Redesigning Specificity in Miniproteins

by

Christina Marie Taylor

B.S. Chemistry
University of Missouri – Rolla, 1999

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

FEBRUARY 2006

Signature of Author .....
Christina M. Taylor
Department of Biology
September 8, 2005

Certified by .....
Amy E. Keating
Assistant Professor of Biology
Thesis Supervisor

Accepted by .....
Stephen P. Bell
Chairman, Department Committee on Graduate Students

Redesigning Specificity in Miniproteins

by

Christina M. Taylor

Submitted to the Department of Biology
on September 8, 2005 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Biology

ABSTRACT

This work focuses on designing specific miniprotein interactions using computational models and then testing these designs with experiments. Miniproteins are small, autonomously-folding proteins that are excellent for testing protein designs because they can be chemically synthesized and computationally modeled. Despite their diminutive size, miniproteins are used as minimal models to discern important features, such as folding and interaction specificity, in natural proteins.

A 21-residue $\beta\beta\alpha$ homotetramer miniprotein (BBA) was computationally redesigned to interact as a heterotetramer. Protein design calculations revealed a large/small pattern of hydrophobic residues in the core and charge complementarity on the surface as a mechanism for attaining heterospecificity. Solution studies showed the designed protein is a tetramer and interacts in the same stoichiometry as its parent homotetramer. The x-ray crystal structure of the heterotetramer revealed a structure very close to the designed structure with near-perfect prediction of core side-chain packing. In a second round of design, the BBA heterotetramer was stabilized to near-native stability.

Next, the coiled-coil region within the Bcr (breakpoint cluster region) oligomerization domain was used to probe antiparallel versus parallel helix-orientation specificity in coiled coils. Based on the Bcr sequence, it is unclear why the oligomerization domain has an antiparallel orientation preference. The isolated Bcr coiled-coil region adopts an antiparallel orientation, so the orientation preference must be encoded in the Bcr coiled-coil sequence itself. Coiled-coil statistics and parallel and antiparallel model structures revealed an alanine and glutamate in the Bcr core as candidates that may be important for helix-orientation specificity. Both residues were mutated to leucine, a common core residue in parallel coiled coils. Based on solution studies of the mutant, both alanine and glutamate play an important role in oligomerization specificity, while glutamate may also be important for orientation specificity in Bcr.

Finally, interaction partners to the Bcr oligomerization domain were computationally designed to act as dominant negative inhibitors. Four interaction partners were designed using different design techniques and energy functions. The inhibitors were expressed in *E. coli* and tested in a pull-down assay.

Thesis Supervisor: Amy E. Keating
Title: Assistant Professor of Biology

DEDICATION

To my parents

ACKNOWLEDGMENTS

The work in this thesis would have been impossible without the help of my advisor, Amy Keating. Amy has been a great advisor, teacher, and role-model; in all of these roles her enthusiasm is always evident. I consider myself very fortunate to have had the opportunity to work in her lab. I appreciate all her efforts to help me in my scientific development. I would also like to thank my thesis committee: Bob Sauer, Barbara Imperiali, Jon King, and Krishna Kumar. My committee has given me many helpful suggestions during my time at MIT. Without funding from Amy's NIH grant (GM67681), National Science Foundation Pre-doctoral Fellowship, and the Anna Fuller Cancer Fellowship, I would not have been able to undertake this work.

The Keating Lab has been an enjoyable and productive environment in which to work. The many insightful discussions and comments during my time in lab have been very helpful. Much of the Bcr expression and purification would not have been done without my UROP, Devdoot Majumdar. Dev also did some of the early work on the Bcr inhibitor design. Over Dev's two-year tenure in the lab, I think I learned as much from him as he did from me! Nora Zizlesperger and I had a good collaboration on the BCR inhibitor project. Jeremy Fisher has helped the lab run very smoothly and has done a wonderful job picking up the Bcr inhibitor pull-down assay. I particularly appreciate him taking over my safety officer responsibilities. Gevorg Grigoryan has been extremely helpful in teaching me how to use some of his protein design code. Shaun Deignan has been helpful in nearly every aspect of lab life, from letting me borrow his TCEP to keeping the computers up and running. Emiko Bare was great company in the wet lab while I was doing experiments, and I'll miss James Apgar's amusing conversations. Thanks to Xioaran Stowell for many helpful discussions about research and post-docs.

During my time at MIT, I was also fortunate to have collaborations and interactions with many people outside the Keating Lab. Mayssam Ali and I had a wonderful collaboration on the BBA project, and Mayssam taught me peptide synthesis. Diren Pamuk helped me with peptide synthesis, and Ginevra Giorgio and I had an interesting collaboration on specificity flouro-peptides in coiled coils. Keith Joung taught me molecular biology, and Scot Wolfe taught me phage display. Koli Taghizadeh, Elizabeth Oakes, and John Newman helped me with LCMS. Debby Pheasant helped me put together numerous AUC cells. Bob Grant kept the anaerobic chamber running, enabling me to do nearly all the experiments in Chapter 3. The Bell, Walker, Baker, and Sauer labs allowed me borrow many materials, particularly during my early days in the lab. The Bell and Walker labs were a lot of fun to have on the sixth floor.

Over the 23 years that I have been in school, I have had many influential science teachers, particularly Mr. Richman, Mr. Westover, and many professors at University of Missouri - Rolla. Dr. Bellone was brave enough to allow me to work in his laboratory as a high school student, and Dr. Ercal, Dr. Rath, Dr. Leventis, Dr. Hogle, and Dr. Wiley were kind enough to allow me to work in their labs during my undergraduate years.

Friends have made my graduate school experience very enjoyable. Xu and Mike Simon have been great friends throughout my time here. Kendra and Caleb Hug were great neighbors and have been wonderful friends. The Ashdown and Eastgate Bible studies and the Graduate Christian Fellowship provided me with much support and many prayers during my time at MIT.

I owe a debt of gratitude to my mom and dad, Karen and Thomas Collins. Without them, I would not be where I am today. They endured many science fair projects through high school and have always been supportive and encouraging in whatever I pursued. Thanks to my sister,

Katie, for being a good friend throughout my time here. The rest of my family, particularly my grandmas, were also very supportive.

Last, but not least, I'd like to thank my husband and best friend, James Taylor, for all his love and support during my graduate school career. He is the best thing that has happened to me, and I look forward to our new life together in Missouri.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

Introduction

This chapter provides an overview of progress in protein design with an emphasis on issues relating to specificity. In some cases, addressing specificity has been critical to success in protein design, and in others, it has been ignored. Developing better methods for addressing specificity will, however, be important for expanding the capabilities of protein design.

Early attempts at protein design are discussed in the first part of the introduction. Initially, proteins were designed for conformational specificity, such that the designed protein had native-like properties, adopted a defined fold, and did not aggregate. A protein lacking conformational specificity sometimes becomes targeted for protein degradation (*1*); this protein occupies several low-energy conformations, which often lead to molten globule states (Figure 1A). In proteins with conformational specificity, a single conformation's energy is lower than alternative states (Figure 1B), allowing a unique structure to form.

Although designing conformational specificity is still an important problem in protein design, scientists have recently begun focusing on designing more complex interactions by building on protein design techniques obtained through conformational studies. The second part of the introduction will focus on designing interaction specificity. Interaction specificity (Figure 1C & D) is defined as interaction with one protein but not a similar protein and is driven by one protein complex having a lower energy than like complexes. Protein-protein interactions are crucial to many biological processes, including transcription, viral entry, and signaling, making interaction specificity a key to cellular life (*2*).

CONFORMATIONAL SPECIFICITY

Conformational specificity was an early problem tackled by protein design. Some of the first designed proteins, and even some proteins designed recently, failed to adopt native-like structures. It is not unusual for designed proteins to form molten globules or aggregate due to

loosely packed hydrophobic cores (*3*). Successful strategies for attaining conformational specificity include incorporating more specific packing interactions in the core (*4-15*), considering forces beyond van der Waals (*16*), and modeling alternative states implicitly to prevent aggregation (*17-19*). These principles provide positive and negative design elements that increase the energy gap between desired and undesired states.

Core packing and conformational specificity

DeGrado and co-workers found core packing to be a critical element, but not sufficient, to attain conformational specificity in a manually designed helical bundle. Linking together elements of secondary structure and incorporating general properties thought to be important for conformational specificity, such as helix propensity, hydrophobic/polar patterning, etc., were insufficient to form a unique, well-folded structure in the α_2B design protein (*6*). Although α_2B associated as a stable four-helix bundle, it adopted a molten globule conformation due to the many different topologies the structure could adopt (*4*). Conformational degeneracy was decreased by introducing steric complementarity into α_2C, the next generation design. The α_2C peptide had many qualities of a native protein, such as a cooperative thermal transition at low temperatures, but had a molten globule state at higher temperatures (*8*). With the α_2C design, the molecule could still adopt two different topologies. By altering two additional interfaces, DeGrado and co-workers introduced a metal-binding site to serve as a negative-design element against alternative states. Even without the metal ions present, α_2D (Figure 2A) had all the hallmarks of a native protein (*4, 5, 9*). However, α_2D was less stable than α_2B (*4*). To gain conformation specificity, alternative states had to be destabilized by utilizing steric matching and incorporating a metal-binding site.

Core packing is important for obtaining conformational specificity, but is difficult to design manually. Desjarlais and Handel developed a program called ROC (repacking of core) to systematically evaluate van der Waals energies of many possible sequences using a genetic algorithm (*11*). Core positions in 434 cro (*11*) (Figure 2B) and ubiquitin (*15*) (Figure 2C) were redesigned using ROC. Within the accuracy of basic low-resolution biophysical tests, the variants were still able to adopt well-defined conformations (*11, 15*). Some of the 434 cro designs were as thermally stable as wild type, but one mutant was unfolded under equivalent conditions and was less stable than a variant with only leucine in the core (*11*). Most of the ubiquitin mutants were more stable than random hydrophobic core variants, but none were more stable than wild-type ubiquitin.

The low stabilities seen in the ubiquitin mutants were puzzling. Johnson and Handel solved the NMR structure of some ubiquitin mutants designed by ROC and found the low stability was due to rotamers in less statistically favorable conformations compared to wild type (*13*). Another ubiquitin mutant was in slow exchange between two conformations due to rotamer strain caused by rare rotamers populating multiple conformations in the structure (*15*). Rotamer strain may be caused by the uniform reduction in van der Waals radii, which may have led to overpacked cores, or by the lack of a torsional strain term in the ROC design program. The addition of a rotamer strain term or evaluating the energies with 100% van der Waals radii also may be important to achieving conformational specificity.

Polar and charged interactions also affect conformational specificity

In the literature, the role of polar and charged residues in determining protein conformational specificity is mixed and is likely to be context dependent. Upon replacing a buried salt-bridge network with hydrophobic residues in the Arc repressor, Waldburger et al.

found conformational specificity was not compromised (*20*), whereas Bolon and Mayo found polar residues in *E. coli* thioredoxin core important for this purpose (*21*). Using a physics-based energy function, an automated design study by Koehl and Levitt demonstrated that energy terms beyond van der Waals forces, including electrostatics and environmental solvation energies, were important for conformational specificity (*16*). Fold-recognition software was used to test for the sequence's conformational specificity. The sequences were threaded onto a library containing the target fold and a large number of non-native folds, and the difference in scores for the target structure and other "decoy" structures was assessed. Each term of Koehl and Levitt's energy function was needed for specificity in design, demonstrating contributions from charged and polar residues are important in defining specificity.

Preventing protein aggregation

Protein design energy functions model physical forces that stabilize proteins, often by increasing the number of hydrophobic residues, for which there is no penalty in many energy functions in the absence of competing aggregation. However, greater hydrophobic content can lead to aggregation and loss of unique structure. Therefore, a negative design component is important to prevent extremely hydrophobic sequences.

The insoluble aggregated state is an important alternate state against which to design. The structure of a protein's aggregated state is not known and is impossible to explicitly model in a calculation. Rather than explicitly modeling the aggregated state, some groups attain conformational specificity by making large approximations to minimize hydrophobics on the surface that could cause aggregation or other undesired states. For example, hydrophobic residues would be excluded from possible amino acids at surface positions. Others have tried to

overcome this problem by adding terms to their energy function to crudely approximate the effect of various states.

The Mayo group used a technique called binary patterning to obtain conformational specificity. Exposed hydrophobic surface area was penalized by partitioning the protein into core, boundary, and surface, allowing hydrophobic residues in the core, charged and polar residues at the surface, and the combined core and surface set at the boundary. Binary patterning introduces negative design against aggregation without designing against explicit states. Not only does binary patterning potentially yield conformational specificity, it also significantly reduces the possible sequences through which to search (*19*). Using this methodology, Dahiyat and Mayo did full-sequence design on the ββα motif, and NMR studies revealed well-ordered secondary structure (Figure 2D).

The number of boundary positions is typically overestimated in the binary patterning method. To improve the assignment of residues to the core, surface, and boundary positions, Marshall and Mayo calculated the solvent accessibility of a protein structure with generic methyl acetylene side chains (*17*). At each site, positions were classified at different cutoffs based on their solvent accessibility. A set of engrailed homeodomain variants was designed with different solvent accessibility cutoffs and tested experimentally for stability and conformational specificity. Variants with fewer hydrophobic residues were destabilized, whereas too many hydrophobics caused aggregation (*17*). This study provided interesting insights into the trade-offs between adding more hydrophobics (and more stability) versus conformational specificity.

Like the Mayo group, the Harbury and Handel groups also had problems with aggregation when negative design against various alternative states was not employed (*18*). Havranek and Harbury modeled an aggregated state by placing their target in a medium with a

dielectric lower than water (*22*). To address aggregation, Pokala and Handel used a crude unfolded-state model with random amino acids threaded onto 13-mer protein fragments extracted from the PDB. Buried polar residues lacking a satisfied hydrogen bond and solvent-exposed hydrophobic residues were given a special reference state to further address conformational specificity, and the 13-mer unfolded state model was used for all other residue reference states (*18*). Although Pokala and Handel's work has not been experimentally tested, studies by Mayo, Handel, and Havranek used crude approximations to prevent aggregation and yield conformational specificity.

INTERACTION SPECIFICITY

Designing interaction specificity is more complex than designing conformational specificity, as both interaction and conformational specificity are required. Interaction specificity is determined by the relative thermodynamic stabilities of all possible protein-protein interactions. To obtain interaction specificity, forces stabilizing the desired protein-protein interaction and forces that destabilize all undesired competing states must be considered. The same general principles important for conformational specificity are also important for interaction specificity. Negative design has been successfully used for some interaction specificity problems where it was possible to model competing states explicitly. In other cases, interaction specificity has been achieved by using positive design alone.

Protein-Protein and Protein-Ligand Specificity

Several interaction specificity designs have been successful without explicitly incorporating negative design, only utilizing positive design to obtain stability and specificity. Naturally promiscuous proteins are optimized to bind many different targets. Each of the targets

have different binding surfaces, so optimizing the promiscuous protein for one target is unlikely to optimize interactions with other targets (Figure 3). Therefore, disregarding negative design in these calculations may be a reasonable approach. Because promiscuous targets are not optimized to specifically bind one target, stability is also increased as a result of optimization for a specific target.

The class I PDZ domain (Figure 4A) (*23*) and calmodulin (Figure 4B) (*24, 25*) are promiscuous proteins that bind with low specificity and high affinity to a number of targets. Reina et al. redesigned a class I PDZ domain to bind novel class I and II sequences without using negative design. Using the ORBIT protein design program, Shifman and Mayo optimized calmodulin's large hydrophobic binding region to increase its specificity for binding a peptide from the myosin light chain kinase (smMLCK) (*24*). Because models of undesired calmodulin complexes contained large hydrophobic clashes, specificity of calmodulin complexes is predominantly due to van der Waals clashes. Shifman and Mayo assumed suboptimal contacts would be produced between calmodulin and other targets by simply optimizing for smMLCK stability. Rather than using negative design to destabilize the undesired calmodulin complexes, the calmodulin-smMLCK complex was stabilized by targeting buried residues on calmodulin near smMLCK. Designed calmodulin's affinity for smMLCK was higher than that of the wild-type calmodulin, whereas affinity for other peptides were reduced by 1.5- to 86-fold (*24*). Binding specificity was further increased up to 155-fold by redesigning boundary and surface positions on calmodulin in a subsequent calculation (*25*).

In their calculations, both Reina et al. and Shifman and Mayo introduced negative design implicitly by increasing the weight of the protein-ligand interaction terms, relative to intramolecular ligand and protein interactions. Mayo and Shifman changed the distance-

dependent dielectric constant from 40r to 4r to improve long-range electrostatic interactions and introduced a β factor to weight the intermolecular calmodulin/target energies more heavily than intramolecular interactions (*25*). The Coulombic term became much more important when designing the boundary and surface, as salt bridges in calmodulin play a special role in target recognition. Decreasing the dielectric helped to increase the Coulombic energy contribution and the number of salt bridges selected. Part of the design process used by Reina relied on human intervention to choose design sites and sequences that formed inter- rather than intramolecular interactions (*23*). However, human intervention in picking the designed sequences that form intermolecular interactions is a subjective process where other types of negative design could also have been incorporated.

Negative design was also not necessary to redesign ligand-binding specificity of a non-promiscuous enzyme. Looger et al. successfully redesigned various sugar-binding proteins, including the ribose-binding protein (RBP) (Figure 4C), such that they bound specifically to trinitrotoluene (TNT), L-lactate, or serotonin (*26*). Although the enzyme was only optimized for a single target ligand, the enzyme was specific for the target ligand over the wild-type sugar and similar ligands. Negative design was not critical because the ligand had many steric restraints in the core, preventing similar ligands from binding with high specificity. The repulsive term in the Lennard-Jones potential was reduced by 35%, yielding a tightly-packed binding site. In addition, the distance-dependant dielectric of 8 for all Coulombic calculations and the explicit hydrogen-bonding potential yielded solutions with unique charge complementarity to the ligand.

Negative design may be necessary when redesigning proteins with high-affinity and specificity. Kortemme and co-workers redesigned specificity at the interface of DNase (colicin E7) and its tight binding inhibitor, immunity protein (I7), using a computational second-site

suppressor strategy to ensure the designed proteins would have specificity for each other and not the wild-type protein. Given partner A and B, the second-site suppressor design strategy is a negative design strategy that finds sequences for partner A in the complex that would destabilize the interface, but would be compensated for by redesigning the interface of partner B. Unlike Shifman and Mayo and Reina et al., negative design is more critical for the DNase (colicin E7) / I7 interface because both DNase (colicin E7) and the immunity protein (I7) are very specific and bind with high affinity to each other, suggesting a highly optimized binding interface. If the interface was merely optimized during design, sequences close to, or identical to, wild-type would probably be selected. Both van der Waals and electrostatic interactions were critical to obtaining the following two modes of specificity: i) polarity switch that replaced hydrogen bonding at the interface with hydrophobics ii) steric switch where steric packing was altered. The designed protein-protein interaction with the highest predicted affinity was found to have good specificity in *in vitro* biophysical and *in vivo* assays, and many of the predicted side-chain interactions were present in the designed protein x-ray crystal structure (Figure 4D) (*27*).

Hetero- vs. Homooligomeric Specificity

Explicit negative design has been used a great deal and found to be very important in designing hetero- and homooligomeric specificity. Many experimental studies have led to underlying positive and negative design "rules" for specificity that can be introduced into proteins, particularly coiled coils. A coiled coil, shown in Figure 5A, consists of two or more α-helices intertwined, containing a heptad repeat, HPPHPPP, where H is a hydrophobic residue, and P is a polar residue. The heptad repeat, denoted by letters **a-g**, is shown in helical wheel form in Figure 5B and C. Computational design programs have also independently found these "rules" and used them to introduce interaction specificity.

Interactions at **a**, **d**, **e**, **f**, and **g** coiled-coil positions can help define hetero- versus homooligmeric interaction specificity through both van der Waals and electrostatic interactions. In coiled coils, steric matching at **a** and **d** positions is a useful technique to give hetero- versus homooligomeric specificity (*28, 29*). Steric matching involves packing a small amino acid across from a large amino acid. A study by Tanaka and co-workers made a specific ABC heterotrimer by incorporating steric matching in the core (via Trp/Ala), charge complementarity at **e** and **g**, and an Ala at the **f**-position amino acid to destabilize certain states. Charge complementarity can be used as both a negative and positive design element for and against homo- or heterooligomers. The specificity of the Fos-Jun heterodimer (Figure 6A) is driven by electrostatic clashes at **e** and **g** (*30*) and two **a**-position lysines (*31*) that destabilize the Fos homodimer (*32*). Protein-protein interactions have been designed based on an understanding of specificity gained from natural homo- versus heterooligomeric interactions. O'Shea et al. used electrostatic clashes as an explicit negative design element to destabilize homodimeric states (Figure 6B) (*33*).

Steric matching and complementary charge motifs have been found using computation to design hetero- versus homooligomers. Using computational design, these techniques emerged as a means for introducing heterospecificity into the BBA homotetramer in Chapter 2. Summa et al. used charge complementarity to devise a very simple scoring function that rewarded +/-, +/+, and -/- interactions with -1, +2, and +3 scores, respectively, to computationally design an A_2B_2 heterotetrameric four-helix diiron bundle (*10*). A similar positive and negative design strategy was also used to design a heterotrimer (*34*).

Havranek and Harbury designed coiled coils to favor homodimers that did not cross hybridize and heterodimers that did not self-associate by explicitly designing against undesired

states in their calculations (*22*). The calculation also simultaneously optimized for stability. A genetic algorithm was used to maximize the transfer energy from the homodimer target state to the various competing states: heterodimer, unfolded, and aggregated. To include all of the states explicitly in the design calculation, they maximized the probability of occupying the desired state over undesired states using Boltzman probabilities. Havranek and Harbury found omitting the aggregated state yielded sequences with fewer charged residues, making the designed protein more prone to aggregation. If the unfolded and aggregated states were omitted, the sequences were computationally predicted to be specific, but unstable due to polar groups in the core. Further, if the homotetramer state was omitted from the heterotetramer design (and vice versa), calculations suggested that specificity would be lost. The top one-hundred sequences were clustered and examined for sources of specificity. The following four sources of specificity emerged from the clustering: steric complementarity between Trp and Gly side chains, poor packing of Leu at **a**, isoleucine at **g'** and **a'**, and Glu at position **d** favoring basic side chains at **e'** over hydrophobics. Interestingly, some sources of specificity seen in the selection are not found in nature. The designs were biophysically characterized and were found to be specific for their desired state, as well as stable with respect to the unfolded state.

Bolon et al. also explicitly designed against undesired states to change SspB from a homodimer to a heterodimer (Figure 4E) (*35*). Repacking the SspB structure in an asymmetric fashion, the ORBIT design code was modified to optimize the energy of the heterodimeric state over that of the two homodimeric states. The results revealed a steric matching pattern predicted to provide heterodimeric specificity. If SspB was designed only for stability without regard for heterooligomeric specificity, it formed heterodimers and homodimers. Designing for specificity

resulted in solely heterodimeric complexes (*36*). In the Havranek, Bolon, and BBA studies stability was compromised to obtain the desired specificity.

