992. **225**: p. 577-583.

42.   Rottgen, P. and J. Collins. A human pancreatic secretory trypsin inhibitor presenting a hypervariable highly constrained epitope via monovalent phagemid display. Gene, 1995. **164**: p. 243-250.

43.   Wang, C., Q. yang, and C.S. Craik. Isolation of a High Affinity Inhibitor of Urokinase-type Plasminogen Activator by Phage Display of Ecotin. J. Biol. Chem., 1995. **270**: p. 12250-12256.

44.   Markland, W., A.C. Ley, and R.C. Ladner. Iterative Optimization of High-Affinity Protease Inhibitors Using Phage Display. 2. Plasma Kallikrein and Thrombin. Biochemistry, 1996. **35**: p. 8058-8067.

45.   Dennis, M.S. and R.A. Lazarus. Kunitz Domain Inhibitors of Tissue Factor-Factor VIIa. J. Biol. Chem., 1994. **35**: p. 22137-22144.

46.   vanMeijer, M., *et al.* Selective Screening of a Large Phage Display Library of Plasminogen Activator Inhibitor 1 Mutants to Localize Interaction Sites with Either Thrombin or the Variable Region 1 of Tissue-type Plasminogen Activator. J. Biol. Chem., 1996. **271**: p. 7423-7428.

47.   Tanaka, A.S., *et al.* Functional Display and Expression of Chicken Cystatin Using a Phagemid System. Biochem. Biophys. Res. Comm., 1995. **214**: p. 389-395.

48.   Roberts, B.L., *et al.* Directed evolution of a protein: Selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 2429-2433.

49.   Toniatti, C., *et al.* Engineering human interleukin-6 to obtain variants with strongly enhanced bioactivity. EMBO J., 1996. **15**: p. 2726-2737.

50.   Saggio, I., I. Gloaguen, and R. Laufer. Functional phage display of ciliary neurotrophic factor. Gene, 1995. **152**: p. 35-39.

51.   Bass, S., R. Greene, and J.A. Wells. Hormone Phage: An Enrichment Method for Variant Proteins With Altered Binding Properties. Proteins Struct. Funct. Genet., 1990. **8**: p. 309-314.

52.   Laird-Offringa, I.A. and J.G. Belasco. Analysis of RNA-binding proteins by *in vitro* genetic selection: Identification of an amino acid residue important for locking U1A onto its RNA target. Proc. Natl. Acad. Sci. U.S.A., 1995. **92**: p. 11859-11863.

53. Wu, H., W. Yang, and C.F. Barbas. Building zinc fingers by selection: Toward a therapeutic application. Proc. Natl. Acad. Sci. U.S.A., 1995. **92**: p. 344-348.

54. Choo, Y. and A. Klug. Toward a code for the interactions of zinc fingers with DNA: Selection of randomized fingers displayed on phage. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 11163-11167.

55. Jamieson, A.C., S. Kim, and J.A. Wells. *In Vitro* Selection of Zinc Fingers with Altered DNA-Binding Specificity. Biochemistry, 1994. **33**: p. 5689-5695.

56. Widersten, M. and B. Mannervik. Glutathione Transderases withNovel Active Sites Isolated by Phage Display from a Library of Random Mutants. J. Mol. Biol., 1995. **250**: p. 115-122.

57. McCafferty, J., R.H. Jackson, and D.J. Chiswell. Phage-Enzymes: expression and affinity chromatography of functional alkaline phosphatase on the surface of bacteriophage. Protein Eng., 1991. **4**: p. 955-961.

58. Corey, C.R., *et al.* Trypsin display on the surface of bacteriophage. Gene, 1993. **128**: p. 129-134.

59. Garrard, L.J. and D.J. Henner. Selection of an anti-IGF-1 Fab from a Fab phage library created by mutagenesis of multiple CDR loops. Gene, 1993. **128**: p. 103-109.

60. Gram, H., *et al. In vitro* slelection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 3576-3580.

61. Hoogenboom, H.R. and G. Winter. By-passing Immunisation: Human Antibodies from Synthetic Repertoires of Germline VH Gene Segments Rearranged *in Vitro*. J Mol Biol, 1992. **227**: p. 381-388.

62. Chen, H. and D.A. Kendall. Artificial Transmembrane Segments. J. Biol. Chem., 1995. **270**: p. 14115-14122.

63. Davis, N.G. and P. Model. An Artificial Anchor Domain: Hydrophobicity Suffices to Stop Transfer. Cell, 1985. **41**: p. 607-614.

64.   Gailus, V. and I. Rasched. The adsorption protein of bacteriophage fd and its neighbour minor coat protein build a structural entity. Eur. J. Biochem., 1994. **222**: p. 927-931.

65.   Goldsmith, M.E. and W.H. Konigsberg. Adsorption Protein of the Bacteriophage fd: Isolation, Molecualr Properties, and Location in the Virus. Biochemistry, 1977. **16**: p. 2686-2693.

66.   Glaser-Wuttke, G., J. Keppner, and I. Rasched. Pore-forming properties of the adsorption protein of filamentous phage fd. Bichim. Biophys. Acta, 1989. **985**: p. 239-247.

67.   Smith, G.P. and J.K. Scott. Libraries of Peptides and Proteins Displayed on Filamentous Phage. Methods Enzymol., 1993. **217**: p. 228-257.

68.   McConnell, S.J. and R.H. Hoess. Tendamistat as a Scaffold for Conformationally Constrained Phage Peptide Libraries. J. Mol. Biol., 1995. **250**: p. 460-470.

Table 1:   Detergent effects on zif phage plating recoveries

Zif phage were preincubated with the biotinylated DNA duplex shown in figure 3 at the indicated concentrations. Phage were then diluted 5-fold into a buffer containing 0.5% (w/v) of the indicated detergents, applied to streptavidin - coated wells of a microtiter plate, washed with buffer containing 0.5% (w/v) of the indicated detergents, eluted, and quantitated via transduction of *E. coli* to tetracycline resistance. The % of applied phage recovered in the eluates is shown. Reported values represent the average of duplicate samples. All duplicate samples agreed within a factor of 2.

Procedures: Binding reactions were prepared which contained Zif phage ($\sim 6 \times 10^{10}$/ml (= 0.1 nM), quantitated as described above), the biotinylated target DNA fragment shown in figure 3 (at the indicated concentrations), and VCSM13 phage ($\sim 6 \times 10^{12}$/ml) in 10 µl of binding buffer. Reactions were equilibrated 90'. Next, each reaction was diluted into 40 µl of 0.05 M NaCl wash buffer containing either no detergent, 0 .5% (w/v) triton x-100, or 0.5% (w/v) of NP-40, and applied immediately to a streptavidin - coated well of a microtiter plate. Reactions were equilibrated in the wells for 30', and then the wells were rinsed with 10 washes of 250 µl 0.05 M NaCl wash buffer containing either no detergent, 0 .5% (w/v) triton x-100, or 0.5% (w/v) of NP-40. The total wash time was 25'. Next, 40 µl of elution buffer was added to each well and incubated for 30'. Eluted phage were quantitated essentially as described above.

Table 1: Detergent effects on Zif phage plating recoveries

$\begin{bmatrix} \text{biotinylated} \\ \text{target DNA} \end{bmatrix}$ (nM)	% phage recovered		
	no detergent	triton x-100	NP-40
---	0.22	0.004	0.003
23	0.30	0.73	0.57
918	0.30	0.35	0.27

Table 2:   Specificity of Zif phage binding to plates

Zif phage were preincubated with the biotinylated DNA duplex shown in figure 3 and the indicated competitors. Phage were then diluted 5-fold, applied to streptavidin - coated wells of a microtiter plate, washed, eluted, and quantitated via transduction of *E. coli* to tetracycline resistance. The % of applied phage recovered in the eluates is shown.

Duplicate samples were tested for each condition. For successful conditions (recoveries of ~1%), both samples agreed to within a factor of 1.5 and were averaged. For unsuccessful conditions, quantitation was more difficult due to low colony counts. Therefore, an upper bound is provided (0.03%) which equals the highest apparent recovery observed in this set of samples.

Procedures: Binding reactions were prepared which contained Zif phage ($\sim 2 \times 10^{10}$/ml (= 0.033 nM)), the biotinylated target DNA fragment shown in figure 3 (15 nM), and VCSM13 phage ($\sim 3 \times 10^{12}$/ml) in 16 µl of binding buffer. As indicated, several reactions also contained one of the following competitors: nonbiotinylated target DNA (from figure 4, 1.5 µM); calf thymus DNA (47 µg/ml = 7.2 µM, assuming 1 binding site every 10 base pairs); streptavidin (1.5 µM of biotin - binding activity); or biotin (577 µM). Reactions were equilibrated 50', diluted into 40 µl of 0.05 M NaCl wash buffer, and applied immediately to a streptavidin - coated wells of a microtiter plate (except for the no-streptavidin control sample, which was applied to an untreated well). Reactions

were equili. .ed in the wells for 40', and then the wells were rinsed with 10 washes of 250 µl 0.05 M NaCl wash buffer. The total wash time was 20'. Next, 40 µl of elution buffer was added to each well and incubated for 30'. Eluted phage were quantitated as described above.

Table 2: Specificity of Zif phage binding to plates:

Streptavidin coating	Binding reaction components		% Phage recovered
	Biotinylated target DNA	Competitor (molar excess over target DNA)	
+	+	--	1.1
+	--	--	<.03
--	+	--	<.03
+	+	target DNA (no biotin)(100x)	<.03
+	+	calf thymus DNA (480x)	1.1
+	+	streptavidin (100x)	<.03
+	+	biotin (38500x)	<.03

Table 3:   Plating behavior of Zif phage on various surfaces

Zif phage were preincubated with the biotinylated DNA duplex shown in figure 3. Phage were then diluted 5-fold, applied to streptavidin - coated wells of the indicated plates, washed, eluted, and quantitated via transduction of *E. coli* to tetracycline resistance. The ratio of % phage recovered in the presence and absence of biotinylated DNA target site is indicated.

Duplicate samples were tested for each condition, and similar results were obtained. Reported results are the average of each duplicate pair.

Procedures: A binding reaction was prepared which contained Zif phage ($\sim 6 \times 10^{10}$/ml (= 0.1 nM), quantitated as transducing units), the biotinylated target DNA fragment shown in figure 3 (21 nM), and VCSM13 phage ($\sim 3 \times 10^{12}$/ml), in 90 μl of binding buffer. A second reaction was also prepared without DNA. Both reactions were equilibrated 60' and diluted into 360 μl of 0.05 M NaCl wash buffer. 50 μl aliquots of the diluted binding reactions were applied to streptavidin - coated wells of the indicated microtiter plates and equilibrated in the wells for 35'. The wells were then rinsed with 10 washes of 250 μl 0.05 M NaCl wash buffer. The total wash time was 40'. Finally, 40 μl of elution buffer was added to each well and incubated for 30'. Eluted phage were quantitated as described above.

Table 3: Plating behavior of Zif phage on various surfaces

Plating surface		$\left[\dfrac{\text{\% recovery (+ DNA)}}{\text{\% recovery (no DNA)}}\right]$
Vendor	Product	
Dynatech	Immulon-2	50
	Immulon-3	30
	Immulon-4	53
Corning	# 25860	1
Falcon	# 3915	77
Falcon	# 3072	1

Table 4:   DNA-binding specificity of Zif phage

Zif phage were preincubated with the indicated biotinylated DNA duplexes. Phage were then diluted 5-fold, applied to streptavidin - coated wells of a microtiter plate, washed, eluted, and quantitated via transduction of *E. coli* to tetracycline resistance. The % of applied phage recovered in the eluates is shown.

Duplicate samples were tested for each condition. Reported results are the average of each duplicate pair.

Procedures: Binding reactions were prepared which contained Zif phage ($\sim 5 \times 10^{10}$/ml (= 0.083 nM)), VCSM13 phage ($\sim 3 \times 10^{12}$/ml), and either of the biotinylated target DNA fragments shown (43 nM), in 44 μl of binding buffer. A control binding reaction was also prepared which contained no DNA site. Reactions were equilibrated 55', diluted into 160 μl of 0.05 M NaCl wash buffer. 40 μl of each diluted reaction was then applied to a streptavidin - coated well of a microtiter plate. Reactions were equilibrated in the wells for 40', and then the wells were rinsed with 10 washes of 250 μl 0.05 M NaCl wash buffer. The total wash time was 29'. Finally, 40 μl of elution buffer was added to each well and incubated for 30'. Eluted phage were quantitated as described above.

Table 4: DNA-binding specificity of Zif phage:

biotinylated duplex		% phage recovered
Zif site (overlined)	5 'CTGAGCCGTGGGCGTAGTGATCGATC --biotin GACTCGCACCCGCATCAC	0.22
'CCTG' site (changes in grey)	5 'CTGAGCCGTGGCGCTGAGTGATCGATC --biotin GACTCGCACCGGACTCAC	0.017
no DNA		0.007

Figure 1A: Sketch of fd phage

Fd is shaped like a long, flexible cylinder. As indicated, several copies of the pIII coat protein protrude from one end. The remainder of the phage body is composed of four other proteins which are not shown.

Figure 1B: Sketch of pIII protein

PIII contains four functional regions: a signal sequence, which is clipped off during secretion and is not present on the mature protein; a large globular domain; a flexible spacer region; and a carboxy terminal segment which inserts into the phage body.

Figure 1A: fd phage

pIII proteins
(3 shown)
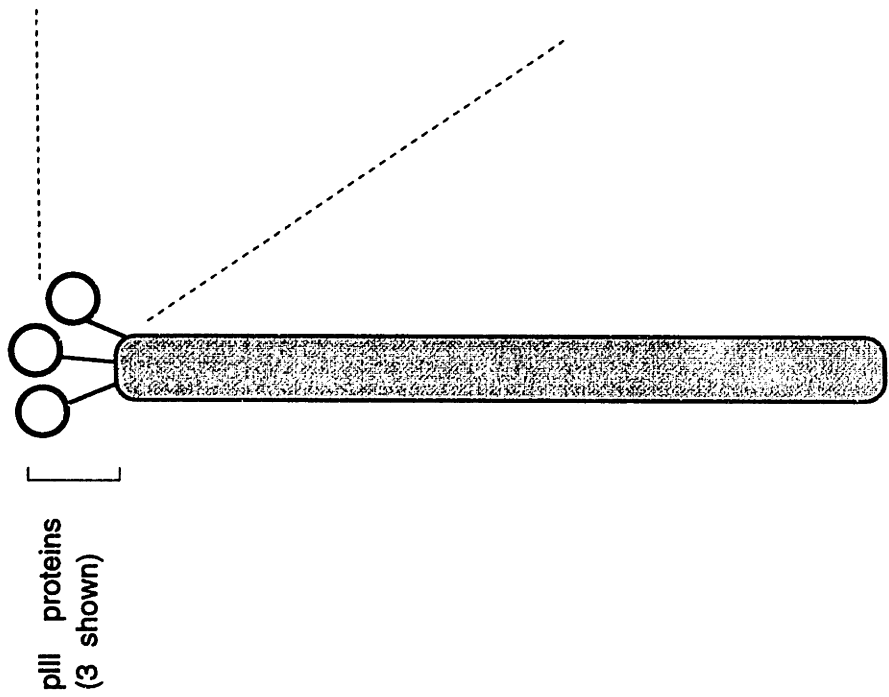
Figure 1B: pIII structure

signal sequence
(clipped off
during secretion)

NH2

globular
domain

spacer

carboxy terminal segment

COOH
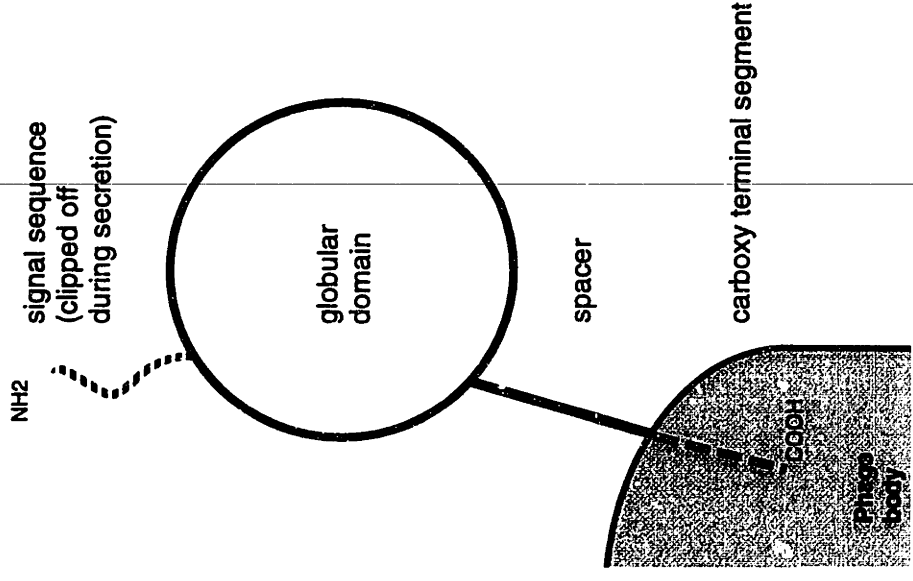
Phage
body

Figure 2A: Construction of the Zif phage vector fd-tet.Zif

The fd phage vector fUSE3 was converted into fd-tet.Zif in two steps: (i) A polylinker was inserted into the XhoI site of fUSE3. (ii) A PCR-amplified fragment of Zif268 cDNA encompassing bases 1287-1585[17], was cut with Apa1 and Xba1 and then ligated into the Apa1 and Spe1 sites of the polylinker.

fUSE3 was a gift of G. P. Smith, and is identical to fUSE2[32], except that the 2 vectors differ in a small region near the 5' end of the pIII gene. The DNA sequences of this region in fUSE2 and fUSE3 are, respectively: 5'-GCTGAAGATCTTGAAAGT TGT and 5'-GCTGAAACTCGAGTTGT. Differences are underlined. Bold letters show the Xho1 site of fUSE3. The PCR primers used to clone the Zif fingers were: 5'-GGATCGATTCCATGGGGCCCCATGAACG CCCATATGC and 5'-GGGTCGACTGCAG ATCTAGAGGCCACCACACTTTTGTC. [The Apa 1 and Xba 1 sites are underlined.]

Figure 2B: Sketch of the Zif phage:

The Zif268 zinc finger peptide, which contains three fingers (denoted by the numbered circles), is fused to the amino terminus of the globular domain of the phage coat protein pIII. Three to five copies of this fusion protein should be present at one end of the virion.

Figure 2A: Construction of fd-tet.Zif

Sph I

Zif268 DNA bases 1287-1585

Apa I   Spe I

TCGATCAGGGCCCTTCGAAACTAGTGCACCTGCAGAGGATCTT
AGTCCCGGGAAGCTTTGATCACGTGGACGTCTCCTAGAAAGCT

Xho I

pIII gene

fUSE3:

Tet^R
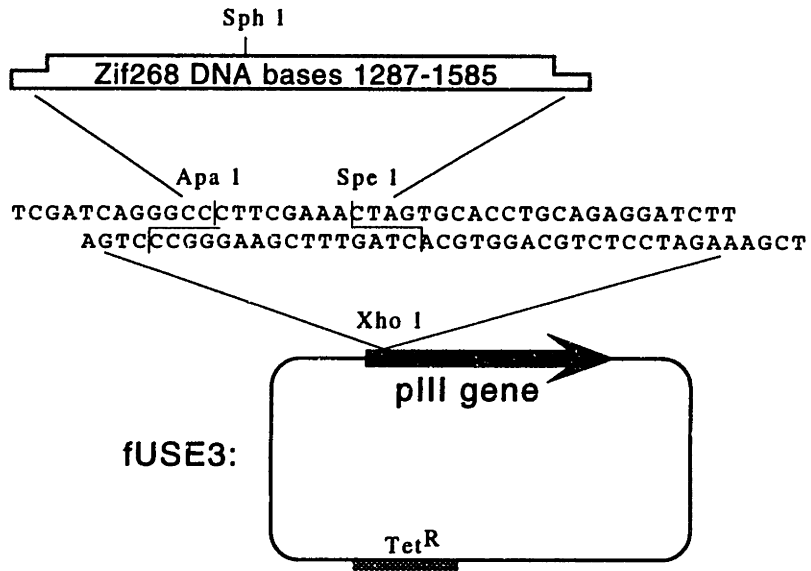
Figure 2B: Sketch of Zif phage particle

Phage body

pIII—③②①
pIII—③②①
pIII—③②①

Zif268-pIII
fusion proteins
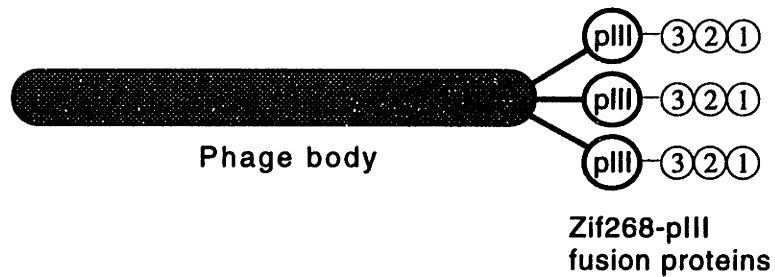
Figure 3: Gel shift assay for Zif phage DNA-binding activity

Binding reactions were prepared which contained Zif phage, DNA probe, and various competitors. Phage-bound and unbound probe were separated via electrophoreses through a 1% agarose gel. Autoradiography revealed bands corresponding to the wells, phage-bound probe, and free probe as indicated. The DNA probe used for this study is shown at bottom and contains four Zif sites (overlined). The 'b' in the top strand denotes the location of a single biotin moiety. Note that the amounts of competitor are provided as molar ratios of the labeled probe concentration, since probe was present in excess over Zif phage. This experiment was performed twice, using Zif phage derived from two different constructs, and similar results were obtained.

Procedures: Five binding reactions were prepared which contained DNA probe (90 pM), Zif phage (~25 pM, quantitated as infectious units), and dithiothreitol (DTT, 1mM) in 22 µl of binding buffer. Binding reactions also contained various competitors as shown. Competitor concentrations were: Zif protein: 18 nM; cold probe: 9 nM; and calf thymus DNA: 9 µg/ml or 0.45 µg/ml (~1400 nM or ~70 nM, assuming one site for every ten base pairs). Note that Zif phage was added last to all binding reactions except the one which contained competitor Zif protein. For this reaction, labeled probe was added last.

Samples were equilibrated 15', loaded onto a 1% agarose gel, and electrophoresed at 100 V for approximately 1 hour. Gel running buffer was 15 mM Hepes pH 7.8. Gel was dried and exposed to film.

Figure 3: Gel shift assay for Zif phage DNA-binding activity:

competitor (molar ratio to probe)   :

no competitor		+			
zif protein (200x)			+		
calf thymus DNA: (780x)				+	
(15600x)					+
cold probe (100x)						+

wells

bound

free
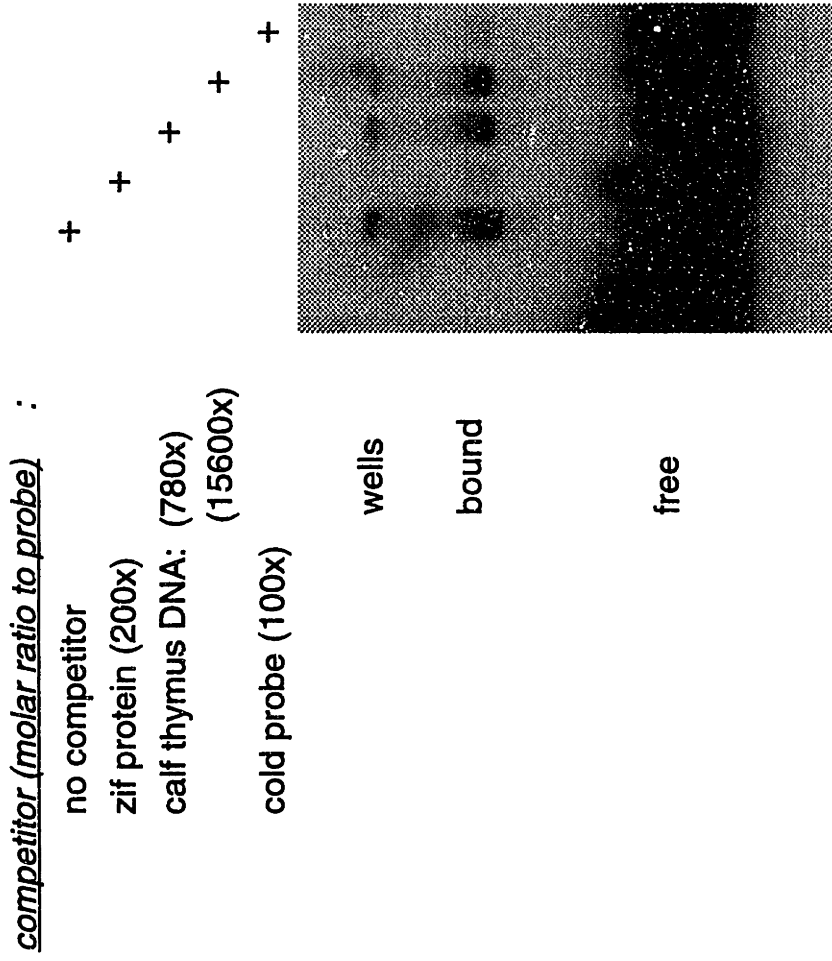
Zif sites
_____   _____

DNA probe:

TGGTAGCGGGGGGCGTGGGCGGCGGGCGGCGTGGGCGGACTCCAAATAbAA 3'
ACCATCGCCCCCGCACCCGCCGCCCCCGCACCCGCCGCCCTGAGG 5'

Figure 4:   Detergent effects on Zif phage gel shift behavior

Binding reactions were prepared which contained Zif phage, DNA probe, and various detergents or chaotropic agents. Each additive was tested at two concentrations: KSCN: 2 M or 0.5 M; o-nitrophenyl ß-D-galactopyranoside: 1% or .1% (w/v); all others: 0.5% or 0.05% (w/v). These reactions were then electrophoresed through a 1% agarose gel. Autoradiography of the dried gel revealed bands corresponding to the wells, phage-bound probe, and free probe as indicated. Additionally, a series of faint bands were observed which ran between the free and phage-shifted probe. These bands may be due to presence of phage fragments or proteolytic pIII-Zif products in this phage prep.

The DNA probe used for this study was similar to that shown in figure 3. Its exact sequence was:

```
5'     AATTCCATGGCGGGGGCGTGGGCGGCGGGGGCGTGGGCGGG
3'          GGTACCGCCCCCGCACCCGCCGCCCCCGCACCCGCCCCTAG
```

Zif sites are overlined and underlined.

Procedures: Binding reactions were prepared which contained DNA probe (0.25 nM) and Zif phage (~0.2 nM), in 14 µl of binding buffer. Binding reactions also contained the detergents or chaotropic agents at the following concentrations: KSCN: 2 M or 0.5 M; o-nitrophenyl ß-D-galactopyranoside: 1% or .1% (w/v); all others: 0.5% or 0.05% (w/v). Samples were equilibrated 40', loaded onto a 1% agarose gel, and electrophoresed at approximately 120 V for 80 minutes. Gel running buffer was 15 mM Hepes pH 7.8. Gel was dried and exposed to film.