Explicitly representing certain competing states for hetero- and homooligomerization design appears to be very important for much the same reason negative design was important for Kortemme et al's second-site suppressor strategy. Many homo- and heterooligomers have interfaces that are already optimized for binding. Further, homo- and heterooligomeric states are very similar; sequences optimized to favor one state will likely favor the other state, making negative design against undesired states important to prevent the wild-type or nonspecific sequences (as in the experiment by Bolon et al.) from being selected. Designing homo- and heterooligomeric states makes negative design easier because the structures are already known and each state can be explicitly represented. Steric matching, charge complementarity, and hydrogen bonding all play important roles in favoring hetero- versus homooligmerization.

Oligomerization State Specificity

Most studies of oligomerization specificity have been done experimentally and not computationally. Oligomerization specificity determines whether a monomer, dimer, trimer, etc. is formed. Determinants of oligomerization specificity are still being extracted from new mutational studies. Ideally, alternative oligomerization states would be explicitly considered in any protein design. However, including all competing states is difficult and computationally intractable, particularly for coiled-coil helical bundles, due to the large number of very similar competing oligomerization states. In this section, three influences on oligomeric specificity in core coiled-coil residues will be discussed: hydrophobic, polar, and charged interactions within core residues of coiled coils.

Hydrophobic core packing

Hydrophobic core packing has been found to influence the oligomerization state of coiled coils. Minimization of cavity size in protein cores may be the driving force behind oligomerization state specificity in some proteins. Figures 6C and 7A illustrate how the location of Ala can lead to oligomerization specificity, consistent with minimization of cavities in the core (*37*). Mutating an alanine to a leucine in a coiled coil led to loss of oligomerization specificity in Chapter 3.

Steric clashes can also influence oligomerization state. Changing certain leucine residues in bZIP transcription factors to other hydrophobic residues can interfere with dimerization and DNA binding (*38, 39*). Biophysical characterization of certain non-functional GCN4 mutants has shown that hydrophobic packing can lead to different oligomerization states (*40-42*). The type of hydrophobic amino acids at **a** and **d** positions determine whether dimers, trimers, tetramers, or mixtures of oligomerization states form (*43*). For instance, isoleucine at **a** and leucine at **d** favor dimeric coiled coils, but isoleucine at **d** and leucine at **a** favor trimers. Crystal structures of a coiled-coil dimer (*44*), trimer (*40*), and tetramer (*43*) with these sequence motifs revealed that steric restrictions on packing were a negative design element, causing one oligomerization state to form while excluding others (Figure 8). β-branched amino acids at **d** could not pack into a dimeric structure, explaining why coiled-coil trimers form instead (*40, 41, 43*). The results of these studies have helped design specific oligomerization states (*34*).

Oligomerization specificity is a particularly hard problem to tackle computationally due to the large number of alternative states that must be considered. Changes in coiled-coil oligomerization state can inadvertently cause changes in helix-orientation preference. Two studies have successfully designed one oligomerization state, while using negative design to

explicitly disfavor others. In a study by Park and Keating, amino acids at **a** and **d** positions were targeted to design a coiled coil that had dimer over trimer specificity (or vice versa) by maximizing the energy difference between the two states ($\max f, f = -\frac{1}{2}E_d + \frac{1}{3}E_t$ or $\max f, f = -\frac{1}{3}E_t + \frac{1}{2}E_d$) (*45*). β-branched amino acids at the **d** position were found to favor trimers over dimers in the computational results, confirming the experiment by Harbury. In another study, computer-generated coiled-coil backbones with a right-handed superhelical twist and various oligomerization states were generated using a method developed by Francis Crick (*46*). Right-handed coiled coils have a repeating pattern with 11 residues, compared to 7 for common left-handed coiled coils. Sequences were optimized for oligomeric specificity and stability in the target conformation by i) expressing stability as the standard deviation from the mean of all possible sequences and ii) finding sequences that ranked high on the target backbone and low on other backbones by expressing the stability difference between states as a standard deviation. The designed trimer and tetramer were well-folded based on biophysical experiments. The crystal structure of the tetramer matched the designed structure well (*47*), but some differences were seen in the crystal structure of the trimer (*48*).

Core polar interactions

Polar interactions within protein cores have also been found to be very important in directing oligomerization state. About 20% of amino acids at **a** and **d** positions in coiled coils are charged or polar (*49*). Polar interactions can be manually designed to provide specificity. However, computationally selecting polar interactions is still very difficult, as hydrogen bonds within protein cores are challenging to model with current energy functions.

The position of Asn in the leucine zipper is critical to providing oligomerization specificity. Asn residues are often found in one of the **a** positions of dimeric, parallel coiled coils (*50*). Upon changing the **a**-position Asn in GCN4 to a leucine, the coiled-coil stability increases, but a tetramer is formed, rather than a dimer (*51*). As seen in previous studies, Asn decreases stability to provide oligomeric specificity. Using computational protein design, Park and Keating also found preference for Asn in the **a** position of dimers versus trimers (*45*). Another study moved Asn from **a**-position 16 to **a**-position 9 in the coiled coil and found the thermal denaturation curve switched from two-state to three-state. The additional state is a partially unfolded state that is a mixture of dimers and trimers (*52*). When Asn in GCN4 is changed to aminobutyric acid (Appendix A, Figure 2), dimers and trimers form due to very similar packing surfaces, as revealed by crystal structures of the two states (*53*).

Not all polar side chains promote oligomerization specificity in GCN4. Both dimers and trimers formed when Gln was substituted for Asn (*54*). The crystal structure of the Gln mutant was solved to explain the structural differences that resulted from the two very similar amino acids. In the dimer and trimer structures, Gln 16 makes different interactions, which must yield similar stabilities to enable both oligomerization states to form. Akey et al. substituted various polar amino acids, Asn, Gln, Ser, and Thr, into the **a** and **d** positions of GCN4 to determine if they yield oligomerization specificity (*55*). A threonine at **d** gave a specific trimer, and an asparagine at **a** gave a specific dimer in solution. However, all other combinations of Asn, Gln, Ser, and Thr at **a** and **d** did not yield oligomerization specificity. Four of the mutants formed trimeric x-ray crystal structures, despite three of these not having a specific oligomerization state in solution (*55*).

Using a different GCN4 variant, two studies by the Hodges group substituted all 20 amino acids in **a** (*42*) and the **d** (*41*) positions and found slightly different results compared to Akey et al. (*55*). Only a few polar residues at **a** and **d** positions were able to impart oligomerization specificity. In the experiment by Akey et al., Gln at **a** did not yield oligomerization specificity (*55*), but yielded trimeric oligomerization specificity in Hodge's studies (*42*). Thus, the influence of polar residues in the core on oligomerization may be context dependent.

Core charged residues

Charged residues in the coiled-coil cores have been found to impart oligomerization specificity. In fact, charged residues are found with greater frequency in coiled-coil dimers versus trimers (*56*). Higher-order oligomers impose a larger desolvation penalty on charged residues in the core due to the larger buried surface area at **a** and **d**. In a synthetic parallel coiled coil, a buried lysine in the **a** position, that could make interhelical contacts with a Glu at **g'** on the opposite helix (Figure 6D), provided dimeric oligomerization specificity (*57*). A buried charged group can also impart oligomerization specificity in a designed dimeric antiparallel coiled coil. An Arg at a **d** position and a Glu at the **g'** position on the opposite helix was sufficient to provide dimeric specificity on the designed peptide (*58*). At the **d** position, charged residues (Lys, Arg, Ornithine (Orn), Glu, and Asp) are able to specify the dimeric oligomerization state (*41*), whereas Lys, Orn, and Arg are able to provide dimerization specificity if they are present at the **a** position (*42*). The structure of Orn is shown in Figure 4 in Appendix A.

Like polar residues in the core, charged residues may also direct oligomerization in a context-dependant fashion. In Chapter 3, changing a core glutamate residue to leucine caused

the oligomerization state to change from a coiled-coil dimer to trimer. The Fos leucine zipper (Figure 6A) has an unusual core, containing Lys or Thr at four out of five **a** positions. The two lysine groups at **a** were substituted with norleucine (Figure 2, Appendix A), both independently and together, and the peptides were biophysically characterized (*31*). One of the mutated lysines changed the oligomerization specificity from dimer to a dimer-tetramer equilibrium, whereas the other lysine retained its dimeric state when mutated. However, if both lysines are changed to norleucine, a stable homotetramer forms.

Helix-Orientation Specificity

Coiled coils and helical bundles have been used as the primary means for studying orientation specificity, thus far. The helix orientation specificity problem in coiled coils is further complicated by oligomers with similar competing states. Understanding orientation specificity can aid in understanding the function and mechanistic details of a coiled-coil protein (*59*). Only recently has there been a large enough body of antiparallel coiled-coil crystal structures to obtain general structure and sequence information (*59*).

Current understanding of orientation specificity has mainly come from experiments. Computational design has not aided this part of the field yet, but offers a fruitful area of research. Three main aspects of antiparallel structure have been probed for their role in determining antiparallel orientation specificity: core hydrophobic packing, core polar and charged residues, and charged residues at **e** and **g** positions. From these studies, general principles that direct helix orientation have emerged and can be used to design antiparallel coiled-coils (*60-64*), but not predict orientation preference in natural coiled coils.

Hydrophobic core packing

Hydrophobic packing in the core of coiled coils has led to orientation preference. Parallel versus antiparallel orientations yield very different packing interactions at **a**, **d**, **e**, and **g** positions, with parallel coiled coils making **a**-**a**' and **d**-**d**' interactions and antiparallel coiled coils making **d**-**a**' and **a**-**d**' interactions (Figure 5B and C) (*59*). In parallel coiled coils, **a** and **d** residues align directly across from the opposite helix, whereas residues can align either directly across from the opposite helix or stagger to varying degrees in antiparallel coiled coils.

Steric matching (*65-68*) in the core, pairing a small residue with a large residue, is the major way packing can direct orientation preference. Alanine is found more often in the **d** position of antiparallel than parallel coiled coils (*50*). The relative position of alanine in the core can yield orientation specificity. Figure 7B and C illustrate how the position of alanines in the core can direct orientation preference (*66*). In Chapter 3, changing an alanine to leucine in the core of a natural antiparallel coiled coil did not affect orientation specificity, but changed the oligomerization specificity, however. Steric matching has been used to successfully control helix orientation in an antiparallel trimer by placing combinations of alanine and cyclohexylalanine at **a** on two helices and at **d** on the other helix (*68*). The position of the cyclohexylalanine was staggered on each of the helices, such that the layers of the three helices would have the pattern XAA, AXA, and AAX, where X is the cyclohexylalanine.

Core polar / charged residues

Buried polar residues in the coiled-coil core can control helix orientation, whereas charged residues in the core have not been found to have an effect. At the **e** and **g** positions, Oakley et al. designed a GCN4 variant with Lys on one helix and Glu on the other. On one of the helices, the position of Asn was moved such that the two asparagines could only interact in

the antiparallel orientation. Changing the position of Asn was sufficient to change the helix orientation from parallel to antiparallel (Figure 6E) (*69*). However, the Asn-Asn interaction is not found in natural antiparallel coiled coils (*59*). Charged residues in the core are found in antiparallel coiled coils, and **d** to **g'** or **a** to **e'** saltbridges may be important for controlling the orientation. Using a model antiparallel coiled-coil system, a potential interhelical interaction between a **d**-position Arg and a **g'**-position Glu was not enough to control helix orientation, but did control the oligomerization state (*58*). Charged residues in the core may play a context-dependent role in determining helix-orientation preference, preventing studies on model coiled-coil systems from capturing an effect. Natural proteins provide an opportunity to study these effects. Statistically, both parallel and antiparallel coiled coils have an equal number of Lys, Arg, and Asp at **a** and **d** positions. However, glutamate at a **d** position is much more prevalent in antiparallel coiled coils (*50*). In Chapter 3, a glutamate in the core of a natural antiparallel coiled coil was found to potentially play a role in determining the coiled-coil orientation.

*Charged residues at **e** and **g***

Charged residues at **e** and **g** positions in coiled coils and helical bundles can also influence helix orientation (*70-72*). If a coiled coil has many potential salt bridges at **e** and **g** in one orientation and many repulsive interactions at **e** and **g** in the other, the orientation with favorable electrostatics will form. In one study, when **e** and **g** had equally attractive interactions, the antiparallel orientation was more stable, indicating there may be an intrinsic preference for coiled coils to associate in the antiparallel orientation (*72*). Another study demonstrated that Coulombic interactions at **e** and **g** and a buried polar Asn-Asn interaction are nearly equal in their importance for determining orientation specificity (Figure 6F) (*70*).

CONCLUSION

The ability to design proteins with conformational and interaction specificity has improved greatly over the past few decades. General principles regarding how charged and hydrophobic residues influence interaction specificity have been discerned, using both experimental and computational techniques. However, designing conformational and interaction specificity is still very difficult.

In manual and computational protein design, both conformational and interaction specificity design are challenging due to the inability to predict and model alterative states. If alternative states are known, negative design elements can be incorporated to make the alternative states unfavorable, increasing the energy between undesired and desired states. Hetero- versus homooligomerization specificity is an example of where this can be done well. However, oligomerization specificity (monomer, dimer, trimer, etc) is more difficult to design because the number of possible alternative states is large. Further, oligomerization and helix-orientation specificity are related, in that changing oligomerization specificity could also change the helix-orientation. When designing for conformational specificity, particularly preventing aggregation, it is hard to determine exactly what alternative models to use *a priori* because proteins can associate in many different ways. Modeling the unfolded state is even more difficult, as the structure is not known and is likely to be very sequence specific. Modeling all possible alternative protein states becomes computationally infeasible quickly, especially as more complex protein-protein interactions are designed. Despite these challenges, approximations for aggregation and unfolded-state models have been reasonably successful.

The necessity to incorporate alterative models into designs seems to be dependent on the biological function of the protein. For promiscuous proteins that interact with many different

proteins in the cell, negative design may not be critical. By stabilizing promiscuous proteins for one target, other targets are destabilized. However, redesigning a protein interface that already has high affinity and specificity will probably require negative design elements to produce a protein that is specific for a different interaction. Thus, considering the biology and affinity of the protein target could prove useful for protein design.

Experimental studies have probed the isolated effects of hydrophobic packing and polar and charged residues on conformational and interaction specificity, particularly in coiled coils. Protein design studies have incorporated these into automated algorithms, and all the forces appear to play an important role in both types of specificity. Designing specificity using computational methods is difficult due to the many approximations that must be made in the energy function, particularly for electrostatics and solvation. For instance, interaction of water molecules with surface residues is important for interaction specificity. However, water is computationally infeasible to model explicitly, as most of the computational time would be spent in calculating water-water interactions. Instead, continuum solvation models are used to model the solvent as a dielectric, rather than as explicit molecules. The distance-dependent dielectric function is an approximation of Coulomb electrostatics in a polar environment. In some of the design calculations, the distance-dependent dielectric constant had to be changed to obtain more intermolecular interactions.

Two more approximations made in protein design include modeling side chains using discrete rotamer conformations and modeling fixed backbones. Side chains in high-probability rotamers are important for conformational and interaction specificity. The rotamer approximation can be improved by adding more rotamers, often yielding more exact results (*73, 74*). However, increasing the number of rotamers increases the computational time. Design

using a fixed backbone is another approximation, as protein backbones have the ability to relax in solution. The fixed backbone approximation often leads to small steric clashes and artificially strained rotamers.

Improvements to energy functions, as well as faster computers will make current design techniques more accurate. The ability to model alternative targets explicitly and implicitly will enhance conformational and interaction specificity design. There are still many very exciting avenues for computational protein design to follow, and improvements in designing both conformational and interaction specificity will continue to be made.

Summary of Thesis Work

The design of a heterotetrameric BBA protein from a homotetrameric BBA is described in Chapter 2. Experimental characterization of the BBA heterotetramer in solution and x-ray crystallography techniques revealed that the structure is tetrameric and has a structure very close to the designed structure. In Chapter 3, the orientation specificity of the coiled-coil region from the breakpoint cluster region (Bcr) oligomerization domain was explored. Unusual **d**-position residues were changed to leucine and were biophysically characterized to elucidate their role in determining oligomerization and orientation specificity in Bcr. The computational design of an antiparallel coiled-coil interaction partner for the Bcr oligomerization domain is reported in Chapter 4.

REFERENCES

1. Maurizi, M. R. (2002) Love it or cleave it: tough choices in protein quality control. *Nat Struct Biol 9*, 410-2.
2. Nooren, I. M., and Thornton, J. M. (2003) Diversity of protein-protein interactions. *Embo J 22*, 3486-92.
3. DeGrado, W. F., Raleigh D.P., and Handel, T. (1991) *De novo* protein design: what are we learning? *Curr. Opin. Struct. Biol. 1*, 984-993.
4. DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F., and Lombardi, A. (1999) De novo design and structural characterization of proteins and metalloproteins. *Annu Rev Biochem 68*, 779-819.
5. Hill, R. B., DeGrado, W.F. (1998) Solution Structure of α_2D, a Nativelike de Novo Designed Protein. *J. Am. Chem. Soc. 120*, 1138-1145.
6. Ho, S. P., DeGrado, W.F. (1987) Design of a 4-Helix Bundle Protein: Synthesis of Peptides Which Self-Associate into a Helical Protein. *J. Am. Chem. Soc. 109*, 6751-6758.
7. Luque, I., Mayorga, O. L., and Freire, E. (1996) Structure-based thermodynamic scale of alpha-helix propensities in amino acids. *Biochemistry 35*, 13681-8.
8. Raleigh, D. P., DeGrado, W.F. (1992) A deNovo Designed Protein Shows a Thermally Induced Transition from a Native to a Molten Globule-like State. *J. Am. Chem. Soc. 114*, 10079-10081.
9. Raleigh, D. P., Betz, S.F., DeGrado, W.F. (1995) A de Novo Designed Protein Mimics the Native State of Natural Proteins. *J. Am. Chem. Soc. 117*, 7558-7559.
10. Summa, C. M., Rosenblatt, M. M., Hong, J. K., Lear, J. D., and DeGrado, W. F. (2002) Computational de novo design, and characterization of an A(2)B(2) diiron protein. *J Mol Biol 321*, 923-38.
11. Desjarlais, J. R., and Handel, T. M. (1995) De novo design of the hydrophobic cores of proteins. *Protein Sci 4*, 2006-18.
12. Desjarlais, J. R., and Handel, T. M. (1999) Side-chain and backbone flexibility in protein core design. *J Mol Biol 290*, 305-18.
13. Johnson, E. C., Lazar, G. A., Desjarlais, J. R., and Handel, T. M. (1999) Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Structure Fold Des 7*, 967-76.
14. Lazar, G. A., Desjarlais, J. R., and Handel, T. M. (1997) De novo design of the hydrophobic core of ubiquitin. *Protein Sci 6*, 1167-78.
15. Lazar, G. A., Johnson, E. C., Desjarlais, J. R., and Handel, T. M. (1999) Rotamer strain as a determinant of protein structural specificity. *Protein Sci 8*, 2598-610.
16. Koehl, P., and Levitt, M. (1999) De novo protein design. I. In search of stability and specificity. *J Mol Biol 293*, 1161-81.
17. Marshall, S. A., and Mayo, S. L. (2001) Achieving stability and conformational specificity in designed proteins via binary patterning. *J Mol Biol 305*, 619-31.
18. Pokala, N., and Handel, T. M. (2005) Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *J Mol Biol 347*, 203-27.
19. Dahiyat, B. I., Sarisky, C. A., and Mayo, S. L. (1997) De novo protein design: towards fully automated sequence selection. *J Mol Biol 273*, 789-96.

20. Waldburger, C. D., Schildbach, J. F., and Sauer, R. T. (1995) Are buried salt bridges important for protein stability and conformational specificity? *Nat Struct Biol 2*, 122-8.
21. Bolon, D. N., and Mayo, S. L. (2001) Polar residues in the protein core of Escherichia coli thioredoxin are important for fold specificity. *Biochemistry 40*, 10047-53.
22. Havranek, J. J., and Harbury, P. B. (2003) Automated design of specificity in molecular recognition. *Nat Struct Biol 10*, 45-52.
23. Reina, J., Lacroix, E., Hobson, S. D., Fernandez-Ballester, G., Rybin, V., Schwab, M. S., Serrano, L., and Gonzalez, C. (2002) Computer-aided design of a PDZ domain to recognize new target sequences. *Nat Struct Biol 9*, 621-7.
24. Shifman, J. M., and Mayo, S. L. (2002) Modulating calmodulin binding specificity through computational protein design. *J Mol Biol 323*, 417-23.
25. Shifman, J. M., and Mayo, S. L. (2003) Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc Natl Acad Sci U S A 100*, 13274-9.
26. Looger, L. L., Dwyer, M. A., Smith, J. J., and Hellinga, H. W. (2003) Computational design of receptor and sensor proteins with novel functions. *Nature 423*, 185-90.
27. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L., and Baker, D. (2004) Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol 11*, 371-9.
28. Schnarr, N. A., and Kennan, A. J. (2002) Peptide tic-tac-toe: heterotrimeric coiled-coil specificity from steric matching of multiple hydrophobic side chains. *J Am Chem Soc 124*, 9779-83.
29. Kashiwada, A., Hiroaki, H., Kohda, D., Nango, M., and Tanaka, T. (2000) Design of a Heterotrimeric Alpha-Helical Bundle by Hydrophobic Core Engineering. *J Am Chem Soc 122*, 212-215.
30. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1992) Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. *Cell 68*, 699-708.
31. Campbell, K. M., Sholders, A. J., and Lumb, K. J. (2002) Contribution of buried lysine residues to the oligomerization specificity and stability of the fos coiled coil. *Biochemistry 41*, 4866-71.
32. O'Shea, E. K., Rutkowski, R., Stafford, W. F. d., and Kim, P. S. (1989) Preferential heterodimer formation by isolated leucine zippers from fos and jun. *Science 245*, 646-8.
33. O'Shea, E. K., Lumb, K. J., and Kim, P. S. (1993) Peptide 'Velcro*': design of a heterodimeric coiled coil. *Curr. Biol. 3*, 658-667.
34. Nautiyal, S., Woolfson, D. N., King, D. S., and Alber, T. (1995) A designed heterotrimeric coiled coil. *Biochemistry 34*, 11645-51.
35. Bolon, D. N., Wah, D. A., Hersch, G. L., Baker, T. A., and Sauer, R. T. (2004) Bivalent tethering of SspB to ClpXP is required for efficient substrate delivery: a protein-design study. *Mol Cell 13*, 443-9.
36. Bolon, D. N. (2005) Specificity versus stability in computational protein design (unpublished).
37. Monera, O. D., Sonnichsen, F. D., Hicks, L., Kay, C. M., and Hodges, R. S. (1996) The relative positions of alanine residues in the hydrophobic core control the formation of two-stranded or four-stranded alpha-helical coiled-coils. *Protein Eng 9*, 353-63.

38. Hu, J. C., O'Shea, E. K., Kim, P. S., and Sauer, R. T. (1990) Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. *Science 250*, 1400-3.
39. Kouzarides, T., and Ziff, E. (1988) The role of the leucine zipper in the fos-jun interaction. *Nature 336*, 646-51.
40. Harbury, P. B., Kim, P. S., and Alber, T. (1994) Crystal structure of an isoleucine-zipper trimer. *Nature 371*, 80-3.
41. Tripet, B., Wagschal, K., Lavigne, P., Mant, C. T., and Hodges, R. S. (2000) Effects of side-chain characteristics on stability and oligomerization state of a de novo-designed model coiled-coil: 20 amino acid substitutions in position "d". *J Mol Biol 300*, 377-402.
42. Wagschal, K., Tripet, B., Lavigne, P., Mant, C., and Hodges, R. S. (1999) The role of position a in determining the stability and oligomerization state of alpha-helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins. *Protein Sci 8*, 2312-29.
43. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science 262*, 1401-1407.
44. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-ray Structure of the GCN4 Leucine Zipper, a Two-Stranded, Parallel Coiled Coil. *Science 254*, 539-544.
45. Park, D. J. (2002) in *Electrical Engineering and Computer Science* pp 73, Massachusetts Institute of Technology, Cambridge.
46. Crick, F. H. C. (1953) The packing of a-helices: simple coiled coils. *Acta Cryst. 6*, 689-697.
47. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) High-resolution protein design with backbone freedom. *Science 282*, 1462-1467.
48. Plecs, J. J., Harbury, P. B., Kim, P. S., and Alber, T. (2004) Structural test of the parameterized-backbone method for protein design. *J Mol Biol 342*, 289-97.
49. Lupas, A., vanDyke, M., and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science 252*, 1162-1164.
50. Walshaw, J., and Woolfson, D. N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol 307*, 1427-50.
51. Lumb, K. J., and Kim, P. S. (1995) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry 34*, 8642-8648.
52. Zhu, H., Celinski, S. A., Scholtz, J. M., and Hu, J. C. (2001) An engineered leucine zipper a position mutant with an unusual three-state unfolding pathway. *Protein Sci 10*, 24-33.
53. Gonzalez, L., Jr., Brown, R. A., Richardson, D., and Alber, T. (1996) Crystal structures of a single coiled-coil peptide in two oligomeric states reveal the basis for structural polymorphism. *Nat Struct Biol. 3*, 1002-1009.
54. Gonzalez, L., Jr., Woolfson, D. N., and Alber, T. (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat. Struct. Biol. 3*, 1011-1018.
55. Akey, D. L., Malashkevich, V. N., and Kim, P. S. (2001) Buried polar residues in coiled-coil interfaces. *Biochemistry 40*, 6352-6360.
56. Woolfson, D. N., and Alber, T. (1995) Predicting oligomerization states of coiled coils. *Protein Sci 4*, 1596-607.
57. Campbell, K. M., and Lumb, K. J. (2002) Complementation of buried lysine and surface polar residues in a designed heterodimeric coiled coil. *Biochemistry 41*, 7169-75.

58. McClain, D. L., Gurnon, D. G., and Oakley, M. G. (2002) Importance of potential interhelical salt-bridges involving interior residues for coiled-coil stability and quaternary structure. *J Mol Biol 324*, 257-70.
59. Oakley, M. G., and Hollenbeck, J. J. (2001) The design of antiparallel coiled coils. *Curr Opin Struct Biol 11*, 450-7.
60. Betz, S. F., Liebman, P. A., and DeGrado, W. F. (1997) De novo design of native proteins: characterization of proteins intended to fold into antiparallel, rop-like, four-helix bundles. *Biochemistry 36*, 2450-8.
61. McClain, D. L., Woods, H. L., and Oakley, M. G. (2001) Design and characterization of a heterodimeric coiled coil that forms exclusively with an antiparallel relative helix orientation. *J Am Chem Soc 123*, 3151-2.
62. Gurnon, D. G., Whitaker, J. A., and Oakley, M. G. (2003) Design and characterization of a homodimeric antiparallel coiled coil. *J Am Chem Soc 125*, 7518-9.
63. Myszka, D. G., and Chaiken, I. M. (1994) Design and characterization of an intramolecular antiparallel coiled coil peptide. *Biochemistry 33*, 2363-72.
64. Ghosh, I., Hamilton, A.D., and Regan, L. (2000) Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *J Am Chem Soc 122*, 5658-5659.
65. Gernert, K. M., Surles, M. C., Labean, T. H., Richardson, J. S., and Richardson, D. C. (1995) The Alacoil: a very tight, antiparallel coiled-coil of helices. *Protein Sci 4*, 2252-60.
66. Monera, O., Zhou, N., Lavigne, P., Kay, C., and Hodges, R. (1996) Formation of parallel and antiparallel coiled-coils controlled by the relative positions of alanine residues in the hydrophobic core. *J Biol Chem 271*, p3995-4001.
67. Betz, S. F., and DeGrado, W. F. (1996) Controlling topology and native-like behavior of de novo-designed peptides: design and characterization of antiparallel four-stranded coiled coils. *Biochemistry 35*, 6955-62.
68. Schnarr, N. A., and Kennan, A. J. (2004) Strand orientation by steric matching: a designed antiparallel coiled-coil trimer. *J Am Chem Soc 126*, 14447-51.
69. Oakley, M. G., and Kim, P. S. (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry 37*, 12603-10.
70. McClain, D. L., Binfet, J. P., and Oakley, M. G. (2001) Evaluation of the energetic contribution of interhelical Coulombic interactions for coiled coil helix orientation specificity. *J Mol Biol 313*, 371-83.
71. Monera, O. D., Zhou, N. E., Kay, C. M., and Hodges, R. S. (1993) Comparison of antiparallel and parallel two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization. *J Biol Chem 268*, 19218-27.
72. Monera, O., Kay, C., and Hodges, R. (1994) Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils. *Biochemistry 33*, p3862-71.
73. Xiang, Z., and Honig, B. (2001) Extending the accuracy limits of prediction for side-chain conformations. *J Mol Biol 311*, 421-30.
74. Peterson, R. W., Dutton, P. L., and Wand, A. J. (2004) Improved side-chain prediction accuracy using an ab initio potential energy function and a very large rotamer library. *Protein Sci 13*, 735-51.

FIGURES

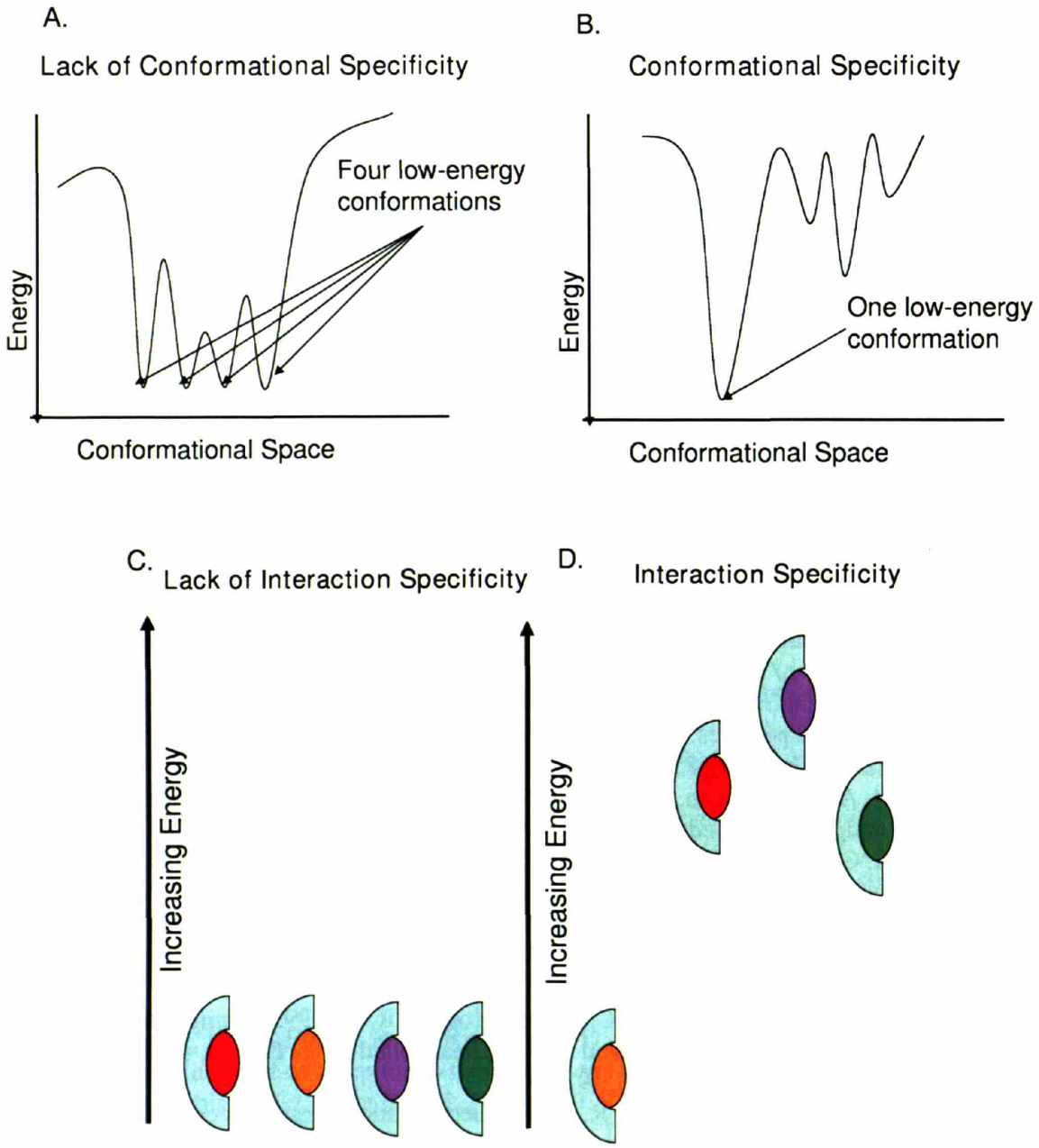


Figure 1. Conformational and Interaction Specificity. (A) Lack of conformational specificity is caused by several conformations with similar folded energies. (B) A unique structure is formed when there are no kinetically accessible competing conformations with similar low energies. (C) If several interactions yield the same favorable energy upon binding, all the interactions will form. (D) Interaction specificity occurs when there is one interaction with a much lower energy than other potential interactions.

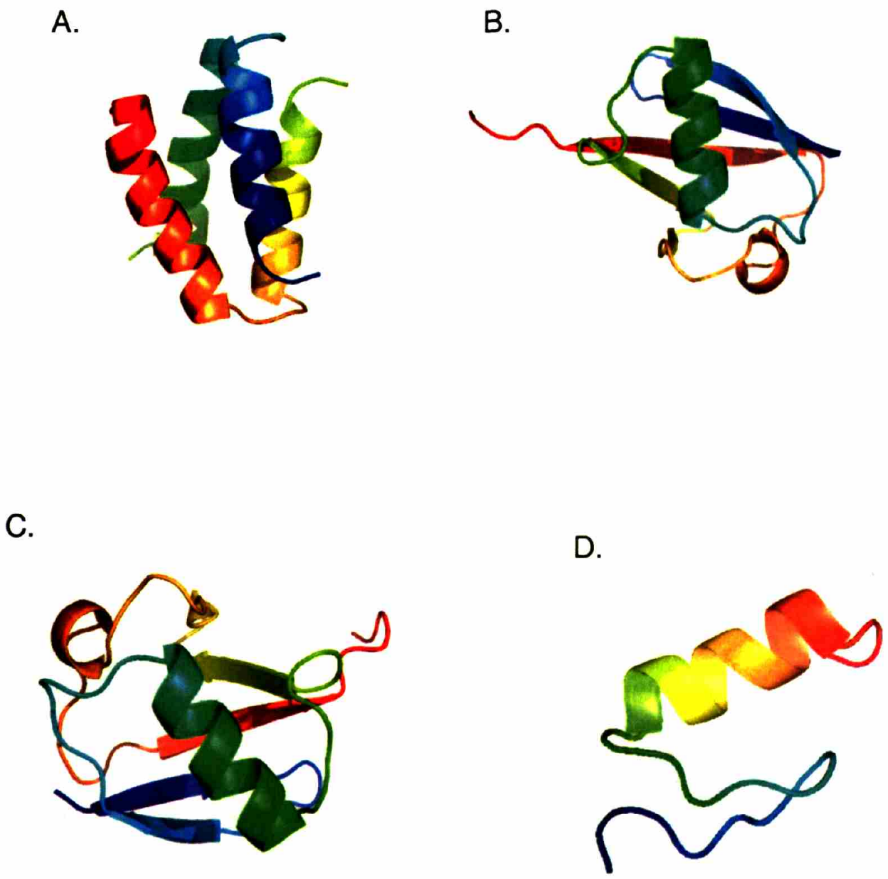


Figure 2: Structures of proteins designed or redesigned for conformational specificity. (A) α_2D (1QP6). (B) phage 434 cro protein (1UBI) (C) ubiquitin (1UD7) (D) $\beta\beta\alpha$ (1FSD).

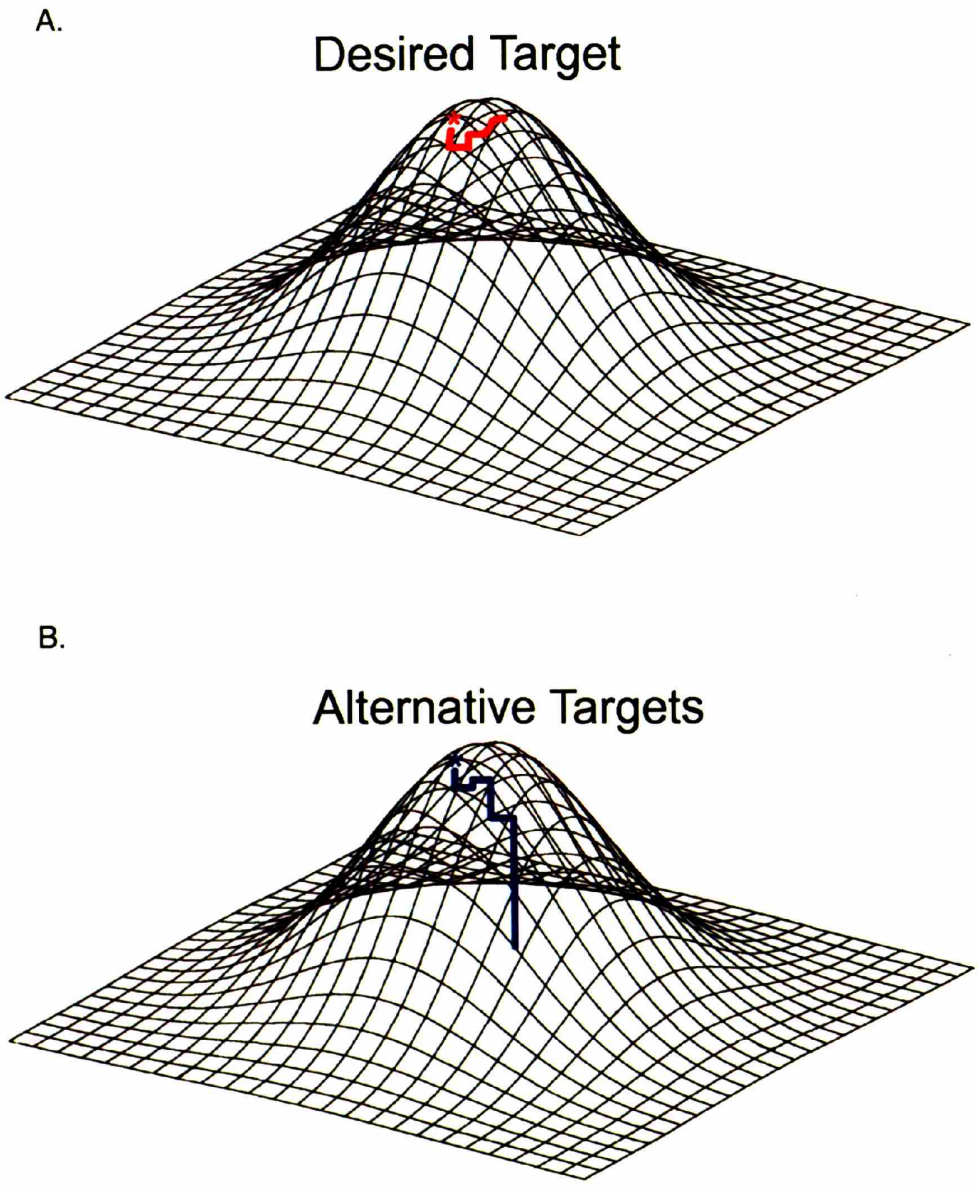


Figure 3. "Fitness" surface. "Fitness" surface for (A) a desired target and (B) alternative or undesired targets (adapted from (*25*)). A high "fitness" indicates the protein and target are compatible, whereas a decrease in "fitness" indicates the protein and target have become less compatible. The asterisks indicate the starting sequence, and lines on the charts represent changes in sequence through the optimization process. For the desired target, the sequence changes, but the "fitness" remains the same. For the alternative target, the sequence changes and the "fitness" decreases.

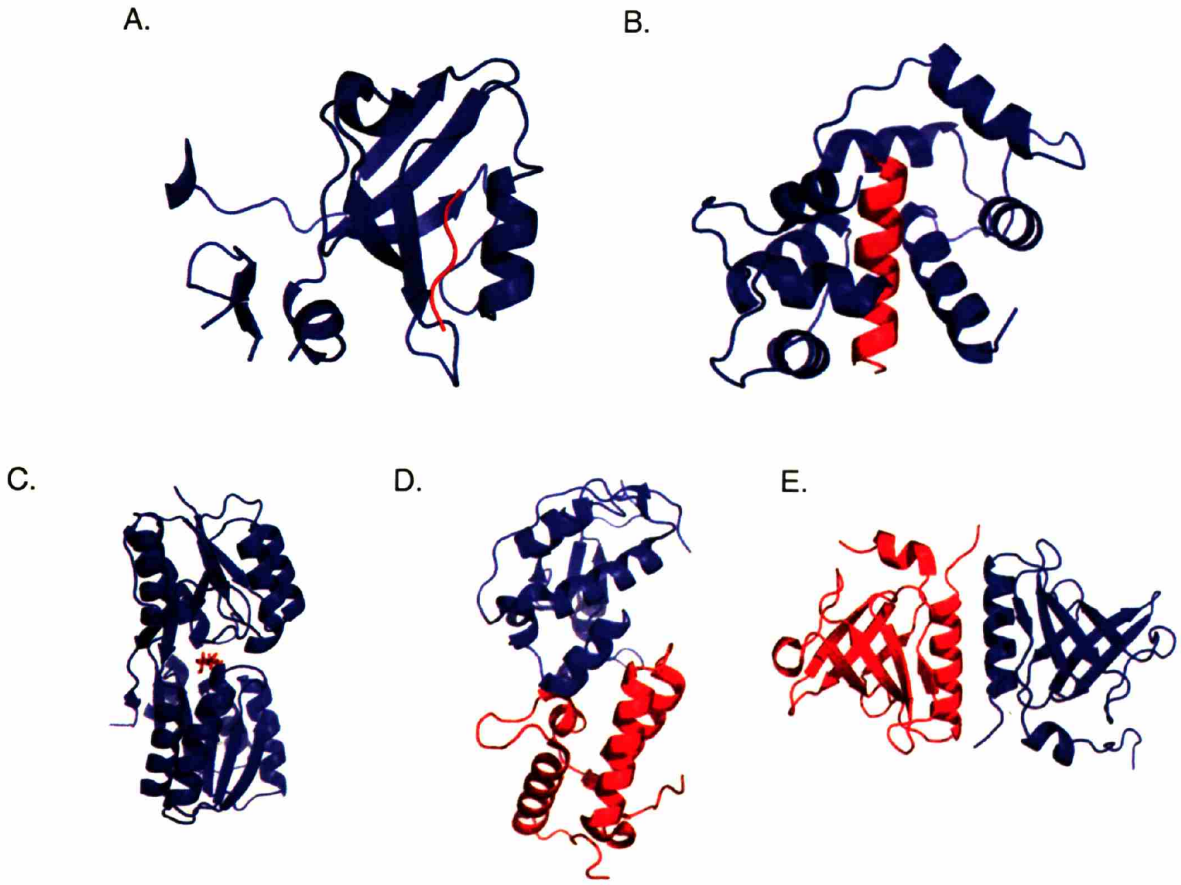


Figure 4. Structures of proteins used to design interaction specificity. (A) PDZ domain (1BE9) (B) calmodulin in complex with smMLCK (1CDM) (C) ribose-binding protein (2DRI) (D) the designed DNase (colicin E7) (1UJZ) (E) SspB homodimer (1OU9). The blue and red colors are used to differentiate different domains or proteins in protein-protein interactions, as well as protein versus ligand interactions.