Figure 4: Detergent effects on Zif phage gel shift behavior:
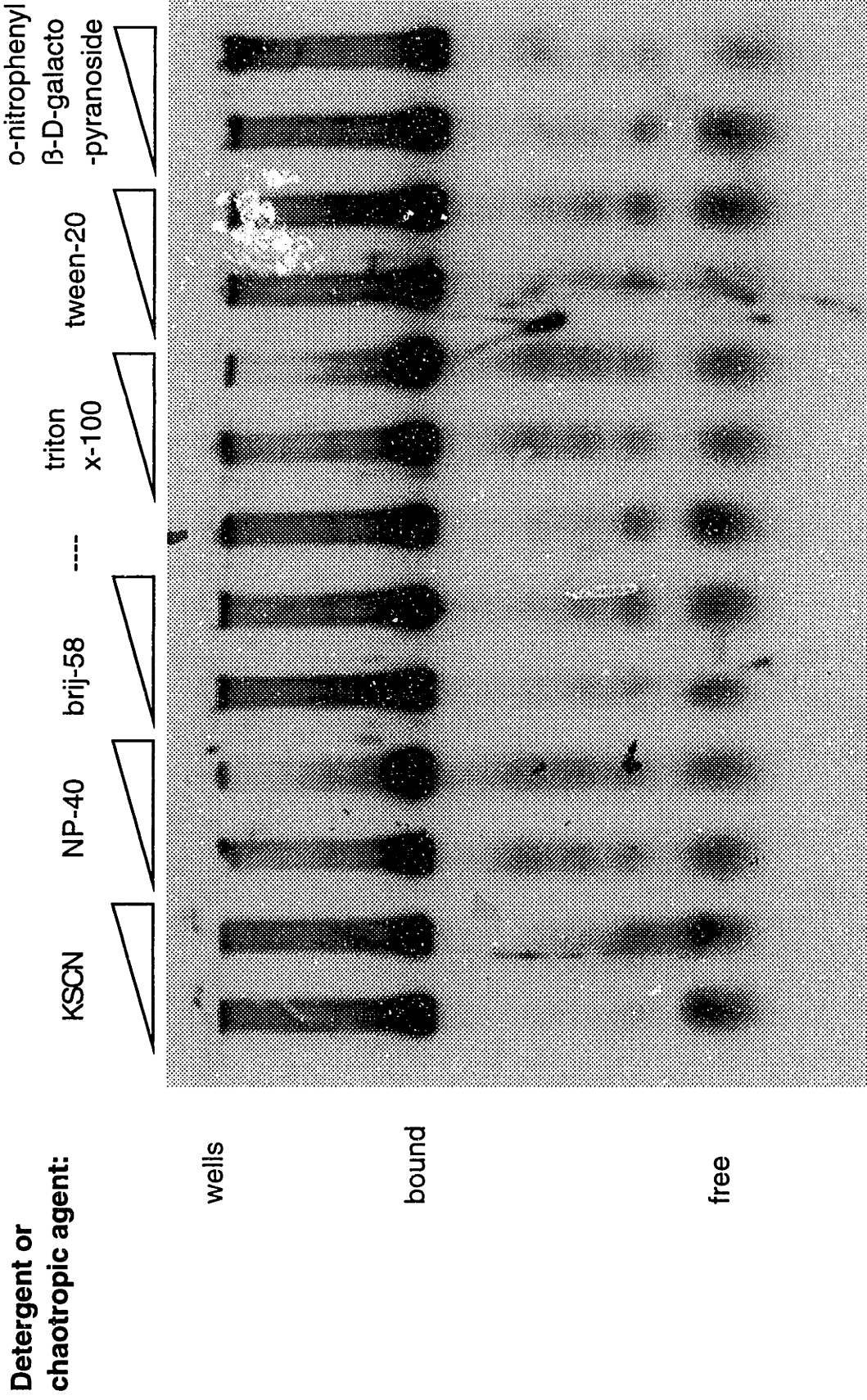
Figure 5:   Zif phage plating behavior with multi-site and single site biotinylated DNA fragments.

Zif phage were preincubated with various concentrations of the multisite biotinylated DNA duplex shown in figure 3 or with the following single site duplex: CTGA<u>GCGTGGGCGT</u>AGTGATCGATC-b
                                                                    GACTCGCACCCGCATCAC

The Zif site is underlined, and the biotin is indicated by 'b'. Phage were then diluted 5-fold, applied to streptavidin - coated wells of the indicated plates, washed, eluted, and quantitated via transduction of *E. coli* to tetracycline resistance. The % of applied phage recovered in the eluates is plotted.

Plotted values are the average of duplicate samples, which in all cases agreed to within a factor of 2.

Procedures: Binding reactions were prepared which contained Zif phage ($\sim 3 \times 10^{10}$/ml (= 0.05 nM)), VCSM13 phage ($\sim 6 \times 10^{12}$/ml), and the indicated concentrations of biotinylated DNA fragments in 12 µl of binding buffer. Reactions were equilibrated 50', diluted into 40 µl of 0.05 M NaCl wash buffer, and applied to streptavidin - coated wells of a microtiter plate. The reactions were equilibrated in the streptavidin - coated wells for 25' and then rinsed with 10 washes of 250 µl 0.05 M NaCl wash buffer (total wash time: 20'). Next, 40 µl of elution buffer was added to each well and incubated 60'. Eluted phage were quantitated as tetracycline transducing units essentially as described above.

Figure 5: Zif phage plating behavior with multisite and single site DNA fragments
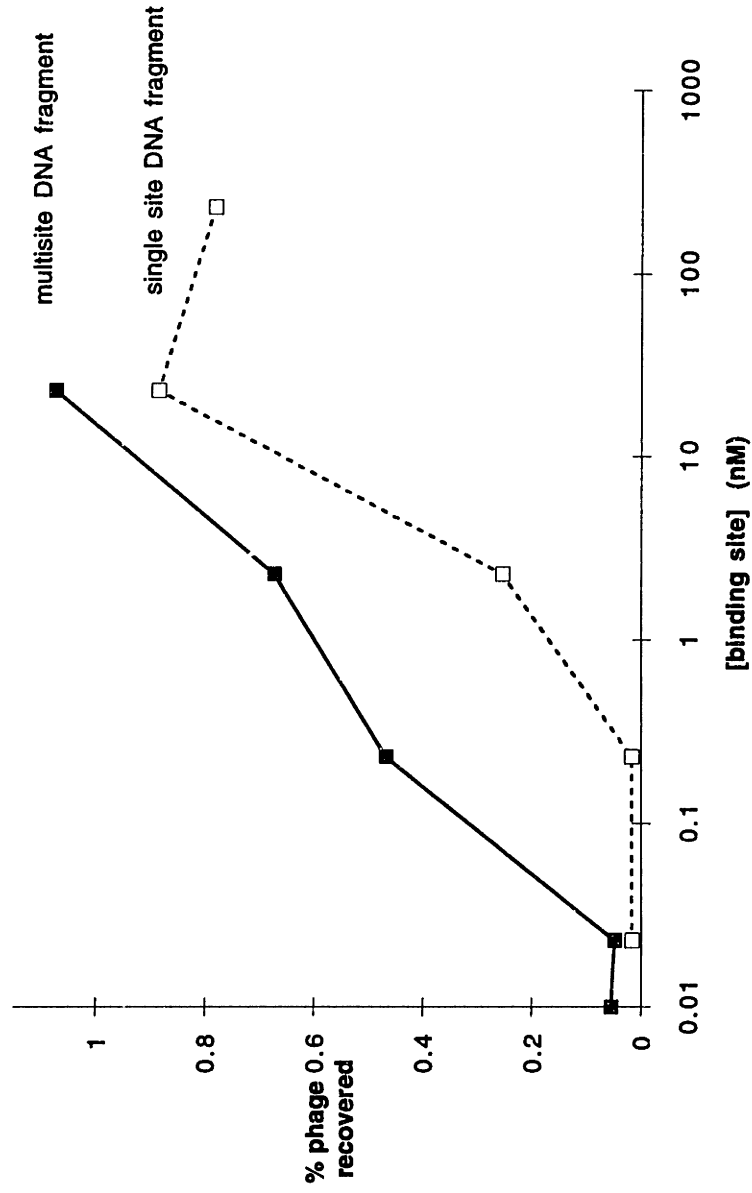
CHAPTER 3

ZINC FINGER PHAGE: AFFINITY SELECTION OF FINGERS WITH NEW DNA-BINDING SPECIFICITIES

Abstract

A phage display system was developed and used to select zinc finger proteins with altered DNA-binding specificities. The three zinc fingers of the Zif268 protein were expressed on the surface of filamentous phage, and a library of variants was prepared by randomizing critical amino acids in the first zinc finger. Affinity selections, using DNA sites with base changes in the region recognized by the first finger, yielded Zif268 variants that bound tightly and specifically to the new sites. This phage system provides a new tool for the study of protein-DNA interactions and may offer a general method for selecting zinc finger proteins that recognize desired target sites on double-stranded DNA.

Designing and selecting proteins with new DNA binding specificities can test and extend our understanding of protein-DNA interactions. The zinc finger motif, of the type first discovered in transcription factor IIIA (1), offers an attractive framework for these studies. This zinc finger is one of the most common eukaryotic DNA-binding motifs (2), and this family of proteins can recognize a diverse set of DNA sequences (3). Zinc finger proteins also exhibit a modular organization which suggests that it may be possible to "mix and match" fingers to obtain proteins with novel DNA-binding specificities (4, 5). Crystallographic studies of the Zif268-DNA complex (4) and other zinc finger-DNA complexes (3, 6) show that residues at four positions make most of the base contacts, and there has been some discussion about rules or codes that may explain zinc finger-DNA recognition (7). However, the recently reported structures of the GLI-DNA complex (3) and of the Tramtrack-DNA complex (6) show that zinc fingers can dock against the DNA in a variety of slightly different ways. This complexity makes model building and rational design more difficult, but it also reemphasizes the versatility of the zinc finger motif.

Phage display systems have provided powerful selection methods for many studies of peptides and proteins (8, 9). To explore the usefulness of phage display for studying zinc finger-DNA interactions, we first expressed the three Zif268 zinc fingers (10) on the surface of filamentous phage. The resulting construct - fd-tet.Zif - produced useful titers of "Zif phage" and these phage bound specifically to the nine base pair site recognized by Zif268. [Details of the construction of fd-tet.zif and testing of Zif phage are provided in chapter 2.] We then created a library of Zif variants by randomizing the four positions

of the first finger that appear most important for making base contacts (*3, 4, 6*). These randomized positions include the residue immediately preceding the α helix and include the second, third, and sixth residues of this helix (*11*).

Affinity selection methods were then used to search the library for phage that would recognize altered binding sites. In each round of affinity selection, phage were equilibrated with biotinylated target DNA and then applied to streptavidin-coated microtiter wells. After washing, the retained phage were eluted in high salt, amplified in *Escheria coli*, and purified to prepare for the next cycle. The target DNA duplexes for these selections contained modified Zif268 binding sites with changes in the region recognized by finger one (*4*), and we refer to each duplex by giving the sequence of this region (Fig. 1A). Initially, we performed five rounds of selection with each of the target sites (Fig. 1B) (*12*). During these initial selection series, retention efficiencies in the GACC and GCAC selected phage pools increased about 100 times, while retention efficiencies for the CCTG pool remained low (*13*). We then used these enriched GACC and GCAC pools as a starting point for additional, more stringent, selection cycles (Fig. 1B) (*12*). The CCTG pool was not studied further.

Phage pools from critical stages of the GACC and GCAC selections were characterized by sequencing (Fig. 1B), and clear amino acid preferences were apparent in each pool. For the GACC pool, sequencing after the initial selection series showed that all of the phage (12/12) could be characterized by the consensus sequence (S/D/T) _ N R (Table 1). Three additional rounds of selection using high salt washes did not substantially change this consensus (Table 2). For the GCAC selections,

sequencing revealed interesting changes in the later pools. After the initial selection series, many of the phage belonged to a group characterized by the consensus sequence R _ D R (18/22), but there also was a group characterized by the sequence _ G (S/T) R (4/22) (Table 1). After additional rounds of selection with high salt washes, a single sequence - R A D R - from the first group predominated (Table 2). However, when the additional rounds of selection used both high salt washes and competitor Zif268 site in the binding reactions, we found that a single sequence - Q G S R - from the second group predominated.

Three Zif268 variants were studied in more detail by recloning and overexpressing them in *E. coli*, then purifying the resultant peptides (*14*) and measuring DNA-binding affinities. We studied the predominant Zif268 variants obtained in each of the three later selection series - 'D S N R', 'R A D R' and 'Q G S R' (Table 2) - and also included the wild-type peptide 'R D E R' as a control. Peptide affinities for each of three binding sites ['GACC', 'GCAC' and 'GCGC' (wild-type)] were determined by quantitative gel shift analysis (Table 3) (*15*). Each of the variant peptides binds with high affinity to the site used for its selection (Table 3 - boxed entries). Moreover, the DSNR and QGSR peptides exhibit new specificities in that they bind to these new sites substantially better than they bind to GCGC. The RADR peptide (unlike the QGSR peptide) does not discriminate well between GCAC and GCGC. It is interesting to note that the only difference in the selection conditions for these two variants was the use of competitor Zif268 site in the selections which yielded QGSR.

Although our main goal was to test the feasibility of this selection system, these experiments have several interesting

implications: 1) Our results show how this phage system can be used to help identify critical protein-DNA contacts. Thus the arginine selected at position 6 of the α helix was present in each of the selected phage (72/72), and the crystal structure of the Zif268 complex readily explains this preference (4). We presume that this arginine makes a pair of hydrogen bonds with the guanine at the 5' end of the GACC and GCAC subsites. Likewise, every Zif variant selected with GACC has an asparagine at position 3, and, as seen in the Tramtrack-DNA complex (6), this asparagine could make critical contacts with the adenine. 2) The apparent failure of the CCTG selections raises some important questions. Is this sequence inherently more difficult to recognize with a zinc finger? Would further variation (for example, randomizing additional residues to change base contacts and phosphate contacts) have allowed recognition of this subsite? 3) Results from the GCAC selections suggest that adding specific competitor DNA can aid in the recovery of variants with altered specificities, and we presume that the specific competitor screened out variants which still recognized the wild type site.

The experiments reported in this paper demonstrate that the phage display system can be used to select zinc fingers with novel DNA-binding specificities, and extensions of these strategies will allow us to explore the limits of zinc finger design: Is it possible to select a zinc finger peptide that recognizes any desired sequence on double-stranded DNA? What are the limits of affinity and specificity in DNA binding? The phage display system offers a new tool for addressing fundamental questions about protein-DNA recognition and

may also provide a means for generating new DNA-binding proteins that can be used for research, diagnosis, and therapy.

References and Footnotes

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

2. G. H. Jacobs, *ibid.* **11**, 4507 (1992); G. R. Pellegrino and J. M. Berg, *Proc. Natl. Acad. Sci. USA* **88**, 671 (1991).

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

4. N. P. Pavletich and C. O. Pabo, *ibid.* **252**, 809 (1991).

5. J. R. Desjarlais and J. M. Berg, *Proc. Natl. Acad. Sci. USA* **90**, 2256 (1993).

6. L. Fairall, J. W. R. Schwabe, L. Chapman, J. T. Finch, D. Rhodes, *Nature* **366**, 483 (1993).

7. J. R. Desjarlais and J. M. Berg, *Proc. Natl. Acad. Sci. USA* **89**, 7345 (1992); R. E. Klevit, *Science* **253**, 1367 (1991).

8. J. K. Scott and G. P. Smith, *Science* **249**, 386 (1990); R. H. Hoess, *Curr. Opin. Struct. Biol.* **3**, 572 (1993).

9. G. P. Smith and J. K. Scott, *Methods Enzymol.* **217**, 228 (1993).

10. B. A. Christy, L. F. Lau, D. Nathans, *Proc. Natl. Acad. Sci. USA* **85**, 7857 (1988).

11.   To construct the library, 2 oligonucleotides were synthesized: 5'
GGAATCGATTCCATG<u>GGGGCCCC</u>ATGAACGGCCGTACGCTTGCCCTGTCGAGTCCT
GCGATCGTCGATTTTCG and 5' CCATCTCGATC<u>GCATGC</u>ATATTCGACACTGG
AAGGGCTTCTGGCCT GTGTGGATCCGGATATG**SNN**GGTGAGS**NNS**NNAGAS**NN**
CGAAAATCGACG (**N** = A, T, G and C; **S** = G and C), with complementary
12-base 3' ends.   These were annealed and then extended using
sequenase 2.0 (United States Biochemical).   The resulting duplex was
digested with Apa 1 and Sph 1 [sites are underlined] and ligated with
the large Apa 1/Sph 1 fragment of fd-tet.Zif.   Ligation products were
electroporated into MC1061 cells (9) and this yielded ~2.8 x 10^7
independent transformants.

This library was grown as described in the chapter 3 appendix.
Phage were purified by ultracentrifugation (171,000x G, 4°C, 6 hours)
and phage pellets were resuspended in ~1/100 volume of binding buffer
(50 mM NaCl, 5 mM $MgCl_2$, 10 µM $ZnCl_2$, 5% glycerol, 0.1 mg/ml BSA, and
15 mM Hepes pH 7.8).   This final library phage prep (~4.7 x 10^{11} TTU)
was stored anaerobically (<1 ppm O_2) on ice.   Due to concerns about
oxidation, all phage manipulations were performed so as to minimize
exposure to O_2.

To estimate library complexity we sequenced 20 unselected
clones.   [Single-stranded templates were sequenced using sequenase
2.0 and protocols from United States Biochemical.]   Three corresponded
to the parent construct (fd-tet.Zif) and appear to have resulted from
the reinsertion of the fragment excised during library construction.
Seventeen phage contained the correct library insert, but there was a
significant cytosine bias at the randomized codons.   Base ratios were

C:A:T:G=48:19:19:15 at the first 2 codon positions and C:G=74:26 at position 3.

12. Our selection protocol is based on the 'biopanning' procedure (9). The first round in each initial selection series (the leftmost arrow in each of the 3 pathways in Fig. 1B) was performed as follows: Binding reactions (121 µl) were made which contained ~3.5 x 10^{10} TTU of library phage, 39 nM of biotinylated target DNA (GACC, GCAC or CCTG (Fig.1A)), and 0.059 mg/ml sheared calf thymus DNA in .9X binding buffer (11). Each sample was preincubated for 50 min, diluted into 3.6 volumes of 0.05 M NaCl wash buffer (0.05 M NaCl with 5 mM $MgCl_2$, 10 µM $ZnCl_2$, 5% glycerol, 0.5% w/v triton X-100, and 15 mM Hepes pH 7.8) and applied to streptavidin-coated wells (6 wells, 30 µl/well) of a Pro-Bind™ plate (Becton Dickinson). [These wells had been prepared as described in the appendix to chapter 3.] After 50 min the samples were removed from the wells and then: i) over a period of 35 min, the wells were rinsed 10 times with 0.25 ml of 0.45 M NaCl wash buffer (identical to 0.05 M NaCl wash buffer except for the higher NaCl concentration), and ii) 40 µl of elution buffer (binding buffer (11) with 4 M NaCl) was added to each well. After eluting for 2 hours, each set of 6 eluates was pooled, titered, and used to infect K91 cells. Transduced cells were incubated for 1 hour at 37°C in 5 ml of LB broth containing 0.2 µg/ml tetracycline, pelleted (15 min,1600x G) and resuspended in 50 ml of degassed Zif phage broth. Each culture was then grown anaerobically in a 50 ml centrifuge tube and purified essentially as described (11). Other rounds of selection in the initial series were similar except that, starting at round 3, sonicated salmon sperm DNA

was substituted for sheared calf thymus DNA in the binding reactions. Selections in the later series were similar except that 0.75 M NaCl washes were used and the binding reactions in one of the GCAC selection series included a nonbiotinylated Zif268 binding site (0.36 µM) as a specific competitor. All phage manipulations, except for elution and infection of K91 cells, were performed in an anaerobic chamber with < 1 ppm O_2. All buffers used in the anaerobic chamber had been degassed (by bubbling with helium for 1-2 hours) and then equilibrated with the chamber atmosphere for at least 1 day.

13. In the first round of selection, < 0.009% of library phage applied to the streptavidin-coated wells was recovered in the eluates. By the fifth round this retention efficiency had risen to 0.6% - 0.8% for the GACC and GCAC phage pools, but was <0.001% for the CCTG pool. For comparison, control experiments using Zif phage and a biotinylated wild type Zif site typically yielded retention efficiencies of .5% - 1.0%.

A plot of retention efficiency versus selection cycle is provided in figure 3 of the appendix to this chapter.

14. The zinc finger regions from the phage variants (corresponding to residues 333-421 of Zif268 (10)) were subcloned into the T7 expression vectors pET-3d or pET-21d (Novagen). To do this, we i) used PCR to retrieve the zinc finger-encoding regions of the phage variants; ii) digested the PCR fragments with Nco 1 and Hind 3; and iii) ligated these fragments into the Nco 1 and Hind 3 sites of the T7 expression vectors pET-3d (for RADR and QGSR) or pET-21d (for DSNR) (Novagen). PCR primers were: 5'-GTACGAATTC <u>AAGCTT</u>ACTAGTCCTTCTGTCTTAAAT

and 5'-GATGCTAGCG<u>CCATGG</u>AACGGC CGTACGCTTGC. Restriction sites are underlined.

These expression constructs were transformed into BL21 cells containing the pLysS plasmid and then induced as recommended (Novagen). Additionally, the corresponding wild type peptide [RDER] was expressed as described (4). Zinc finger peptides were purified as follows: i) the peptides were released from the cells, solubilized, and reduced as described (4); ii) peptides were extracted from the reduction solution using C18 sep-packs (Millipore), eluted into 40% $CH_3CN/.1\%$ TFA, and lyophilized; iii) peptides were reduced a second time and purified by reversed phase HPLC as described (4), and lyophilized; iv) the final peptide preparations were reconstituted in water in an anaerobic chamber and adjusted to 2.75 mM $ZnSO_4$ and 50 mM bis-tris propane (pH 6.8). Peptide aliquots were stored at -80°C. To estimate purity, peptides were subjected to SDS-PAGE and silver stained. No impurities staining as intensely as 2% of the purified peptide were observed in any preparation (data not shown).

To ensure peptide identity, we i) removed an aliquot of each culture immediately before induction; ii) used PCR to create single stranded template DNA from each aliquot; and iii) sequenced the entire zinc finger region of each template (*11*).

15. To derive apparent K_d's we: i) used quantitative gel shift analysis to determine the fraction of DNA fragment bound at a series of peptide concentrations; ii) estimated the K_d at each point in the transition region of the resulting "binding curve"; and iii) averaged these K_d's. We used those points for which 0.1≤ fraction DNA bound ≤ 0.9 (6 or 7

points). Standard deviations were always $<K_{d_{avg}}/4$. To estimate K_d's, we used the approximation $K_d \approx ([P_{tot}])([D]/[PD])$ (where $[P_{tot}]$ is the total protein concentration and $[D]/[PD]$ is the ratio of free to bound DNA).

Binding reactions contained radioactive DNA fragment (~2.5 pM or ~25 pM), peptide (from a 2-fold dilution series prepared in deoxygenated gel shift buffer) and 14.7 µg/ml poly (dI-dC)-poly (dI-dC) (Pharmacia) in degassed gel shift buffer (50 mM NaCl, 5 mM $MgCl_2$, 10 µM $ZnSO_4$, 5% glycerol, 0.1 mg/ml BSA, 0.1% NP-40, and 15 mM Hepes pH 7.8). Binding reactions were prepared in wells of a 96-well plate (Nunc) outside of the anaerobic chamber. Binding reactions were equilibrated at room temperature for either 30 min (for RDER) or 4 hours (for the variant peptides), and electrophoresed at 150 V on 10% polyacrylamide gels in 0.03 M Tris-Hepes, pH 7.8. [Note: control experiments showed that the variant peptides required longer equilibration times (data not shown)]. Dried gels were quantitated using a PhosphorImager system (Molecular Dynamics).

A freshly thawed aliquot of peptide was used for each set of gel shift experiments, and the binding activity was determined by titrating a portion of each aliquot against a defined concentration of binding site (150 µM or 300 µM). Each aliquot was titrated twice, using 2 different DNA fragments (Table 3), and the calculated activities always agreed within 20%.

DNA fragments for the binding studies were prepared by annealing two synthetic oligonucleotides. [They were heated at 85°C for 10 min and cooled to about 30°C in 1 hour.] Oligonucleotides had been prepared essentially as described in footnote (11), except i) a 12 % denaturing polyacrylamide gel was used, and ii) after the final ethanol

precipitation, the oligonucleotides were resuspended in 10 mM tris pH 8. [We did not use TE due to concerns that the EDTA would chelate zinc.] The final oligonucleotide preparations were quantitated by spectrophotometry. Extinction coefficients were calculated from base sequence as described in [J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).]

Radioactive DNA fragments were prepared by i) end-labeling using g^{32}P-ATP and polynucleotide kinase ii) phenol extraction and iii) gel filtration through a spin column containing G-25 Sephadex (Boehringer Manneheim).

Table 1. Amino acid sequences of phage from the GACC and GCAC pools after the initial selection series (Fig. 1B). The four randomized positions in the a helical region of finger one are denoted as '-1', '2', '3', and '6'. Consensus sequences are indicated in bold. An underscore (_) indicates that there is no clear preference at the corresponding position. The numbers in parentheses indicate the total number of times this amino acid sequence was recovered and the number of distinct DNA sequences that encoded this amino acid sequence.

Table 1. Amino acid sequences of phage from the GACC and GCAC pools after the initial selection series

GACC	GCAC
-1 2 3 6	-1 2 3 6
SQNR (4,2)	RSDR (4,2)
DANR (2,1)	RPDR (3,2)
DRNR	RGDR (3,1)
DSNR	HSDR (2,2)
SSNR	RVDR (2,2)
STNR	AADR
TANR	KSDR
TPNR	RADR
	RAER
	R_DR
	NGSR (2,2)
	SGSR
	TGTR
s/d/t_N R	**_G s/t R**

Table 2. Amino acid sequences in the final phage pools (after the later selection series shown in Fig. 1B). The designation '+ competitor' indicates that competitor DNA [non-biotinylated wild type Zif268 site] was added to the binding mixes during the later selection series (12). For an explanation of symbols, see the legend to Table 1.

Table 2. Amino acid sequences in the final phage pools

GACC		GCAC		GCAC (+ competitor)	
-1 2 3 6		-1 2 3 6		-1 2 3 6	
DSNR	(8,4)	RADR	(7,4)	QGSR	(16,3)
SSNR	(4,3)				
DRNR	(2,1)				
NSNR					
D/S S N R		R A D R		Q G S R	

Table 3.   Apparent Kd's for the binding of zinc finger peptides to DNA fragments containing the 'GACC', 'GCAC' and 'GCGC' (wild type) forms of the Zif268 binding site.   Each peptide is specified by giving the amino acid residues at the four positions of finger one that were randomized in the library (-1 , 2, 3 and 6).   RDER is wild type.   The three DNA duplexes share the sequence: AGCAGCTGA[GCGTGG_ _ _]_ AGTGAGCT and are specified by giving the bases at the four underscored positions. [The bracketed region marks the position of the Zif268 binding site (GCGTGGGCG (4)).]   Boxes mark the interaction of each variant with the site used for its selection.

Table 3. Apparent K_d's for the binding of zinc finger peptides to DNA fragments containing the 'GACC', 'GCAC' and 'GCGC' (wild type) forms of the Zif268 binding site.

Finger one sequence	Apparent K_d (nM)		
-1 2 3 6	GACC	GCAC	GCGC (w.t.)
DSNR	0.019	2.5	1.8
RADR	9.3	0.068	0.035
QGSR	1.8	0.055	0.54
RDER (w.t.)	33.	5.6	2.7

Figure 1. (A) The three biotinylated DNA sites used for affinity selections. The sequences of the underscored region were GACC, GCAC or CCTG [where xxxx indicates the appropriate complementary sequence] and these duplexes are referred to as 'GACC', 'GCAC' and 'CCTG'. Zif normally recognizes the consensus sequence GCGTGG<u>GCG</u> (with the first finger contacting the underlined 'GCG' subsite) (4), and the box marks the corresponding region of the duplexes. (B) Overview of selections. Aliquots from the phage library were subjected to five rounds of selection using the biotinylated DNA duplexes GACC, GCAC or CCTG. The GACC and GCAC pools were then used in additional rounds of selection under more stringent conditions. [The washes contained more salt and, for one of the GCAC selections, the binding reactions contained non-biotinylated Zif268 binding site as a specific competitor.] (12) Pools were characterized at the indicated stages by sequencing randomly chosen phage. The 'X' indicates that there were no further selections with the CCTG pool.

Figure 1A: Target sites for affinity selections

GACC
GCAC
CCTG

CTGA|GCGTGG_ _ _|AGTGATCGATC-biotin
GACT|CGCACCxxxx|TCAC
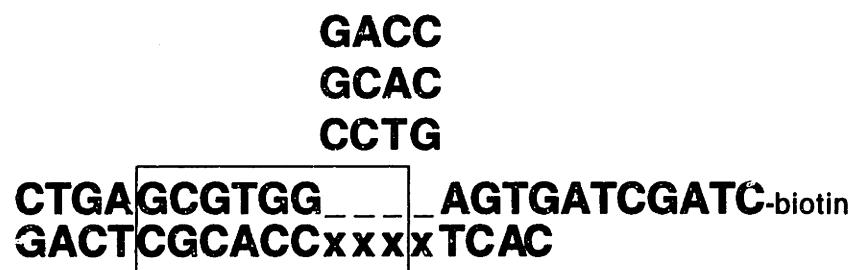
Figure 1B: Selections overview
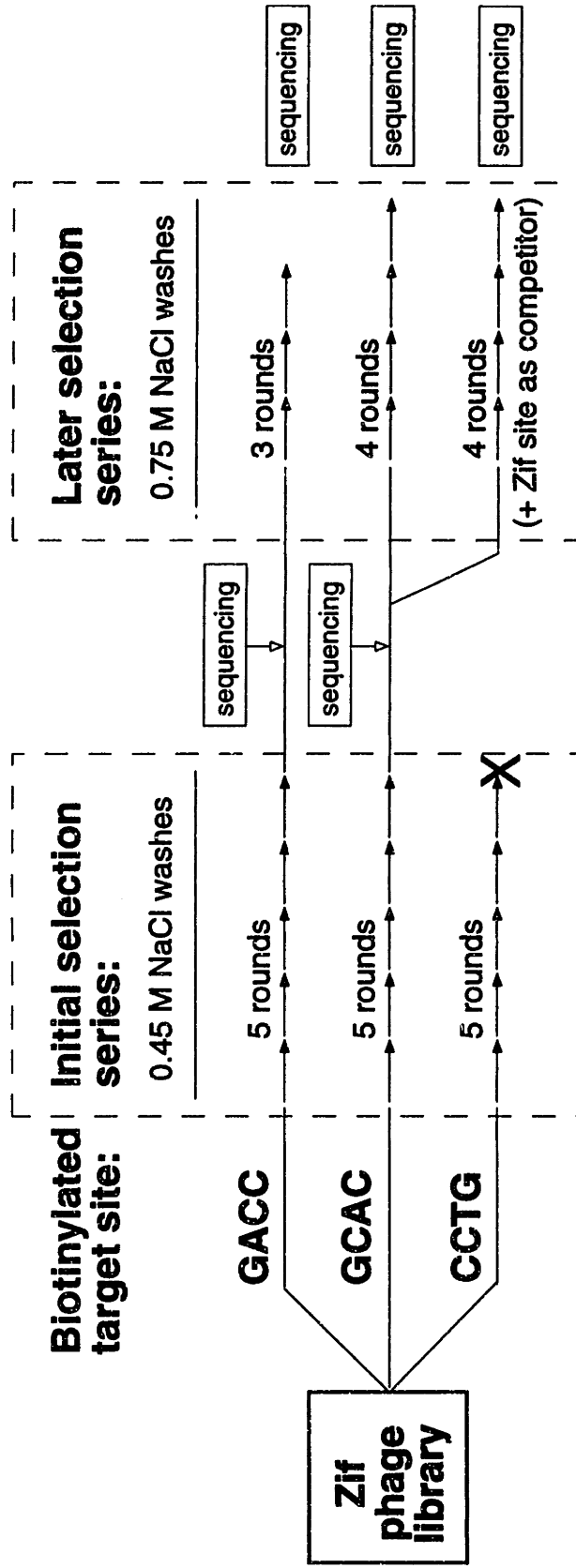
[CHAPTER 3 APPENDIX]

Expanded protocols for, and discussion of, the preparation
and selection of zinc finger phage libraries

<u>Overview</u>:

This appendix provides expanded protocols for preparing zinc finger phage libraries and selecting phage with new DNA-binding specificities. A variety of standard procedures and reagents are also described. In addition, this appendix discusses several key zinc finger phage selection issues, including library design, library construction, choice of affinity purification conditions, and strategies for characterizing selected phage.

<u>Preparing Zinc Finger Phage Libraries</u>:

Zinc finger phage vectors:

Zinc finger phage vectors - like other phage display vectors - can be designed in a variety of ways. For example, such vectors may be phagemids or phage, and they may use different zinc finger/pIII hybrid genes to express zinc fingers on the phage surface[1-5]. These options can be helpful in tailoring a zinc finger phage vector to the requirements of a particular selection experiment. [For an informative review of many fusion phage design considerations, see Smith and Scott[6].] The variety of successful vectors that have been reported suggests that there are no special restrictions on the design of zinc finger phage vectors, relative to the design of phage display vectors for other proteins.

The experiments and protocols described in this appendix use the phage vector fd-tet.Zif (fig. 1), although similar protocols have been used with two other phagemid vectors, p12[2] and p1212 (see chapter 4).

Strategies for preparing randomized vector:

Combinatorial cassette mutagenesis: The first step in making a zinc finger phage library is to prepare vector DNA which contains a suitably randomized zinc finger gene. One efficient way to do this is by combinatorial cassette mutagenesis[7]. In this procedure, the region of interest is removed from the zinc finger phage vector by restriction digests and is replaced with a DNA 'cassette' containing random sequences at key codons of the zinc finger gene. For instance, to produce a library in fd-tet.Zif, a randomized DNA cassette is substituted for the Apa 1/Sph 1 fragment of the fusion gene (fig. 1). The DNA cassette can be prepared using chemical DNA synthesis and standard molecular biology techniques. An excellent guide to combinatorial cassette mutagenesis has recently appeared in this series[7].

Library design: Several fundamental choices must be made before constructing the randomized cassette. One basic issue is how many fingers to vary. Many zinc finger proteins contain several tandem fingers which bind to adjacent DNA subsites (fig. 2A)[8-11], and it might seem reasonable to simultaneously randomize several or all fingers of a multifinger protein. However, reported zinc finger phage selections

have randomized only one finger at a time[1-5]. The key advantage of this approach is that the target site for affinity selection is clearly defined, since binding of the other fingers positions the randomized finger at a specific site on the DNA (fig. 2B). [If all fingers of a protein were randomized at once, it would be more difficult to direct selection to a specific target site.] An additional advantage is that this approach generally yields libraries of lower complexity (since fewer residuec are randomized), and this permits a more thorough sampling of the possible amino aciu combinations at the randomized positions.

Deciding which residues to vary is another important issue, and the choices here obviously depend on the goals of the experiment. Since our goal is to select zinc fingers with new DNA sequence specificities, one reasonable approach is to vary amino acids at positions which can make base contacts. Structural studies of zinc finger- DNA complexes have identified six such positions in the zinc finger motif and most zinc finger phage libraries have randomized at least some of these positions[1-5]. [The vast majority of base contacts originate from four of these positions (see chapter 1 figure 6). These four 'base contact' positions appear to be particularly important for determining the specificity of the zinc finger motif.] It also may be useful to vary residues which can make phosphate contacts[5] or finger-finger contacts[1]. Substitutions at these positions might modify the range of accessible base contacts by changing the precise way that a zinc finger docks against the DNA.

One must also decide how many different amino acids to test at each position. In many phage display libraries the codons of interest

are randomized to give all twenty amino acids. This typically is achieved by using the 32-fold degenerate codons N N g/c or N N g/t[6]. For the base contact positions in zinc finger phage libraries, a reasonable alternative may be the 24-fold degenerate codon g/a/c N g/c, which avoids stop codons and encodes all amino acids except Phe, Tyr, Trp and Cys. These amino acids rarely occur at the base contact positions of zinc fingers[1], and we assume that their absence has little effect on the range of sequence specificities encoded in a zinc finger phage library. At other positions it may be desirable to try fewer amino acid choices. For instance, the last position in the second strand of β-sheet, which is usually an important core residue, could be encoded as T a/t C (Tyr or Phe) which would allow or prevent a DNA phosphate contact from this position (see figure 6, chapter 1). Similarly, the ninth position of the α-helix has been encoded as A g/a G (Lys or Arg) in an attempt to modulate interfinger contacts[1].

Protocols for preparing zinc finger phage libraries:

As with most types of fusion phage, zinc finger phage libraries are prepared by isolating and concentrating phage that have been secreted by growing E. coli. The protocol that we have used to prepare the fd-tet.Zif zinc finger phage library is provided below. In this protocol, E. coli are electroporated with vector that has been suitably randomized by insertion of a library cassette. After a brief growth period cultures are grown to saturation in rich media supplemented with buffer, zinc and suitable antibiotics. Phage are then recovered

from culture supernatants by ultracentrifugation and resuspended in a small volume of buffer. Due to concerns about oxidation of the critical cysteines of the zinc fingers (which coordinate the zinc and stabilize the structure) our protocols are designed to minimize the exposure of zinc finger phage to oxygen. Throughout growth and purification the zinc finger phage are maintained in deoxygenated solutions, and all critical manipulations of phage samples are performed in an anaerobic chamber (Coy, type B; atmosphere: ~97% N_2, ~3% H_2, <1 part per million (ppm) O_2).

The anaerobic conditions used in preparing zinc finger phage libraries (and similar conditions used during the selections) are designed to minimize any selection bias that might be caused by oxidation. These measures may seem excessive, since it is clear that many zinc finger proteins (including Zif268) retain activity when produced and purified under less stringent conditions. However, our concern is that many variants in a zinc finger phage library might not be as robust, so we have used these conditions to protect against any loss of diversity due to oxidation of variant fingers. It is possible that less rigorous safeguards could suffice for this purpose, and other successful selections have used precautions such as excess zinc and/or antioxidants such as dithiothreitol (DTT)[1, 3, 5]. Further experience will be needed to see whether these simpler measures are adequate or whether anaerobic conditions are required for obtaining optimal results from zinc finger phage selections.

fd-tet.Zif library protocol: A library cassette was prepared which corresponded to the small Apa 1/Sph 1 fragment of fd-tet.Zif (fig. 1). This cassette contained the random sequence N N g/c at codons encoding residues '-1', '2', '3', and '6' of the α-helix of finger 1. [The crystal structure of the Zif-DNA complex suggested that these were the most important base contact positions[8].] The library cassette (~3 picomoles) was ligated to the large Apa 1/Sph 1 fragment of fd-tet.Zif (~0.4 picomoles). Ligation products were ethanol precipitated, resuspended in 20 μl of TE, and transformed into MC1061 cells[6] (using 23 electroporations). This yielded ~2.8 x 10^7 transformants.

After a short growth period in SOC (1 hour, 37°, with aeration), transformants were pooled, titered, and divided into four samples. Each sample was pelleted (15 min, 1600 g), resuspended in 50 ml of deoxygenated Zif phage broth, and transferred to the anaerobic chamber. After 15 hours of growth (room temperature, no agitation), cultures were transferred to dialysis tubing (Spectrapor 7, 50 kd cut-off) and the dialysis tubes were suspended in 1 liter of deoxygenated Zif phage broth. Cultures were then grown under anaerobic conditions for 25 hours (room temperature, slow stirring). [We performed this extra growth in dialysis tubing in an attempt to enhance phage titers, reasoning that this arrangement would permit the exchange of nutrients (and thus continued growth of phage cultures) without allowing phage to escape.] This yielded a phage titer of 8.4 x 10^9 tetracycline transducing units (TTU)/ml.

For harvesting, phage cultures were first cleared of cells by centrifugation (12000 g, 4°, 80 min). Culture supernatants (108 ml

total) were then placed in Quick Seal tubes (Beckman) and ultracentrifuged (171000 g, 4°, 6 hours). Phage pellets were resuspended in 1 ml deoxygenated binding buffer and these samples were microfuged briefly to remove insoluble matter. The final library phage prep (0.85 ml; 4.7 x 10^{11} TTU/ml) was stored on ice in the anaerobic chamber. Control experiments suggest that zinc finger phage may be stored this way for short periods, but the DNA binding activity of samples seems to decay over the course of several weeks.

Due to the concerns about oxidation, preparation of the library was performed so as to minimize exposure of phage to oxygen. After the SOC growth period, phage solutions were never exposed to air except during two brief procedures: the transfer of phage cultures to dialysis tubing and the balancing of ultracentrifuge tubes. For all other steps of growth and purification, phage solutions were manipulated either in the anaerobic chamber or in containers which were sealed before removal from the anaerobic chamber.

Selecting Phage with New DNA-Binding Specificities

Affinity purification

The power of phage display stems from the fact that phage which exhibit desired binding properties can be separated from the vast excess of library phage by 'affinity purification'[12]. The desired phage are selected by their ability to bind ligand which is attached to an insoluble support. We affinity purify zinc finger phage using a

variation of the 'biopanning' procedure developed by Smith and Scott[6]. Our affinity purification procedure involves four steps: i) binding, in which library phage are incubated in solution with a biotinylated DNA target site; ii) plating, in which the phage/target site mixes are applied to streptavidin-coated wells of a microtiter plate; iii) washing, in which unbound phage are rinsed from the plates; and iv) elution, in which bound phage are recovered from the wells. Control experiments indicate that our affinity purification can give a 100- to 1000- fold enrichment of high affinity phage.

Since a zinc finger phage library may contain millions of variants, several sequential affinity purifications may be needed to recover rare high affinity phage. After each affinity purification, phage are amplified in E. coli and isolated essentially as described in previous sections. [We use the term 'selection cycle' to refer to one such iteration of phage affinity purification, amplification and isolation.] To monitor the progress of a particular selection we determine the phage retention efficiency in each selection cycle (i.e., the percentage of applied phage recovered in the elutions). In a successful selection, the retention efficiency will rise over the course of several selection cycles to a level of .5% to 5%. We typically stop these selections once the retention efficiencies plateau, and this usually occurs after about five selection cycles (fig. 3). If the retention efficiency does not eventually rise above background levels (such as for the selections using the 'CCTG' site in fig. 3), we assume that no enrichment has occurred and we discontinue the selection.

Selection strategies

Affinity purification conditions: When planning affinity purification conditions, it is useful to consider the goals of the particular study. Does one want to recover only the highest affinity proteins from a library? Or is it better to obtain a larger set of proteins and thus include some that do not bind as tightly? Are there specific target sites which the selected proteins must not bind? Answers to these questions will determine the best selection strategy. For example, if it is important for the selected proteins to discriminate against particular target sequences, then it may be useful to include these specific sites as competitors in the binding mixes. [We have successfully employed this strategy using a nine-fold molar excess of competitor to target site[4].]

If one wishes to identify the very highest affinity proteins in a library, then it may be safest to perform several parallel selections using a range of different stringencies. This is because it is impossible to predict which selection conditions will eliminate all but the best binders (and these conditions may differ for each target site). Moreover, no suitable phage will be recovered if selection conditions are too severe. In our experiments, we have varied the selection stringency by changing the ionic strength of the washes and/or the concentration of nonspecific competitor DNA in the binding reactions. Control experiments suggest that increasing either component generally raises the stringency of affinity purification (however, for reasons which are not entirely clear, our attempts to perform

selections in the complete absence of competitor have been unsuccessful[13]). We have tested a wide range of both components. Successful selections - as judged by the retention efficiencies of selected phage pools - have used washes with total salt concentrations ranging from ~0.05 to ~0.75 M and binding reactions containing 0.06 to 3.0 mg/ml of nonspecific competitor DNA[2, 4].

Strategies for processing multiple samples: Since one may be interested in testing a variety of target sites and a range of affinity purification conditions, an experiment will often involve a substantial number of selections (for example: 4 target sites x 4 affinity purification conditions = 16 independent selections). Given this, it is usually worthwhile to incorporate multiplexing into the experimental strategies. In particular, we use multichannel pipettes and 96-well plates for many steps of our most tedious procedures (affinity purification and titering), and this allows us to perform many more selections in parallel.

Selection protocol

Provided below are protocols which were used to select zinc finger phage from the fd-tet.Zif library.

fd-tet.Zif library selections: The first selection cycle was performed as follows: Binding reactions (121 μl) were made which contained ~3.5 x 10^{10} TTU of phage from the fd.tet-Zif library, 39 nM

of biotinylated target DNA (GACC, GCAC, or CCTG (fig. 3)), and 0.06 mg/ml sheared calf thymus DNA in 0.9X binding buffer. Binding was allowed to proceed for 50 min. Samples were then diluted into 3.6 volumes of 0.05 M NaCl wash buffer and applied to streptavidin-coated wells (6 wells, 30 µl/well) of a Pro-Bind™ plate (see procedure below for preparing these plates). After 50 min the samples were removed and the wells were washed 10 times with 0.45 M NaCl wash buffer (0.25 ml/wash, ~3 min/wash). [Note: Each wash is removed from the wells by shaking the plate upside-down over fresh paper towels.] Forty microliters of elution buffer was then added to each well. After eluting for 2 hours, each set of 6 elutions was removed from the anaerobic chamber, pooled and titered[4].

Eluted phage (0.15 ml) were then mixed with 0.2 ml freshly thawed starved K91 E. coli (prepared as described below). After 10 min, 5 ml of LB broth containing 0.2 µg/ml tetracycline was added. Cells were incubated for 1 hour (37°, with aeration), pelleted (15 min, 1600 g) and resuspended in 50 ml of deoxygenated Zif phage broth. Each culture was then grown anaerobically in a 50 ml centrifuge tube and phage was purified by ultracentrifugation (using essentially the same conditions as those described in the fd-tet.Zif library protocol). The phage prep was stored in the anaerobic chamber on ice[4].

Subsequent selection cycles used similar procedures. [After the second cycle sheared salmon sperm DNA was substituted for sheared calf thymus DNA in the binding reactions, but we believe that this change had little effect on the selection outcome.] By completion of

the fifth selection cycle, high retention efficiencies were observed in the GACC and GCAC phage pools (fig. 3)[4].

Characterizing phage obtained in selections

Depending on the stage of a project and the information needed we have used several different strategies to characterize phage obtained in selections. Typically, our first step involves testing phage pools for sequence-specific DNA binding. This is usually accomplished by determining the retention efficiency of a phage pool using its selection target, and comparing this with retention efficiencies obtained using other, unrelated sites. [It may be convenient to use the target sites from other selections that have been run in parallel, as shown in Table 1] This quick test can confirm that a pool exhibits weak and/or nonspecific binding (such as the CCTG pool of Table 1), and such pools are usually not further characterized. This test can also reveal whether a pool contains phage that bind to other target sites. If this occurs, it may be useful to perform further selection cycles using specific competitor binding sites[4].

If these initial tests indicate that the DNA-binding properties of a phage pool are acceptable, then the next step is to sequence random members of the pool. Suitable template DNA for sequencing may be prepared from phage or phagemids using a variety of methods[14, 15] and this template can be sequenced using convenient commercial kits. Typically a dozen or more phage are sequenced in order to obtain a clear picture of the zinc finger sequences that have been selected.

Once interesting and/or representative sequences are identified, the corresponding zinc finger peptides are expressed and prepared for binding studies. A variety of strategies may be used to accomplish this. For example, the zinc finger region of the fusion gene may be recovered using PCR and then subcloned into other expression vectors (such as those of the Novagen pET series). Overexpressed zinc finger peptides can then be purified from inclusion bodies and refolded[4, 8].

Once purified protein is obtained, binding studies are used to characterize the affinity and specificity of DNA binding. Typically, dissociation constants (K_d's) are determined for binding to the target site and to any other relevant site (such as the target site of the wild type protein). It is important to use a sensitive assay for this purpose, since selected zinc finger proteins may bind with very high affinity to target DNA (Kd <50 pM)[2, 4]. Methods which are suitable for such studies - such as the DNase I footprint[16] and gel shift[17] assays - have been described in previous volumes of this series.

Concluding Remarks

We have used our protocols to select zinc finger variants that bind novel DNA sequences[2, 4]. Our experience indicates that these methods are quite robust, and we have highlighted key variables in order to emphasize the adaptability of this system. Other groups have also performed zinc finger phage selections[1, 3, 5], and their methods further emphasize the range of protocols that may be successful. In particular, these groups have used simpler and less stringent

conditions to combat oxidation, and it appears that such conditions may be useful for performing selections when an anaerobic chamber is not available.

These other groups also have demonstrated interesting and useful variations of zinc finger phage display. Choo and Klug, for instance, randomized finger 2 of Zif268. In an extensive set of experiments, they performed successful selections using 16 different three-base-pair subsites (of the 28 they tested)[1]. In other studies, Jamieson *et al.*[3] and Wu *et al.*[5] performed successful selections using monovalent zinc finger phage. Two of these groups[1, 5] also reported specific DNA-binding by phage carrying the first three fingers of TFIIIA. Taken together, these results suggest that the protocols described in this chapter can be extended to other zinc finger proteins and perhaps even to different DNA-binding motifs.

While all of these phage display strategies have yielded variant zinc finger domains that bind novel DNA sequences, it is important to remember that each strategy (including our own) has failed in selections with some DNA sequences. We do not know whether these failures reflect imperfections in experimental protocols, inappropriate or insufficient randomization, or some inherent limitation of the zinc finger motif. In light of these open questions, we have highlighted variables that might be optimized in planning selection strategies. The other zinc finger phage reports suggest additional important variables, including the solid phase support used for affinity purification, the temperature at which selections are carried out, and the conditions for eluting selected phage from the support[1-3]. Further work will be

needed before we know the limits of selection and design strategies based on the zinc finger motif.

Finally, the construction of so many novel DNA-binding variants from the Zif268 zinc fingers confirms not only the power of phage display but also the robustness and adaptability of the zinc finger. As anticipated from the structural studies of the Zif complex[8], the zinc finger appears to be a very powerful motif for the design of novel DNA-binding proteins.

Solutions and Related Procedures

In this section we describe solutions and procedures which are related to the protocols provided above.

Deoxygenated solutions

After each solution is made, it is degassed by bubbling with helium (0.5 to 2 hours) and placed inside the anaerobic chamber. Components that might lead to excessive foaming during helium bubbling (for instance, the triton x-100 in certain buffers) are not added until degassing is completed.

Zif phage broth: 32 g/liter tryptone, 20 g/liter yeast extract, 5 g/liter NaCl, 8 g/liter glucose, 1 g/liter casamino acids, 20 mg/liter tetracycline, 0.5 mg/liter thiamine, 100 μM $ZnCl_2$, 50 μM DTT, and 40 mM Hepes pH 7.8.

To make one liter of anaerobic Zif phage broth we: i) autoclaved 900 ml of media containing: 35.6 g/l tryptone, 22.2 g/l yeast extract, 95.6 mM NaCl, 111 μM $ZnCl_2$, pH 7.8; ii) added 40% glucose (20 ml), 10% casamino acids (10 ml), 10 mg/ml thiamine (.05 ml), 1M Hepes pH 7.8 (40 ml), and water to 1 liter; iii) bubbled helium through the mix (1-2 hours); iv) added 1M DTT (.05 ml) and 5 mg/ml tetracycline in EtOH (4 ml); and v) equilibrated this final broth for at least 1 day with the atmosphere of a Coy anaerobic chamber: (~97% N_2, ~3% H_2, and <1 ppm O_2).

5x PEG solution: 17.5% PEG 8000 (w/v), 12.5% NaCl (w/v) in water;

binding buffer: 50 mM NaCl, 5 mM $MgCl_2$, 10 μM $ZnCl_2$, 5% glycerol (v/v), 0.1 mg/ml acetylated bovine serum albumin (BSA) (Sigma product # B-8894), and 15 mM Hepes pH 7.8.

elution buffer: identical to binding buffer, except that the concentration of NaCl is 4 M;

0.05 M NaCl wash buffer: 0.05 M NaCl, 5 mM $MgCl_2$, 15 mM Hepes, .01 mM $ZnCl_2$, 5% glycerol (v/v), 0.5% triton X-100 (w/v), pH 7.8;

0.45 M NaCl wash buffer: identical to 0.05 M NaCl wash buffer except that the concentration of NaCl is 0.45 M;

Other solutions:

SOC: 20 g/liter tryptone, 5 g/liter yeast extract, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 10 mM NaCl, 2.5 mM KCl, 20 mM glucose, .2 mg/liter tetracycline;

TE: 10 mM tris, 1 mM EDTA, pH 8.0;

VCSM13 helper phage: VCSM13 phage is obtained from Stratagene (product # 200251) and is amplified essentially as described[18]

starved K91 E. coli: Starved K91 cells are prepared essentially as described in Smith and Scott[6], and these are stored at -80° in a buffer containing: 67 mM NaCl, 42 mM $NH_4H_2PO_4$, and 14% glycerol (v/v).

LB broth: 10 g/liter tryptone, 5 g/liter yeast extract, 10 g/liter NaCl;

Related procedures:

Preparation of streptavidin-coated Pro-Bind^TM wells: Pro-Bind^TM plates are available from Becton Dickinson (product # 3915). To coat the wells of these plates with streptavidin, we: i) place 30-60 μl of fresh 0.1 M $NaHCO_3$, in each well; ii) add 1-2 μl of streptavidin stock [1 mg/ml streptavidin (Pierce product # 21125) in 0.1 M $NaHCO_3$] to each well and incubate the plate in a humidified container (overnight, 4°), iii) replace the streptavidin solution with 0.25 ml of blocking solution (5 mg/ml BSA (Sigma product # A-3912), 1 μg/ml streptavidin, 25 μg/ml sheared nonspecific competitor DNA, 15 mM Hepes pH 7.8) and block the wells for >2 hours; and iv) immediately before use, rinse the wells four times with 0.25 ml of 0.05 M NaCl wash buffer or pZif12 wash buffer. Our procedure is a variation of that described in Smith and Scott[6].

Design and preparation of biotinylated target DNA fragments:

The design of our target DNA fragments includes several basic precautions. First, the binding site is oriented so that the pIII protein

should be projected away from the biotin, to avoid potential steric clashes with the solid support. Second, the length of the DNA fragment is kept short, to reduce the chance that two zinc finger proteins from the same phage could simultaneously bind to the same DNA fragment (even nonspecific binding could present a serious problem). Third, the sequence of the fragment is checked to ensure that there are no other binding sites for the unmodified fingers. [Such sites might direct the selection of the randomized zinc finger for binding to an unwanted target sequence.]

We use an Applied Biosystems DNA synthesizer and standard DNA synthesis chemistry to make the target site oligonucleotides. Biotinylated oligonucleotides were prepared using biotin solid support from Glen Research Inc. For purification, the oligonucleotides were: i) electrophoresed in a 20% denaturing polyacrylamide gel (Sequagel) and eluted into TE; ii) batch extracted using a Sep-Pak C_{18} reversed-phage cartridge (Millipore) and eluted into 70% CH_3CN (v/v); iii) dried and resuspended in TE; and iv) ethanol precipitated and resuspended in 10 mM Tris pH 8.0. Oligonucleotides are quantitated by spectrophotometry[15] and hybridized using a 20% - 50% molar excess of the non-biotinylated oligonucleotide.

Sheared nonspecific competitor DNA: Nonspecific competitor DNA used in our experiments [either salmon sperm DNA (Sigma product # D-1626) or calf thymus DNA (Sigma product # D-1501)] is prepared as follows: Solutions containing dissolved DNA (0.3 mg/ml - 10 mg/ml) are chilled on ice and sonicated. [This typically yields a

fragment size distribution centered on ~1000 base pairs.] If necessary, the DNA is then concentrated to >1.5 mg/ml by isopropanol precipitation. The DNA is extracted with phenol and chloroform, precipitated with either ethanol or isopropanol, dissolved in a dilute Hepes buffer (5 - 10 mM, pH ~7.9), and quantitated by spectrophotometry. [Final stock concentrations are 7 - 15 mg/ml.] Aliquots are stored at -20°.

Titering phage: The following procedure is a rapid titering protocol that is useful for estimating the relative phage concentrations of many different samples. It is straightforward to titer several dilutions of 50-100 different phage samples using this protocol. Our protocol involves the following steps:

(1) Serial dilutions of phage in 2xYT are prepared in 'V'-bottomed 96-well plates (Nunc product # 245128) (2) A fresh culture of XL1-Blue (grown overnight at 37° in 2xYT containing 12.5 μg/ml tetracycline) is dispensed into the wells of another 96-well plate (50 μl/well). (3) Ten microliters of the relevant phage dilutions are added to the side of each well and the plate is tapped gently until phage and cells are mixed. (4) The phage/DNA mixes are incubated in the plate at room temperature for 10 min. (5) After 190 μl of 2xYT is added to each well, the plate is transferred to 37° for 30 min. (6) Twenty five microliters of each mix are then spotted onto LB agar plates containing 1% glucose and the appropriate drugs (5 spots of 5 μl each), and the spots are allowed to dry. (7) Plates are incubated at 37° until

colonies are large enough to count under a dissecting microscope (8-16 hours for XL1 Blue cells).