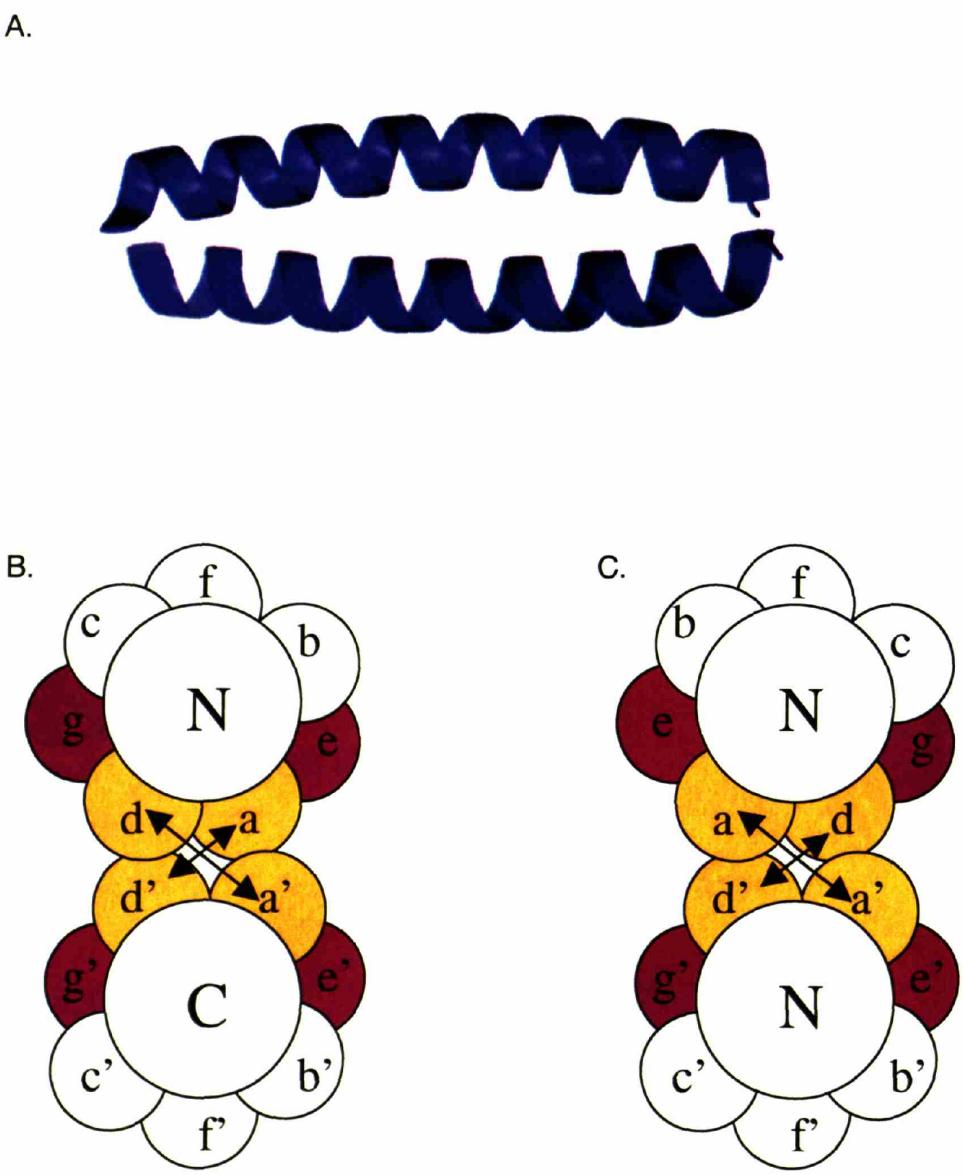


Figure 5. Coiled-coil structure and helical wheel diagram. (A) X-ray crystal structure of the GCN4 leucine zipper. Depicting coiled coils in the helical wheel format shown above can give a general picture regarding interactions between **a**, **d**, **e**, and **g** positions in (B) antiparallel and (C) parallel coiled coils. Typically, **a** and **d** positions are hydrophobic, whereas **e** and **g** positions are polar and charged.

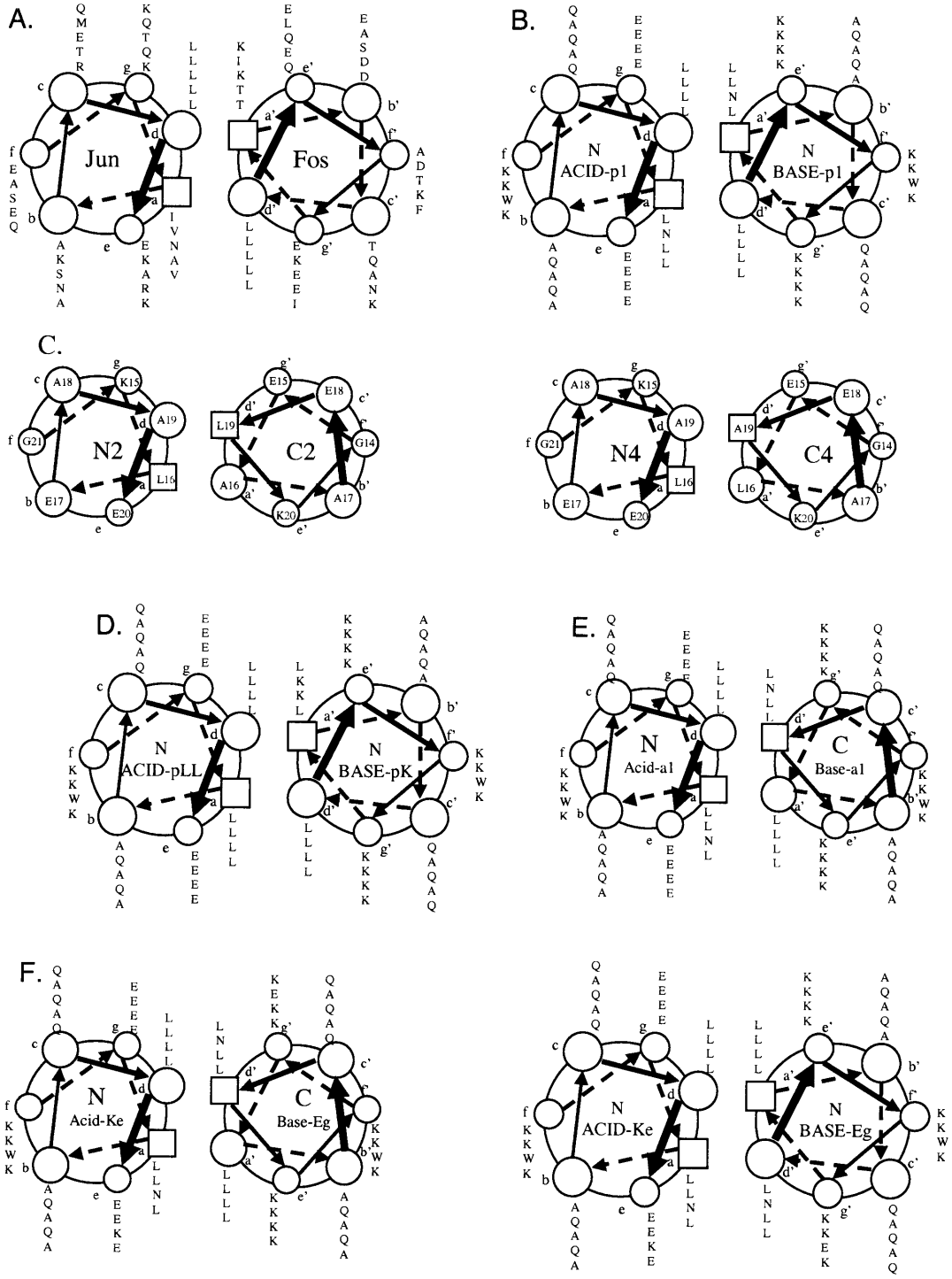


Figure 6. Helical wheel diagrams of coiled coils used in previous studies. (A) Fos/Jun (B) ACID-p1 and BASE-p1 (Velcro) (C) N2-C2 form dimers, whereas N4-C4 form tetramers. (D) ACID-pLL and BASE-pK (E) ACID-a1 and BASE-a1 (F) ACID-Ke and BASE-Eg

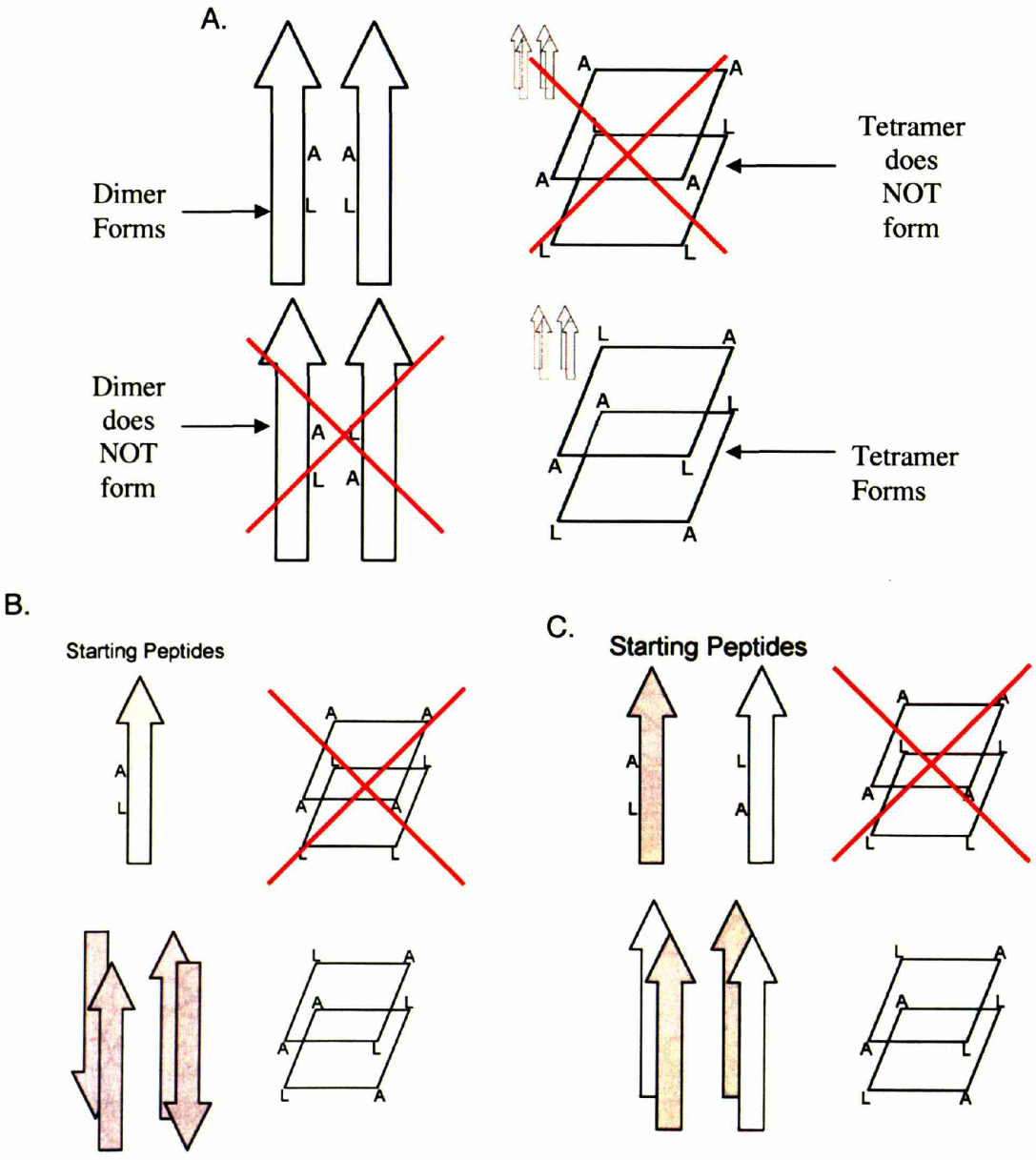
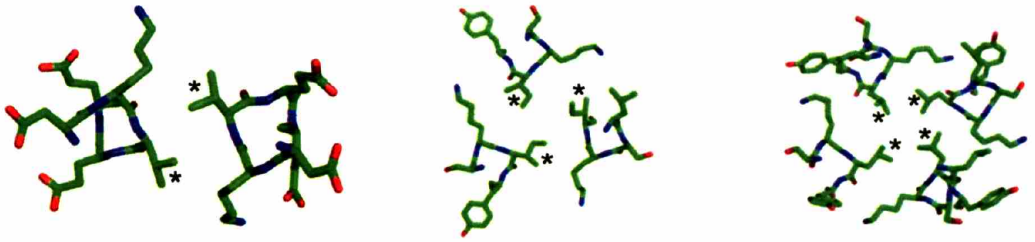
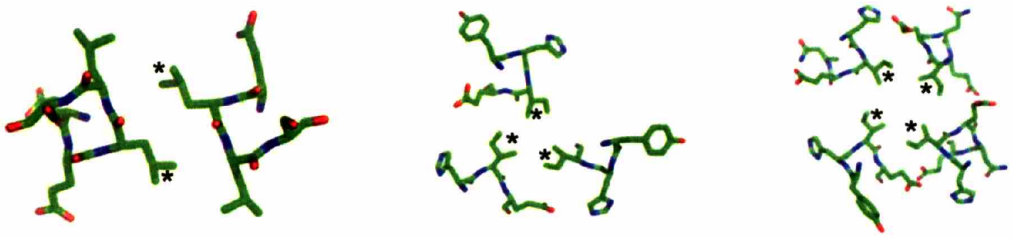


Figure 7. Role of alanine in oligomerization and helix orientation specificity. (A) Alanine steric matching in determining oligomerization state. When Ala is on the same layer, a dimer forms. When the alanines are on adjacent layers, a tetramer forms. (B) and (C) Helix orientation is driven by position of alanines in the core. The helices never orient such that all the alanines are on the same layer.



a layer



d layer

Figure 8. The a- and d- layer packing in X-ray crystal structures of dimer, trimer, and tetramer GCN4 variants. The asterisks on the core side chains indicate the residues important for determining oligomerization state.

CHAPTER TWO

Design of a Heterospecific, Tetrameric, 21-Residue Miniprotein with Mixed α/β Structure

Reprinted from:

Ali, M.H., Taylor, C.M., Grigoryan, G., Allen, K.N., Imperiali, B., Keating, A.E., *Design of a heterospecific, tetrameric, 21-residue miniprotein with mixed alpha/beta structure.* Structure (Camb), 2005. **13**(2): p. 225-34.

Collaborator Note:
Mayssam Ali solved the crystal structure of **BBAhetT1**, made the peptides, and performed some of the experiments.

ABSTRACT

The study of short, autonomously folding peptides, or "miniproteins," is important for advancing our understanding of protein stability and folding specificity. Although many examples of synthetic α-helical structures are known, relatively few mixed α/β structures have been successfully designed. Only one mixed-secondary structure oligomer, an α/β homotetramer, has been reported thus far. In this report, we use structural analysis and computational design to convert this homotetramer into the smallest known α/β-heterotetramer. Computational screening of many possible sequence/structure combinations led efficiently to the design of short, 21-residue peptides that fold cooperatively and autonomously into a specific complex in solution. A 1.95 Å crystal structure reveals how steric complementarity and charge patterning encode heterospecificity. The first- and second- generation heterotetrameric miniproteins described here will be useful as simple models for the analysis of protein-protein interaction specificity and as structural platforms for the further elaboration of folding and function.

Abbreviations: DapBz, benzoylated L-α,β-diaminopropionic acid; Abu, L-α-aminobutyric acid; AUC, analytical ultracentrifugation

INTRODUCTION

Miniproteins are short polypeptides, typically having fewer than 40 residues, that adopt specific, discrete folds in aqueous solution (*1*). Many miniproteins have been described in the literature, including α-helical coiled coils and helical bundles (*2*), mixed α/β motifs (*3, 4, 5*), predominantly-β motifs (*6-8*), and other folds (*9, 10*). Some of these miniproteins are derived by reducing naturally occurring protein folds to a minimal folding core, whereas others have been designed *de novo*, either by visual inspection or with the use of computational methods. Miniproteins have served as scaffolds for ligand and metal binding, as well as for the introduction of catalytic activity (*11-13*). Miniproteins, and coiled-coil miniproteins in particular, have been successfully utilized in materials science for the introduction of nano-scale structure and organization, often with "switchable" physical properties (*14, 15*). Heterooligomeric peptides are particularly useful in these contexts, as they provide a mechanism for integrating units with distinct properties. We are interested in designing heterooligomeric miniproteins both for potential applications and for fundamental investigations of how protein-protein interaction specificity is encoded in sequence and structure.

The small size of miniproteins provides several advantages for studying protein folding and association. First, the sequence-structure relationship is simplified. There are fewer variables available in miniproteins to encode properties of interest, and as a consequence these variables can be more systematically dissected. Second, miniproteins can be synthesized chemically, providing a straightforward method to test sequence variants including both natural and non-natural amino acids. Chemical synthesis also enables the facile introduction of biophysical probes such as fluorophores. Finally, the small size of miniproteins renders them

amenable to computational analyses, including structure design and the simulation of protein folding (*4, 16-18*).

Numerous short homooligomeric miniproteins have been described, the majority of which are coiled coils and helical bundles. There are fewer examples of heterospecific systems, although a number of heterodimeric (*19-23*), heterotrimeric (*24-26*), and heterotetrameric (*27, 28*) coiled-coil peptides have proven useful for protein engineering applications (*14, 29*). Coiled coils consist of a variable number of α-helices associated in a bundle with a slight superhelical twist. An "acid/base" strategy, in which heterospecifcity is obtained by patterning the residues at the hydrophobic/hydrophilic interface of paired helices with complementary charged residues, has been shown to have great utility for encoding specific structure in coiled coils (*20, 28*). Steric complementarity between juxtaposed residues in the hydrophobic core has also been used to impart heterospecificity (*30, 31*).

Studies of designed heterooligomeric coiled-coil miniproteins have enriched our understanding of how specificity is achieved in naturally occurring proteins such as the transcription factors Fos and Jun (*32*). The shortest heterooligomeric miniprotein previously reported in the literature consists of 21 amino acids and forms a tetrameric coiled coil in solution (*28*). It was used to test the influence of side-chain length on charge-charge interactions. At this time, the structure of this complex has not been reported. In fact, relatively few designed heterospecific miniproteins have been characterized by NMR or x-ray crystallography, resulting in an incomplete understanding of the structural basis of heterospecificity (*22, 33, 34*).

The miniproteins **BBAT1** and **BBAT2** (Figure 1B) are 21-residue homotetramers in which each monomer adopts a mixed ββα (BBA) secondary structure (*5, 35*). **BBAT1** and **BBAT2** are derived from the monomeric α/β miniprotein **BBA5** (Figure 1A), a *de novo* designed

metal ion-independent zinc finger (*3, 36*). **BBAT1** was selected from a small library of **BBA5** derivatives by means of a fluorescence based quenching assay. A shortened loop between α and β subunits in **BBAT1** precludes intramolecular burial of the hydrophobic surfaces and results in self-association to a homotetramer. **BBAT2** is a more stable derivative of **BBAT1** having D-alanine in place of the "hinge" glycine at position 9. The x-ray crystal structure of **BBAT2** was recently reported (*37*), revealing a novel architecture with certain structural elements reminiscent of a tetrameric coiled coil. As in coiled coils, the central core of the **BBAT2** tetramer is formed by association of the hydrophobic face of each monomer, including significant contributions from the helical portion. The hydrophobic core consists of five palindromic layers, each layer composed of one side chain from each monomer, similar to the core layers found in coiled coils. Furthermore, both apolar and polar residues are located along the intermonomer interfaces. Unlike a typical four-stranded coiled coil, however, the superhelical twist of **BBAT2** is right-handed, and the "knobs-into-holes" packing that characterizes coiled coils (*38*) is not observed. Despite these differences, we anticipated that strategies employed in the design of heterooligomeric coiled coils would be applicable to the design of a heterotetrameric BBA complex.

Herein we describe the computational design and characterization of two miniprotein complexes, **BBAhetT1** and **BBAhetT2**. Our goal was to design pairs of short peptides that would adopt the overall backbone structure of the BBA homotetramer, but in a heterospecific A_2B_2-type complex with C_2 or pseudo-C_2 symmetry. We sought a stable and highly specific motif that could be used to probe sequence determinants of interaction specificity and that could serve as a scaffold for further elaboration of structure or function. The design strategy was motivated by the structure of **BBAT2** and used computational methods to identify and rank

mutations likely to impart both stability and heterospecificity. Solution characterization and a high-resolution x-ray structure confirmed the success of our design. The evolution of the BBA family from monomer to heterotetramer is summarized in Figure 1.

RESULTS

Computational Design

Homotetramer **BBAT2** is very sensitive to mutation. Many residues in this small motif have multiple roles in determining specificity and stability, and seemingly minor sequence changes can lead to loss of structure or aggregation (*39*). We therefore adopted a stepwise strategy whereby the effects of mutations to core and surface sites were modeled and evaluated independently. An empirical, molecular mechanics-based energy function was used to identify suitable sites for mutation and to evaluate sequence substitutions at the chosen sites. Each sequence was modeled as a heterotetramer (giving E_{ABAB}), as two homotetramers (giving E_{AAAA} and E_{BBBB}) and as an unfolded monomer (giving E_{unfold}). We sought sequences with large values for both stability ($E_{unfold} - E_{ABAB}$) and specificity ($E_{AAAA} + E_{BBBB} - 2E_{ABAB}$).

The hydrophobic core of **BBAT2** consists of five palindromic layers. The three unique layers are layer A (composed of residues 3 and 20), layer B (composed of residues 8 and 16), and layer C (composed of residues 12 and 12'). To identify sites for mutation, layers B and C were analyzed in detail. Calculations were carried out to select residues at positions 12, 8 and 16 that would impart heterospecificity. The largest predicted specificities came from substitutions at position 12. More qualitative analyses also indicated that position 12 was a good candidate for mutation. For example, two leucines at this site in **BBAT2** were modeled with alternate conformations in the three crystal structures, suggesting a non-optimal fit of the side chains at this site in the homotetramer.

The most heterospecific sequences from the position 12 calculations systematically suggested a "large/small" design strategy, with two monomers contributing a bulky residue that could not be accommodated in the core of a homotetramer and two others contributing a small side chain that could not form good packing interactions in a homotetramer. Models of many different "large/small" combinations (e.g. Trp/Ala, Trp/Ser, Trp/Thr, Phe/Ala, Tyr/Val) indicated that these residues could be accommodated in a heterotetramer. Two different backbone structures for **BBAT2** were used in the design calculations (see Methods). On one backbone, Trp/Ala combinations were predicted to give the highest heterospecificity at positions 12 and 12'; on the other, Phe/Ala pairs were the most heterospecific. Phe/Ala pairs were computed to be more stable than Trp/Ala on both backbones. Furthermore, core side chains in the predicted L12F/L12A heterotetramer could be accommodated in statistically common rotamers (*40, 41*). A peptide pair having Phe at position 12 in one chain and Ala at position 12 in the other chain was therefore selected for experimental analysis (Figure 2B).

Structural analysis, as well prior studies of **BBAT2** (*37*), were used to select surface sites for mutation. At the hydrophobic/hydrophilic interface, positions 13 and 18 (alanine and glutamine, respectively, in **BBAT2**) were previously mutated to methionine selenoxide without structural perturbation, suggesting that the introduction of a charged residue would be tolerated. Position 13 is directly across from position 13 of an adjacent monomer, and distances in **BBAT2** indicated that an interchain interaction would be possible at this site between residues with long side chains. Position 18 is opposite position 11 (glutamate) of an adjacent monomer in the homotetramer; these two sites were also selected for redesign.

Computational selection for heterospecificity at the surface sites suggested placing residues of opposite charge at structurally opposed positions, reminiscent of similar interactions

found in heterospecific coiled coils. The highest specificity scores were obtained for combinations having Glu and/or Asp at positions 11, 13, and 18 of one monomer, and Arg and/or Lys at opposing sites of adjacent monomers. Computed specificity rankings, in conjunction with a visual examination of the structures generated by side-chain repacking, led to the selection of Glu/Lys or Asp/Lys pairs at position 13/position 13 and position 11/position 18 sites (Figures 2D and 2E).