It is very important that the plates used for spotting be dried for the right length of time. If the plates are too dry then the plate surface is uneven and the spots run into each other. If the plates are not dry enough, then the spots take too long to dry, and this may yield inaccurate colony counts. Typically, we dry the plates (with their lids on) for ~24 hours at 37°.

We use a Rainin M8 multichannel pipettor for all phage manipulations involved in this procedure. The M8 is especially useful in this protocol because its mix function allows one to alternate between small volume pipetting and large volume mixing without tedious reprogramming of the pipettor. This feature makes the serial dilution and phage addition steps much less tedious.

References:

1. Choo, Y. and A. Klug. Toward a code for the interactions of zinc fingers with DNA: Selection of randomized fingers displayed on phage. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 11163-11167.

2. Greisman, H.A. and C.O. Pabo. A General Strategy for Selecting High-Affinity Zinc Finger Proteins for Diverse DNA Target Sites. Science, 1997. **275**: p. 657-661.

3. Jamieson, A.C., S. Kim, and J.A. Wells. *In Vitro* Selection of Zinc Fingers with Altered DNA-Binding Specificity. Biochemistry, 1994. **33**: p. 5689-5695.

4. Rebar, E.J. and C.O. Pabo. Zinc Finger Phage: Affinity Selection of Fingers with New DNA-Bindig Specificities. Science, 1994. **263**: p. 671-673.

5. Wu, H., W. Yang, and C.F. Barbas. Building zinc fingers by selection: Toward a therapeutic application. Proc. Natl. Acad. Sci. U.S.A., 1995. **92**: p. 344-348.

6. Smith, G.P. and J.K. Scott. Libraries of Peptides and Proteins Displayed on Filamentous Phage. Methods Enzymol., 1993. **217**: p. 228-257.

7. Reidharr-Olson, J.F., *et al.* Random Mutagenesis of Protein Sequences Using Oligonucleotide Cassettes. Methods Enzymol., 1991. **208**: p. 564-586.

8. Pavletich, N.P. and C.O. Pabo. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. Science, 1991. **252**: p. 809-817.

9. Pavletich, N.P. and C.O. Pabo. Crystal Structure of a Five-Finger GLI-DNA Complex: New Perspectives on Zinc Fingers. Science, 1993. **261**: p. 1701-1707.

10. Houbaviy, H.B., *et al.* Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. Proc. Natl. Acad. Sci. U.S.A., 1996. **93**: p. 13577-13582.

11.   Fairall, L., *et al.* The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature, 1993. **366**: p. 483-487.

12.   Scott, J.K. and G.P. Smith. Searching for Peptide Ligands with an Epitope Library. Science, 1990. **249**: p. 386-390.

13.   Greisman, H.A. and C.O. Pabo. (unpublished data). .

14.   Ausubel, F.M., *et al.*, eds. *Current Protocols in Molecular Biology.* . 1994, Current Protocols.

15.   Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratory Manual.* 1989, Cold Spring Harbor, NY: Cold Spring Harbor Press.

16.   Hochschild, A. Detecting Cooperative Protein-DNA Interactions and DNA Loop Formation by Footprinting. Methods Enzymol., 1991. **208**: p. 343-361.

17.   Carey, J. Gel Retardation. Methods Enzymol., 1991. **208**: p. 103-117.

18.   Vieira, J. and J. Messing. Production of Single-Stranded Plasmid DNA. Methods Enzymol., 1987. **153**: p. 3-11.

Table 1: Phage pools after selection cycle 4 of the fd-tet.Zif initial selections were tested for binding to each of the sites shown in figure 3. Binding sites are identified by giving the four-base sequence of the underscored region in figure 3. These tests were performed using affinity purification conditions which were essentially the same as those described in the fd-tet.Zif library selections protocol.

TABLE I

DNA-Binding Specificity of Phage Pools
from the fd-tet.Zif Library Selections

Target site used for selection of phage pool	Retention efficiency (%)		
	Target site used for test:		
	GACC	GCAC	CCTG
GACC	0.61	0.03	<0.01
GCAC	0.03	0.91	<0.01
CCTG	0.01	0.02	0.01

Fig. 1.   The zinc finger phage vector fd-tet.Zif with the randomized cassette ('R$_4$') used to make the library.   The thick bar indicates the zinc finger/pIII hybrid gene.   White, numbered boxes encode zinc fingers (Zif268 fingers 1, 2 or 3) and the black bar represents the pIII gene. For simplicity, some features of the hybrid gene have been omitted, including the region encoding the signal peptide (necessary for membrane export) and short linkers between different gene segments. In the cassette, 'R$_4$' encodes a zinc finger with 4 randomized residues. To make the zinc finger phage library, this cassette is inserted using the indicated restriction sites.

The parent construct for fd-tet.Zif is fd-tet (for details of construction, see chapter 2).

Figure 1:

randomized insert

fd-tet.Zif
(phage)
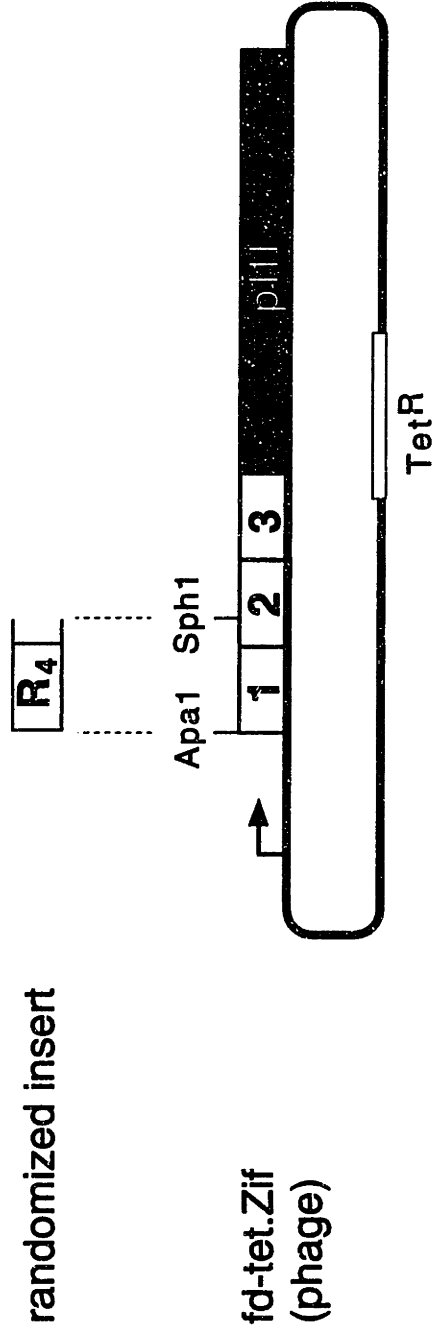
Fig. 2. (A) Sketch of the zinc finger protein Zif268 bound to its target site[8]. Each numbered circle represents the corresponding finger of Zif. Brackets highlight the core DNA subsite of each finger. Only one strand of the DNA is shown (5' is on the left). (B) Sketch of the fd-tet.Zif library (R_4 = finger with 4 random residues) and the target sites used for the selection of new zinc fingers from this library. As shown, the sequences of the underscored region were GACC, GCAC, or CCTG. Our target DNA sites are designed to select fingers for binding to subsites containing four (instead of three) changed base pairs because it appears that zinc fingers can interact with four base pairs of DNA[8-11]. Note that the interaction of fingers 2 and 3 with their subsites ensures that the randomized finger is properly aligned for binding to the new DNA sequence.

Figure 2:

A   Sketch of Zif268
     interacting with its
     binding site

$$3 - 2 - 1$$

GCG TGG GCG
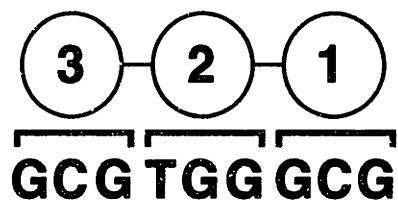
B   fd-tet.Zif library
     and selection target
     sequences

$$3 - 2 - R_4$$

GCG TGG _ _ _ _
           GACC
           GCAC
           CCTG
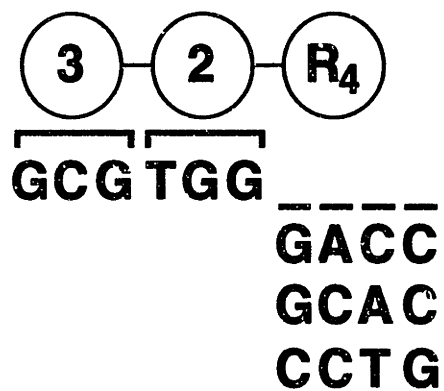
Figure 3: Plot of phage retention efficiency versus selection cycle for the fd-tet.Zif library selections. The inset shows the biotinylated DNA fragments used for the selections ('b' represents biotin). The region of each fragment that corresponds to the Zif268 binding site is in bold letters. As shown, the sequences of the underscored region were GACC, GCAC or CCTG.

Figure 3: Overview of fd-tet.Zif initial selections

GACC
GCAC

5' CTGA**GCGTGG**___AGTGATCGATC-b
3' GACT**CGCACC**xxxTCAC

CCTG

Selection Cycle

% Phage Retained

1.000
0.100
0.010
0.001

1   2   3   4   5
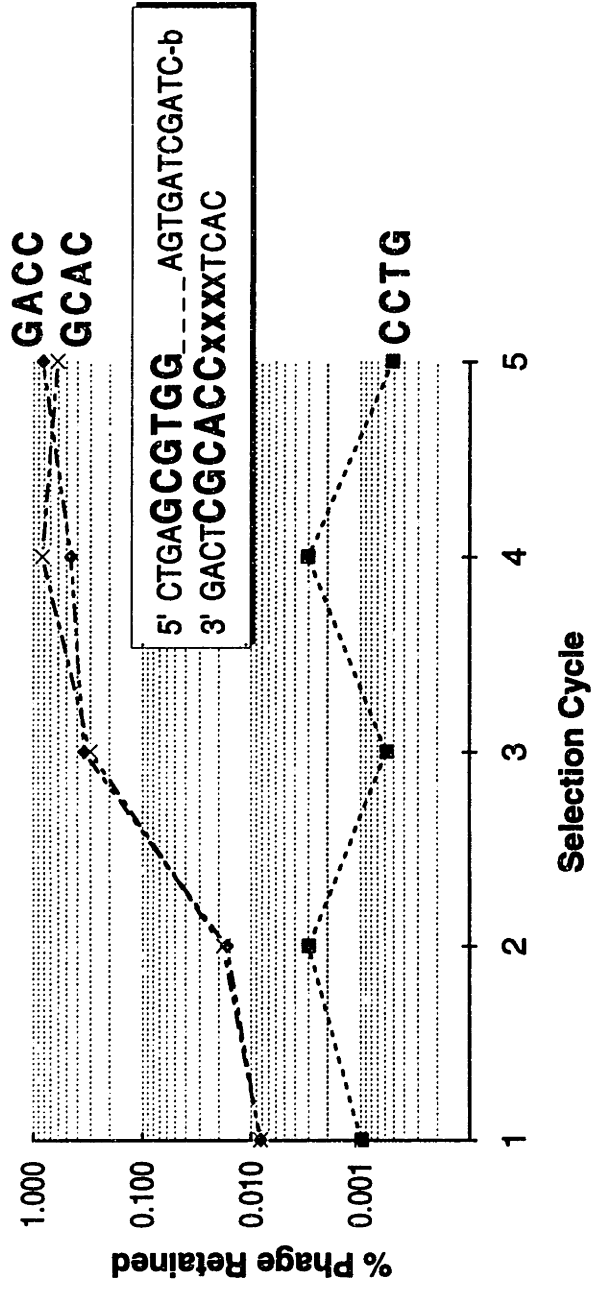
CHAPTER 4

PREFERRED BASE CONTACTS OF THE ZINC FINGER MOTIF

Introduction:

Although there is no global 'code' for protein-DNA recognition, structural and biochemical studies suggest the existence of motif-dependent preferences for particular base contacts from key amino acid positions[1-5]. Discovering these preferences can help us to understand mechanisms of protein-DNA recognition and might also yield a practical strategy for designing DNA-binding proteins with desired specificities. Insights into these preferences can be gained through a variety of methods, including phylogenetic comparisons[6, 7], design efforts[8-10], and structural studies. Perhaps the most direct approach, however, is to use affinity and specificity - based selection methods to isolate variants which can bind to new DNA sequences. By randomizing key base - contacting positions in the motif of interest, and selecting the resulting library for variants that bind to new DNA sites, one may directly determine motif-dependent 'answers' to the problem of recognizing different DNA sequences. From the frequency and types of these answers one may infer the adaptability and base-contact preferences of a given motif.

A particularly attractive motif for such studies is the zinc finger (of the type first discovered in TFIIIA). The zinc finger appears to be one of the most common eukaryotic DNA-binding motifs. Zinc finger proteins have been identified in diverse organisms from yeast to mammals, and many of these organisms contain hundreds of zinc finger genes[11, 12]. Because of its pervasiveness throughout eukaryotic biology, insights into the function of the zinc finger motif should be broadly useful for understanding numerous cellular processes. The zinc

finger is also a very adaptable DNA-binding motif. Binding sites for zinc finger proteins encompass a diversity of different sequences, and range in composition from AT-rich to GC-rich sequences (examples are discussed in Pavletich and Pabo[13]). Since the zinc finger may be adapted to bind so many different DNA sequences, it is a particularly rich subject for studies of DNA-binding specificity. Finally, the zinc finger is proving to be useful as a module for the construction of DNA-binding proteins with designed specificities. For example, zinc fingers have been 'mixed and matched' with other fingers[10, 14] and grafted onto other DNA-binding domains[15, 16] to create hybrid proteins with composite sequence specificities. Additionally, selection methods have been used to construct multifinger proteins which bind to completely novel DNA sequences[17]. These studies require fundamental decisions regarding target site choice and experimental strategy, and such decisions could benefit from insights concerning the adaptability and preferred contacts of the zinc finger motif.

An additional attractive feature of the zinc finger is that it is structurally well - characterized. To date, four x-ray cocrystal structures (Zif268[4, 18], GLI[13], tramtrack[19] and YY1[20]) and one NMR solution structure(the GAGA protein[21]) have been determined for zinc finger proteins complexed with their target sites. [In addition, the structure of a designed zinc finger protein ·· DNA complex has also been reported[22].] In almost all of these structures, tandem fingers of a multifinger protein bind in the major groove and interact with a series of adjacent DNA subsites[4, 13, 18-20]. The periodicity of fingers in these complexes is typically one finger every three base pairs. In one structure, a peptide containing just a single finger binds to DNA[21]. In

this structure, interactions from adjacent, non-finger regions are required for high-affinity DNA binding. Together, these structures contain thirteen fingers which make base contacts and which appear to contribute to DNA sequence specificity (figure 6, chapter 1).

These fingers exhibit several conserved features which provide a framework for planning and interpreting selection experiments. First, the vast majority of base contacts originate at four positions within these fingers, and it appears that these 'base contact positions' play an especially important role in determining the DNA-binding specificity of the zinc finger motif. These positions include the residue immediately preceding the α helix and include the second, third, and sixth residues of this helix (positions '-1', '2', '3' and '6', figure 6, chapter 1). Second, alignment of these fingers reveals that the vast majority of base contacts occur within a four - base pair span of DNA (figure 7, chapter 1). Thus, it is most appropriate to consider the zinc finger motif as having a four - base pair binding site. Finally, each of the four base contact positions in the zinc finger tends to interact with a particular, preferred base position in the target sequence (figure 7, chapter 1). This consistent tendency for interactions between particular pairs of finger and base positions simplifies the interpretation of selection outcomes, since it suggests that one may often appropriately interpret the selection of a particular amino acid at positions '-1', '2', '3' or '6' as being due to the interaction with its preferred base position.

In recent years powerful phage display systems have been developed for selecting zinc fingers with new specificities. In initial studies, the three fingers of the protein Zif268 were expressed on the surface of fd phage, and the base contact positions of finger one were

randomized to encode all possible amino acids. The resulting libraries were then selected for binding to variations of the Zif site containing new sequences opposite finger one, and these selections yielded fingers with desired sequence specificities[5, 23]. In several subsequent studies, each finger of Zif has been randomized and selected for binding to new DNA sequences[17, 24, 25]. These studies have confirmed the importance of the base contact positions for determining finger specificity and the effectiveness of phage display methods for affinity - selecting new fingers[26]. In addition, these studies have demonstrated that phage display methods may be used to make multifinger proteins which bind to completely novel DNA sequences [17]. However, none of these studies has provided a comprehensive assessment of the preferred amino acid - base interactions at the four base contact positions. Moreover, it has not been possible to reliably estimate the DNA-binding adaptability of the zinc finger motif, since none of these studies selected fingers for binding to a suitably large and unbiased panel of four - base pair target sites.

In this report, phage display methods are used to select a library of zinc fingers for binding to twenty different four - base pair target sites. These targets include every possible four - base pair sequence from a key region of the HIV-1 promoter. In sixteen of these selections, pools of fingers are obtained which exhibit specificity and affinity for the selection targets, and this high success rate suggests that the zinc finger motif may be adapted to recognize most four - base pair DNA sequences. Sequencing of these selected phage pools reveals distinct consensus sequences which, when combined with structural data and the results of previous selections, suggest base contact

preferences for the zinc finger motif. Several of these preferences are quite strong, however, it is clear that these preferences exhibit too much dependence on context to constitute any general zinc finger - DNA recognition 'code'. Finally, an analysis of selection results suggests two key context effects which may be important for understanding zinc finger - DNA recognition.

Results:

Choice of selection targets:

Our selection targets included every possible four - base pair sequence in a region of the HIV-1 promoter surrounding the TATA box (figures 1 and 3). This region contains 20 distinct four - base pair sequences. Two key features of this region make it especially attractive as a source of selection targets. First, this region is functionally important for HIV-1 transcription[27] and is highly conserved [28]. Therefore, fingers or multifinger proteins which could recognize sequences in this region might find practical application as components of HIV therapeutic strategies or research tools. Second, this region is naturally defined and features a relatively even base composition. Thus the target sequences from this region would appear to provide a relatively unbiased test of zinc finger adaptability.

Choice of 'host' protein for these selections:

To date, all zinc finger selections have used a single randomized finger in a multifinger 'host' protein[5, 17, 23-25]. This arrangement is desirable because the extra fingers of the host protein help to fix the randomized finger over the target sequence and also provide energy required for high affinity binding and efficient affinity selection of phage[26]. In choosing a host protein for our selections we were guided by several considerations. First, we wanted to ensure that our selections would yield fingers which bound to the desired target sequences, and not to 'frameshifted' variations. Although multifinger proteins typically exhibit a periodicity of one finger every three base pairs, the GLI structure makes it clear that four - base pair spacings are also possible[13]. Selections which do not exclude this arrangement run the risk of yielding fingers which bind to an undesired sequence shifted by one base pair. Such outcomes might be difficult to detect, and would complicate any analysis of zinc finger adaptability and preferred base contacts. In addition, this problem could make it more difficult to use these fingers as subunits for the construction of multifinger proteins with desired specificities. To help to avoid this possibility, we wanted to flank our randomized finger on both sides with sequence specific, base-contacting zinc fingers. Thus we desired to perform our selections using the 'interior' finger of a multifinger protein. However, we also wanted to be able to change all four base pairs in the subsite of our randomized finger without disrupting critical contacts of neighboring fingers with the DNA. This can be difficult to achieve in an interior finger of a multifinger protein, since the subsites of adjacent fingers typically overlap by one base pair.

Therefore, we wanted to perform our selections using an interior finger whose subsite lacked critical contacts with neighboring fingers.

Given these considerations, we chose to perform our selections using the 'interior' finger 1 of the protein 1212 (figures 2 and 3). 1212 is a designed four finger protein which consists of Zif fingers one and two in the indicated arrangement. Our understanding of Zif268 - DNA binding energetics suggests that this choice satisfies the concerns discussed above. Using an interior finger helps to ensure selection for binding to the correct DNA sequence, since frameshifting forces adjacent fingers to bind to suboptimal sequences. Also, the adjacent fingers should make only relatively weak contacts with the subsite of our chosen finger[4, 18, 29], and so it should be possible to select this finger for binding to most four - base pair sequences.

Since 1212 is a designed zinc finger protein, it was important to confirm key features of its binding which were critical for our selections. In particular, we wanted to ensure that the carboxy - terminal finger 2 (the 'edge' finger 2 - figure 2) could contribute to the specificity and affinity of 1212, since we were counting on this finger to prevent the potential frameshifting problems described above. We tested this two ways. First, a gel shift assay was used to measure the affinity of 1212 for its target sequence and for a disrupted variation containing mutations in the subsite for the carboxy - terminal finger 2 (figure 4). The affinity of 1212 for the disrupted site was about 35-fold lower than for the correct sequence. [In contrast, Zif interacts with a site adjacent to the disrupted region, and its binding was hardly affected by these changes (figure 4).] Next, we expressed 1212 on the surface of phagemid particles as a fusion with the pIII protein, and

then performed a test plating of the resulting phagemids using the 1212 target site and also a series of variant sequences containing mutations in the subsites for each finger. All tested mutations abolished binding (figure 5). These results confirmed the functional importance of the carboxy - terminal finger 2 and also indicated that all fingers of 1212 participate in DNA recognition.

Library construction:

We wanted our randomized fingers to encompass as much of the potential DNA-binding repertoire of the zinc finger motif as possible. Therefore, we wished to suitably vary any parameter which could influence zinc finger specificity, including base contacts, interfinger contacts, phosphate contacts and the spacing of the zinc-coordinating histidines. However, we were also worried that the variation of too many important positions could functionally inactivate a large portion of our library. Because of these concerns, we constructed two libraries for these selections (figure 6). In our first library, 'library 1', we randomized six positions: the four base contact positions ('-1', '2', '3', and '6') and two other positions which occasionally make base contacts ('1' and '5') (figure 6). To accomplish this, we synthesized the codon for each of these positions as the combinatorial mixture AGC / AGCT / GC, which encodes all amino acids except cysteine, tryptophan, tyrosine, or phenylalanine. We chose this randomization scheme because it eliminates stop codons, and because the excluded amino acids are rarely observed in this region of the zinc finger motif[24]. In our second library, 'library 2', we introduced limited variation at six additional

positions of the zinc finger motif (in addition to the combinatorial mutagenesis described above) (figure 6). In four cases we sought to modulate phosphate contacts by encoding two alternative residues at each varied position. Thus we encoded position '-5' as lysine or arginine, '-3' as tyrosine or phenylalanine, '4' as leucine or arginine, and '10' as isoleucine or threonine. In one case, we used a similar strategy to modulate interfinger contacts, encoding position '9' as serine or arginine. In the last case, we varied the histidine - histidine spacing to be either three or four residues by insertion of an asparagine residue between positions '8' and '9'. In designing library 2, we hoped that by varying key phosphate contacts, interfinger contacts, and the histidine - histidine spacing, we could expand the number of zinc finger - DNA docking arrangements present in our library and thus the possibility of finding new base contacts.

Our two libraries were constructed as described in the materials and methods section. Titering of transformation mixes suggested that each library contained approximately 5×10^9 clones.

Selections:

The two libraries were pooled for experimental convenience and the combined library was selected for binding to the 20 sites shown in figure 3. Over the course of five selection cycles the retention efficiencies rose to $\geq 0.1\%$ for 18 of these sites (data not shown). These pools were chosen for further characterization.

Specificity of selected phage pools:

To probe the specificity of the selected phage, we grouped the 18 high affinity pools into three sets of six, and plated every combination of phage and target site within each set (tables 1A-1C). We also included 1212 phage as positive controls. In defining these sets, we grouped phage pools having similar target sites. For example, one set contained five pools whose targets differed by a series of single base-pair changes (TAAG --> TATG --> TATA --> CATA --> CTTA) (table 1B), and a second set contained a similar series of four phage pools (GCAG --> GCTG --> GCTT --> GCAT) (table 1C). This arrangement enabled us to test each phage pool using a range of sequences which differed from the selection target site at 0, 1, 2, 3 or 4 base pairs.

The results of these platings, listed in tables 1A - 1C, show that each pool binds to DNA in a sequence - specific fashion. In particular, each pool i) yields a relatively high plating efficiency (0.1 % - 1.0%) when tested using its selection target site (tables 1A - 1C, gray boxes); ii) yields a much lower background efficiency (0.0004 % - 0.004 %) when plated without DNA (tables 1A - 1C, right columns); and iii) exhibits clear sequence preferences, since some sites yield much higher plating efficiencies than others. Together, these features suggest that the phage pools contain fingers with new, energetically significant base contacts.

In addition to these features, most pools exhibit patterns of site discrimination which further suggest sequence - specific DNA recognition. This is highlighted in tables 2A - 2C, which show the observed DNA-binding 'selectivities' of each phage pool (the ratios of plating efficiencies obtained using target and nontarget sites). Most

pools exhibit relatively modest levels of selectivity (1x to 23x) against sites with single base changes. These same pools are much more selective (usually > 80x, and as high as 700x) when tested using more distinct sequences. Thus for many pools a clear trend is apparent, with stronger discrimination against sites which contain more base changes relative to the selection target site. This trend, which is summarized in table 3, provides further evidence that most pools contain fingers which make new base contacts.

Although most phage pools behave as described above, some exhibit features which are distinct enough to warrant special consideration. These pools are highlighted in two groups at the bottom of table 3. The first group contains phage pools selected for binding to sites of the form GCxx (GCAG, GCTG, GCAT, GCTT). These pools discriminate poorly among sites within this group, and so may appear relatively nonspecific. However, each of these pools discriminates strongly against two other test sites (ATGC and AAGC, table 2C). This behavior suggests that these pools may make their most important DNA contacts to the guanine and/or cytosine bases in the GCxx sequences, and that the apparent lack of specificity may be an accident of the phage grouping arrangement used for this test. This possibility is supported by sequencing results. As described in the next section, the consensus sequence of each pool features a basic residue at position '6' and an aspartate at position '3' which are predicted to interact with the guanine and cytosine of the GCxx sites.

The second questionable set includes phage pools selected for binding to the AAGC and CTGC sites. Unlike all other selected phage, these pools exhibit poor selectivities against several highly variant

sequences. For example, the 'AAGC' phage pool discriminates against the sequences GCAT and GCGT by factors of 6 and 7 (table 2C), and the 'CTGC' phage pool discriminates against the sequence TATA by a factor of 5 (table 2B). Additionally, these pools exhibit the lowest average selectivities against sites with 3 or 4 base changes (table 3). Since this behavior raises serious questions about the specificity of these phage we have chosen to exclude these two pools from further consideration in this chapter.

Sequences of selected phage pools:

To further characterize the sixteen well-behaved phage pools, we sequenced randomly picked clones and derived consensus sequences (figures 7A - 7C). [Note that in these figures the target sequences are listed in reverse (3' --> 5'), in order to align each target sequence with consensus features it selects.] Several features of the sequences are worth noting. First, at the six randomized positions, we usually observe from 2 to 4 consensus features in each phage pool, for a total of 45 such features. The vast majority of these features (40) are observed at the base-contact positions ('-1', '2', '3', and '6'). Second, most of these consensus sequences are uniquely associated with their particular binding site, and this is consistent with the variety of DNA-binding specificities observed in our plating experiments. The only consensus sequences which are not unique are those for the GCxx phage pools, and as noted previously, these pools all behaved similarly in our specificity test.

Third, each consensus is derived from a variety of different clones, and we never observe repeated sequences. Apparently, our library is sufficiently complex to encode a diverse set of solutions to the challenge of recognizing each sequence. Because of these results, it is likely that each consensus assignment reflects a true preference for the indicated residue or residue group. In contrast, when selection experiments yield only one or a few successful clones, it can be difficult to distinguish between residues which are required for function and those which are present due to chance association with an important consensus feature.

Finally, in almost all cases we observe no strong preferences for any of the alternatives encoded by library 2. [At each varied position, half or less of each set of sequenced clones contain the alternative residue or His-His spacing (data not shown).] Evidently, most of these alternatives are not critical for efficient binding under our selection conditions. Although we cannot conclude that these factors are unimportant for zinc finger - DNA recognition (for example, they might play a role if other target sequences were used or if the selections were performed using conditions of higher stringency), we have elected to exclude most of these results in order to simplify figure 7, with the following exceptions: i) position 4 results are included because this position lies at the center of our randomized codons; and ii) three pools (CTTA, CATA, AGCA) include results from positions between the histidines (positions 9 and 10), because there appears to be a preference for a histidine-histidine spacing of four residues in these pools (figure 7C).