Solution and Structural Characterization

Peptides **A-Ala** and **B-Phe** were synthesized, incorporating the core and surface changes suggested by the computational analysis (Table 1). The two peptides individually exhibited very weak circular dichroism (CD) spectra between 200 and 300 nm at 50 μM, indicating they have little secondary structure. By contrast, an equimolar mixture of **A-Ala** and **B-Phe** revealed an increase in ellipticity and a qualitatively different spectrum indicative of interhelical association (*42*), as observed for homotetramers **BBAT1** and **BBAT2**. The change in secondary structure upon mixing (Figure 3A) strongly supports the formation of a heterooligomeric complex. Thermal melts of the **A-Ala/B-Phe** complex, termed **BBAhetT1**, showed a cooperative-unfolding transition (Figure 3B). Furthermore, heterooligomerization was supported by a fluorescence-quenching assay (Figure 4). The molecular weights of **BBAhetT1** and of the individual components were determined by sedimentation equilibrium analytical ultracentrifugation (AUC) experiments (Table 2). Peptides **A-Ala** and **B-Phe** were found to be monomeric at low concentrations, but **BBAhetT1** was best described as a single tetrameric species. A tracer sedimentation equilibrium experiment (*43*) further confirmed the heterospecificity of the interaction.

The designed combination of a "large/small" core packing motif and charge complementarity was successful in imparting heterospecificity to **BBAhetT1**. However, this complex is significantly destabilized relative to the parent homotetramer **BBAT2**. An analysis of computationally predicted structures indicated that a larger side chain could be accommodated in place of Ala in the "small" partner at position 12. Serine was predicted to result in more favorable van der Waals interactions relative to alanine at this site, but also in an overall reduction in stability. The non-natural amino acid L-α-aminobutyric acid (Abu) was predicted to stabilize the tetramer by ~ 4 kcal/mol, based on estimates of van der Waals and solvation contributions, and to retain high specificity for heterotetramerization. Peptide **A-Abu**, with an Ala to Abu substitution at position 12 (Table 1), was synthesized. The CD spectrum revealed that **A-Abu** has little secondary structure alone. An equimolar mixture of **A-Abu** and **B-Phe** gave a spectrum very similar to that observed for **BBAhetT1**, but with greater ellipticity at 208 and 222 nm. A complex of **A-Abu** and **B-Phe**, termed **BBAhetT2**, exhibited a cooperative thermal unfolding transition, and was considerably more stable to thermal denaturation than **BBAhetT1** (Figure 3B). Heterooligomerization was confirmed by fluorescence quenching (Figure 4). AUC experiments indicated that equimolar mixtures of **A-Abu** and **B-Phe** are best described as a single tetrameric species (Table 2). By contrast, peptides **A-Abu** and **B-Phe** are both monomeric at 25 μM.

The 1.95 Å crystal structure of **BBAhetT1** was solved by molecular replacement, using two monomers from **BBAT2** as a search model (Figure 5). The overall fold (Figure 1C) is very similar to that of **BBAT2**, with an all-atom rms deviation of 1.76 Å and an all-backbone atom rms deviation of 0.71 Å. The similarity of the backbone structures indicates that the calculations were successful in indentifying an alternative sequence compatible with the precise geometry of

the homotetramer fold. The structure confirms that **BBAhetT1** is a C_2-symmetric heterotetramer, as intended. The tetramer axis coincides with a perfect two-fold screw axis in the crystallographic symmetry. The conformations of the designed side chains and their surrounding residues in the core are in excellent agreement with the calculated predictions (Figure 6). On the surface, two designed salt bridges between Glu 11 on **A-Ala** and Lys 18 on **B-Phe** are formed, with slightly different side-chain conformations than predicted. Four other designed salt bridges are not present, but in each of these cases the designed residues participate in inter-tetramer crystal contacts that were not represented in the calculations (Figure 6).

DISCUSSION

The design of miniproteins is challenging. Many residues in the BBA peptides play multiple structural roles, and single point mutations can affect the overall fold or solution properties. In our work, computational methods proved valuable for rapidly identifying mutations that would change the interaction specificity of **BBAT2** while maintaining its overall structure. Calculations were used to identify possible sites for mutation as well as to find the best combination of residues for stabilizing a heterotetramer relative to competing homotetramers and the unfolded state. Despite the very large number of possible sequences, the first set of designed peptides that we tested exhibited the desired properties. The efficiency of this structure-based computational approach can be compared to the much slower process of performing an experimental selection or the iterative process of testing sequences suggested by visual inspection.

Computational protein design does have significant limitations. Modeling structural relaxation, particularly backbone flexibility, is challenging (*44-46*). In the absence of a realistic model to describe relaxation, our method predicted that sequences with large groups in the core

would be highly destabilized and not fold as homotetramers. By contrast, our experiments revealed that **B-Phe**, an example of such a peptide, nevertheless associates to some extent in solution at high concentrations (data not shown). The strategy of designing holes in the core, rather than steric bulk, was significantly more effective at destabilizing the homotetramer state. Relaxation of side-chain conformations in the calculations, and the use of multiple tetramer structures as templates in the design process, alleviated the problem of large steric clashes to some extent.

BBAhetT1 and **BBAhetT2** are both less stable than **BBAT2**, indicating that specificity in this system comes from destabilization of the competing homotetramer states rather than stabilization of the heterotetramer. Charge patterning is an effective negative design strategy in some heterospecific coiled coils (*20*). However, we found that BBA peptides in which only the surface side chains are altered (**A-Leu** and **B-Leu**, Table 1) self-associate in solution (data not shown). Charge-charge repulsion in these systems is not sufficient to prevent the undesired states from forming under the conditions studied. **A-Leu** and **B-Leu** do, notably, preferentially form heterooligomers when mixed, as demonstrated by fluorescence quenching and CD experiments (data not shown). But a significant amount of additional heterospecificity in **BBAhetT1** and **BBAhetT2** is apparently derived from the core redesign and its role in disfavoring homotetramerization.

We observed excellent agreement of the backbone and core structure of **BBAhetT1** with the computationally predicted design. The behavior of the designed charged residues in the crystal structure is more complex than predicted by the calculations. Notably, all but one of the computationally predicted surface residues were observed to have high temperature factors or multiple conformations in the **BBAT2** or **BBAhetT1** crystal structures, suggesting that the

conformational preferences at these sites are not absolute. In addition, most of the designed side chains in the **BBAhetT1** structure are involved in inter-tetramer interactions. At sites where **BBAhetT1** surface residues do not participate in crystal contacts, the designed salt bridges form as predicted. Charged surface residues on **BBAhetT1** are probably dynamic in solution and influenced by lattice contacts in the crystal. Information about whether the designed salt bridges form in solution would allow a more direct comparison with the calculations.

We used the native **BBAhetT1** backbone to model the difference in energy between the predicted and the experimental structures of the heterotetramer. We approximated the experimental structure using native side-chain chi angles but bond lengths and angles from CHARMM. The energy function used for design strongly favored the formation of salt bridges between acidic and basic surface residues, as observed in the sequence-selection calculations. Interestingly, when the side-chain conformations were relaxed by minimization and the energies re-evaluated using more accurate electrostatics functions (see Methods), the difference between salt-bridge forming and non-salt bridge forming configurations decreased dramatically. These calculations support a small energy gap between structures with different numbers of salt bridges, consistent with the alternative conformations seen in the BBA homo and heterotetramer crystal structures.

Differences in the placement of residue Tyr 6, and other very minor differences between the design and the **BBAhetT1** structure, resulted from simplifying assumptions of the computational method. For example, the incorrect prediction of Tyr 6 arose because the rotamer library used did not contain all of the experimentally observed side-chain chi angles. In addition, even slight changes in backbone geometry (e.g. the 0.5 Å difference between the **BBAT2** and the **BBAhetT1** backbones) made significant differences in the correct placement of rotamers and

in salt-bridge formation. Predictive performance for Tyr 6 was improved when wild-type rotamers were included and the sequence was modeled using the correct heterotetramer backbone.

In conclusion, we have designed and characterized two compact, heterotetrameric, mixed α/β miniproteins, **BBAhetT1** and **BBAhetT2**. Both were derived from a homotetrameric precursor using computational screening of many possible sequence/structure combinations. These tetramers constitute the first reported heterooligomeric α/β miniproteins, and thus arguably the most complex miniproteins designed thus far. The BBA family of peptides has proven to be quite remarkable. The power of small sequence changes to encode monomeric, homotetrameric and heterotetrameric BBA variants, and to tune stability and specificity within each class, makes this system ideal for studying basic principles of protein structure. The study described here has established several ways that oligomerization specificity can be manipulated in the BBA peptides. Future work is likely to suggest others, and to lead to yet more novel architectures, activities and functions, all specified by a sequence of only 21 amino acids.

EXPERIMENTAL PROCEDURES

Computational Design

All calculations, unless stated otherwise, used symmetry-generated tetramers from the crystal structures of two selenomethionine mutants of **BBAT2** (1SNA and 1SNE). In the case of a residue having alternate conformations, the first listed conformation was used. Methionine selenoxide residues were modeled as alanines. Allowed side-chain conformations were defined by the backbone-dependent rotamers of Dunbrack (*40, 41*), using default bond lengths and angles from CHARMM param19 (*47*). Non-standard amino acids added were: D-alanine (D-

Ala), D-proline (D-Pro), benzoylated L-α,β-diaminopropionic acid (DapBz), and L-α-aminobutyric acid (Abu). The angle CH1E-CH2E-NH1, with a force constant of 100.0 and equilibrium geometry of 109.5, and the improper dihedral NH1-H-C-CH2E, with a force constant of 750.0, multiplicity of 0, and minimum geometry of 0.0, were added to the CHARMM 19 parameter set for DapBz (*37*). Rotamers of Abu were modeled using rotamers of serine (*40*).

Design energies were defined as the sum over all self energies plus the sum over all unique residue-residue energies for all flexible sites. The self energies include intra-side chain interactions, as well as interactions with the backbone and non-designed side chains; the pair-wise energies include the interactions of a particular side chain with another side chain. The van der Waals radii from CHARMM param19 were scaled by 90% to accommodate discrete rotamer conformations. The total van der Waals energy was also scaled by 90%. Torsion energies were computed using param19. The pair-wise electrostatic energy was calculated using a coulombic function with $\varepsilon = 4r$ for polar-polar interactions, 8r for polar-charged interactions, and 16r for charged-charged interactions; for self energies $\varepsilon = 4r$ was used. Both the self and pair-wise solvation energies were calculated using an Effective Energy Function (EEF1), (*48*) with reference/free energies as published except for NH3 (-10/-10), NC2 (–7.5/–7.5), OC (–5.33/–5.85). The unfolded state was modeled by treating each residue in turn as the central residue, X, in a Gly-Gly-X-Gly-Gly pentapeptide. The backbone structure employed was that of residue X and the two residues preceding and following X in the crystal structure. The side chain of residue X was modeled as an ensemble average of all possible rotamer states, employing the same energy function used for the folded states.

Sequences were either enumerated or sampled using a Monte Carlo search algorithm. The Monte Carlo search algorithm employed 64 cycles of 1500 steps with a linear temperature

gradient from 300 to 200 K. For each sequence, the side chains were placed onto one or more backbones in an optimal configuration using a DEE/A* algorithm (*49, 50*). Two quantities were evaluated: the stability (E_{unfold} - E_{ABAB}) and the specificity (E_{AAAA} + E_{BBBB} – $2E_{ABAB}$). Sequences were ranked according to these two scores, and those with high stability and specificity on both of the backbones used were considered for further analysis. For selected models, the energies were re-evaluated with a function including the following terms: CHARMM param19 van der Waals energy (100% radii), Coulomb energy evaluated with a dielectric constant of 4, polarization energy for transfer from a protein/environment dielectric of 4/4 to 4/80, and a surface tension term computed as the solvent accessible surface area multiplied by 7 cal/mol•Å^2. The polarization energy was computed with a Generalized Born model (*51*) that used PEP to solve for Born radii (*52*). A version of the energy function in which the Generalized Born reaction field energies were substituted by EEF desolvation energies was also used. Models were evaluated both before and after 10 steps of steepest descent minimization (maintaining a fixed backbone) to relieve side-chain steric clashes.

Peptide Synthesis

Peptides were prepared by standard Fmoc-based solid-phase peptide synthesis as in Ali et al. (*37*). Identity was confirmed by electrospray mass spectroscopy (PerSeptive Biosystems) and purity by analytical HPLC (>95%).

Circular Dichroism

CD spectra from 300 to 200 nm were collected at 25 °C in duplicate or triplicate on an Aviv circular dichroism spectrometer Model 202 using strain-free quartz cells having a path length of 0.1 cm and an averaging time of 5 seconds. Peptides were dissolved in degassed buffer

(50 mM sodium phosphate, 100 mM NaCl, pH 7.2), and concentrations were determined using the method of Edelhoch (*53*). Melting experiments involved monitoring $[\theta]_{222}$ using a 30 second averaging time, 90 second equilibration time, and temperature increments of 2 °C from 5 °C to 80 °C.

Fluorescence Quenching

Peptides were synthesized with a 3-nitrotyrosine quencher at position 6 (**A-Ala** and **A-Abu**) or an anthranilamide fluorophore at position 20 (**B-Phe**). 12.5 μM samples were prepared of fluorophore- or quencher-labeled peptides in 10 mM phosphate buffer, pH 7.2. Combination samples comprised 12.5 μM fluorophore-labeled peptide plus 12.5 μM quencher-labeled peptide. Data were collected at 25 °C with a Jobin Yvon Horiba FluroMax-P fluorescence spectrometer using strain-free quartz cells having a path length of 1 cm. Samples were excited at 315 nm and emission spectra recorded from 350 to 550 nm. Comparisons were made of fluorescence at 412 nm.

Analytical Ultracentrifugation

Peptides or mixtures of peptides at equimolar concentrations, dialyzed against reference buffer (50 mM sodium phosphate, 100 mM NaCl, pH 7.2), were spun at 25 °C in a Beckman XL-I analytical ultracentrifuge at 40,000, 45,000 and 50,000 rpm for approximately twenty-four hours at each speed. The following concentrations were used: **A-Ala**, 40, 100, 266 μM; **B-Phe**, 25 μM; **A-Abu,** 50, 100, 220 μM; 1:1 **A-Ala/B-Phe**, 50, 150, 320 μM; **A-Abu/B-Phe**, 50, 150, 320 μM. The contents of each cell were confirmed to be at equilibrium using WINMATCH prior to increasing the speed. Data were analyzed using the programs NONLIN (*54*) and SEDPHAT (*55, 56*). Several association models were fit, including a single ideal species, a single non-ideal species

and different monomer-oligomer equilibria. The results reported in Table 2 are from the model that best fits the data, as indicated. Molecular weights were determined using a partial specific volume, $\bar{\nu}$, calculated by SEDNTERP (*57*).

Crystallization

Crystals were grown using vapor diffusion with hanging-drop geometry from an equimolar mixture of peptides **A-Ala** and **B-Phe** (11 mg/ml in 10 mM phosphate at pH 7.2) by mixing 1.5 μl of protein with an equal volume of reservoir solution (100 mM Na HEPES buffer pH 7.5, 10% *v/v* *i*-propanol, 20% w/v PEG 4000). Bipyramidal crystals grew after approximately one week.

X-Ray Data Collection and Phasing

Crystals were frozen in a stream of N_2 gas cooled to -180 °C using FMS oil (Hampton Research) as a cryoprotectant. A 1.95 Å data set was collected at the Boston University Core Facility for Macromolecular Crystallography using a Rigaku RU-H3RHB x-ray generator with an MSC R-Axis IV++ area detector and 2-theta stage. The DENZO and SCALEPACK packages were used for data indexing, reduction, and scaling (*58*). Molecular replacement was performed with MOLREP (*59*). The search model comprised a dimer generated by applying the crystallographic symmetry operators of the **BBAT2** structure (PDB ID 1SNE).

Refinement

Manual fitting was performed using SigmaA weighted $2F_o$ - F_c composite-omit and F_o - F_c electron density maps (*60*) in the graphics program O (*61*). Refinement using an MLF target consisted of iterative rounds of minimization and simulated annealing (3000-5000 K) using slow-cool torsional molecular dynamics followed by individual B-factor refinement and manual

rebuilding, and was performed until R_{free} ceased to decrease. Water molecules added to the structure were checked by visual inspection of the map at each cycle of refinement. For statistical cross-validation purposes 10% of the data was excluded from refinement as a test set (*62, 63*). Topology and parameter files were created for nonstandard groups using bond lengths and angles from the literature (*37*). Values for D-Ala and D-Pro were derived from their L-enantiomers. Analysis of the Ramachandran plot defined by PROCHECK (*64*) showed a good final model with 88% of residues in the most favorable regions and 12% of residues in additionally allowed regions. Wilson plot values were calculated using the CCP4 program TRUNCATE (*65*) with resolution limits 2.5 – 1.95 Å. The refined structure contained 371 protein atoms and 52 solvent atoms per asymmetric unit (two monomers). Final model statistics are summarized in Table 2. All figures were created using MOLSCRIPT (*66*).

ACKNOWLEDGMENTS

We acknowledge the use of the MIT Computational and Systems Biology Initiative High Performance Computing and Proteomics/Structural Biology cores, and we thank Ezra Peisach for helpful discussions, Deborah J. Pheasant for experimental assistance, and Jiangang Chen for contributions to the energy design package. We acknowledge NSF support (CHE 9996335) to B.I. and NIH support to A.E.K. (GM67681) and K.N.A. (GM61099). C.M.T is an Anna Fuller predoctoral fellow.

REFERENCES

1. Imperiali, B., and Ottesen, J. J. (1999) Uniquely folded mini-protein motifs. *J Pept Res 54*, 177-84.
2. Hill, R. B., Raleigh, D. P., Lombardi, A., and DeGrado, W. F. (2000) De novo design of helical bundles as models for understanding protein folding and function. *Acc Chem Res 33*, 745-54.
3. Struthers, M. D., Cheng, R. P., and Imperiali, B. (1996) Design of a Monomeric 23-Residue Polypeptide with Defined Tertiary Structure. *Science 271*, 342-345.
4. Dahiyat, B. I., and Mayo, S. L. (1997) De novo protein design: fully automated sequence selection. *Science 278*, 82-7.
5. Mezo, A. R., Ottesen, J. J., and Imperiali, B. (2001) Discovery and Characterization of a Discretely Folded Homotrimeric Peptide. *J. Am. Chem. Soc. 123*, 1002-1003.
6. Ottesen, J. J., and Imperiali, B. (2001) Design of a Discretely Folded Mini-Protein Motif with Predominantly Beta-Structure. *Nat. Struct. Biol. 8*, 535-539.
7. Kortemme, T., Ramirez-Alvarado, M., and Serrano, L. (1998) Design of a 20-Amino Acid, Three-Stranded Beta-Sheet Protein. *Science 281*, 253-6.
8. Cochran, A. G., Skelton, N. J., and Starovasnik, M. A. (2001) Tryptophan zippers: stable, monomeric beta -hairpins. *Proc Natl Acad Sci U S A 98*, 5578-83.
9. Neidigh, J. W., Fesinmeyer, R. M., and Andersen, N. H. (2002) Designing a 20-residue protein. *Nat Struct Biol 9*, 425-30.
10. Zondlo, N. J., and Schepartz, A. (1999) Highly specific DNA recognition by a designed miniature protein. *Journal of the American Chemical Society 121*, 6938-6939.
11. Ghirlanda, G., Lear, J. D., Lombardi, A., and DeGrado, W. F. (1998) From Synthetic Coiled Coils to Functional Proteins: Automated Design of a Receptor for the Calmodulin-bonding Domain of Calcineurin. *J. Mol. Bio. 281*, 379-391.
12. Lombardi, A., Summa, C. M., Geremia, S., Randaccio, L., Pavone, V., and DeGrado, W. F. (2000) Inaugural article: retrostructural analysis of metalloproteins: application to the design of a minimal model for diiron proteins. *Proc Natl Acad Sci U S A 97*, 6298-305.
13. Moffet, D. A., Certain, L. K., Smith, A. J., Kessel, A. J., Beckwith, K. A., and Hecht, M. H. (2000) Peroxidase activity in heme proteins derived from a designed combinatorial library. *Journal of the American Chemical Society 122*, 7612-7613.
14. Ryadnov, M. G., and Woolfson, D. N. (2003) Engineering the morphology of a self-assembling protein fibre. *Nat Mater 2*, 329-32.
15. Petka, W. A., Harden, J. L., McGrath, K. P., Wirtz, D., and Tirrell, D. A. (1998) Reversible hydrogels from self-assembling artificial proteins. *Science 281*, 389-92.
16. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) High-resolution protein design with backbone freedom. *Science 282*, 1462-1467.
17. Snow, C. D., Nguyen, H., Pande, V. S., and Gruebele, M. (2002) Absolute comparison of simulated and experimental protein-folding dynamics. *Nature 420*, 102-6.
18. Zagrovic, B., and Pande, V. S. (2003) Structural correspondence between the alpha-helix and the random-flight chain resolves how unfolded proteins can have native-like properties. *Nat Struct Biol 10*, 955-61.

19. Zhou, N. E., Kay, C. M., and Hodges, R. S. (1994) The Role of Interhelical Ionic Interactions in Controlling Protein Folding and Stability. *De Novo* Designed Synthetic Two-Stranded α-Helical Coiled Coils. *J. Mol. Biol 237*, 500-512.
20. O'Shea, E. K., Lumb, K. J., and Kim, P. S. (1993) Peptide 'Velcro': Design of a Heterodimeric Coiled Coil. *Curr. Biol. 3*, 658-667.
21. Moll, J. R., Ruvinov, S. B., Pastan, I., and Vinson, C. (2001) Designed heterodimerizing leucine zippers with a ranger of pIs and stabilities up to 10(-15) M [In Process Citation]. *Protein Sci 10*, 649-55.
22. Keating, A. E., Malashkevich, V. N., Tidor, B., and Kim, P. S. (2001) Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. *Proc Natl Acad Sci U S A 98*, 14825-30.
23. McClain, D. L., Binfet, J. P., and Oakley, M. G. (2001) Evaluation of the Energetic Contribution of Interhelical Coulombic Interactions for Coiled Coil Helix Orientation Specificity. *J. Mol. Biol 313*, 371-383.
24. Nautiyal, S., Woolfson, D. N., King, D. S., and Alber, T. (1995) A Designed Heterotrimeric Coiled Coil. *Biochemistry 34*, 11645-11651.
25. Lombardi, A., Bryson, J. W., and DeGrado, W. F. (1996) *De Novo* Design of Heterotrimeric Coiled Coils. *Biopolymers 40*, 495-504.
26. Schnarr, N. A., and Kennan, A. J. (2003) Specific Control of Peptide Assembly with Combined Hydrophilic and Hydrophobic Interfaces. *J. Am. Chem. Soc. 125*.
27. Sia, S. K., and Kim, P. S. (2001) A Designed Protein with Packing between Left-Handed and Right-Handed Helices. *Biochemistry 40*, 8981-8989.
28. Fairman, R., Chao, H.-G., Lavoie, T. B., Villafranca, J. J., Matsueda, G. R., and Novotny, J. (1996) Design of Heterotetrameric Coiled Coils: Evidence for Increased Stabilization by Glu^--Lys^+ Ion Pair Interactions. *Biochemistry 35*, 2824-2829.
29. Hodges, R. S. (1996) *De Novo* Design of α-Helical Proteins: Basic Reseach to Medical Applications. *Biochem. Cell Biol. 74*, 133-154.
30. Kashiwada, A., Hiroaki, H., Kohda, D., Nango, M., and Tanaka, T. (2000) Design of a heterotrimeric alpha-helical bundle by hydrophobic core engineering. *Journal of the American Chemical Society 122*, 212-215.
31. Monera, O., Sonnichsen, F., Hicks, L., Kay, C., and Hodges, R. (1996) The relative positions of alanine residues in the hydrophobic core control the formation of two-stranded or four-stranded alpha-helical coiled-coils. *Protein Eng 9*, p353-63.
32. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1992) Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. *Cell 68*, 699-708.
33. Nautiyal, S., and Alber, T. (1999) Crystal Structure of a Designed, Thermostable, Heterotrimeric Coiled Coil. *Protein Sci. 8*, 84-90.
34. Marti, D. N., Jelesarov, I., and Bosshard, H. R. (2000) Interhelical ion pairing in coiled coils: solution structure of a heterodimeric leucine zipper and determination of pKa values of Glu side chains. *Biochemistry 39*, 12804-18.
35. Mezo, A. R., Cheng, R. P., and Imperiali, B. (2001) Oligomerization of Uniquely Folded Mini-Protein Motifs: Development of a Homotrimeric Beta Beta Alpha Peptide. *J. Am. Chem. Soc. 123*, 3885-3891.
36. Struthers, M. D., Ottesen, J. J., and Imperiali, B. (1998) Design and NMR Analyses of Compact, Independently Folded BBA Motifs. *Fold. Des. 3*, 95-103.