Discussion

How adaptable is the zinc finger motif?

Given the current interest in using fingers for DNA-binding protein design, it would be useful to know whether there are any intrinsic limits to the DNA-binding adaptability of the zinc finger motif. Can fingers be adapted to recognize any possible four - base pair sequence? Or are certain sites refractory to efficient recognition by zinc fingers? If so, then what fraction of sites can be recognized by a suitably modified zinc finger? Answers to these questions obviously depend in part on one's requirements for recognition. Demanding higher levels of affinity or specificity makes it more difficult to find suitable fingers for any given sequence[5], and also increases the fraction of sites for which no suitable fingers can be discovered. Despite these complications, however, insights into these questions can help us to understand the function the zinc finger motif and might also provide useful information for DNA-binding protein design efforts.

The most direct way to gauge the adaptability of the zinc finger motif is to select a suitably randomized finger for binding to a large, unbiased panel of four - base pair sequences. [Obviously, the most thorough approach would select fingers for binding to *every* possible target site, but the large number of such sites ($4^4 = 256$) makes this a more difficult undertaking.] If this experiment is performed properly, then the proportion of successful selections provides a basis for estimating the fraction of all sequences for which one may find specifically-binding zinc fingers. The usefulness of this approach

depends critically on the choice of selection targets. In particular, any compositional bias would tend to reduce our confidence in extrapolating from the selection results to the adaptability of the zinc finger motif. In previous studies, zinc fingers have been selected for binding to a variety of different sequences [5, 17, 23-25]. However, the sets of four - base pair target sequences used in each of these studies either have been relatively limited in size or have featured disproportionate base ratios, and so none of these studies has provided a suitable basis for estimating zinc finger adaptability.

Our results allow us to estimate the DNA-binding adaptability of the zinc finger motif. In this study, we have selected a library of fingers for binding to 20 different four - base pair sequences. These sites would appear to qualify as a reasonable test set for gauging zinc finger adaptability, since they represent every possible four - base pair site from a natural DNA sequence (the HIV-1 promoter), and possess a relatively even base composition. Our results (16 successes in 20 attempts) suggest that the zinc finger may be adapted to recognize most four - base pair DNA sequences under the conditions of our selections. [If, for example, the true proportion of DNA sequences which may be recognized by the zinc finger motif is 50%, then the odds that we would randomly chose 16 or more winners in a set of 20 sequences is <.007.] These results illustrate the remarkable adaptability of the zinc finger motif.

Importance of positions '-1', '2', '3' and '6' in DNA recognition:

These results also reemphasize the importance of the base contact positions for sequence recognition by the zinc finger motif. Of the 45 different consensus features recovered at our six randomized codons, 40 are located at base contact positions (positions '-1', '2', '3' and '6'). In addition, the consensus features recovered at the remaining positions ('1' and '5') exhibit relatively limited diversity (three arginines, one histidine, and the set [arginine + lysine]) which further suggests inherent limitations to the possibilities of base recognition by these positions. Thus these experiments provide further evidence that one may appropriately focus on the identity and function of the four key base contact positions when designing new fingers and when studying mechanisms of finger - DNA recognition.

Preferred contacts of the zinc finger motif

The consensus sequences in figures 7A to 7C also reveal correlations between the features selected at a particular base contact position and the identity of its preferred base position. For example, when the second base from the 5' end of the target site is adenine, three of four successful selections yield an asparagine at position '3' (the 3' GTAT, 3' GAAT and 3' ATAT selections, figure 7A), while successful selections involving thymine often yield a nonpolar residue (the 3' CGTA, 3' AATA and 3' TATA selections, figure 7A) and selections involving a cytosine yield exclusively aspartate (the 3' GACG, 3' GTCG, 3' TACG and 3' TTCG selections, figure 7B). Interactions between base contact position '-1' and its preferred base also feature trends. Two of three target sites which contain a cytosine at the appropriate base

select a glutamate at position '-1' (3' TC̲GT and 3' AC̲GT, figure 7A), and three of six selections involving an adenine at this base yield glutamine (3' AA̲TA, 3' TA̲TT and 3' GA̲AT, figure 7A). These trends and correlations, combined with structural studies demonstrating the tendency of each base-contact position to interact with a single preferred base position (figure 7, chapter 1), suggest that the zinc finger often uses similar mechanisms to recognize the same base pair in the context of different sites. Therefore, learning these trends should provide a fruitful approach for understanding the adaptability and function of the zinc finger motif.

Our selected sequences, combined with the results of other recent selection experiments and structural studies, provide a rich database of observed and inferred finger - DNA contacts which may be examined for such trends. To date, 69 selections have been performed for binding to 64 different four - base pair sequences, and 47 of these selections have yielded pools of fingers which exhibit new consensus sequences and which contain fingers with new specificities (this study, and previous studies[5, 14, 17, 23]). Table 4 shows the consensus sequences and target sites of these selected fingers, and also highlights unsuccessful selections (in gray). In addition, structural studies have observed thirteen natural fingers making base contacts with eleven different subsites [4, 13, 18-21]. Table 5 lists these fingers with their subsites, and highlights residues which contact their preferred base position (in gray). The consensus sequences obtained in selections and the base contacts observed in structural studies exhibit similar correlations between the residues at base contacting positions

and the identities of contacted bases. These correlations are combined and summarized in figures 8A - 8D.

These correlations suggest clear, position-dependent base-contact preferences. Furthermore, the large size of the database suggests that these preferences may transcend particular experimental or structural contexts and may apply to all Zif-like zinc fingers. For example, position '6' of the zinc finger motif exhibits a dramatic preference for an arginine - guanine contact (figure 8D). This contact has been directly observed in five zinc finger - DNA complexes [4, 18, 19, 21], and has been selected 16 different times, in the context of 15 different sites (table 4, top row), in five independent selection studies (this work and previous studies[5, 17, 23, 24]). In contrast, position '6' appears to have problems recognizing cytosine. Eleven of thirteen selections involving possible contacts between position '6' and cytosine either failed to yield tight binding fingers or yielded no consensus feature at position '6' (table 4, bottom row). These failed selections occurred in four different studies (this work and previous studies[5, 23, 24]).

The other base contact positions also exhibit clear preferences. For example, position '-1' often uses arginine-guanine, glutamine-adenine and aspartate (or glutamate)-cytosine interactions (figure 8A), and position '3' often uses asparagine to interact with adenine, and either aspartate or threonine to interact with cytosine (figure 8C). To a lesser extent, position '3' also appears to recognize guanine using histidine, and thymine using nonpolar residues (figure 8C). Finally, position '2' exhibits weaker preferences. However, aspartate - cytosine interactions are observed, and there also may be a preference

for the interaction of the aliphatic portion of arginine with thymine (e.g., as proposed in a recent structural study[21]) (figure 8B). To date, however, too few studies have involved this possibility to say for certain whether this is a generally preferred interaction of the zinc finger motif.

At a broad level, most of the observed preferences are explainable in terms of potential residue-base contacts and the general docking arrangement of fingers against DNA. For example, at position '-1', the preference for guanine-arginine and adenine-glutamine interactions can be rationalized in terms of the especially favorable hydrogen bonding geometries of these arrangements[30] (see chapter 1, figure 3A). Likewise, at position '3', preferences for histidine-guanine and asparagine-adenine interactions can be similarly explained. Indeed, each of these interactions have been observed directly in structural studies[4, 18-21]. Furthermore, the differences between positions '-1' and '3' can also be explained by the fact that the α–carbon of position '-1' lies further away from the bases than the α-carbon of position '3'. [Arginine and glutamine possess longer side chains than histidine and asparagine, respectively.] Corresponding arguments can be made about other conserved residues.

Implications for design

The preferences highlighted here appear to comprise key determinants of the adaptability of the zinc finger motif, and as such are useful to consider when designing fingers with new specificities. For example, Berg and coworkers have used rules similar to the

observed preferences to make fingers with a variety of specificities[9, 10]. However, it would appear that these preferences do not constitute a 'recognition code' in any strict sense, because factors outside of a particular residue - base interaction can clearly influence its usefulness for DNA recognition. This has been previously suggested by mutational studies which show that the substitution of amino acids at a single base contact position can change DNA-binding specificity at multiple base pairs[31, 32] and that changing multiple base contact positions simultaneously can often yield fingers which bind DNA poorly and/or exhibit unexpected specificities[33]. This possibility is further underscored by the relatively large differences that we observe in the consensus sequences selected for several pairs of similar target sites. For example, our selections for the target site 3' $\underline{T}$CGT yielded E/$\underline{Q}$/H/T (positions '-1'/'$\underline{2}$'/'3'/'6'), while selections for 3' $\underline{A}$CGT yielded E/$\underline{R}$/x/x (where x indicates no preference) (figure 7A). The changed base in these sites (underlined) should interact with position '2' (also underlined), and so we would expect a modification in the consensus feature at this position. However, we also see differences at positions '3' and '6'. A second example involves our selections using 3' TAT$\underline{T}$, which yielded Q/G/E/$\underline{x}$, and 3' TAT$\underline{A}$, which yielded x/x/G/$\underline{Q}$ (symbols as above). Again, we see changes at several base-contact positions, not just the one expected. These complexities suggest the existence of additional factors which modulate the base contact preferences that we have highlighted above. In order to fully understand the function and adaptability of the zinc finger motif, it will be necessary to identify and characterize these factors.

A careful inspection of tables 4 and 5 suggests what some of these factors may be. For example, data in these tables suggests the conditions which are required for the effective recognition of guanine by arginine at position '-1'. In 8 of 8 selections involving sites of the form 3' GGxx (x = any base, and the target base for position '-1' is underlined), an arginine at position '-1' was co-selected with an aspartate at position '2' (table 4). In contrast, none of the selections involving sites of the form 3' CGxx yielded arginine at position '-1'. [Of these nine selections, five failed, two yielded no consensus feature, and two yielded alternate residues.] These results suggest that arginine at position '-1' effectively recognizes guanine only when suitably buttressed by an aspartate at position '2' (as observed in previous structural studies[4, 18, 19]), and that aspartate is effective in this role only when it productively interacts with its target base. Thus it appears that an arginine at position '-1', combined with an aspartate at position '2', effectively functions as a single, indivisible recognition unit which can be used for binding sequences of the form 3' GGxx (and possibly 3' TGxx, see the structural studies cited above). There is little evidence that either of these amino acids work by themselves. An additional example involves the recognition of guanine by histidine at position '3'. Histidine has been obtained in 3 of 4 selections involving sites of the form 3' xxGT, but only in 1 of 9 selections using sites of the form 3' xxGA, 3' xxGC, or 3' xxGG (x = any base, and the target base for position '3' is underlined). Swirnoff et. al. have pointed out that the histidine at position '3' in Zif finger 2, in addition to its expected guanine contact, also derives extra binding energy from stacking interactions with the adjacent thymine[29]. The

trends that we observe are consistent with this proposal, and further suggest that this interaction may be required for histidine to function effectively in guanine recognition at position '3'.

Conclusions:

Our results reemphasize the extraordinary adaptability of the zinc finger motif and the power of phage display to find fingers with desired specificities. They also suggest that the zinc finger motif may be adapted to exhibit specificity and affinity for most four - base pair sequences under our selection conditions. Our results, coupled with the results of previous studies, demonstrate clear, position - dependent base contact preferences for the zinc finger. However, the behavior of our selected consensus sequences also underscores the difficulty (suggested by previous studies [31-33]) of using these preferences in any strictly 'code' - like fashion to make fingers with new specificities. Context effects can clearly modulate and/or negate the usefulness of any particular contact for base recognition, and we identify and discuss two such effects here. Understanding the preferred base contacts of the zinc finger motif, and the contexts under which they are effective for base recognition, will be required if we are ever to achieve a thorough understanding of the function and adaptability of the zinc finger motif.

Materials and Methods:

Construction of the 1212 expression vector pET21d-1212:

The region of Zif268 cDNA[34] encoding fingers 1 and 2 was amplified using PCR and two different sets of primers: i) 5' GATGCTAG CG**CCATGG**AACGCTATGCTTGCCCTGT + 5' CCGGGATC<u>CACCGGT</u> <u>G</u>TGGGTGCG GATGTGGGT; and ii) 5' GCCA rGGA<u>CGCCGGTG</u>AAAAGCCATATG CTTGCCCTGTCGAGT + 5' CCG**GGATCC**TCATTCTCCTGTATGGGTGCGGATGTG GGT. PCR products were digested with the restriction endonuclease SgrA1 (sites underlined) and ligated together. The resulting DNA fragment was then cut with Nco1 and BamH1 (sites in bold) and ligated into the corresponding restriction sites of the expression vector pET 21d (Novagen), yielding the plasmid pET21d-1212. The 1212 region of this plasmid was sequenced using standard methods[35, 36].

Expression, purification, and gel shift studies of 1212 and Zif268:

pET21d-1212 was transformed into BL21 cells and induced as recommended (Novagen). Zif 268 was induced using the expression constructs, cells, and conditions previously described[4]. Both proteins were purified, reconstituted, and stored essentially as described[5]. Peptide identities were confirmed by mass spectroscopy (data not shown). Apparent dissociation constants were derived using gel shift studies essentially as previously described[5]. Each Kd was measured twice, using two different peptide preparations, and all duplicate

measurements agreed to within a factor of 1.8. Reported values are the averages of both measurements.

Construction of the 1212 phagemid vector p1212:

The phagemid vector p1212 was constructed in two steps (figure 9):

First, the region of fd-tet.Zif[5] encoding the entire pIII-Zif fusion gene was amplified using PCR, digested with the enzymes Kpn 1 and Sac 1, and ligated with the large Kpn 1 / Sac 1 fragment of pBluescript KS+ (Stratagene). Primers used for PCR were: 5' GCTAGAAT TCTGCA<u>GGTACC</u>TGATTTTTGGAGATTTTCAACGTG and 5' CGCATTCTAGA<u>GAGC TC</u>AATAA TAACGGAATACCCAAAAGAACTGG. The Kpn 1 and Sac 1 sites are underlined.

Next, a synthetic DNA fragment encoding the 1212 gene was assembled using oligonucleotides and standard molecular biology techniques[35, 36](data not shown). The sequence of this DNA fragment was: 5' GACCGATTC GAAAGTCGACCA <u>GGGCCC</u>CATGAACGCCCATATGCTTGCCCGGTCGAATCCTGCGATCGCCGCTTCT CGCGAAGCGATGAGCTTACCCGCCATATCCGCATTCATACAGGTCAAAAACCGT TTCAGTGCCGTATCTGCATGAGAAATTTTAGCAGATCTGACCACCTTACCACCC ACATCCGGACTCATACAGGTGAAAAACCTTACGCGTGT<u>CCAGTGGAGAGCTGCG</u> <u>ACCGCCGCTTCTCGCGCAGCGATGAGCTTACCCGCCATATCCGCATCCACACTGG</u> <u>C</u>CAGAAGCCGTTTCAGTGTCGTATATGCATGCGCAACTTCTCGAGAAGCGATCA CCTTACCACCCACATTCGCACTCATACAGGTGAAAAAGACAAGAAAGCTAGCA AAAGCGTGGTGGCCTCTAGCGCAC<u>CTGCAG</u>AGTTCGAATAGACTAGTG. This fragment was digested with Apa 1 and Pst 1 (sites underscored with

double lines) and used to replace the corresponding Apa 1 - Pst 1 fragment of pZif - pIII fusion gene, yielding p1212. The 1212 region of this construct was sequenced using standard methods[35, 36].

Construction of the Library vector p1212-BstX:

The parent vector for library construction - p1212-BstX 1- was assembled in parallel with the p1212 vector using identical procedures, except that the underscored region of the 1212 gene was replaced with the following sequence: CCAGTGGAGTGGTGCGACGCACGCT TGTCGGCAAGCGCAGAGCTTACCGCACATATCCGCCACCACACTGGC. This new sequence encodes two BstX restriction sites (underlined) and also contains other changes.

Expression, purification, and plating studies of p1212 phage:

p1212 phagemid particles were produced using protocols similar to those described for the library production (below), except cultures were inoculated using single colonies of cells containing the p1212 phagemid construct (instead of transformation mixes), and final phage pellets were resuspended and stored in the following buffer: 50 mM NaCl, 5 mM MgCl2, 10 μM ZnCl2, 5% glycerol (v/v), 0.1 mg/ml acetylated bovine serum albumin (BSA) (Sigma product # B-8894), and 15 mM Hepes pH 7.8.).

The plating experiment shown in figure 5 was performed as follows:

Binding reactions were prepared which contained 1212 phagemid ($\sim 5 \times 10^9$/ml (= 8.3 pM)) and one of the biotinylated target DNA fragments shown (37 nM), in 48 μl of the storage buffer described above. A control binding reaction was also prepared which contained no DNA site. Reactions were equilibrated 45' and diluted into 192 μl of 0.05 M NaCl wash buffer (storage buffer without BSA). 50 μl of each diluted reaction was then applied to a streptavidin - coated well of a microtiter plate[26]. Reactions were equilibrated in the wells for 40', and the wells were rinsed with 6 washes of 125 μl 0.05 M NaCl wash buffer. The total wash time was 40'. Finally, 240 μl of elution buffer (storage buffer containing 4 M NaCl) was added to each well and incubated for 2 hr. Eluted phage were quantitated as infectious units essentially as described[5].

Preparation of randomized library cassettes:

As described in the text, we prepared randomized cassettes for two different phage libraries. For library 1, we constructed cassette '#1':

```
5' GAGTTGCGACAGACGCTTCTCCVNSVNSVNSVNSCTCVNSVNSCATATCCGTATCCACA
  ACCTCTCAACGCTGTCTGCGAAGAGG 5'                    GTATAGGCATAG 5'
```

mixtures: N: A, T, G, C;   V: A, G, C;   S: G, C;

This cassette contains six codons which are randomized to encode all amino acids except cysteine, tryptophan, tyrosine and phenylalanine. To construct this cassette, the three required oligos were synthesized using

standard chemistry and equipment (Applied Biosystems). The N̲, V̲, and S̲ positions were synthesized using premixed phosphoramidites (Glenn research). The oligos were gel purified, phosphorylated, and annealed using standard protocols[35, 36].

For library 2, we constructed the cassettes '#2A' (top) and '#2B' (bottom):

```
5' GAGTTGCGACARACGCTWCTCCVNSVNSVNSVNSCKCVNSVNSCATATCMGTAYCCACA
   ACCTCTCAACGCTGTYTGCGAWGAGG 5'                    GTATAGKCATRG 5'
```

```
5' GAGTTGCGACARACGCTWCTCCVNSVNSVNSVNSCKCVNSVNSCATATCAACMGTAYCCACA
   ACCTCTCAACGCTGTYTGCGAWGAGG 5'                      GTATAGTTGKCATRG 5'
```

mixtures: W̲: A, T;   K̲: G, T;   M̲: A, C;   Y̲: C, T;   R̲: G, A;

These cassettes are identical to each other except that '#2B' contains an extra asparagine codon (underlined). As described in the text, these cassettes together encode a library which includes variation at additional positions. These cassettes were constructed exactly as described for '#1', except that the W̲, K̲, M̲, V̲, and R̲, positions were synthesized using line mixing.

Preparation of phage libraries:

We prepared our two libraries separately. To prepare library 1, cassette #1 (17 pmol) was ligated to p1212-BstX 1 (3.3 pmol) which had been digested with BstX 1 and gel purified. Ligation products were ethanol precipitated, resuspended in 30 µl of 0.1x TE, and transformed

into K91 cells[37] using 12 electroporations. This yielded approximately ~5 x 10^9 transformants.

Transformants were incubated for a short growth period (1 hour, 37°, with aeration) in 12 ml of SOC (20 g/liter tryptone, 5 g/liter yeast extract, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 10 mM NaCl, 2.5 mM KCl, and .4 % (w/v) glucose). We then added VCSM13 helper phage (~6 x 10^{11} phage) and fresh glucose (1.2 ml of 4.2 % (w/v) in 2xYT) and allowed infection to proceed (1.5 hour, 37°, with aeration). A large fraction (greater than 50% as estimated by conversion to kanamycin resistance) of the transformants were infected with helper phage.

Our transformant culture (~13 ml) was placed in an anaerobic chamber, diluted into 1200 ml of deoxygenated growth media (2xYT containing 40 mM Hepes, 0.1 mM $ZnSO_4$, 0.5 µg /mi thiamine, 100 µg/ml ampicillin and 70 µg/ml kanamycin), and grown to saturation (room temperature, 24 hours, no agitation) under anaerobic conditions. Final culture titers were ~1.0 x 10^{11} ampicillin transducing units (ATU)/ml.

Phage were harvested from the large volume of culture supernatant using two sequential PEG precipitations. The cultures were first cleared of cells by centrifugation (11000 g, 4°, 45 min) and 960 ml of the supernatants were placed in fresh centrifuge tubes. Phage were precipitated by adding deoxygenated 5X PEG solution (240 ml), inverting the tubes 40 times, incubating at 0° (9 hours), and centrifuging (12200 g, 4°, 2 hr). Phage pellets were resuspended in 32 ml of deoxygenated growth media and reprecipitated by adding deoxygenated 5X PEG solution (8 ml), inverting the tubes 40 times, incubating at 0° (2 hours), and centrifuging (30600 g, 4°, 1 hr). Phage pellets were resuspended in 1 ml deoxygenated pZif12 wash buffer (25

mM potassium phosphate, 60 mM potassium acetate, 60 mM potassium glutamate, 2 mM $MgCl_2$, 0.02 mM $ZnSO_4$, 5% glycerol (v/v), 0.5% triton x-100 (w/v), pH 7.8)[17] containing 0.1 mg/ml bovine serum albumin (BSA). Due to concerns about oxidation, all steps after the addition of helper phage were performed in an anaerobic chamber (<1 ppm O_2), except for the centrifugations and 0° incubations. For these steps, zinc finger phage samples were sealed in screw-capped centrifuge tubes before they were removed from the anaerobic chamber.

Our second library was prepared in parallel with library 1 using the same procedures, except that the ligations used cassettes #2A and #2B. A similar number of transformants and total purified ATU were obtained.

Our final library phage preps (each about 3 ml and 6 x 10^{13} ATU/ml) were mixed together, sealed in screw-capped microfuge tubes (0.1 ml aliquots), removed from the anaerobic chamber, and frozen at -80°.

Selections:

Our first selection cycle was performed as follows: Binding reactions (240μl) were made which contained library phage (6.5x 10^{11} ATU), biotinylated target site (40 nM) sheared salmon sperm DNA (0.65 mg/ml), and BSA (0.1 mg/ml) in .9x pZif12 wash buffer. Binding was allowed to proceed for 110 min. Aliquots (10 μl) of each binding reaction were applied to 16 streptavidin-coated wells of a Pro-Bind™ plate containing pZif12 wash buffer (40 μl), and the plates were agitated on a rotary shaker [Labline model # 4625, setting 6 (600 rpm),

90 min]. The samples were then shaken from the wells, and the wells were washed 5 times with 125 µl of pZif12 wash buffer. [Each wash was agitated for ~4 min on the rotary shaker (setting 6) and then removed from the wells by shaking the plate upside-down over fresh paper towels.] Elution buffer (30 µl) was then added to each well. After 90 min, each set of elutions was removed from the anaerobic chamber, pooled and titered.

Pooled eluted phage (0.48 ml) were mixed with a freshly saturated overnight culture of K91 *E. coli* (8 ml). After 80 min of infection (37°, with aeration), we added 0.25 ml of VCSM13 helper phage solution (2xYT containing ~3x10^{11} phage/ml, 5% glucose (w/v) and 500 µg/ml ampicillin) and allowed helper phage infection to proceed (90 min, 37°, with aeration). A large fraction (greater than 50% as estimated by conversion to kanamycin resistance) of the ampicillin-resistent cells were infected with helper phage. After transfer to the anaerobic chamber, deoxygenated growth media (400 ml) was added and each culture was grown to saturation (room temperature, 24 hours, no agitation) under anaerobic conditions. Cells were cleared by centrifugation and phage were purified by a single PEG precipitation and stored frozen at -80°.

Subsequent selection cycles used similar procedures.

Specificity platings:

Binding reactions were prepared which contained phagemid (~3x10^9/ml (= 5 pM)), biotinylated target DNA fragment (40 nM), sonicated salmon sperm DNA (0.65 mg/ml) and BSA (0.1 mg/ml) in 15

μl of .9x pZif12 wash buffer. Reactions were equilibrated 45', diluted into 40 μl pZif12 wash buffer, and applied to streptavidin - coated wells of a microtiter plate. Reactions were equilibrated in the wells for 90' and the plates were agitated on a rotary shaker [Labline model # 4625, setting 6 (600 rpm), 90 min]. The samples were shaken from the wells, and the wells were washed 5 times with 125 μl of pZif12 wash buffer. [Each wash was agitated for ~4 min on the rotary shaker (setting 6) and then removed from the wells by shaking the plate upside-down over fresh paper towels.] Elution buffer was added to each well and incubated 70 min. Eluted phage were quantitated as infectious units essentially as described in previous studies[26], except freshly saturated K91 was used instead of frozen starved cells.

Sequencing

Single - stranded sequencing templates were prepared using commercially available kits (Qiagen) and sequenced using standard methods (Applied Biosystems).

Assignment of consensus residues or residue sets:

We scanned our selected sequences for 23 possible consensus features: 16 amino acids and 7 amino acid sets. In organizing our amino acid sets, we grouped residues which could potentially serve similar functions in protein-DNA recognition. These sets include acidic residues [D+E], basic residues [R+K], nonpolar residues [G+M+V+A+I+L], and pairs of residues that can position hydrogen bond donors and/or

acceptors at identical locations relative to the C-α carbon ([Q+E], [D+N], [S+T]). In addition, we included the set [N+T], because asparagine and threonine feature almost identical volumes (117.7 $Å^3$ and 116.1 $Å^3$) and can position hydrogen-bonding residues at similar locations relative to the α-carbon. The fact that we found this set on two occasions, at the same base-contact position, and apparently recognizing the same base (thymine) suggests that this last grouping was reasonable.

In assigning each consensus feature, we were careful to require enough occurrences to avoid problems with 'false positives'. Specifically, each individual residue (except arginine) was assigned as a consensus feature if it is observed at a given position in half or more sequences, and each pair of residues (except [R+K]) was noted if the relevant amino acids are observed in more than half of all sequences. For arginine and the pair [R+K], we used a higher threshold (an occurrence in ≥ 2/3 and ≥ 3/4 sequences, respectively), in order to account for the greater random frequency of arginine codons in our libraries. Finally, we required that nonpolar residues occur in all sequences in order to assign a given position as hydrophobic ('o'), since this residue group [G+A+I+L+M+V] encompasses an especially large proportion of our random codons (10 of 24).

References:

1.   Ades, S.E. The Engrailed Homeodomain: Determinants of DNA-Binding Affinity and Sepecificity. Ph. D. thesis., 1995.

2.   Damante, G., *et al.* A molecular code dictates sequence-specific DNA recognition by homeodomains. EMBO J., 1996. **15**: p. 4992-5000.

3.   Kissinger, C.R., *et al.* Crystal Structure of an engrailed Homeodomain-DNA Complex at 2.8 Å Resolution: A Framework for Understanding Homeodomain-DNA Interactions. Cell, 1990. **63**: p. 579-590.

4.   Pavletich, N.P. and C.O. Pabo. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. Science, 1991. **252**: p. 809-817.

5.   Rebar, E.J. and C.O. Pabo. Zinc Finger Phage: Affinity Selection of Fingers with New DNA-Bindig Specificities. Science, 1994. **263**: p. 671-673.

6.   Jacobs, G.H. Determination of the base recognition positions of zinc fingers from sequence analysis. EMBO J., 1992. **11**: p. 4507-4517.

7.   Nardelli, J., *et al.* Base sequence discrimination by zinc-finger DNA-binding domains. Nature, 1991. **349**: p. 175-178.

8.   Lehming, N., *et al.* Mutant lac repressors with new specificities hint at rules for protein-DNA recognition. EMBO J., 1990. **9**: p. 615-621.