37. Ali, M. H., Peisach, E., Allen, K. N., and Imperiali, B. (2004) X-ray structure analysis of a designed oligomeric miniprotein reveals a discrete quaternary architecture. *Proc. Natl. Acad. Sci. USA 101*, 12183-12188.
38. Walshaw, J., and Woolfson, D. N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol 307*, 1427-50.
39. McDonnell, K. A., and Imperiali, B. (2002) Oligomeric Beta Beta Alpha Miniprotein Motifs: Pivotal Role of Single Hinge Residue in Determining the Oligomeric State. *J. Am. Chem. Soc. 124*, 428-433.
40. Dunbrack, R. L., Jr., and Karplus, M. (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol. 230*, 543-574.
41. Dunbrack, R. L., Jr., and Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci 6*, 1661-81.
42. Zhou, N. E., McKay, C., and Hodges, R. S. (1992) Synthetic Model Proteins: The Relative Contributions of Leucine Residues at the Nonequivalent Positions of the 3-4 Hydrophobic Repeat to the Stability of the Two-Stranded α-Helical Coiled Coil. *Biochemistry 31*, 5739-5746.
43. Rivas, G., and Minton, A. P. (2003) Tracer Sedimentation Equilibrium: A Powerful Tool For The Quantitative Characterization of Macromolecular Self- and Hetero-Associations in Solution. *Biochem. Soc. Trans. 31*, 1015-1019.
44. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) A Switch Between Two-, Three-, and Four-Stranded Coiled Coils in GCN4 Leucine Zipper Mutants. *Science 262*, 1401-1407.
45. Desjarlais, J. R., and Handel, T. M. (1999) Side-chain and backbone flexibility in protein core design. *J Mol Biol 290*, 305-18.
46. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science 302*, 1364-8.
47. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM. *J. Comput. Chem. 4*, 187-217.
48. Lazaridis, T., and Karplus, M. (1999) Effective energy function for proteins in solution. *Proteins 35*, 133-52.
49. Desmet, J., Maeyer, M. D., Hazes, B., and Lasters, I. (1992) The Dead-End Elimination Theorem and Its Use in Protein Side-Chain Positioning. *Nature 256*, 539-542.
50. Goldstein, R. F. (1994) Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys. J. 66*, 1335-1340.
51. Dominy, B. N., and Brooks, C. L. (1999) Development of a generalized born model parametrization for proteins and nucleic acids. *Journal of Physical Chemistry B 103*, 3765-3773.
52. Beroza, P., and Case, D. A. (1998) Calculations of proton-binding thermodynamics in proteins. *Methods Enzymol 295*, 170-89.
53. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry 6*, 1948-1954.
54. Johnson, M. L., Correia, J. C., Yphantis, D. A., and Halvorson, H. R. (1981) Analysis of Data from the Analytical Ultracentrifuge by Nonlinear Least-Squares Techniques. *Biophys. J. 36*, 575-588.

55. Schuck, P. (2003) On the Analysis of Protein Self-Association by Sedimentation Velocity Analytical Ultracentrifugation. *Anal. Biochem. 320*, 104-124.
56. Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R. A., Roualt, T. A., and Schuck, P. (2004) Sedimentation Equilibrium Analysis of Protein Interactions with Global Implicit Mass Conservation Constraints and Systematic Noise Decomposition. *Anal. Biochem. 326*, 234-256.
57. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer-Aided Interpretation of Analytical Sedimentation Data for Proteins, in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., and Horton, J. C., Eds.) pp 90-125, Royal Society of Chemistry, Cambridge.
58. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol. 276*, 307-326.
59. Vagin, A., and Teplyakov, A. (1997) MOLREP: An Automated Program For Molecular Replacement. *J. Appl. Cryst. 30*, 1022-1025.
60. Read, R. J. (1997) Model Phases: Probabilities and Bias. *Methods Enzymol. 278*, 110-128.
61. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models. *Acta Crystallogr., Sect. A 47*, 110-119.
62. Brünger, A. T. (1992) Free R Value: A Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures. *nature 355*, 472-475.
63. Brünger, A. T. (1997) Free R Value: Cross-Validation in Crystallography. *Methods Enzymol 277*, 366-396.
64. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck - A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Cryst. 26*, 283-291.
65. French, G. S., and Wilson, K. S. (1978) On The Treatment of Negative Intensity Observations. *Acta Crystallogr., Sect. A 34*, 517-525.
66. Kraulis, P. J. (1991) MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J. Appl. Cryst. 24*, 946-950.

ACCESSION NUMBERS

Coordinates of the refined structure of **BBAhetT1** have been deposited with the Protein Data Bank (accession code 1XOF).

TABLES

Table 1. Sequences of designed peptides

		hairpin								*helix*												
Peptide		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
BBAT2	Ac	Y	R	I	p	S	Y	D	F	a	D	**E**	**L**	**A**	K	L	L	R	**Q**	A	Z	G	NH_2
A-Ala	Ac	Y	R	I	p	S	Y	D	F	a	D	**E**	**A**	**E**	K	L	L	R	**D**	A	Z	G	NH_2
B-Phe	Ac	Y	R	I	p	S	Y	D	F	a	D	**K**	**F**	**K**	K	L	L	R	**K**	A	Z	G	NH_2
A-Abu	Ac	Y	R	I	p	S	Y	D	F	a	D	**E**	**B**	**E**	K	L	L	R	**D**	A	Z	G	NH_2
A-Leu	Ac	Y	R	I	p	S	Y	D	F	a	D	**E**	**L**	**E**	K	L	L	R	**D**	A	Z	G	NH_2
B-Leu	Ac	Y	R	I	p	S	Y	D	F	a	D	**K**	**L**	**K**	K	L	L	R	**K**	A	Z	G	NH_2

BBAhetT1 = **(A-Ala)**$_2$**(B-Phe)**$_2$

BBAhetT2 = **(A-Abu)**$_2$**(B-Phe)**$_2$

a = D-Ala, p = D-Pro, Z= DapBz, B=Abu

Table 2. AUC Results

Peptide	**A-Ala**	**B-Phe**	**BBAhetT1**	**A-Abu**	**BBAhetT2**
$\bar{\nu}^2$ (cm^3/g)[a]	0.7234	0.7510	0.7372[1]	0.7248	0.7379[1]
MW_{calc}	2561	2648	10418	2575	10446
NONLIN					
Best Model[b]	a	a	a	b	b
Sigma[c]	0.5012	0.5493	1.6758	0.4452	1.8010
MW_{obs}	2616	3208	9253	2336	9938
Stoichiometry[d]	1.0	1.2	3.6	0.9	3.8
SEDPHAT					
MW_{obs}	2703	3193	9393	2631	9491
Stoichiometry	1.1	1.2	3.6	1.0	3.6

a. Partial specific volumes were calculated using the program SEDNTERP (*57*). Partial specific volumes of complexes were approximated as the average of the partial specific volumes of the individual components.

b. Best model: a: Single species, B = 0. b: Single species, B ≠ 0.

c. Sigma is related to the molecular weight by the equation $M = \frac{\sigma RT}{(1-\bar{\nu}\rho)\omega^2}$.

d. Stoichiometry is defined as $\frac{MW_{obs}}{MW_{calc}}$. In the case of a mixture, MW_{calc} is the average of the calculated molecular weights of the components.

Table 3. X-ray statistics

Data Collection Statistics[1]	
Unit cell (Å, °)	a = b = 41.70 c = 51.33 α = β = 90 γ = 120
Space group	$P3_121$
Wavelength (Å)	1.5418
Resolution (Å)	∞ – 1.95
Total/ unique reflections	32,475 3,881
Completeness (%)	95.7 (100.0)
I/σ(I)	24.9 (7.1)
R_{merge}[2] (%)	5.8 (38.9)
Refinement Statistics	
Resolution (Å)	17.03 – 1.95
No. of reflections working/ test set	3369 380
R_{work}[3]/R_{free}[4] (%)	22.2 / 24.0
Average B-factors (Å^2)	
Wilson plot	32.4
Amino acids	37.8
Water	55.5
RMS Deviations from Ideality	
Bond lengths (Å)	0.009
Angles (°)	1.5
Dihedral angles (°)	21.7

[1] Values for the outermost shell (2.02 -1.95 Å) are shown in parentheses.

[2] $R_{merge} = \sum_{hkl} \sum_i | I_{hkl,\,i} - \langle I_{hkl} \rangle I | / \sum_{hkl} \sum_i | I_{hkl,\,i} |$, where $\langle I_{hkl} \rangle$ is the mean intensity of the multiple $I_{hkl,\,i}$ observations for symmetry related reflections.

[3] $R_{work} = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} |F_{obs}|$.

[4] $R_{free} = \sum_{hkl\ \Sigma T} |F_{obs} - F_{calc}| / \sum_{hkl} |F_{obs}|$, where the test set T includes 10% of the data.

FIGURES

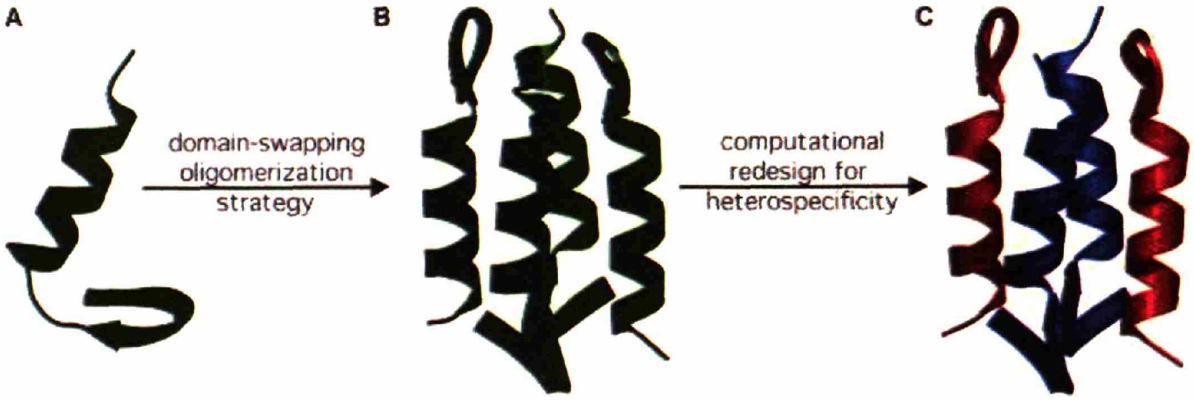


Figure 1. The design history of the BBA heterotetramers. (**A**) NMR structure of **BBA5**, an autonomously folding 23-mer designed to adopt the α/β fold of a zinc finger (*3, 36*). (**B**) X-ray crystal structure of **BBAT2**, a homotetrameric derivative of **BBA5** (*37*). (**C**) X-ray crystal structure of **BBAhetT1**, a heterotetrameric miniprotein derived from **BBAT2** by computer-aided design based on the structure of **BBAT2** (this study).

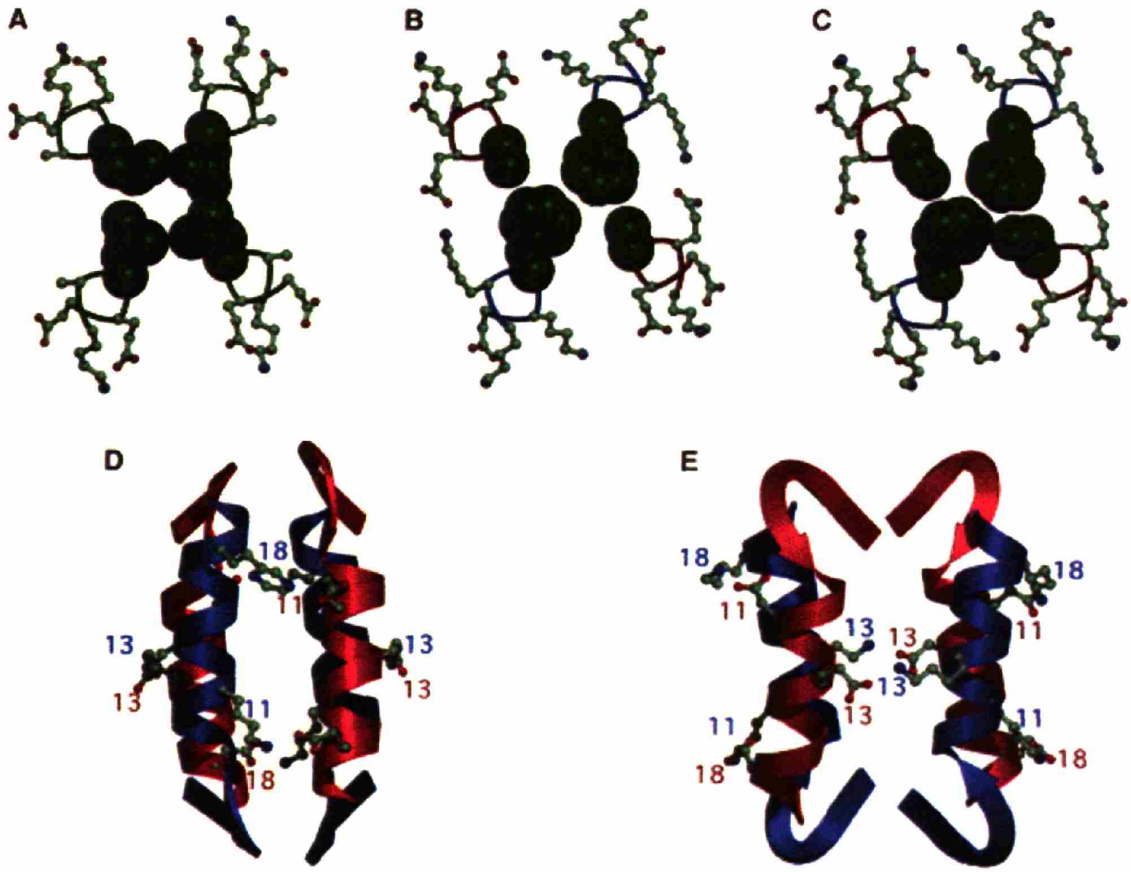


Figure 2. Design of heterospecificity. (**A-C**) Core redesign. (**A**) Four leucines at position 12 in **BBAT2** pack poorly in the hydrophobic core (*37*). (**B**) Ala was predicted to pack well against Phe in the core of a heterotetramer, and to disfavor homotetrameric states. (**C**) The stability of the heterotetramer was improved and specificity was retained when Abu was substituted for Ala in **BBAhetT2**. (**D-E**) Surface redesign. Positions 11, 18 and 13 of **BBAT2** are Glu, Gln and Ala, respectively. Calculations suggested that Asp or Lys at position 18 could form salt-bridging interactions with Glu or Lys at position 11, and that Lys or Glu residues at position 13 on adjacent subunits could also interact favorably, as shown. (**D**) and (**E**) are two views of the designed complex, rotated 90° around the tetramer axis with respect to one another.

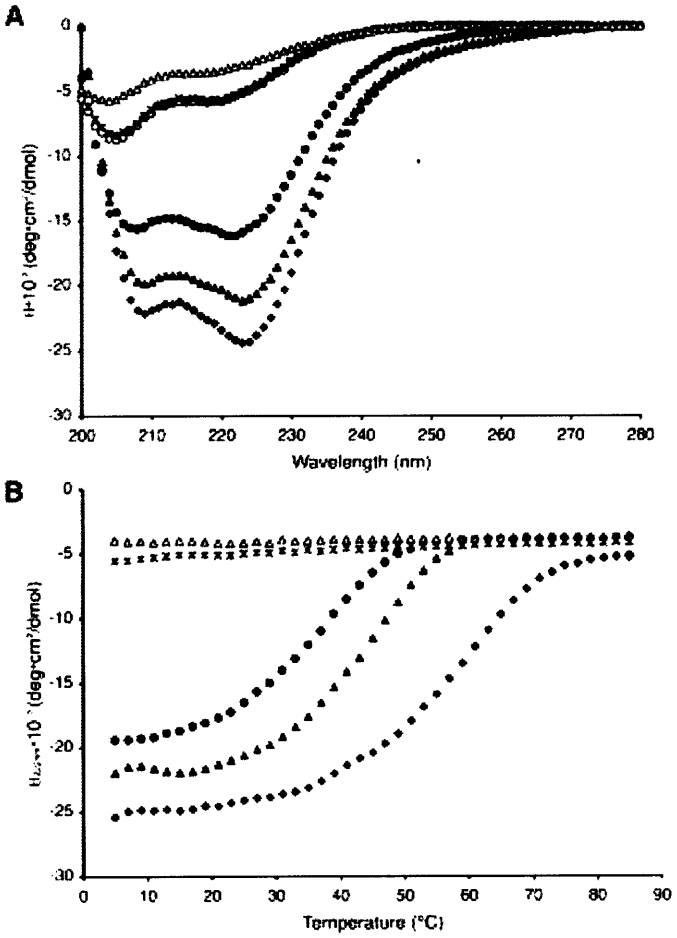


Figure 3. Solution characterization of BBA oligomers. (**A**) Circular dichroism spectra of **A-Ala** (o), **A-Abu** (*), **B-Phe** (Δ), **BBAhetT1** (•), **BBAhetT2** (▲), and **BBAT2** (◆) with total peptide concentration of 50 μM. (**B**) Thermal denaturation of **A-Abu**, **B-Phe**, **BBAhetT1**, **BBAhetT2**, and **BBAT2** at 50 μM total peptide concentration; symbols as in (**A**).

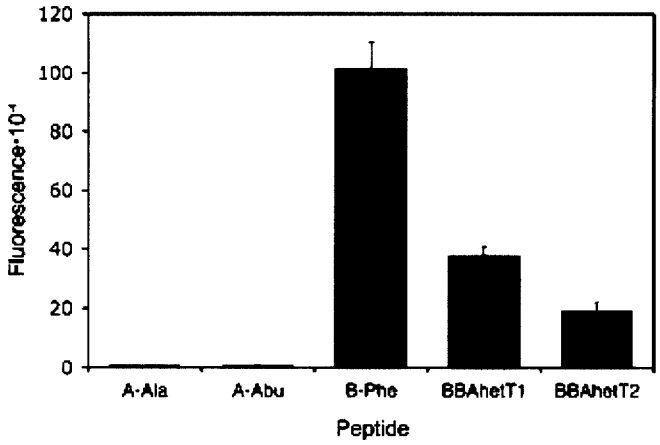


Figure 4. Fluorescence quenching experiments demonstrating heterospecific interactions between **A-Ala** and **B-Phe** and **A-Abu** and **B-Phe**. Peptides were synthesized with a quencher (**A-Ala** and **A-Abu**) or fluorophore (**B-Phe**) label. Combinations of **A-Ala** and **B-Phe** (**BBAhetT1**) and **A-Abu** and **B-Phe** (**BBAhetT2**) exhibit fluorescence quenching.

Figure 5. Stereo view of layer C of BBAhetT1 with composite omit map contoured at 1.0 σ.

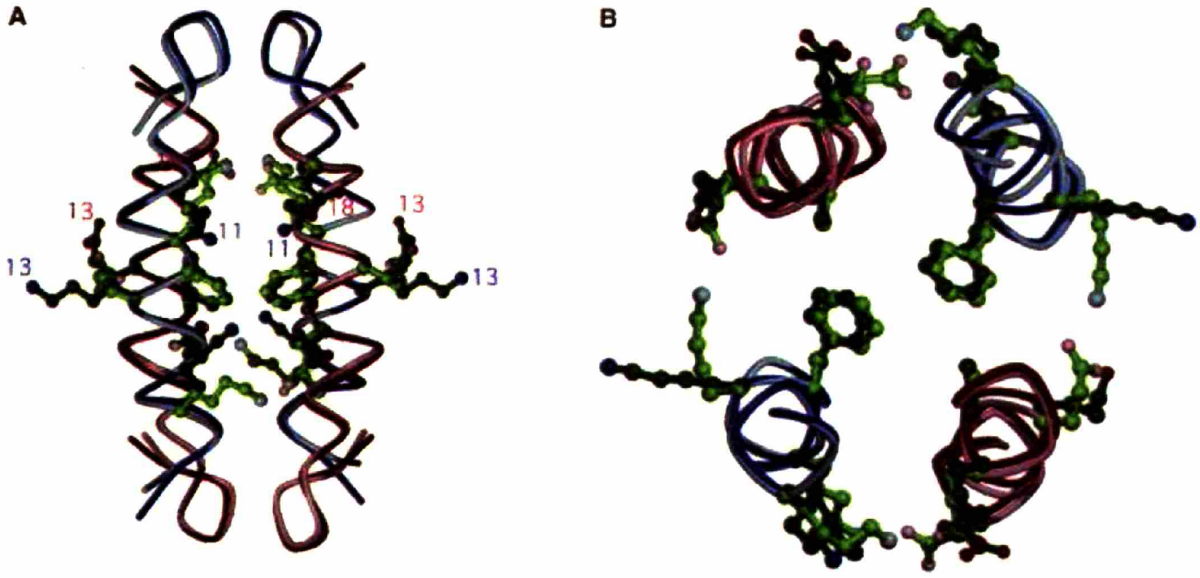


Figure 6. Crystal structure of BBAhetT1 (dark) superimposed on the predicted designed structure (light). Superimposed backbones are depicted as ribbons, with side chains at the designed positions 11, 12, 13 and 18 rendered as ball-and-stick models. Residues involved in inter-tetramer contacts are indicated with red arrows. (**A**) Entire backbone structure with designed sides chains. (**B**) The helical region of **BBAhetT1** showing layer 12 and the designed surface residues.

CHAPTER THREE

Orientation Specificity of the Bcr Coiled-Coil Oligomerization Domain

This work has been submitted to Biochemistry

ABSTRACT

The Bcr oligomerization domain, from the Bcr-Abl oncoprotein, homodimerizes via an antiparallel coiled coil with an adjacent short, helical swap domain. Inspection of the coiled-coil sequence does not reveal obvious determinants of helix-orientation specificity, raising the possibility that the antiparallel orientation preference is due to interactions of the swap domains. Coiled-coil constructs containing either an N- or C-terminal cysteine were synthesized without the swap domain. When crosslinked to adopt exclusively parallel or antiparallel orientations, these showed similar circular dichroism spectra. The antiparallel construct was ~16 °C more stable than the parallel to thermal denaturation. Equilibrium disulfide-exchange studies confirmed that the isolated coiled-coil homodimer shows a very strong preference for the antiparallel orientation. We conclude that the orientation preference of Bcr is not caused by the presence of the swap domains, but rather is directly encoded in the coiled-coil sequence. We further explored possible determinants of orientation specificity by mutating residues in the **d** position of the coiled-coil core. Some of the mutations caused a change in orientation specificity, and all of the mutations led to the formation of higher-order oligomers. In the absence of the swap domain, these residues play an important role in disfavoring alternate states and are especially important for encoding dimeric oligomerization specificity.

INTRODUCTION

The α-helical coiled coil is among the most common protein motifs found in nature and forms the basis of a wide range of protein-protein interactions. The motif consists of two or more α-helices packed together with a left-handed superhelical twist; the helices can associate in either a parallel or antiparallel orientation. Coiled coils play a fundamental functional role in many different proteins, including transcription factors, SNARE complexes, and proteins that mediate viral membrane fusion (*1, 2*).

Knowledge of whether the helices in a coiled coil are arranged in a parallel or antiparallel orientation can be critical in determining how it functions. For example, the bZip transcription factors contain a parallel coiled-coil oligomerization domain (*2, 3*), but there are other DNA binding proteins that dimerize in an antiparallel orientation (*4*). Determining the orientation of coiled coils in DNA binding proteins can help in elucidating function and can establish what surfaces are available for other proteins to bind. Structural maintenance of chromosomes (SMC) proteins, which are involved in chromosome condensation, sister chromatid cohesion, gene dosage, and DNA recombination, were originally expected to contain a parallel coiled-coil motif (*5-7*). After electron micrographs and crystallographic analyses revealed that the helices are actually antiparallel, however, proposed mechanisms had to be re-evaluated (*8, 9*). In another example, models of coiled coil-mediated membrane fusion are strongly dependent on the orientation of the proteins involved. Mitochondrial fusion proteins contain antiparallel coiled coils (*10*), whereas SNARE coiled coils are parallel (*11, 12*). Viral membrane fusion proteins contain a parallel coiled-coil trimer that forms antiparallel interactions with additional helices (*13*). Understanding the orientation of the coiled coils in each of these cases has precipitated models of membrane fusion (*10, 11, 13*).

Coiled coils have been extensively studied because of their simple repeating sequence and structure, and knowledge obtained from coiled-coil studies has been useful for understanding general protein structure principles. The motif contains a characteristic repeating heptad pattern of amino acids, $(\mathbf{abcdefg})_n$. Hydrophobic **a** and **d** residues create a stripe down one side of the α-helix and form the core of a coiled-coil oligomer. Frequently, charged residues at **e** and **g** form inter- and/or intrahelical salt bridges (*2*). In parallel coiled coils, charge complementarity at the **e** and **g** positions can impart heterooligomeric specificity (*14, 15*), and this general principle has been expanded in the design of non-coiled coil proteins (*16-18*). Similarly, polar residues at **a** positions in parallel coiled-coil cores, although destabilizing, can provide orientation and oligomerization specificity (*19-22*). Following this finding, polar residues have been found to impart specificity in globular folds (*23-26*).

Sequence elements that influence coiled-coil helix-orientation preference are not well understood (*27*). Parallel and antiparallel coiled coils have very different interactions in the core, with **a** to **a'** and **d** to **d'** for parallel, versus **a** to **d'** for antiparallel coiled coils (where the prime indicates a residue on an opposing helix). Parallel and antiparallel coiled coils also have different interactions on the surface, with **g** to **e'** interactions in parallel versus **e** to **e'** and **g** to **g'** in antiparallel structures (Figure 1A and B) (*27*). Despite an abundance of both parallel and antiparallel x-ray crystal structures, antiparallel examples are frequently short and intramolecular (*28*), making it unclear whether the coiled-coil sequence actually encodes helix orientation in these cases.

A handful of studies have addressed the determinants of helix-orientation specificity and established that the **a**, **d**, **e**, and **g** residues can each play a role (*27*). One theory suggests antiparallel coiled coils favor small hydrophobics in the **a** and **d** positions because they allow

tighter van der Waals packing between the α-helices (*29*), and another suggests a "steric matching" argument in which helix orientation can be established by the juxtaposition of small and large residues (*30-33*). Polar residues in the core may participate in specific hydrogen bonding interactions in one state preferentially to another, leading to an orientation preference (*34*). Charge complementarity of the surface **e** and **g** positions can also play a role in determining helix orientation (*27, 35-37*). Antiparallel coiled coils have been successfully designed by exploiting general principles, such as steric matching and charge complementarity (*35, 36, 38, 39*). Nevertheless, orientation cannot be accurately predicted from amino-acid sequence for most coiled coils.

The Bcr oligomerization domain is an example of an important natural protein for which orientation cannot be predicted from the amino-acid sequence. The Bcr-Abl oncoprotein is responsible for ~95% of chronic myeloid leukemias (CML) (*40*) and 17-30% of acute lymphoblastic leukemias (ALL) (*41*). The oncoprotein results from a reciprocal translocation event between *bcr*, on chromosome 22, and *abl*, on chromosome 9 (*40*). Oligomerization of Bcr relieves the autoinhibition of Abl, leading to aberrant downstream signaling (*42*). The Bcr oligomerization domain contains a thirty-six residue, antiparallel coiled coil, shown in Figure 2. Each Bcr monomer consists of a short helix, a loop, and then a long coiled-coil helix. The short helix associates with the long coiled-coil helix on the other monomer, also in an antiparallel fashion, forming a swap domain (*43*). Two such dimers further associate as a tetramer. The oligomerization domain is required for activation of the transforming function of Abl kinase (*44*) and may be an interesting target for treating disease.

The determinants of helix orientation in the Bcr coiled-coil dimers are not obvious in the structure (*43*) and cannot easily be explained using known general principles. It is possible that

the N-terminal swap domain establishes the geometry of the oligomerization interface. It is also possible that the coiled-coil domain adopts a different structure in solution than is observed in the crystal structure. Indeed, solution characterization and crystallographic studies of coiled coils can sometimes yield different results, as for spectrin (*45*), Coil-Ser (*46, 47*), and an alanine-zipper peptide (*48*). Domain swapped proteins have also been reported to change oligomerization state and/or swap region depending upon crystallization conditions (*49*).

In the first part of this paper, we investigate the intrinsic orientation preference of the Bcr coiled coil in solution and find that it is strongly antiparallel, even without the swap domain. To explore the origins of this preference, computational methods, statistical analysis, and structural evaluation were employed. In the second part, we mutated several residues in **d** positions that are non-optimal for parallel coiled coils and found that these play an important role in establishing the interaction specificity of Bcr.

MATERIALS AND METHODS

Modeling of Parallel and Antiparallel Bcr and Mutants

The Bcr crystal structure (1K1F, shown in Figure 2) was used for general structural examination. Additionally, the Bcr sequence and numerous mutants were modeled as dimers in both parallel and antiparallel orientations. Several different parallel and antiparallel backbones were created using Crick's parameterization of coiled coils (*50, 51*). Native side chains were placed onto these backbones in their optimal conformations with a DEE/A* algorithm, using energies calculated with a molecular mechanics energy function, as in a previous study (*17*). Structures for the mutants A38L and E52L were based on model Bcr parallel and antiparallel structures, but side chains near the mutation site were reoptimized.

Peptide Design, Synthesis, and Cleavage

Native and mutant Bcr peptides were synthesized using standard Fmoc synthesis and consisted only of the Bcr coiled-coil region (residues 30-65 1K1F). As in the crystal structure, the "native" sequence had several mutations from the wildtype: Cys 38 was mutated to alanine to prevent undesired disulfide bonds; Ile 57 was changed to alanine to disfavor the formation of Bcr tetramers; Phe 54 was changed to serine to eliminate hydrophobic surface exposed as a result of removing the swap domains. All mutant peptides were purchased from Bio-Synthesis, Inc, Lewisville, TX. The peptides were acetylated and amidated to eliminate charges on the termini, and N-terminal KWCGG or C-terminal GGC were added (as shown in Figure 1C). Reduced peptides are indicated with "C" in their name. For example, C-BCRA38L is a peptide with a reduced N-terminal cysteine and position 38 mutated from Ala to Leu. All mutant peptides have Ile 31 changed to Leu. Disulfide-bonded peptides are indicated with a superscript P or AP to indicate parallel or antiparallel helix orientation, respectively. C-BCRA38L^{P} is the oxidized form of C-BCRA38L, whereas BCRA38L^{AP} indicates a construct in which C-BCRA38L and BCRA38L-C are disulfide bonded. Cap-C-BCRA38L indicates C-BCRA38L with the N-terminal cysteine alkylated.

Alkylation of Cysteine Thiol groups

For some experiments, thiol groups were alkylated with iodoacetamide. Peptides were dissolved in 50 mM sodium phosphate buffer, pH 7.2, 2 mM TCEP for 30 minutes. At least a 10-fold molar excess of iodoacetamide was added to the reduced peptide. The solution was stirred in the dark at room temperature for 2 hours. Peptides were HPLC purified following the reaction, and the presence of the acetamide group was confirmed with electrospray mass spectrometry to 2 Da.

Purification and Handling

Prior to purification, peptides were either reduced in 0.2 M Tris/HCl and 100 mM DTT, pH 8.8, for 30 minutes or oxidized in 0.2 M Tris/HCl overnight to form disulfide-linked dimers. Both the reduction and oxidation reactions were quenched by adding acetic acid to a final concentration of 5%, yielding a final pH around 2. Reduced or oxidized peptides were purified on a C18 reverse phase HPLC column using a 0.1%/min acetylnitrile/water gradient with 0.1% trifluoroacetic acid. Samples were immediately frozen with liquid nitrogen after eluting from the HPLC, then lyophilyzed. All peptide solutions were made in an anaerobic chamber with degassed solvents to minimize oxygen exposure. The solutions were kept in the anaerobic chamber until immediately before use. This careful handling procedure eliminated most methionine oxidation problems. The purity of each peptide was confirmed to be greater than 95 percent by analytical HPLC. Masses for the native Bcr peptides were correct to within 1 Da using electrospray mass spectrometry, and the masses for the mutant Bcr peptides were verified by the supplier, Bio-Synthesis, Inc. Concentrations were determined using the method of Edelhoch (*52*).

Circular Dichroism Spectroscopy

Circular dichroism (CD) spectra from 300 to 200 nm were collected at 25 °C in triplicate on an Aviv circular dichroism spectrometer Model 202 using strain-free quartz cells with a path length of 0.1 cm and an averaging time of 5 seconds. Disulfide-linked peptides were dissolved in degassed buffer (50 mM sodium phosphate, 150 mM NaCl, pH 7.2) in an anaerobic chamber. Thermal unfolding experiments involved monitoring θ_{222} using a 30 second averaging time, 90 second equilibration time, and temperature increments of 2 °C from 5 to 85 °C. Melts were done in the presence of 2M GdnHCl. Several consecutive melts were done on the same sample and

compared. The T_m, the midpoint of the thermal unfolding curve, was estimated by extrapolating the pre- and post-transition baselines and then determining the temperature for which the CD signal was half of the difference. The T_m for each peptide was reproducible within ~ 1 to 2 °C.

Analytical ultracentrifugation

Alkylated and disulfide-linked peptides, dialyzed against reference buffer (50 mM sodium phosphate, 150 mM NaCl, pH 7.2) in an anaerobic chamber, were spun at 25 °C in a Beckman XL-I analytical ultracentrifuge at 17,000, 20,000, and 23,000 rpm for $BCRA38L^{AP}$, BCRA38L-C-cap, and cap-C-BCRE52L and 28,000, 31,000, and 34,000 rpm for BCR^{AP} and $BCR\text{-}C^{P}$ for approximately 24 hours at each speed. The following concentrations were used: BCR^{AP} and $BCR\text{-}C^{P}$ (5 μM , 25 μM, and 50 μM), $BCRA38L^{AP}$ (25 μM and 50 μM), BCRA38L-C-cap (30 μM, 50 μM , and 100 μM), cap-C-BCRE52L (30 μM , 50 μM , and 100 μM). The contents of each cell were confirmed to be at equilibrium prior to increasing the speed. Data were analyzed using the programs NONLIN (*53*) and SEDPHAT (*54, 55*). Various association models were fit, including a single, ideal species and monomer-oligomer equilibria. The results reported are from single-species fits. Partial specific volumes were calculated from the amino-acid sequence (*56*). Solvent density was calculated by SEDNTERP from its composition (*56*).

Disulfide-Exchange Experiment

Helix orientation was determined using an equilibrium disulfide-exchange assay (*22, 34, 57*), shown in Figure 5A. Different starting reactants were used to ensure that the products were not kinetically trapped. Peptides were allowed to equilibrate in an anaerobic chamber at room temperature in 50 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA, pH 7.2. The total

monomer peptide concentration was 50 μM. Samples were quenched by adding acetic acid to a final concentration of 5% after the reaction reached equilibrium. Samples were considered to be at equilibrium when consecutive HPLC traces showed no change. C-BCRA38L^{P} was sparingly soluble in phosphate buffer, so the assay was performed in the presence of 0.25 M urea and 5% acetonitrile. For the BCRA38L disulfide-exchange experiment, acetonitrile was added to a final concentration of 20% immediately before acetic acid was used to quench the reaction. The products were run on an analytical HPLC, monitored at 229 nm, and the identities of the peaks were confirmed by liquid chromatography-mass spectrometry. Adjustments to the absorbance were made for the difference in the number of amide bonds in the peptides, and an equilibrium constant was calculated.

RESULTS

Biophysical Characterization of the Bcr coiled coil

To probe determinants of helix-orientation specificity in the Bcr oligomerization domain, peptides corresponding to Bcr residues 30-65 were constructed with either GGC on the C-terminus (BCR-C) or KWCGG on the N-terminus (C-BCR) (*57*), as shown in Figure 1C. The glycine residues on the peptides provide flexibility in disulfide bond formation, while the trptophan and lysine residues provide a means to separate and identify the various peptides by HPLC. These constructs contain only the coiled-coil portion of the oligomerization domain, lacking the linker and short "swap domain" helix. The cysteine residues on the termini allow the peptides to be constrained in a parallel or antiparallel orientation via a disulfide bond.

Both the parallel and antiparallel disulfide-constrained peptides are highly helical in phosphate buffer, as shown in Figure 3A. The ratio of the minima at 208 and 222 nm indicates

that the helices are associating and is typical of spectra observed for coiled coils (*58*). The antiparallel orientation is slightly more helical than the parallel orientation and is much more stable to thermal denaturation (Figure 3B). The T_m of the antiparallel construct is ~53 °C, whereas that of the parallel construct is ~37 °C. Equilibrium sedimentation experiments run at and around concentrations used for other characterization studies indicated that both the disulfide-constrained parallel and antiparallel peptides had the molecular weight expected for a two-stranded coiled coil (Figure 4A, 4B, and Table 1).

To confirm that the Bcr coiled-coil peptide prefers an antiparallel orientation, oxidized and reduced peptides were mixed together and allowed to equilibrate in an anaerobic atmosphere. A clear antiparallel preference was evident from two experiments with different starting conditions. The first experiment was performed by mixing BCR^{AP} (12.5 μM), C-BCR (12.5 μM), and BCR-C (12.5 μM) at pH 7.2 (Figure 5B). Exchange took place within a few minutes and showed a preference for an antiparallel orientation with $K_{eq} \sim 1.3x10^{-3}$. To ensure that the reactants were not kinetically trapped, the disulfide-exchange experiment was repeated by mixing $BCR\text{-}C^{P}$ (12.5 μM) with C-BCR (25 μM) (Figure 5C). This experiment gave $K_{eq} \sim 1.6x10^{-3}$. These experiments establish that although the Bcr peptides can, in fact, fold as both parallel and antiparallel two-stranded coiled coils, the antiparallel orientation is preferred and helix-orientation preference in full-length Bcr is not determined by the swap domain.

Sequence-based and structural analysis of Bcr

The role of charged residues at **e** and **g** positions has been extensively studied as a source of interaction specificity for coiled coils. In Bcr, however, inspection of the **e** and **g** position residues does not suggest any strongly preferred helix orientation. A helical-wheel diagram (Figure 1A and B) can be used to assess potential interactions in each state. In the parallel

orientation, salt bridges typically occur between the **g** position of one helix and the next **e** position of a partner helix (**e'**$^{+}$). Four such interhelical salt bridges are possible in a parallel Bcr structure, and there are no putative **g** to **e'**$^{+}$ electrostatic repulsions in the parallel state (Figure 1B). The antiparallel x-ray structure of Bcr also contains four interhelical salt bridges and no potentially repulsive **g** to **g'**$^{+}$ or **e** to **e'**$^{+}$ interactions (Figure 1A). Thus, a simple sequence-based analysis of charged surface residues does not strongly favor either orientation.

Coiled-coil orientation specificity is determined by the relative stability of the parallel and antiparallel states. Several **d**-position residues in the Bcr crystal structure, Ile 31, Ala 38, and Glu 52, are unusual for parallel, dimeric coiled coils, suggesting that these residues may specify an antiparallel state by destabilizing the parallel one. We analyzed the unusual **d**-position residues by constructing computational models of the native and mutant Bcr sequences on both parallel and antiparallel backbones and by examining the SOCKET database of coiled coils with greater than 15 amino acids. SOCKET identifies coiled coils in the PDB automatically by detecting packing interactions and can be used to derive the frequency with which different amino acids occur in certain heptad positions (*28*).

The Bcr crystal structure contains an isoleucine (Ile 31) at a **d** position near both ends of the coiled coil. Harbury et al. have demonstrated that β-branched residues, such as isoleucine, do not pack well at the **d** position in a dimeric parallel orientation and typically lead to the formation of higher-order oligomers (*22*). Thus, Ile 31 could be an element of negative design, favoring an antiparallel orientation of Bcr by destabilizing the parallel state. On the other hand, the effect of Ile 31 may be mitigated by its location at the end of a helix, where it is relatively solvent exposed and not restricted to as stringent packing requirements as it would be in the middle of the coiled coil. To test its role in determining the orientation specificity of Bcr,

position 31 was mutated to a leucine in all mutants made. Leucine was chosen because it is found more frequently than any other residue in the **d** positions of dimeric parallel coiled coils (*28, 59*). Modeling on parallel and antiparallel backbones showed that leucine is easily accommodated in the core of both parallel and antiparallel Bcr dimers.

Ala 38 is another unusual residue at a **d** position in Bcr that may provide a natural negative design element disfavoring the parallel orientation. In the parallel orientation, two alanines would be directly across from one another, creating a large cavity in the core. The residues above and below the pair of alanines, Ile 31 and Leu 45, are not large enough to reach into the cavity and fill the void space. Analysis of coiled coils in the PDB supports the idea that Ala is highly unfavorable at the **d** position of parallel coiled coils and may be better accommodated in antiparallel structures. In dimeric coiled coils with greater than 15 amino acids, there are five times more alanines at **d** positions in antiparallel coiled coils compared to parallel in the SOCKET database (*28*). Further, our molecular mechanics calculations suggest that although an Ala to Leu mutation at position 38 stabilizes both orientations, it stabilizes the parallel orientation more. Thus, we chose to mutate Ala 38 to Leu to test whether this would stabilize the parallel orientation preferentially to the antiparallel. Leu was again chosen because it is so common in the **d** position of parallel dimers; it is found five times more frequently at the **d** position in parallel dimeric coiled coils than antiparallel (*28*).

Glu 52 is a third **d**-position residue that may provide an element of negative design disfavoring the parallel orientation. In the crystal structure of Bcr, the glutamates at position 52 reach into solvent, leaving very little hydrophobic side-chain packing in the core. In a model parallel structure, the glutamates behave similarly, but leave a larger cavity in the core because they are positioned directly across from each other. Interestingly, there are no glutamates in **d**

positions in parallel coiled coils within the SOCKET database, but glutamates are found nearly as often as hydrophobic residues such as Ile, Ala, Met, and Tyr at **d** positions in antiparallel coiled coils. In our models, changing Glu 52 to Leu improved core packing in the parallel orientation, but was not as easily accommodated in an antiparallel orientation. Using idealized backbones, a leucine at position 52 can be modeled in parallel structures with rotamer conformations that are common in the PDB (*60*). However, the most energetically favorable packing of leucine in antiparallel structures forces at least one side chain to adopt a statistically uncommon rotamer conformation, indicating that leucine cannot be accommodated without some strain. Therefore, Glu 52 may favor the antiparallel orientation of Bcr via greater destabilization, relative to Leu, of the parallel vs. the antiparallel state.

Helix-orientation specificity and characterization of the mutants

To investigate the role of **d**-position residues Ile 31, Ala 38, and Glu 52, two variants of C-BCR and BCR-C were synthesized. All the mutants replaced Ile 31 with Leu. C-BCRA38L and BCRA38L-C additionally contained Leu at position 38, and C-BCRE52L and BCRE52L-C contained Leu at position 52. Sequences are given in Figure 1C.

BCRA38L remained antiparallel, with an equilibrium constant of ($7.8x10^{-5}\pm3.2x10^{-5}$) (Figure 6A) in a disulfide-exchange assay. Because BCRA38L exhibited such a strong antiparallel preference, we characterized the disulfide-linked peptide $BCRA38L^{AP}$ and compared it with BCR^{AP}. Figure 7A shows that $BCRA38L^{AP}$ is slightly more helical than BCR^{AP} at 25°C, and the ratio of mean residue ellipticities at 208 and 222 nm is typical of coiled coils (*58*). $BCRA38L^{AP}$ is more stable than BCR^{AP} to thermal denaturation (Figure 7B). The T_m for $BCRA38L^{AP}$ is about 17 °C higher than that for BCR^{AP}.