9.   Desjarlais, J.R. and J.M. Berg. Toward rules relating zinc finger protein sequences and DNA binding site preferences. Proc. Natl. Acad. Sci. U.S.A., 1992. **89**: p. 7345-7349.

10.   Desjarlais, J.R. and J.M. Berg. Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. Proc. Natl. Acad. Sci. U.S.A., 1993. **90**: p. 2256-2260.

11.   Pellegrino, G.R. and J.M. Berg. Identification and characterization of "zinc finger" domains by the polymerase chain reaction. Proc. Natl. Acad. Sci. U.S.A., 1991. **88**: p. 671-675.

12.   Bray, P., *et al.* Characterization and mapping of human genes cncoding zinc finger proteins. Proc. Natl. Acad. Sci. U.S.A., 1991. **88**: p. 9563-9567.

13.   Pavletich, N.P. and C.O. Pabo. Crystal Structure of a Five-Finger GLI-DNA Complex: New Perspectives on Zinc Fingers. Science, 1993. **261**: p. 1701-1707.

14.   Choo, Y., I. Sánchez-García, and A. Klug. *In vivo* repression by a site-specific DNA-binding protein designed against an oncogenic sequence. Nature, 1994. **372**: p. 642-645.

15.   Pomerantz, J.L., P.A. Sharp, and C.O. Pabo. Structure-Based Design of Transcription Factors. Science, 1995. **267**: p. 93-96.

16.   Kim, J.S., *et al.* Design of TATA box-binding protein / zinc finger fusions for targeted regulation of gene expression. Proc. Natl. Acad. Sci. U.S.A., 1997. **94**: p. 3616-3620.

17.   Greisman, H.A. and C.O. Pabo. A General Strategy for Selecting High-Affinity Zinc Finger Proteins for Diverse DNA Target Sites. Science, 1997. **275**: p. 657-661.

18.   Elrod-Erickson, M., *et al.* Zif268 protein-DNA complex refined at 1.6A: a model system for understanding zinc finger-DNA interactions. Structure, 1996. **15**: p. 1171-1180.

19.   Fairall, L., *et al.* The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature, 1993. **366**: p. 483-487.

20.   Houbaviy, H.B., *et al.* Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. Proc. Natl. Acad. Sci. U.S.A., 1996. **93**: p. 13577-13582.

21.   Omichinski, J.G., *et al.* The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. Nat. Struct. Biol., 1997. **4**: p. 122-132.

22.   Kim, C.A. and J.M. Berg. A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. Nature Struct. Biol., 1996. **3**: p. 940-945.

23. Jamieson, A.C., S. Kim, and J.A. Wells. *In Vitro* Selection of Zinc Fingers with Altered DNA-Binding Specificity. Biochemistry, 1994. **33**: p. 5689-5695.

24. Choo, Y. and A. Klug. Toward a code for the interactions of zinc fingers with DNA: Selection of randomized fingers displayed on phage. Proc. Natl. Acad. Sci. U.S.A., 1994. **91**: p. 11163-11167.

25. Wu, H., W. Yang, and C.F. Barbas. Building zinc fingers by selection: Toward a therapeutic application. Proc. Natl. Acad. Sci. U.S.A., 1995. **92**: p. 344-348.

26. Rebar, E.J., H.A. Greisman, and C.O. Pabo. Phage Display Methods for Selecting Zinc Finger Proteins with Novel DNA-Binding Specificities. Methods Enzymol., 1996. **267**: p. 129-149.

27. Zeichner, S.L., J.T. Kim, and J.C. Alwine. Linker-Scanning Mutational Analysis of the Transcriptional Activity of the Human Immunodeficiency Virus Type 1 Long Terminal Repeat. J. Virology, 1991. **65**: p. 2436-2444.

28. Myers, G., *et al.*, eds. *Human Retroviruses and AIDS 1885.* . 1995, Los Alamos National Laboratory: Los Alamos, New Mexico.

29. Swirnoff, A.H. and J. Milbrandt. DNA-Binding Specificity of NGF1-A and Related Zinc Finger Transcription Factors. Mol. Cell. Biol., 1995. **15**: p. 2275-2287.

30. Seeman, N.C., J.M. Rosenberg, and A. Rich. Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl. Acad. Sci. U.S.A., 1976. **73**: p. 804-808.

31. Kim, C.A. and J.M. Berg. Serine at Position 2 in the DNA Recognition Helix of a Cys2-His2 Zinc Finger Peptide is Not, in General, Responsible for Base Recognition. J. Mol. Biol., 1995. **252**: p. 1-5.

32. Cheng, C. and E.T. Young. A Single Amino Acid Substitution in Zinc Finger 2 of Adr1p Changes its Binding Specificity at two Positions in UAS1. J. Mol. Biol., 1995. **251**: p. 1-8.

33. Thukral, S.K., M.L. Morrison, and E.T. Young. Alanine scanning site-directed mutagenesis of the zinc fingers of transcription factor ADR1:

Residues that contact DNA and that transactivate. Proc. Natl. Acad. Sci. U.S.A., 1991. **88**: p. 9188-9192.

34.   Christy, B.A., L.F. Lau, and D. Nathans. A gene activated in mouse 3T3 cells by serum growth factors encodes a protein with "zinc finger" sequences. Proc. Natl. Acad. Sci. U.S.A., 1988. **85**: p. 7857-7861.

35.   Ausubel, F.M., et al., eds. *Current Protocols in Molecular Biology.* . 1994, Current Protocols.

36.   Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratcry Manual.* 1989, Coid Spring Harbor, NY: Cold Spring Harbor Press.

37.   Smith, G.P. and J.K. Scott. Libraries of Peptides and Proteins Displayed on Filamentous Phage. Methods Enzymol., 1993. **217**: p. 228-257.

38.   White, S. and C.O. Pabo. unpublished data. 1996.

Tables 1A-1C: Phage pool plating efficiencies with target and nontarget sites

The indicated combinations of phage pools and biotinylated selection target sites (figure 3) were preincubated, applied to streptavidin - coated microtiter wells, washed, eluted, and quantitated via transduction of *E. coli* to ampicillin resistance. As a positive control, 1212 phage was also tested using its target site (the GCGT site). The % of applied phage recovered in the eluates is shown.

The phage pools and selection targets are referred to by the four - base pair selection target sequences.

The plating efficiency of a phage pool with its selection target (or 1212 phage with its target site) is highlighted in gray.

Table 1A: Phage pool plating efficiencies with target and nontarget sites

Plating Efficiency (%)

Phage Pool	sites							
	TTAT	ATAT	ATAA	AGCA	TGCA	TGCT	GCGT	(no site)
TTAT	.22	.028	.011	.0003	.0006	.0011	.0007	.0004
ATAT	.26	.25	.055	.0014	.0013	.0028	.0040	.0005
ATAA	.032	.030	.12	.0004	.0010	.0007	.0006	.0005
AGCA	.0041	.0026	.0074	.11	.040	.0035	.0013	.0006
TGCA	.0029	.0006	.0017	.0057	.13	.025	.0008	.0005
TGCT	.0023	.0009	.0008	.0027	.14	.36	.0019	.0007
1212	.0005	.0011	.0012	.0013	.0007	.0005	.86	.0008

- plating efficiency of phage pool with its target site

Table 1B: Phage pool plating efficiencies with target and nontarget sites

Plating Efficiency (%)

Phage Pool	sites							
	TAAG	TATG	TATA	CATA	CTTA	CTGC	GCGT	(no site)
TAAG	.48	.097	.054	.0053	.0012	.0010	.0029	.0011
TATG	.028	.36	.22	.012	.0010	.0006	.0080	.0006
TATA	.0063	.11	.38	.055	.0020	.0023	.0059	.0006
CATA	.027	.26	.76	.98	.21	.0068	.013	.0039
CTTA	.0008	.0023	.023	.20	.49	.0031	.014	.0011
CTGC	.0088	.029	.094	.0092	.095	.42	.014	.0009
1212	.0006	.0007	.0006	.0008	.0008	.0008	.74	.0004

▨ - plating efficiency of phage pool with its target site

Table 1C: Phage pool plating efficiencies with target and nontarget sites

Plating Efficiency (%)

Phage Pool	sites	ATGC	AAGC	GCAG	GCTG	GCTT	GCAT	GCGT	(no site)
ATGC		.13	.012	.0006	.0007	.0015	.0045	.0020	.0017
AAGC		.037	.12	.0056	.0025	.0039	.020	.018	.0020
GCAG		.0008	.0004	.30	.14	.16	.41	.15	.0006
GCTG		.0005	.0007	.26	.17	.19	.18	.12	.0004
GCTT		.0006	.0007	.27	.18	.31	.27	.28	.0005
GCAT		.0008	.0014	.20	.078	.13	.42	.13	.0006
1212		.0009	.0009	.026	.0042	.0016	.013	.84	.0008

- plating efficiency of phage pool with its target site

Tables 2A-2C: Selectivity of Phage Pools Against Nontarget Sites

The data in tables 1A-1C is presented in a different form in order to highlight the selectivity of each phage pool for its selection target site (the ratio of plating efficiencies obtained using target and nontarget sites).   The interaction of each pool with its selection target is indicated by a '1' enclosed in a gray box.   For reference, the % plating efficiency of each phage pool with its selection target is also provided [under the heading Plating Efficiency (target site)].

Table 2A: Selectivity of phage pools against nontarget sites

Selectivity Against Nontarget Sites:

$$\frac{\text{plating efficiency (target site)}}{\text{plating efficiency (nontarget site)}}$$

Phage Pool	Plating Efficiency $\left(\dfrac{\text{target}}{\text{site}}\right)$	sites TTAT	ATAT	ATAA	AGCA	TGCA	TGCT	GCGT	(no site)
TTAT	0.22 %	**1**	8	20	643	354	193	337	598
ATAT	0.25 %	1	**1**	5	179	186	87	62	459
ATAA	0.12 %	4	4	**1**	266	115	167	201	230
AGCA	0.11 %	26	41	14	**1**	3	30	82	184
TGCA	0.13 %	46	224	78	23	**1**	5	168	291
TGCT	0.36 %	155	381	440	133	2	**1**	188	500
1212	0.86 %	1693	747	715	651	1270	1693	**1**	1081

1 - interaction of phage pool with its target site

Table 2B: Selectivity of phage pools against nontarget sites

Phage Pool	Plating Efficiency $\left(\dfrac{\text{target}}{\text{site}}\right)$	Selectivity Against Nontarget Sites: $\dfrac{\text{plating efficiency (target site)}}{\text{plating efficiency (nontarget site)}}$							
		sites							
		TAAG	TATG	TATA	CATA	CTTA	CTGC	GCGT	(no site)
TAAG	0.48 %	1	5	9	90	393	461	166	428
TATG	0.36 %	13	1	2	31	367	640	45	640
TATA	0.38 %	60	3	1	7	186	165	63	634
CATA	0.98 %	37	4	1	1	5	143	73	250
CTTA	0.49 %	627	211	22	2	1	158	36	445
CTGC	0.42 %	48	14	5	46	4	1	30	461
1212	0.74 %	1215	1133	1261	950	903	924	1	1947

1 - interaction of phage pool with its target site

Table 2C: Selectivity of phage pools against nontarget sites

Phage Pool	Plating Efficiency $\left(\dfrac{\text{target}}{\text{site}}\right)$	Selectivity Against Nontarget Sites: $\dfrac{\text{plating efficiency (target site)}}{\text{plating efficiency (nontarget site)}}$							
		sites							
		ATGC	AAGC	GC/AG	GCTG	GCTT	GCAT	GCGT	(no site)
ATGC	0.13 %	1	11	197	178	87	28	63	76
AAGC	0.12 %	3	1	22	49	31	6	7	62
GCAG	0.30 %	388	701	1	2	2	1	2	477
GCTG	0.17 %	376	233	1	1	1	1	1	438
GCTT	0.31 %	499	452	1	2	1	1	1	583
GCAT	0.42 %	550	311	2	5	3	1	3	726
1212	0.84 %	956	973	33	198	532	66	1	1069

1 - interaction of phage pool with its target site

Table 3:   Summary of Phage Selectivities

This table summarizes the data in tables 2A-2C in a way that highlights the tendency of phage pools to discriminate more strongly against sites which contain more base changes relative to the selection target site.   To make this table, the test sites for each phage pool were grouped according to their degree of difference from the selection target site (e.g.: one, two, three or four base changes), and the selectivities for each of these groups were averaged.   An entry in the table is blank if a phage pool was not tested using a site with the indicated number of changes.

As indicated, a '1' highlighted in gray indicates the interaction of a phage pool with its selection target site.

Table 3: Summary of phage selectivities

Phage Pool	Average selectivity against sites with changes at _ positions					
	0	1	2	3	4	(no site)
TTAT	1	8	107	345	643	598
ATAT	1	3		109	186	459
ATAA	1	4	135	115	184	230
AGCA	1	3	22	41	54	184
TGCA	1	14		62	196	291
TGCT	1	2	144	284	440	500
TAAG	1	5	9	90	340	428
TATG	1	7	31	367	343	640
TATA	1	5	123		114	634
CATA	1	3	4	90	73	250
CTTA	1	2	90	211	332	445
ATGC	1	11		63	123	76
GCxx GCAG	1	1	2		544	477
phage pools GCTG	1	1	1		305	438
GCTT	1	1	1		476	583
GCAT	1	3	5		431	726
discarded AAGC	1	3		7	27	62
phage pools CTGC	1		4	38	22	461
control 1212	1	299	115	1059	1088	1365

1 - interaction of phage pool with its target site

Table 4:   Consensus base contact residues and target sites from zinc

finger selection studies

This table provides consensus sequences obtained at the base contact positions in every zinc finger selection study performed to date. The top of each column indicates the first three base pairs of each selection target site, and the fourth base pair is listed at left. Target sites are shown in the 3' ---> 5' orientation. Protein consensus sequences are provided beneath the first three base-pairs of their selection target and to the right of the fourth base pair of their target The numbers ('-1', '2', '3', and '6') indicate the base-contact positions of the zinc finger motif.

For example: the left-most entry in the top row ('RHHL'), indicates that a selection was performed for binding to the sequence 3'-AGGG-5', and a consensus sequence was obtained which featured arginine at position '-1', histidines at positions '2' and '3', and leucine at position '6'.

Gray entries indicate attempted selections that failed to yield any increase in retention efficiency.

References for entries are as follows: broken underscore: Greisman and Pabo[17]; solid underscore: this report and our previous study[5]; no underscore: either Choo et al.[14] or Jamieson et al.[23].

Table 4: Consensus base contact residues and target sites from zinc finger selection studies

	GGG -1 2 3 6	AGG -1 2 3 6	TGG -1 2 3 6	CGG -1 2 3 6	GAG -1 2 3 6	AAG -1 2 3 6	TAG -1 2 3 6	CAG -1 2 3 6	GTG -1 2 3 6	ATG -1 2 3 6	TTG -1 2 3 6	CTG -1 2 3 6	GCG -1 2 3 6	ACG -1 2 3 6	TCG -1 2 3 6	CCG -1 2 3 6
G						SNRDW	NRDSNRRDo	S	DSNRRDo	S		DSSRRD R		DW	DW	DWDGTR
C								DSNRE	R	R				QE TR / DR RADR / QGSR	TR	TR
T														DW		DR
A			DSKR													IHIR
	BHHL															

	GGA -1 2 3 6	AGA -1 2 3 6	TGA -1 2 3 6	CGA -1 2 3 6	GAA -1 2 3 6	AAA -1 2 3 6	TAA -1 2 3 6	CAA -1 2 3 6	GTA -1 2 3 6	ATA -1 2 3 6	TTA -1 2 3 6	CTA -1 2 3 6	GCA -1 2 3 6	ACA -1 2 3 6	TCA -1 2 3 6	CCA -1 2 3 6
G					RDNAQTNT				RDVN SAQ	GQ / Q GQ			RDVN	QGIR		
C			SRA							QI.Q						
T																
A																

	GGT -1 2 3 6	AGT -1 2 3 6	TGT -1 2 3 6	CGT -1 2 3 6	GAT -1 2 3 6	AAT -1 2 3 6	TAT -1 2 3 6	CAT -1 2 3 6	GTT -1 2 3 6	ATT -1 2 3 6	TTT -1 2 3 6	CTT -1 2 3 6	GCT -1 2 3 6	ACT -1 2 3 6	TCT -1 2 3 6	CCT -1 2 3 6
G	QAHT		MHHEEQHT ER		QDENoTN No / TN N		QENo		RDAS	QGE			KNDK	QHDK		
C																
T																
A																

	GGC -1 2 3 6	AGC -1 2 3 6	TGC -1 2 3 6	CGC -1 2 3 6	GAC -1 2 3 6	AAC -1 2 3 6	TAC -1 2 3 6	CAC -1 2 3 6	GTC -1 2 3 6	ATC -1 2 3 6	TTC -1 2 3 6	CTC -1 2 3 6	GCC -1 2 3 6	ACC -1 2 3 6	TCC -1 2 3 6	CCC -1 2 3 6
G			RDHT				ND				A D		RDTG			
C																
T																
A																

Table 5:   Base contact positions and target sites of structurally studied zinc fingers

This table provides the base contact positions and target site of each finger from figure 6 in chapter 1. Gray is used to highlight residues which interact with the base expected for their position. The abbreviations for each finger are the same as those used in figure 6, chapter 1. Relevant references are provided in the legend to figure 6, chapter 1. Other symbols are identical to those described in table 4.

For example: the left-most entry in the top row ('ESKR') indicates that finger two of YY1 was structurally studied in complex with the target site 3'-GAGG-5', and that the serine at position '2', the lysine at position '3', and the arginine at position '6' interact with their expected base positions (e.g., the arginine at position six interacts with the underlined guanine in 3'-GAG<u>G</u>-5').

Table 5: Base contact positions and target sites of structurally studied zinc fingers

Block 1 (rows G, C, T, A; columns GGG AGG TGG CGG GAG AAG TAG CAG GTG ATG TTG CTG GCG ACG TCG CCG):
- E(S)KR YY1-2 — under AGG
- QRNR GAGA — under GAG
- HSNR TTK-1 — under TAG
- AVMV GLI-2; DSAK YY1-1; D(SS)K GLI-5 — under CAG
- RDER Zif-1; RDER Zif-3 — under GCG

Block 2 (columns GGA AGA TGA CGA GAA AAA TAA CAA GTA ATA TTA CTA GCA ACA TCA CCA):
- RDNA TTK-2 — under GAA
- LFNT YY1-3 — under TAA
- NSDK GLI-4 — under CCA

Block 3 (columns GGT AGT TGT CGT GAT AAT TAT CAT GTT ATT TTT CTT GCT ACT TCT CCT):
- RDHT Zif-2 — under GGT

Block 4 (columns GGC AGC TGC CGC GAC AAC TAC CAC GTC ATC TTC CTC GCC ACC TCC CCC):
- QTNS YY1-4 — under AAC

Figure 1:   Source of selection targets for these studies

This figure shows the region of the HIV-1 promoter which is the source of selection targets for these studies.   '-36' and '-14' denote the distance, in base-pairs, to the start of HIV-1 transcription.   The target site for the TATA binding protein is underlined.

-36   ATGCTGCA<u>TATAA</u>GCAGCTGCTT   -14
           TACGACGTATATTCGTCGACGAA

Figure 2: Sketch of the designed zinc finger protein '1212' bound to its DNA target site. The target sequence shown is presumed to be the optimal binding site for 1212, based on our understanding of the specificities of its component fingers[4, 29, 38]. Only one strand of the DNA is shown (5' is on the left). Each numbered circle represents the corresponding finger of Zif268. Underscores indicate the overlapping four - base pair subsites for the fingers of 1212, and key finger - base contacts are indicated by arrows.

Figure 3: Sketch of the library (R = randomized finger) and the target sites used for the selection of new zinc fingers from this library. As shown, the randomized finger was selected for binding to a total of twenty different four - base pair sequences (in gray). Note that the interaction of the other fingers with their subsites helps to ensure that the randomized finger is properly aligned for binding to the new DNA sequence.

The complete sequence of the biotinylated DNA selection targets was (b = biotin):

```
5'   CCTTTGG____GGGCGTCCTGCAGAC-b
3'   GGAAACC     CCCGCAGGAC
```

Oligonucleotides were synthesized using standard methods (Applied Biosystems) and BioTeg™ controlled pore glass from Glen research. Oligonucleotides were purified essentially as described[26].

Figure 2:

protein '1212'

target site
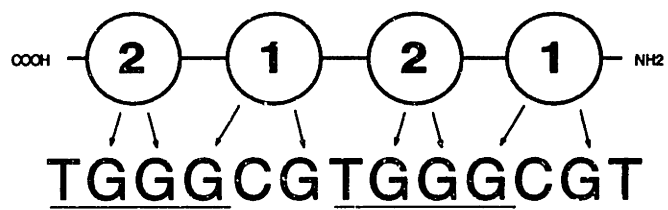
Figure 3:

library

selection
targets
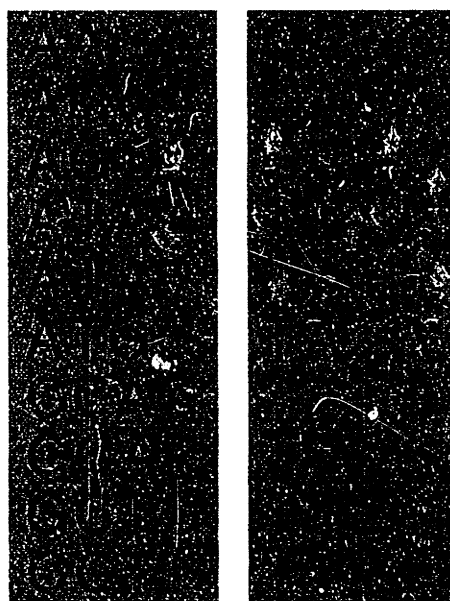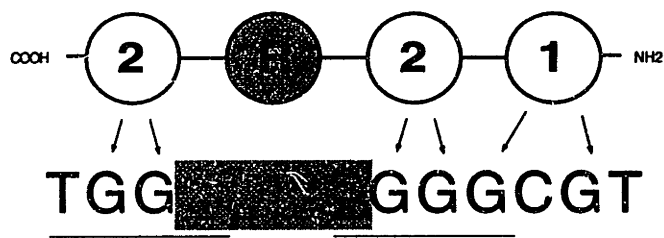
Figure 4:   Affinities of 1212 and Zif268 for the 1212 target site and
           for a variant sequence

Apparent dissociation constants were determined for the binding
of 1212 and Zif268 to the two sites shown. The top site is the target
site for 1212, while the bottom site contains mutations (highlighted by
a gray 'X') which should disrupt binding of the carboxy - terminal finger
2 of 1212. Only one strand of each target site is shown. 5' is on the
left.

The numbered circles represent fingers 1, 2 and 3 of Zif. Both
proteins are positioned over their target sites on the test sequences.

The complete sequences of the DNA fragments used for this study
were: CTAGGCCTCATTGGGCGTGGGCGTCTAGCT
               TCCGGAGTAACCCGCACCCGCAGATCGATT
and: CTAGGCCTCATGACGCGTGGGCGTCTAGCT
               TCCGGAGTACTGCGCACCCGCAGATCGATT

Figure 4: Affinities of 1212 and Zif268 for the 1212 target site and for a variant sequence
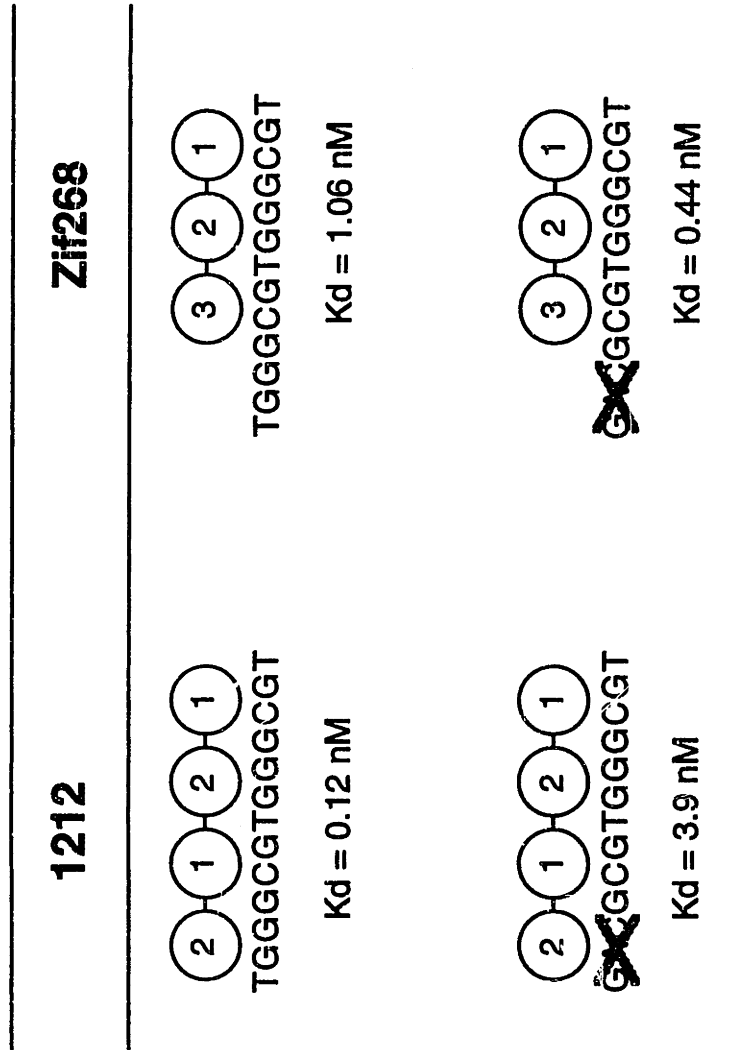
Figure 5:   Mutations throughout the '1212' target site abolish binding
            of '1212' phage

The 1212 phage were preincubated with the biotinylated DNA
duplexes shown.   Phage were then diluted 5-fold, applied to
streptavidin - coated wells of a microtiter plate, washed, eluted, and
quantitated via transduction of *E. coli* to ampicillin resistance.   The %
of applied phage recovered in the eluates is shown.

Reported values are the averages of duplicate samples, which in
all cases agreed to within a factor of 4.

The complete sequence of the biotinylated 1212 target site
duplex was (b = biotin):   CCTTTGGGCGTGGGCGTCCTGCAGAC-b
                                  GGAAACCCGCACCCGCAGGAC

The other duplexes had identical sequences except for the indicated
mutations.

Figure 5: Mutations throughout the 1212 target
site abolish binding of 1212 phage

test sites (mutations in gray)	% recovery of '1212' phage
TGGGCGTGGGGCGT (no mutations)	0.79
TCCGCGTGGGGCGT	0.02
TGGCCGTGGGGCGT	0.01
TGGGGCGTGGGCGT	0.01
TGGGGCGTGGGCCGT	0.01
(no site)	0.01

Figure 6: Randomized libraries

The two randomized libraries are shown. In library 1, six positions ('-1', '1', '2', '3', '5' and '6'; numbering is with respect to the amino terminus of the α-helix) were randomized to encode 16 different amino acids (indicated by the symbol '16'). The remaining positions were unmutated (indicated by dashes) and contained the corresponding amino acids from Zif finger one. In library two, additional variation was introduced at six other positions. Position '-5' was encoded as arginine or lysine; position '-3' as phenylalanine or tyrosine; position '5' as leucine or arginine; position '9' as arginine or serine; and position '10' as isoleucine or threonine. Also, in half of the library, an asparagine was inserted between positions '8' and '9' (indicated by '+N').

For reference, the zinc finger secondary structure and the amino acid sequence of wild - type Zif finger 1 are provided at bottom.

Figure 6:

Randomized positions in libraries 1 and 2:

Positions

	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10
library 1	-	-	-	-	16	16	16	16	-	16	16	-	-	-	-
library 2	-	-	R/K	-	16	16	16	16	L/R	16	16	-	R/S	I/T	-

+N

Zif 1

E R P Y A C P V E S C D R R F S R S D E L T R H I R I H T

Figure 7A: Sequences of Selected Fingers

Individual clones were picked from each selected phage pool and sequenced. Derived protein sequences are shown for the randomized positions (indicated by '-1', '1', '2', '3', '5' and '6') and also for position 4 (indicated by a dash). [Note that position '4' was encoded as either leucine or arginine in library two.] The first zinc - coordinating histidine is also shown as a reference.

The four - base pair DNA target used to select each set of clones is shown at the top of each table. Note that these DNA sequences are shown in a 3' --> 5' orientation. This is done to facilitate comparison of the DNA target with its selected protein sequences. [The 5' base typically interacts with position '6', and the next three bases typically interact with, respectively, positions '3', '-1', and '2'.]

Criteria for deriving consensus sequences are described in the materials and methods section.

Amino acids are designated using standard abbreviations. 'o' indicates that the selections yielded nonpolar residues.

Figure 7B: Sequences of Selected Fingers: GCxx Pools

Nomenclature and symbols are as described for figure 7A.

Figure 7C: Sequences of Selected Fingers: Pools with a four-residue histidine spacing:

Nomenclature and symbols are as described for figure 7A, except that derived protein sequences are shown out to the second histidine. Therefore, sequences are shown for positions '9' and '10' (indicated by dashes), in addition to those described above. [Note that the codons for position '9' were designed to allow arginine or serine, and the codons for position '10' were designed to allow isoleucine or threonine.] Also, the location of the asparagine insert, which was recovered in the vast majority of these clones, is indicated by an asterisk.

Figure 7A: Sequences of selected fingers

TCGT

3′ T C G T 5′

-1 1 2 3	5 6
E V Q H	L Q T H
E R Q H	L A T H
E G G H	L R T H
E R Q H	L L T H
E R Q H	L K T H
E V Q H	L Q T H
E K Q H	L Q T H
E V Q H	L R S H
E Q H	T

ACGT

3′ A C G T 5′

-1 1 2 3	5 6
E R R H	L R D H
V K R V	L E A H
E K R Q	L I Q H
E R S K	L E K H
E S G H	L R T H
G P N G	L T A H
E R R Q	L R D H
V R R V	L S A H
E R K R	

GAAT

3′ G A A T 5′

-1 1 2 3	5 6
Q R E N	L M G H
Q V E N	L R G H
T R D A	L A G H
D A S Q	L K I H
Q L D N	L K M H
Q M E N	L R A H
L K T N	L D M H
T K D G	L A G H
Q DEN	O

GTAT

3′ G T A T 5′

-1 1 2 3	5 6
T G A N	L V G H
N K A N	L E M H
T V G N	L E M H
T V T N	L Q L H
R P N D	L R L H
N A H N	L M A H
N G T A	L S M H
TN N	O

ATAT

3′ A T A T 5′

-1 1 2 3	5 6
V V G N	L M K H
T V G N	L N G H
T I A N	R D G H
G V T A	L A K H
R P V E	L R N H
N Q S A	L L A H
N V D A	L R K H
T E G N	L Q A H
S A G E	L V S H
T L A N	R N G H
TN N	

CGTA

3′ C G T A 5′

-1 1 2 3	5 6
G S N G	L R Q H
G S N G	L A Q H
Q H S A	L R E H
A T Q A	L A A H
A H S A	L R L H
V H S A	L R Q H
HSA	RQ

AATA

3′ A A T A 5′

-1 1 2 3	5 6
Q P T G	L L Q H
Q P T G	L E Q H
Q I T G	L M Q H
Q I G T	L R S H
Q P A P	L Q L H
T R Q A	L E G H
S A A G	R S L H
Q R V L	L R Q H
Q G	Q

TATA

3′ T A T A 5′

-1 1 2 3	5 6
S T D G	L R Q H
S S D G	L R Q H
S A D G	L R Q H
D G T A	L R A H
T G V G	L A Q H
E S T G	L R M H
Q S S A	L R L H
D P V G	L K Q H
G	RQ

TATT

3′ T A T T 5′

-1 1 2 3	5 6
Q H Q E	L Q K H
Q R G E	L A V H
Q A G E	L R V H
Q R G E	L R A H
S G G E	L R T H
Q A I G	L R V H
Q A Q E	L V K H
Q GE	

Figure 7B: Sequences of selected fingers: GCxx pools

3' GACG 5'

-1 1 2 3	5 6
ERTDL	QKH
TREDR	TRH
TGDDL	ERH
SSEDL	QRH
APTDL	ERH
AKADL	DKH
AVADL	DRH
VGTDL	MKH
VGSDL	AKH
VMDDL	TRH
MVADL	DRH
ERSEL	LRH
MDVSL	QRH
EKDML	LRH
RTGDL	EKH
RSSCL	EKH
TLDDL	QRH
QKHDL	VVH
QAEVL	ERH
TGEDL	SRH
VPVDL	RKH
EVDQL	DRH
D	RK

3' GTCG 5'

-1 1 2 3	5 6
LLHDL	RKH
TVDER	TRH
SNPDL	DRH
AAVSL	NKH
QFISDL	NRH
NLPDL	DRH
TLHDR	MRH
NDVDL	KRH
SGKNLA	DH
TATDL	GRH
HVHVL	LKH
RKLDL	DRH
LKVDL	ERH
HARTL	QRH
QRSDL	NRH
RMGDR	NRH
D	RK

3' TTCG 5'

-1 1 2 3	5 6
GHLDL	ARH
NGQTL	AKH
TGVDL	ERH
SKQDL	VRH
QVVDL	RRH
NMSDL	NRH
D	R

3' TACG 5'

-1 1 2 3	5 6
EHASL	MKH
AGIDL	RRH
QTAEL	AKH
TSEDL	DRH
QGSDL	SLH
ALPDL	ERH
DGLGL	ERH
DGVAL	RKH
TDEDL	VRH
GLADL	ERH
IRLDL	ERH
MNVDL	KRH
AVQDL	VKH
RQEDL	MKH
AEQDL	SRH
ATADL	TRH
TPLDL	RRH
GQSDL	LKH
D	RK

Figure 7C: Sequences of selected fingers:

Pools with a four-residue histidine spacing

3' ATTC 5'

-1 1 2 3	5 6 * - -
ARADL	RQHINRIH
VRHDL	RHINRTH
TNHDL	KNHINRIH
ARADL	MVHINRIH
ARADL	MRHINRIH
SNQVL	DAHINRIH
ARPDL	MRHINRTH
AR D	AR

3' AT AC 5'

-1 1 2 3	5 6 * - -
HKADL	RGHINRIH
SMGAL	RAHINSTH
GTQAR	ADHINRIH
MAQNL	RMHINSIH
SAGAR	QAHISTH
NRVDL	RKHINSIH
QMVDL	RKHINRIH
SAGNL	ALHIRIH
DN	

3' ACGA 5'

-1 1 2 3	5 6 * - -
SRRAL	MEHIRIH
SRRAR	MGHINSTH
SRRAL	RDHINSTH
SVRSR	DRHINRTH
AARNL	REHIRIH
TTRSR	KKHINSTH
GRRAL	AVHINSIH
QARAL	KEHIRIH
S RA	S

Figure 8A: Contact preferences of position '-1':

This figure combines and summarizes key information from tables 4 and 5. The gray region of each bar shows the number of times that the indicated base either i) was observed in a structural study contacting the indicated residue at position '-1', or ii) selected the indicated residue at position '-1' by (presumably) making contacts with it. The white region of each bar shows the number of times that the indicated base occurred in selection studies at the target location for position '-1' and either the selection failed or no consensus residue was obtained.

A sketch of a zinc finger subsite is provided at top right, with the most frequently contacted bases highlighted in gray. An arrow indicates the base preferentially contacted by position '-1'.

Figures 8B-8D:

Symbols are identical to figure 8A, except that figure 8B involves position '2', figure 8C involves position '3', and figure 8D involves position '6'.

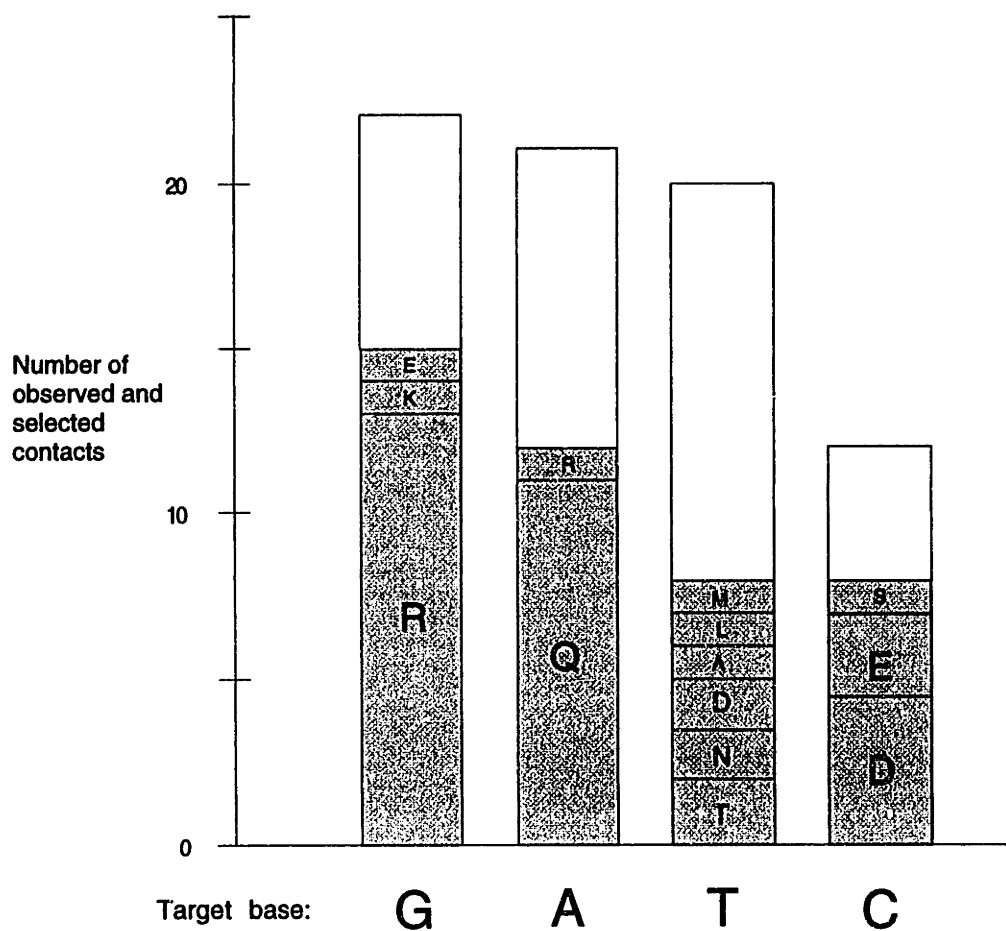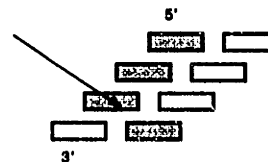# Figure 8A:
Contact preferences of position '-1':

Figure 8B:
Contact preferences of position '2':

Number of observed and selected contacts

Target base: **G** **A** **T** **C**
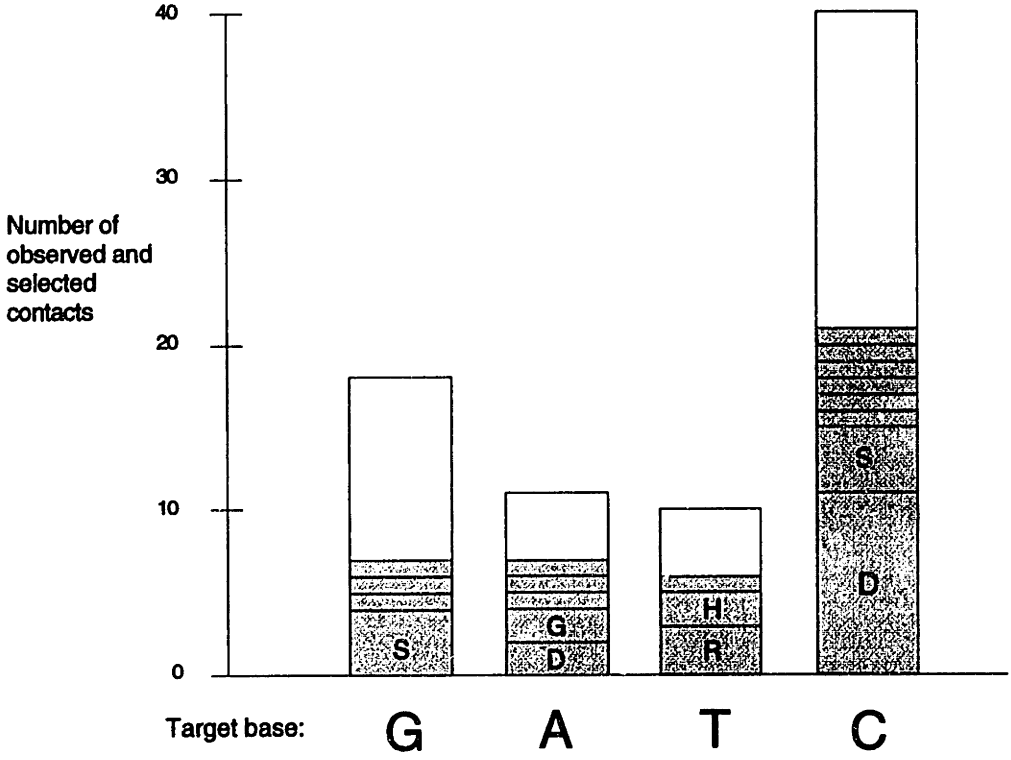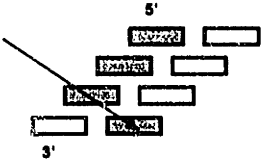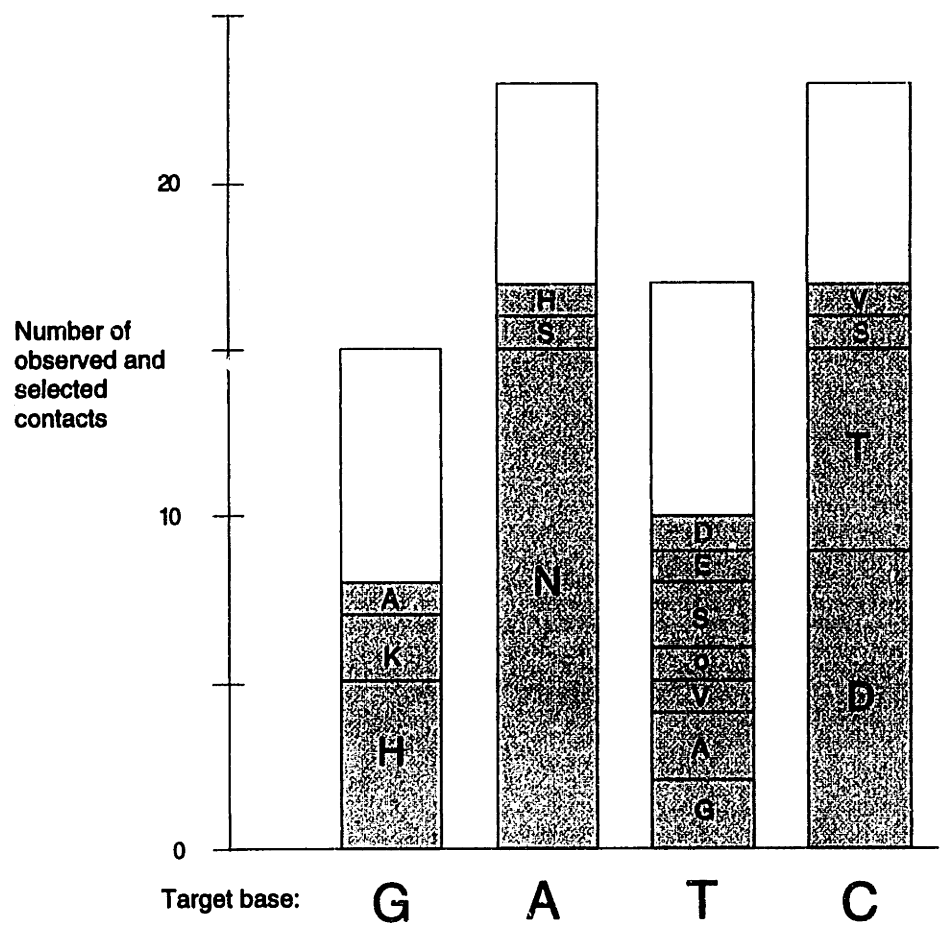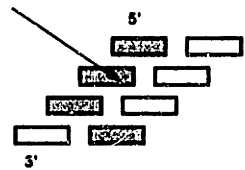
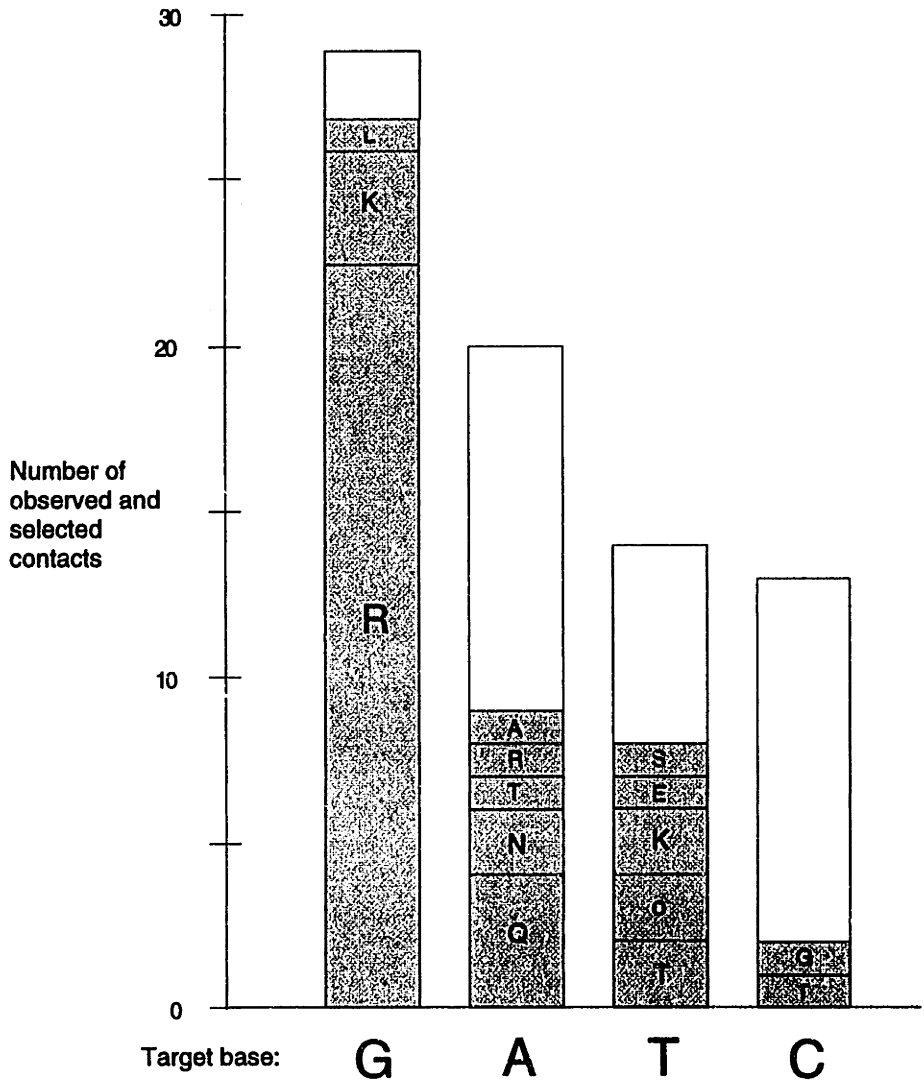**Figure 8C:**

Contact preferences of position '3':

Figure 8D:
Contact preferences of position '6':
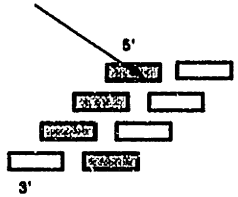
Figure 9:   Construction of p1212

As described in the materials and methods section, the phagemid vector p1212 was constructed in two steps:

First, the region of fd-tet.Zif[5] encoding the entire pIII-Zif fusion gene was amplified using PCR, digested with the enzymes Kpn 1 and Sac 1, and ligated with the large Kpn 1 / Sac 1 fragment of pBluescript KS+ (Stratagene).

Next, a synthetic DNA fragment encoding the 1212 gene was digested with Apa 1 and Pst 1 and used to replace the corresponding Apa 1 - Pst 1 fragment of pZif - pIII fusion gene, yielding p1212.

Figure 9:   Construction of p1212
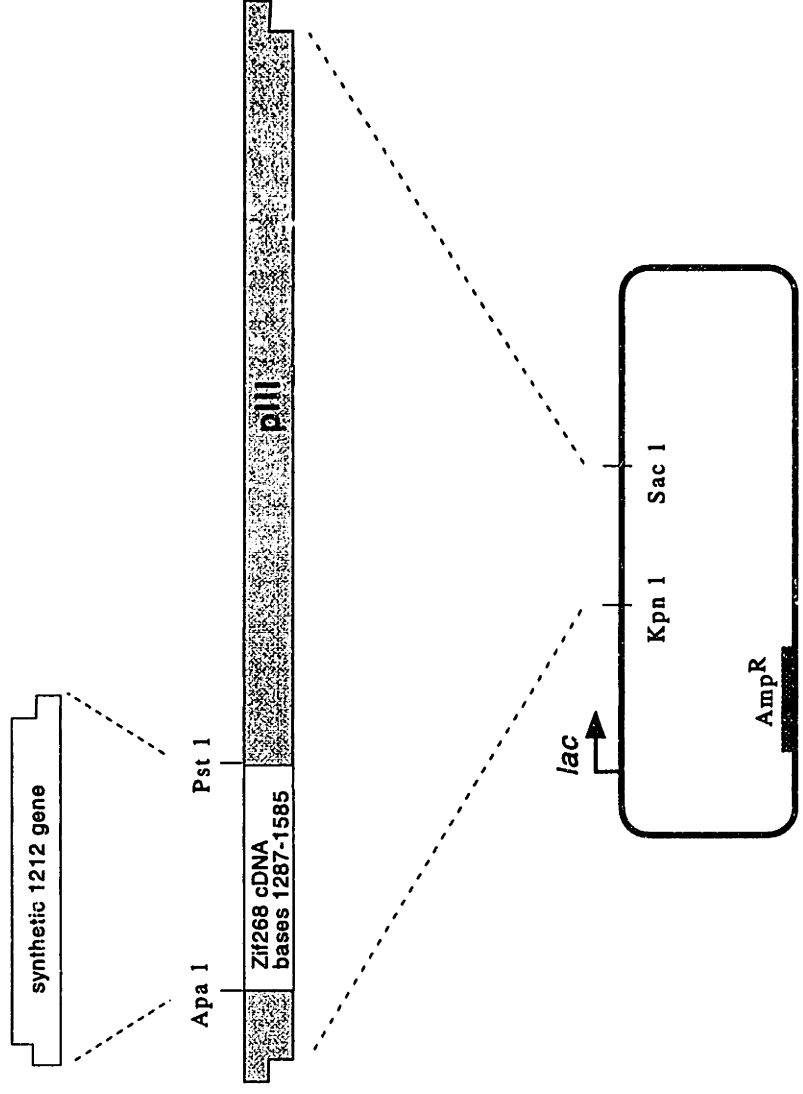
CHAPTER 5

INITIAL TESTS OF A COMBINATORIAL 'MIX AND MATCH' STRATEGY FOR MAKING MULTIFINGER PROTEINS WITH DESIRED SPECIFICITIES

Introduction:

The selections described in chapter 4 were performed as the first step of an attempt to make multifinger proteins with completely novel sequence specificities. Our plan was to link together fingers obtained in the chapter 4 selections to make multifinger proteins which could bind to sites in the '-38' to '-14' region of the HIV-1 promoter. The tandem, periodic binding arrangement of zinc fingers in multifinger protein-DNA complexes and the small number of interfinger contacts suggests that zinc finger-DNA recognition may exhibit a certain degree of functional modularity[1-5]. To the extent that this is true, it should be possible to 'mix and match' fingers to obtain proteins with desired specificities. The behavior of certain biological systems supports this possibility. For example, in *Drosophila*, alternate splicing of mRNA for the CF2 gene yields proteins with different combinations of fingers and alternate DNA-binding specificities[6, 7]. Also, biochemical studies have directly demonstrated that 'mix and match' design approaches may be used with certain combinations of fingers to make proteins with new DNA specificities[8]. It is clear that the zinc finger is not perfectly modular, since similar design efforts have also yielded proteins with unexpectedly weak and/or nonspecific binding[8, 9]. However, these studies suggest that 'mix and match' approaches which generate and test a large number of candidate finger combinations may provide an effective means for making zinc finger proteins with desired specificities.

In this chapter we describe our initial attempts to combine our selected fingers to make multifinger proteins with desired

specificities. Our experiments use a 'mix and match' approach to make multifinger proteins, which is similar to those described in previous studies[8, 9], except that instead of linking together individual fingers, we combine the finger pools obtained from chapter 4. This yields libraries of candidate finger combinations, which are then selected for binding to suitable target DNA sequences. Our initial tests of this method have yielded several multifinger proteins which bind specifically to their target DNA fragments, but only when the target fragments are immobilized on a solid support. We discuss these results, and describe further experiments and strategies for improving this approach.

Results:

Overview and library construction:

Figure 1 provides an overview of our strategy for these studies. In this example, the target sequence is ATATAAGCAG. In the first step (top), a single finger of a multifinger protein is randomized, and selections are performed to obtain pools of successful fingers for the relevant subsites: ATAT, TAAG and GCAG. These pools are then combined as shown, and the resultant three finger library is selected for binding to the entire target sequence. The 'randomization' and 'initial selections' steps in figure 1 were described in chapter 4. In this chapter we discuss the 'combination' and 'target selections' steps, and also describe subsequent characterization of the selected peptides.

These studies targeted a set of ten-base pair sequences from the HIV-1 promoter, which are shown in figure 2. The sites are numbered according to their distance from the -36 base of the HIV-1 promoter region, and also according to their orientation relative to the transcription start site (as '+' or '-'). Thus the sites for this study are: '+3', '+6', '+7', '+8', '+14' and '-1'. Throughout this chapter we will refer to these sites, the associated phage libraries, and the selected peptides by these numbers.

In the first step of these studies, we created three-finger libraries for each of our selection targets. To do this, we first used the polymerase chain reaction (PCR) to retrieve DNA segments encoding the appropriate pools of selected fingers, and then ligated these segments together in the appropriate order. Figure 3 lists the selection targets and indicates the pools used to make each library. Next, we expressed the resultant libraries of three-finger proteins on the surface of filamentous phage (data not shown), and produced, purified and stored phage libraries essentially as described in chapter 4. Titering of transformation mixes suggested that the complexity of each of our six libraries was approximately 1×10^7 clones. Sequencing of 39 unselected phage (7 or 8 per library) suggested that our library genes had assembled properly (all but one clone encoded three fingers in the correct arrangement), and that each library encoded a diverse collection of three finger proteins (all sequences were unique). However, single base pair deletions were observed in a substantial proportion of sequences (~50 %), and so the practical maximum diversity of our libraries was probably about 5×10^6. [Note that this is a maximum diversity because we cannot be sure that all clones of our

library contain unique finger combinations.] All deletions occurred in regions of our library encoded by PCR primers, and so we presume these deletions were due to problems with primer synthesis, as opposed to any sort of toxicity response by *E. coli*.

Selections:

In the next stage of this study, we selected each library for binding to its target site. We used protocols similar to those described in chapter 4, except that in the final two rounds of selection we raised the concentration of nonspecific competitor DNA in the binding reactions to 3.0 mg/ml. Over the course of nine selection cycles, phage plating efficiencies rose to levels of between 0.34 % and 1.66 % for all phage pools except the one selected using the '+8' sequence. Retention efficiencies for the '+8' pool had not risen after 7 selection cycles, and so this pool was discarded.

Characterization of selected phage:

As an initial method of characterizing our selected phage, we tested binding specificity by plating every combination of phage and target site used in these studies (table 1). We also included Zif phage as a positive control. Our selected phage yielded varying results in this test. While the '-1', '+14' and '+7' phage exhibited a high degree of specificity for their targets, the '+3' and '+6' phage discriminated poorly among several test sequences (table 1). Next, we sequenced six

clones from each phage pool. Each pool yielded a single distinct sequence (figure 4).

Binding studies of selected peptides

In preparation for more careful studies, we expressed and purified our selected peptides in free form. Genes encoding these peptides were subcloned into pET expression vectors (Novagen) and overexpressed, and the resultant peptides were isolated at ~99% purity. Our methods were similar to those described in chapter 3.

We then attempted to characterize our peptides using a gel shift assay. However, in a series of initial studies, we were unable to obtain any specific gel shifts. We tested a wide range of parameters, including magnesium chloride concentration (0 mM, 2 mM and 5 mM), potassium chloride concentration (0 mM, 60 mM, 120 mM and 240 mM), equilibration and electrophoresis temperature (20 C and 4C), and the addition of various additives, including streptavidin, dithiothreitol, polyethylene glycol, and 2xYT broth. [The last three additives were probable contaminants of our phage preps, which we tested just in case they might prove crucial for binding.] Throughout these studies we used maximum peptide concentrations in the range of 1 μM to 4 μM. Our default buffers for these studies contained 0.1 mg/ml BSA in either buffer 1 (60 mM potassium glutamate, 60 mM potassium acetate, 25 mM potassium phosphate, 2 mM magnesium chloride, 0.02 mM zinc sulfate, 0.5% triton x-100, 5% glycerol, pH 7.8) or buffer 2 (15 mM Hepes, 50 mM potassium chloride, 50 mM potassium glutamate, 50 mM potassium acetate, 5 mM magnesium chloride, 0.02 mM zinc sulfate, 0.1 % NP-40,

5% glycerol, pH 7.8). In positive control reactions throughout these studies, the Zif three-finger peptide typically shifted the vast majority of its target site, and apparent Zif dissociation constants, when estimated, yielded values in agreement with those published in previous studies[10-12] (from ~0.03 nM to ~4 nM, depending on buffer conditions) (data not shown).

We also attempted to characterize our peptides using a DNAse 1 footprint assay. A template comprising the HIV-1 LTR was prepared using standard methods, and we performed DNAse 1 protection studies using peptide concentrations ranging from ~1.0 to ~10 μM. However, despite testing a wide variety of conditions, we obtained no convincing footprints. As with the gel shift studies, Zif positive control footprint reactions (on the same template) typically yielded strong signals under all conditions (data not shown).

Our problems with the gel shift and footprint assays were puzzling, because most of our peptides exhibited activity in another type of test: a phage competition plating experiment. In this experiment, each phage was plated with its site and with each protein at concentrations of approximately 10 μM to 20 μM. For the '+7' and '+14' peptides, we observed specific disruption by each peptide of its associated phage/DNA complex (table 2). Furthermore, for the '+3' and '+6' peptides, we observed disruption of multiple phage/site combinations, mirroring the relative lack of specificity of these peptides observed in the phage platings. These results suggested the possibility that our peptides might be deriving binding energy from the solid support.

To test this possibility, we developed a DNAse 1 footprint protection assay which used our selection target DNA fragments immobilized on streptavidin - coated plates. We found that several of our peptides ('-1', '+7', and '+14') bound specifically to their target sites in this assay (figure 5). For example, peptide '+7' dramatically perturbed the digest pattern of its site, but not other sites. [Note that the '+7' protein lane of the '+14' site gel is a little light. However, this appears to be due to a global decrease in DNAse activity in this reaction because all bands are lighter, and the relative band intensities are similar to the no-protein control. In contrast, the bands in the '+14' protein lane on this gel are much lighter and also feature changes in the relative intensities of bands at the bottom of the gel.] Also, the peptide '-1' perturbed the cutting pattern of its site, but not the other sites. Finally, the '+14' protein perturbed both its site and the '-1' site, but not the '+7' or Zif sites. This cross - reactivity was not entirely unexpected, since the '+14' phage showed some evidence for binding to the '-1' site in low stringency phage plating tests (data not shown). As a positive control, the Zif peptide protected its site, but not other sites. [Note also, in this and similar experiments, Zif at high concentrations appeared to nonspecifically enhance DNAse 1 cutting efficiency (data not shown).]

The success of this footprint assay required the use of binding conditions similar to those used in our phage plating studies. Figure 6 shows a test of assay requirements using the '+7' peptide. Binding was observed only when the assay was performed using a streptavidin - coated support and a high loading of DNA target site on the support (equivalent to 8 nM in the reaction mix) (left gel). Use of a lower DNA

loading density (equivalent to 0.01 nM DNA in the reaction mix), failure to pre-coat the wells with streptavidin, or the use of streptavidin-bound target DNA fragments in the solution phase all failed to yield protection (remaining three gels). In tests with our other peptides, we observed similar behavior: each peptide required high levels of support loading in order to perturb the DNAse digest patterns in a site - specific fashion. [Note: in control tests, the Zif peptide protected its site under all tested conditions.]

The observation of DNA-binding specificity and a support-loading dependence provides limits on the types of mechanisms which could be governing our site-peptide interactions. One possibility is that our peptides aggregate into higher order complexes, and that these complexes form the DNA-binding species in our studies. In this scenario, the loading effect could be explained by a requirement for the interaction of each peptide aggregate with multiple support - bound sites in order to achieve efficient binding. Consistent with this possibility, peptides '+3' and '+7' aggregate in native gel electrophoresis, although the other peptides do not (data not shown). Another possibility is that each individual peptide interacts with multiple sites. One observation consistent with this scenario is the fact that the DNAse 1 protection effects of each peptide cover a relatively limited portion of its target site, relative to the Zif protection of its site (figure 5). If large protein aggregates were positioning themselves over the target sites, one would expect to see the opposite result (more extensive protection).

Possible future studies:

Given these results, future experiments could take one of several broad directions. One set of possibilities involves attempts to improve our selected peptides. Each peptide exhibits some affinity and specificity for its target site, and so it may be reasonable to attempt to enhance peptide affinities and specificities to higher levels. For example, a fresh set of mutations could be introduced into our selected peptides using either combinatorial cassette mutagenesis or error-prone PCR, and the new libraries could be reselected using affinity methods. Such an approach could be useful if our peptides dock against their sites in a fundamentally correct arrangement, but make poor DNA contacts. The key drawback to this approach, however, is that it would probably not work if our current set of peptides bind to their targets in fundamentally novel arrangements (for example, if they interact with the single - stranded regions or ends of the target fragments). If this is the case, we might obtain peptides with improved binding characteristics for their target fragments, but we would not expect these peptides to recognize their target sequences in other contexts.

Another set of possibilities involves repeating our selections using new conditions designed to provide a better outcome. In particular, it may be useful to decrease the amount of nonspecific competitor DNA in the binding reactions. A remarkable result of these studies is the fact that all selected peptides bind to their sites only when the sites are displayed on a solid support. This suggests that some common feature of our conditions may have driven the selections towards this unusual outcome, and it is possible that our competitor DNA concentrations were too high to permit the selection of any

peptide which interacted with DNA in a normal fashion. Successful selections in our other studies (chapter 3) have used competitor concentrations up to 50-fold lower than those described in this chapter, and so it might be reasonable to reselect our libraries using these reduced competitor concentrations.

Another possible precaution might be to use different site immobilization strategies in alternate selection cycles (for example, streptavidin - coated microtiter wells in odd cycles, and dynabeads (Dynal) in even cycles). Such a strategy would make it much more difficult for a peptide to be selected for any property other than its specific affinity for the target DNA sequence.

Finally, our results suggest that it may be reasonable to test the activity of any affinity - selected phage using a solution - phase assay prior to further characterization of the corresponding peptides. The phage gel shift assay, for example (described in chapter 2), could quickly screen out peptides which require the solid support for efficient site interaction. We note that, when tested, none of our selected phage bound efficiently to their target sites in this assay.

References:

1.   Houbaviy, H.B., *et al.* Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. Proc. Natl. Acad. Sci. U.S.A., 1996. **93**: p. 13577-13582.

2.   Fairall, L., *et al.* The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature, 1993. **366**: p. 483-487.

3.   Elrod-Erickson, M., *et al.* Zif268 protein-DNA complex refined at 1.6A: a model system for understanding zinc finger-DNA interactions. Structure, 1996. **15**: p. 1171-1180.

4.   Pavletich, N.P. and C.O. Pabo. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. Science, 1991. **252**: p. 809-817.

5.   Pavletich, N.P. and C.O. Pabo. Crystal Structure of a Five-Finger GLI-DNA Complex: New Perspectives on Zinc Fingers. Science, 1993. **261**: p. 1701-1707.

6.   Gogos, J.A., *et al.* Sequence Discrimination by Alternatively Spliced Isoforms of a DNA Binding Zinc Finger Domain. Science, 1992. **257**: p. 1951-1955.

7.   Hsu, T., *et al.* Multipl Zinc Finger Forms Resulting from Developmentally Regulated Alternative Splicing of a Transcription Factor Gene. Science, 1992. **257**: p. 1946-1950.

8.   Desjarlais, J.R. and J.M. Berg. Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. Proc. Natl. Acad. Sci. U.S.A., 1993. **90**: p. 2256-2260.

9.   Choo, Y., I. Sánchez-García, and A. Klug. *In vivo* repression by a site-specific DNA-binding protein designed against an oncogenic sequence. Nature, 1994. **372**: p. 642-645.

10.   Greisman, H.A. and C.O. Pabo. A General Strategy for Selecting High-Affinity Zinc Finger Proteins for Diverse DNA Target Sites. Science, 1997. **275**: p. 657-661.

11.   Rebar, E.J. and C.O. Pabo. Zinc Finger Phage: Affinity Selection of Fingers with New DNA-Bindig Specificities. Science, 1994. **263**: p. 671-673.

12.   Jamieson, A.C., S. Kim, and J.A. Wells. *In Vitro* Selection of Zinc Fingers with Altered DNA-Binding Specificity. Biochemistry, 1994. **33**: p. 5689-5695.

13.   Myers, G., *et al.*, eds. *Human Retroviruses and AIDS 1885.* . 1995, Los Alamos National Laboratory: Los Alamos, New Mexico.

Table 1:   Phage pool plating efficiencies with target and
            nontarget sites

The indicated combinations of phage pools and biotinylated
selection target sites (figure 4) were preincubated, applied to
streptavidin - coated microtiter wells, washed, eluted, and quantitated
via transduction of *E. coli* to ampicillin resistance. As a positive
control, Zif phage was also tested using its target site. The % of
applied phage recovered in the eluates is shown.

Both the phage pools and 10-base pair selection targets are
denoted by the numbers '+3', '+6', '+7', '+8', '+14' and '-1' as defined in
figure 2.

The plating efficiency of a phage pool with its selection target
(or Zif phage with its target site) is highlighted in gray.

The protocols used for this plating were essentially the same as
those described for the selection of phage libraries in chapter 4, except
that the <u>wash buffer</u> contained 100 mM NaCl, in addition to its other
components.

Table 1: Phage pool plating efficiencies with target and nontarget sites

Plating Efficiency (%)

Phage Pool	sites							
	+3	+6	+7	+8	+14	-1	zif	(no site)
+3	.46	.37	.18	.04	.22	.08	.01	.02
+6	.18	.15	.15	.09	.08	.1	.006	.007
+7	.006	.006	.42	.005	.008	.007	.006	.007
+14	.004	.003	.004	.003	.32	.007	.003	.004
-1	.003	.005	.005	.003	.004	.11	.002	.009
zif	.002	.003	.002	.003	.0006	.001	.68	.003

indicates plating efficiency of phage with its selection target site

Table 2:   Phage plating efficiencies in the presence of competitor peptides

The indicated phage pools were preincubated with their corresponding biotinylated selection target sites and approximately 10 - 20 μM of the indicated competitor peptides.   These mixes were then diluted five-fold, applied to streptavidin - coated microtiter wells, equilibrated, washed, eluted, and quantitated via transduction of *E. coli* to ampicillin resistance.   The % of applied phage recovered in the eluates is shown.

The competition of each phage pool by its corresponding peptide (or Zif phage with its peptide) is highlighted in gray.

The protocols used for this plating were essentially the same as those described for the selection of phage libraries in chapter 4, except that the binding reactions contained no nonspecific competitor DNA.

Table 2: Phage plating efficiencies in the presence of competitor peptides

Phage Pool	Plating Efficiency (%)							
	competitor peptide						(no competitor)	(no site)
	+3	+6	+7	+14	-1	Zif		
+3	.003	.05	2.42	1.23	.79	2.41	1.12	.01
+6	.005	.005	.78	.13	.16	1.75	.13	.02
+7	.01	.01	.003	.54	.11	1.47	.1	.006
+14	.007	.01	.31	.004	.18	.26	.12	.009
-1	.26	.26	1.14	.69	.25	1.07	.44	.02
Zif	.57	.41	.7	.57	.57	.007	.62	.002

indicates competition of phage by its corresponding peptide

Figure 1:   Experimental strategy

This figure provides an overview of our experimental strategy for making multifinger proteins with desired specificities.   As shown at top, the first step of this approach is the randomization of an individual finger in a multifinger protein and the selection of the resultant library for binding to suitable four - base pair sequences. These experiments were described in chapter 4.

Next, we retrieve DNA segments encoding pools of selected fingers (pools 'X', 'Y' and 'Z'), and ligate these segments together to make genes for libraries of three finger proteins (the 'combination step').   During this ligation, we use asymmetric overhangs to ensure that the DNA segment encoding each finger pool is ligated into a unique position in the three-finger gene.

Finally, we select the resultant libraries for binding to the composite target site.

Figure 1: Experimental strategy
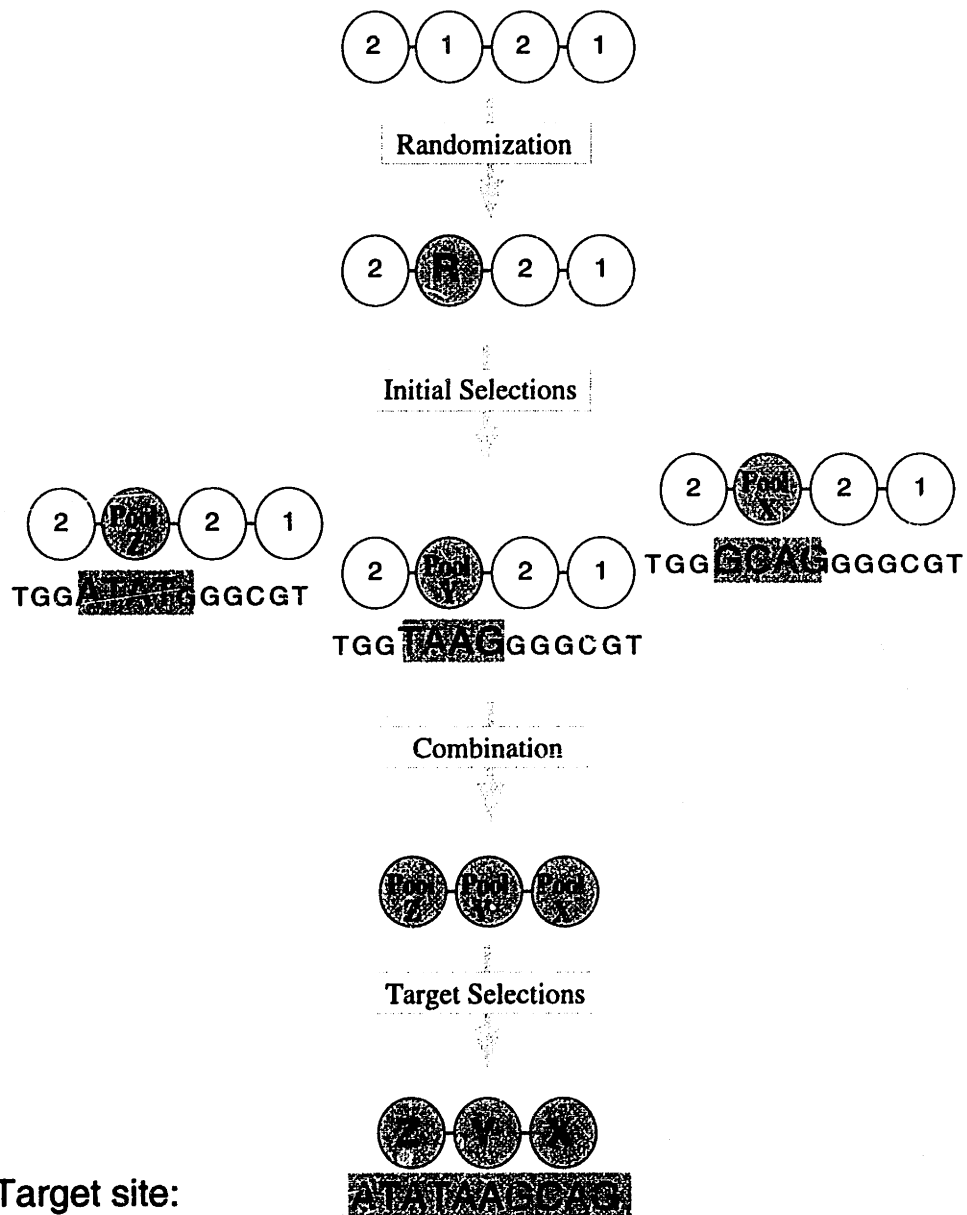
Target site:

Figure 2:   Selection   targets

Each of our six target sites ('+3', '+6', '+7', '+8', '+14' and '-1') is shown aligned over its location in the relevant region of the HIV-1 promoter. The '+3', '+6', '+7', '+8' and '+14' targets are derived from the top DNA strand and the '-1' target is derived from the bottom strand. The HIV-1 sequence is from Myers *et al.*[13].

The complete sequences of the target DNA fragments used for these selections are as follows (target sites are indicated by lines):

Target		Sequence	
+3	5'	CCAT<u>GCTGCATATA</u>AGCCTGAAC-b GGTACGACGTATATTCGG	3'
+6	5'	CCC<u>TGCATATAAGC</u>AGCCTGAAC-b GGGACGTATATTCGTCGG	3'
+7	5'	CCC<u>CATATAAGCA</u>GCCCTGAAC-b GGGGTATATTCGTCGGG	3'
+8	5'	CCC<u>ATATAAGCAG</u>CTCCTGAAC-b GGGTATATTCGTCGAGG	3'
+14	5'	CCAA<u>GCAGCTGCTTTT</u>CCTGAAC-b GGTTCGTCGACGAAAAGG	3'
-1	5'	CCTA<u>TATGCAGCAT</u>CTCCTGAAC-b GGATATACGTCGTAGAGG	3'

Figure 2: Selection targets

Target **Sequence**

+3   5'GCTGCATATA

+6   5'GCATAAGC

+7   5'CATAAGCA

+8   5'ATATAAGCAG

+14   5'GCAGCTGCTT

-1   TACGACGTAT5'

HIV-1 LTR

-36   ATGCTGCATATAAGCAGCTGCTT
      TACGACGTATATTCGTCGACGAA   -14

Figure 3: Finger pools used for library construction

This figure shows the target sequences and the phage pools used as source of fingers for the construction of each library. Overlines highlight the locations of overlapping phage pool target sites in each target sequence. The phage pools are identified using the four - base designations described in chapter 4.

Figure 3: Finger pools used for library construction

Target sequence

+3   5'|G C T G C A T A T A|

+6   5'|G C A T A T A A G C|

+7   5'|C A T A T A A G C A|

+8   5'|A T A T A A G C A G|

+14   5'|G C A G C T G C T T|

-1   5'|T A T G C A G C A T|

pools used for library construction
(from chapter 4 selections)

|GCTG| |GCAT| |TATA|

|GCAT| |TATA| |AAGC|

|CATA| |ATAA| |AGCA|

|ATAT| |TAAG| |GCAG|

|GCAG| |GCTG| |GCTT|

|TATG| |GCAG| |GCAT|

Figure 4:   Sequences of selected proteins

Six individual clones were picked from each selected phage pool and sequenced. For each pool, all six clones yielded a single DNA sequence, and the derived protein sequences for each pool are shown. Boxes and numbers ('-1', '1', '2', '3', '5' and '6') mark the positions of each protein which were randomized in the initial libraries (in chapter 4). Positions which were encoded using lesser amounts of variation are indicated by dashes. The locations of the optional asparagine insertions are indicated by asterisks, and if no asparagine was recovered there is a gap in the sequence.

Figure 4: Sequences of selected proteins

Target   Selected sequence

Target	Selected sequence
+3	PYACPVESCDRRFSRPLDLRIHI RIHTGQKPYACPVESCDRRFSSGQDLQKHI RIHTGQKPYACPVESCDRRFSHVDTLERHINSIHTG
+6	PYACPVESCDRRFSTHNMRAQHI RTHTGQKPYACPVESCDRRFSDRNGLDVHI RIHTGQKPYACPVESCDRRFSTGSDLGKHI RIHTG
+7	PYACPVESCDRRFSAKRARRDHINRTHTGQKPYACPVESCDRRFSQIGTLLQHI RIHTGQKPYACPVESCDKRYSSNQSLIDHINRIHTG
+14	PYACPVESCDRRFSTIHDRERHINSIHTGQKPYACPVESCDRRFSAPIDLARHI RIHTGQKPYACPVESCDRRFSTATDLGKHI RIHTG
-1	PYACPVESCDKRYSLLGDLRRHINSTHTGQKPYACPVESCDRRFSLADDLTRHI RIHTGQKPYACPVESCDRRFSTNDNLRCHI RIHTG

Box position markers: -1123 56

Figure 5:   Specificity of peptides in DNAse 1 protection assay on a
solid support

The indicated (biotinylated) selection target sites were bound to streptavidin - coated wells of a microtiter plate. Proteins were added (approximately 2 µM - 10 µM), and equilibrated with binding site. A limited digest with DNAse was then performed, digestion products were separated on a denaturing polyacrylamide gel, and the gel was exposed using a phosphorimager (molecular dynamics). Arrows indicate lanes which show the interaction of a site with its corresponding peptide. Due to space considerations, the '+' and '-' prefixes have been left off the numbers which refer to the proteins. Note also that parts of each gel have been deleted in order to juxtapose the 'no protein' and '1' lanes.

Procedure: The DNAse 1 protection assay was performed within an anaerobic chamber ($\leq$1 ppm O_2), using deoxygenated buffers, as follows. Wells of a probind plate were coated with streptavidin and blocked essentially as described in the appendix to chapter 3. 30 µl of DNA solution was then added (8 nM of the indicated selection target site (labeled with P-32) and 0.02 mg/ml BSA in buffer 1), and the plate was agitated on a rotary shaker [Labline model # 4625, setting 6 (600 rpm), 60 min]. DNA solutions were then removed from the wells, and replaced with 30 µl aliquots of protein solution (2 to 10 µM of the indicated peptides, 0.02 mg/ml BSA in buffer 1). Protein solutions were agitated in wells [Labline model # 4625, setting 6 (600 rpm), 60 min]. Then, 15 µl of DNAse solution was added (10 ng/µl DNAse (Worthington) in 10

mM Hepes, 1 mM MgCl2, 1 mM CalCl2, 0.1% NP-40, and 20 µg/ml BSA, pH 7.8); reactions were incubated 30 seconds at room temperature, and 45 µl of stop buffer (USB) was added to each sample.

Samples were heated (10 min, 85C), loaded onto 20% denaturing polyacrylamide gels, electrophoresed (350 V, 3.5 hr), dried, and autoradiographed (Molecular dynamics).

Figure 5:   Specificity of peptides in DNAse 1 protection assay on a solid support
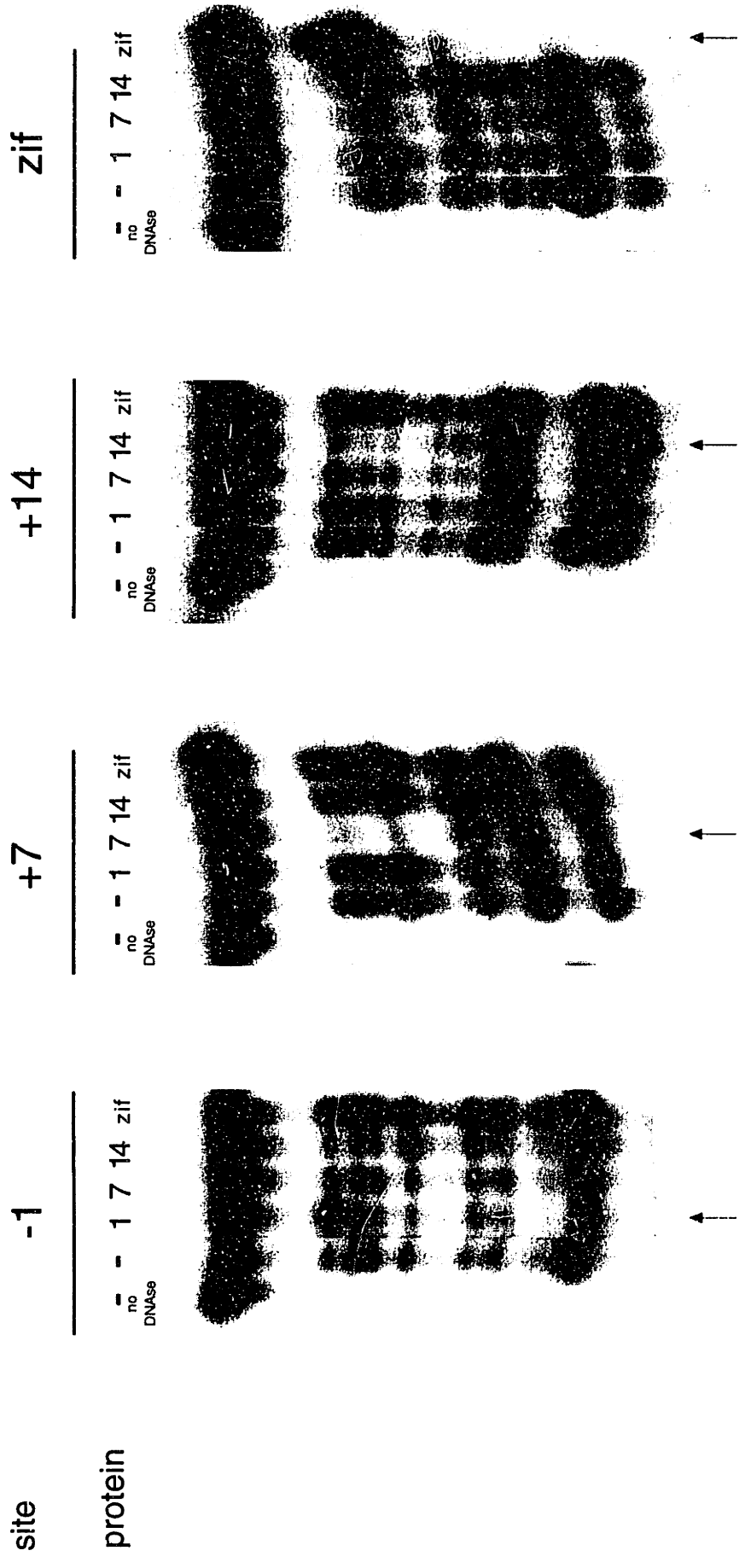
Figure 6:   Test of DNAse 1 protection assay requirements for peptide '+7'

In the leftmost gel, peptide '+7' was assayed for binding activity to its site and the Zif site essentially as described in figure 5. Protection was observed in the indicated lane.

In the left center gel, peptide '+7' was assayed for binding activity exactly as for the leftmost gel, except that the streptavidin - coated wells were loaded with 80-fold less '+7' site and Zif site.   No protection was observed.

In the right center gel, peptide '+7' was assayed for binding activity to its site and the Zif site in a standard footprinting reaction. Sites were not attached to a solid support.   Other details were essentially as described for the leftmost gel.

In the rightmost gel, peptide '+7' was assayed for binding activity to its site and the Zif site in a standard footprinting reaction, except that a molar excess of streptavidin (~1µM, in solution) was added to each reaction.   Sites were not attached to the solid support, but should have been completely complexed with streptavidin through their biotin moiety.   Other details were essentially as described for the leftmost gel.

Figure 6: Test of DNAse 1 protection assay requirements for peptide '+7'

conditions:
solid phase,
8 nM site
solid phase,
0.01 nM site
solution,
8 nM site
solution (+ streptavidin),
8 nM site

'+7' protein: - + - + - + - +

site:
(no DNAse)
+7 zif +7 zif +7 zif
+7 zif +7 zif +7 zif
(no DNAse)
+7 zif +7 zif +7 zif
+7 zif +7 zif +7 zif

protected
lane