The A38L mutation caused oligomerization specificity to be lost. AUC data for

BCRA38L^{AP} did not fit a single-species or two-state model well and gave a weight average close to that of a six-helix species (Table 1). BCRA38L-C-cap still did not fit a single-species model or two-state equilibrium model and had a single-species weight close to BCRA38L^{AP} (Table 1). BCRA38L-C-cap was less helical than BCRA38L^{AP}, shown in Figure 7A, and was much less stable than either disulfide-bonded BCRA38L^{AP} or BCR^{AP} to thermal denaturation.

In disulfide-exchange experiments, BCRE52L showed a loss of orientation specificity (Figure 6B). The equilibrium constant was determined to be 2.2, favoring the parallel orientation slightly. Because there was no clear orientation specificity, the cysteine group was capped for biophysical analysis to preclude the formation of a disulfide bond. Cap-C-BCRE52L was as helical as BCR^{AP}, shown in Figure 7A, and the T_m was ~64 °C in the presence of 2M GdnHCl, indicating this complex was much more stable than BCR^{AP} (Figure 7B). The Glu to Leu mutation also caused a change in oligomerization state; AUC data for cap-C-BCRE52L fit well to a single-species trimer with random residuals (Figure 4C).

DISCUSSION

In this study, the Bcr oligomerization domain was used to study coiled-coil orientation specificity. Analysis of the Bcr sequence using general principles of coiled-coil structure does not reveal why the coiled-coil domain adopts an antiparallel orientation. The short swap helix in the Bcr oligomerization domain could potentially play a role, or the helix orientation could be encoded in the coiled-coil sequence in a way that we do not yet know how to read. To explore how orientation is encoded in this fold, we characterized several model peptides. We established that even in the absence of the N-terminal swap domain, Bcr constructs retain a coiled-coil structure and adopt an antiparallel orientation, indicating that orientation preference is encoded directly in the coiled-coil part of the sequence.

Given that the orientation specificity of Bcr is encoded directly in its sequence, we considered several features that might favor an antiparallel over parallel orientation. First, we examined charged residues at **e** and **g** positions. Surface electrostatics alone have been used successfully to impart orientation specificity (*36, 37*), and charge complementarity is now a reliable technique for designing an antiparallel coiled coil (*31, 32, 38, 39*). According to general principles elucidated in the coiled-coil studies mentioned above, however, the antiparallel and parallel orientations appear roughly equivalent from an electrostatic perspective. Therefore, a role for the **e** and **g** positions in establishing orientation specificity in Bcr is not obvious.

Residues in the **a** and **d** positions of the Bcr core could also be important for encoding an antiparallel orientation preference, and much less is known about the roles of residues at these sites. Some of the **d**-position residues in Bcr do not appear to be optimal for stability, raising the possibility that they instead provide a negative-design element disfavoring the parallel state (*17, 39, 61-64*). Elements of negative design have been used to both manually and computationally design coiled coils and helical bundles that adopt different orientations and oligomerization states (*17, 31, 38, 39, 61, 64, 65*). In this study, the unusual **d**-position residues in Bcr were mutated to determine their role in defining coiled-coil orientation. Based on previous studies and current modeling, isoleucine (or any β-branched amino acid) at **d** cannot be easily accommodated in parallel dimers due to geometric packing restraints (*22*), so Ile 31 was mutated to leucine in all mutants tested.

The alanine at position 38 in the crystal structure, located in the middle of the coiled coil, may also prevent Bcr from adopting a parallel orientation. When electrostatics are equal in both the parallel and antiparallel orientations, placement of large and small core residues in the helix can cause a switch from antiparallel to parallel or vice versa (*33*). An antiparallel coiled-coil

trimer has been designed using steric matching (pairing alanine with a large non-natural residue) to give the desired orientation (*30*). In addition, the SOCKET database (*28*) shows a definite preference for alanines at **d** positions in antiparallel versus parallel dimeric coiled coils. Due to the axial stagger between interacting **a** and **d**-position residues in some antiparallel coiled coils, more interdigitated core packing is possible than in parallel coiled coils. Interdigitation may be a strategy for compensating for holes in the hydrophobic core (*29*).

Mutation of Ala 38 to Leu did not cause a change in helix orientation, but did lead to a change in oligomerization specificity. BCRA38L^{AP} was more helical and stable than BCR^{AP}. Increasing hydrophobic content in proteins is frequently stabilizing (*66*), and indeed, changing an alanine to a leucine at a **d** position in a model coiled-coil homodimer designed by Moitra et al. stabilized the complex by 9.2 kcal/mol and increased the T_m by ~ 30 °C (*67*). In Bcr, an Ala to Leu substitution stabilized antiparallel conformations over parallel ones, as reflected by the increased antiparallel preference measured in the disulfide exchange reactions.

Ala 38 seems to be more important for oligomerization specificity than for helix orientation specificity, as equilibrium centrifugation data for both BCRA38L^{AP} and BCRA38L-C-cap do not fit a single-species or a two-component equilibrium model. BCRA38L^{AP} and BCRA38L-C-cap have a single-species weight average around that of a hexamer. This type of change is not unprecedented. Subtle changes in core packing can cause a change in oligomerization state (*22, 68*). In a study by Monera et al., the relative placement of Ala and Leu residues was important for oligomerization specificity. Two alanine residues on the same layer in an antiparallel coiled coil formed a dimer, but when the positions of the alanines were staggered such that one layer contained Ala-Leu and the other Leu-Ala, a tetramer formed (*68*). This suggests it can be more favorable to form two smaller cavities than one large one. In a

similar manner, Ala at position 38 of Bcr may preclude high-order oligomerization due to an especially large energetic penalty for large cavities in the core. This is consistent with the formation of higher-order oligomers upon mutation of Ala 38 to Leu.

Along with constraints on hydrophobic packing, polar residues in the core can be critical for disfavoring an undesired state. For example, Oakley and Kim showed that if the location of an **a**-position Asn is moved such that a buried Asn-Asn hydrogen bond can form only in the antiparallel orientation, this state is strongly preferred (*34*). However, asparagines are not found in the core of antiparallel dimeric coiled coils often (*28*), and Asn-Asn interactions at the **a** position have not been observed in any antiparallel coiled-coil crystal structures thus far (*27*). Asn-Asn hydrogen bonds are seen very frequently in parallel coiled coils and probably play a large role in giving some coiled coils a parallel orientation.

Polar or charged residues other than Asn could also provide a mechanism for negative design against a particular orientation and, therefore, be important for establishing helix orientation. For example, this idea has been explored by Campbell and Lumb in a model dimeric coiled coil. In their study, a Lys in an **a** position interacting with a charged residue at **g'** gave oligomerization specificity. However, this combination of residues did not impart a specific helix orientation, due to the favorable interactions lysine could make in both the parallel and antiparallel states (*69*). In the context of a different model heterodimer, a buried Arg at a **d** position, with the potential to interact with a Glu at **g',** was not sufficient to specify helix orientation, but did give a specific dimer (*70*). Another study took an antiparallel coiled-coil mitochondrial fusion protein with two glutamates in **d** positions and mutated one at a time to leucine. Membrane fusion was adversely affected by these mutations. The cause of the loss of function was not apparent, however, as no structural analysis of the mutants was done (*10*).

Glutamate is common at the **d** position in antiparallel coiled coils, but is rarely found in parallel dimers; thus, it may be playing a special structural role. Bcr contains a glutamate in a **d** position, and examination of modeled structures suggested that this residue may disfavor the parallel orientation by blocking interhelical salt-bridge interactions between residues Glu 48 and Arg 53, as well as by packing poorly in the core. In C-BCRE52L, Glu 52 was replaced by Leu, and orientation specificity was lost. Although Ile 31 was also changed to Leu in this mutant, the I31L mutation did not alter orientation preference in the context of BCRA38L. Therefore, the loss of orientation specificity seen in BCRE52L is likely due to removal of the core glutamate. The E52L mutant also changed oligomerization state. Cap-C-BCRE52L is a single-species trimer by analytical ultracentrifugation. An up-up-down trimer, in which one α-helix is oriented antiparallel relative to two adjacent parallel helices, is consistent with all of the biophysical data. With a helical bundle in this topology, it would be possible to get nearly equal amounts of parallel and antiparallel helix pairs in the disulfide-exchange experiments.

Thus, Glu 52 plays at least two roles in the Bcr oligomerization domain; it disfavors the formation of higher-order oligomers and simultaneously imparts helix orientation specificity. It is easy to understand why Glu residues at **d** positions might prevent the formation of higher-order complexes. In the dimer, the glutamate carboxyl groups can reach out of the core and interact with water. However, these charged atoms would be almost completely buried in a coiled-coil trimer or tetramer, at a significant cost in solvation energy. Whether or how Glu 52 plays a negative design role in disfavoring a parallel dimer state is less clear. In the x-ray structure of Bcr, Glu 52 interacts with Arg 55 on the same helix. In models of parallel dimeric versions of Bcr, this residue interacts with Arg 53 on the opposite helix. Neither set of interactions is obviously superior, and it is possible that Glu 52 could in fact be accommodated

in a parallel Bcr dimer. Nevertheless, a role for this residue in influencing helix orientation is suggested by the loss of orientational specificity in the disulfide-exchange experiment.

Numerous studies have used protein design as a way to explore determinants of coiled-coil structural specificity, and this has been a powerful and effective approach (*15, 22, 27, 33, 34, 61, 68-70*). However, general principles uncovered in model systems do not explain the specificity observed in many native coiled-coil proteins. Some interactions that are effective in designed proteins are rare or even unprecedented in naturally occurring ones (*34, 61*). In this work, we examined determinants of structural specificity in the Bcr oligomerization domain and found them to be subtle, with some residues playing multiple roles. The influence of charged residues, such as Glu 52, is likely to be highly context dependent, varying according to the precise interactions that can be formed in multiple competing states. Methods for predicting helix-orientation specificity may need to address this complexity explicitly.

ACKNOWLEDGMENTS

We acknowledge the use of the MIT Computational and Systems Biology Initiative Proteomics/Structural Biology Core and High-Performance Computing Facility, and the MIT Center for Environmental Health Sciences Bioanalytical Facility. We thank P. Kim and M. Burgess for BCR-C and C-BCR. K. Taghizadeh, D. Pheasant, M. Ali, J. Glover, D. Pamuk, D. Lee, E. Oakes, G. Grigoryan and D. Mujumdar for experimental and computational assistance and/or helpful discussions. We thank members of the Keating Lab, G. Hersch, T. Keating, and T. Schwartz for helpful comments on the manuscript.

REFERENCES

1. Burkhard, P., Stetefeld, J., and Strelkov, S. V. (2001) Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol 11*, 82-8.
2. Mason, J. M., and Arndt, K. M. (2004) Coiled coil domains: stability, specificity, and biological implications. *Chembiochem 5*, 170-6.
3. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-ray Structure of the GCN4 Leucine Zipper, a Two-Stranded, Parallel Coiled Coil. *Science 254*, 539-544.
4. Bussiere, D. E., Bastia, D., and White, S. W. (1995) Crystal structure of the replication terminator protein from B. subtilis at 2.6 A. *Cell 80*, 651-60.
5. Peterson, C. L. (1994) The SMC family: novel motor proteins for chromosome condensation? *Cell 79*, 389-92.
6. Saitoh, N., Goldberg, I., and Earnshaw, W. C. (1995) The SMC proteins and the coming of age of the chromosome scaffold hypothesis. *Bioessays 17*, 759-66.
7. Hirano, T., and Mitchison, T. J. (1994) A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. *Cell 79*, 449-58.
8. Melby, T. E., Ciampaglio, C. N., Briscoe, G., and Erickson, H. P. (1998) The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. *J Cell Biol 142*, 1595-604.
9. Lowe, J., Cordell, S. C., and van den Ent, F. (2001) Crystal structure of the SMC head domain: an ABC ATPase with 900 residues antiparallel coiled-coil inserted. *J Mol Biol 306*, 25-35.
10. Koshiba, T., Detmer, S. A., Kaiser, J. T., Chen, H., McCaffery, J. M., and Chan, D. C. (2004) Structural basis of mitochondrial tethering by mitofusin complexes. *Science 305*, 858-62.
11. Bonifacino, J. S., and Glick, B. S. (2004) The mechanisms of vesicle budding and fusion. *Cell 116*, 153-66.
12. Sutton, R. B., Fasshauer, D., Jahn, R., and Brunger, A. T. (1998) Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature 395*, 347-53.
13. Eckert, D. M., and Kim, P. S. (2001) Mechanisms of viral membrane fusion and its inhibition. *Annu Rev Biochem 70*, 777-810.
14. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1992) Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. *Cell 68*, 699-708.
15. O'Shea, E. K., Lumb, K. J., and Kim, P. S. (1993) Peptide 'Velcro*': design of a heterodimeric coiled coil. *Curr. Biol. 3*, 658-667.
16. Hendsch, Z. S., Nohaile, M. J., Sauer, R. T., and Tidor, B. (2001) Preferential heterodimer formation via undercompensated electrostatic interactions. *J Am Chem Soc 123*, 1264-5.
17. Ali, M. H., Taylor, C. M., Grigoryan, G., Allen, K. N., Imperiali, B., and Keating, A. E. (2005) Design of a heterospecific, tetrameric, 21-residue miniprotein with mixed alpha/beta structure. *Structure (Camb) 13*, 225-34.
18. Nohaile, M. J., Hendsch, Z. S., Tidor, B., and Sauer, R. T. (2001) Altering dimerization specificity by changes in surface electrostatics. *Proc Natl Acad Sci U S A 98*, 3109-14.

19. Lumb, K. J., and Kim, P. S. (1995) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry 34*, 8642-8648.
20. Gonzalez, L., Jr., Woolfson, D. N., and Alber, T. (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat. Struct. Biol. 3*, 1011-1018.
21. Akey, D. L., Malashkevich, V. N., and Kim, P. (2001) Buried polar residues in coiled-coil interfaces. *Biochemistry 40*, 6352-6360.
22. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science 262*, 1401-1407.
23. Efimov, A. V., and Kondratova, M. S. (2003) [A comparative analysis of interhelical polar interactions of various alpha-helix packings in proteins]. *Mol Biol (Mosk) 37*, 515-21.
24. Bolon, D. N., and Mayo, S. L. (2001) Polar residues in the protein core of Escherichia coli thioredoxin are important for fold specificity. *Biochemistry 40*, 10047-53.
25. Hendsch, Z. S., and Tidor, B. (1994) Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci 3*, 211-26.
26. Rozwarski, D. A., Gronenborn, A. M., Clore, G. M., Bazan, J. F., Bohm, A., Wlodawer, A., Hatada, M., and Karplus, P. A. (1994) Structural comparisons among the short-chain helical cytokines. *Structure 2*, 159-73.
27. Oakley, M. G., and Hollenbeck, J. J. (2001) The design of antiparallel coiled coils. *Curr Opin Struct Biol 11*, 450-7.
28. Walshaw, J., and Woolfson, D. N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol 307*, 1427-50.
29. Gernert, K. M., Surles, M. C., Labean, T. H., Richardson, J. S., and Richardson, D. C. (1995) The Alacoil: a very tight, antiparallel coiled-coil of helices. *Protein Sci 4*, 2252-60.
30. Schnarr, N. A., and Kennan, A. J. (2004) Strand orientation by steric matching: a designed antiparallel coiled-coil trimer. *J Am Chem Soc 126*, 14447-51.
31. Betz, S. F., and DeGrado, W. F. (1996) Controlling topology and native-like behavior of de novo-designed peptides: design and characterization of antiparallel four-stranded coiled coils. *Biochemistry 35*, 6955-62.
32. Gurnon, D. G., Whitaker, J. A., and Oakley, M. G. (2003) Design and characterization of a homodimeric antiparallel coiled coil. *J Am Chem Soc 125*, 7518-9.
33. Monera, O., Zhou, N., Lavigne, P., Kay, C., and Hodges, R. (1996) Formation of parallel and antiparallel coiled-coils controlled by the relative positions of alanine residues in the hydrophobic core. *J Biol Chem 271*, p3995-4001.
34. Oakley, M. G., and Kim, P. S. (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry 37*, 12603-10.
35. Monera, O. D., Zhou, N. E., Kay, C. M., and Hodges, R. S. (1993) Comparison of antiparallel and parallel two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization. *J Biol Chem 268*, 19218-27.
36. Monera, O., Kay, C., and Hodges, R. (1994) Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils. *Biochemistry 33*, p3862-71.
37. McClain, D. L., Binfet, J. P., and Oakley, M. G. (2001) Evaluation of the energetic contribution of interhelical Coulombic interactions for coiled coil helix orientation specificity. *J Mol Biol 313*, 371-83.

38. McClain, D. L., Woods, H. L., and Oakley, M. G. (2001) Design and characterization of a heterodimeric coiled coil that forms exclusively with an antiparallel relative helix orientation. *J Am Chem Soc 123*, 3151-2.
39. Ghosh, I., Hamilton, A, and Regan, L. (2000) Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *J. Am. Chem. Soc. 122*, 5658-5659.
40. Sawyers, C. L. (1999) Chronic myeloid leukemia. *N Engl J Med 340*, 1330-40.
41. Catovsky, D. (1979) Ph1-positive acute leukaemia and chronic granulocytic leukaemia: one or two diseases? *Br J Haematol 42*, 493-8.
42. Smith, K. M., Yacobi, R., and Van Etten, R. A. (2003) Autoinhibition of Bcr-Abl through its SH3 domain. *Mol Cell 12*, 27-37.
43. Zhao, X., Ghaffari, S., Lodish, H., Malashkevich, V. N., and Kim, P. S. (2002) Structure of the Bcr-Abl oncoprotein oligomerization domain. *Nat Struct Biol.*
44. McWhirter, J. R., Galasso, D. L., and Wang, J. Y. (1993) A coiled-coil oligomerization domain of Bcr is essential for the transforming function of Bcr-Abl oncoproteins. *Mol Cell Biol 13*, 7587-95.
45. Yan, Y., Winograd, E., Viel, A., Cronin, T., Harrison, S. C., and Branton, D. (1993) Crystal structure of the repetitive segments of spectrin. *Science 262*, 2027-30.
46. Lovejoy, B., Choe, S., Cascio, D., McRorie, D. K., DeGrado, W. F., and Eisenberg, D. (1993) Crystal structure of a synthetic triple-stranded alpha-helical bundle. *Science 259*, 1288-93.
47. O'Neil, K. T., and DeGrado, W. F. (1990) A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science 250*, 646-51.
48. Liu, J., and Lu, M. (2002) An alanine-zipper structure determined by long range intermolecular interactions. *J Biol Chem 277*, 48708-13.
49. Liu, Y., and Eisenberg, D. (2002) 3D domain swapping: as domains continue to swap. *Protein Sci 11*, 1285-99.
50. Harbury, P. B., Tidor, B., and Kim, P. S. (1995) Repacking protein cores with backbone freedom: structure prediction for coiled coils. *Proc. Natl. Acad. Sci. U S A 92*, 8408-8412.
51. Crick, F. H. C. (1953) The packing of a-helices: simple coiled coils. *Acta Cryst. 6*, 689-697.
52. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry 6*, 1948-1954.
53. Johnson, M. L., Correia, J. C., Yphantis, D. A., and Halvorson, H. R. (1981) Analysis of Data from the Analytical Ultracentrifuge by Nonlinear Least-Squares Techniques. *Biophys. J. 36*, 575-588.
54. Schuck, P. (2003) On the Analysis of Protein Self-Association by Sedimentation Velocity Analytical Ultracentrifugation. *Anal. Biochem. 320*, 104-124.
55. Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R. A., Roualt, T. A., and Schuck, P. (2004) Sedimentation Equilibrium Analysis of Protein Interactions with Global Implicit Mass Conservation Constraints and Systematic Noise Decomposition. *Anal. Biochem. 326*, 234-256.
56. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer aided interpretation of analytical sedimentation data for proteins, in *Analytical*

Ultracentrifugation in Biochemistry and Polymer Science pp 90-125, Royal Society of Chemistry, Cambridge, U. K.

57. O'Shea, E. K., Rutkowski, R., Stafford, W. F. d., and Kim, P. S. (1989) Preferential heterodimer formation by isolated leucine zippers from fos and jun. *Science 245*, 646-8.
58. Zhou, N. E., Zhu, B. Y., Kay, C. M., and Hodges, R. S. (1992) The two-stranded alpha-helical coiled-coil is an ideal model for studying protein stability and subunit interactions. *Biopolymers 32*, 419-26.
59. Hu, J. C., O'Shea, E. K., Kim, P. S., and Sauer, R. T. (1990) Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. *Science 250*, 1400-3.
60. Dunbrack, R. L., Jr., and Karplus, M. (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol. 230*, 543-574.
61. Havranek, J. J., and Harbury, P. B. (2003) Automated design of specificity in molecular recognition. *Nat Struct Biol 10*, 45-52.
62. Pokala, N., and Handel, T. M. (2005) Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *J Mol Biol 347*, 203-27.
63. Park, S., Yang, X., and Saven, J. G. (2004) Advances in computational protein design. *Curr Opin Struct Biol 14*, 487-94.
64. Summa, C. M., Rosenblatt, M. M., Hong, J. K., Lear, J. D., and DeGrado, W. F. (2002) Computational de novo design, and characterization of an A(2)B(2) diiron protein. *J Mol Biol 321*, 923-38.
65. Nautiyal, S., Woolfson, D. N., King, D. S., and Alber, T. (1995) A designed heterotrimeric coiled coil. *Biochemistry 34*, 11645-51.
66. Acharya, A., Ruvinov, S. B., Gal, J., Moll, J. R., and Vinson, C. (2002) A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: amino acids I, V, L, N, A, and K. *Biochemistry 41*, 14122-31.
67. Moitra, J., Szilak, L., Krylov, D., and Vinson, C. (1997) Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. *Biochemistry 36*, 12567-12573.
68. Monera, O., Sonnichsen, F., Hicks, L., Kay, C., and Hodges, R. (1996) The relative positions of alanine residues in the hydrophobic core control the formation of two-stranded or four-stranded alpha-helical coiled-coils. *Protein Eng 9*, p353-63.
69. Campbell, K. M., and Lumb, K. J. (2002) Complementation of buried lysine and surface polar residues in a designed heterodimeric coiled coil. *Biochemistry 41*, 7169-75.
70. McClain, D. L., Gurnon, D. G., and Oakley, M. G. (2002) Importance of potential interhelical salt-bridges involving interior residues for coiled-coil stability and quaternary structure. *J Mol Biol 324*, 257-70.
71. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer-Aided Interpretation of Analytical Sedimentation Data for Proteins, in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., and Horton, J. C., Eds.) pp 90-125, Royal Society of Chemistry, Cambridge.

TABLES

Table 1: Analytical Ultracentrifugation Data Fit to a Single Species Model for Bcr peptides

	BCR^{AP}	$BCR\text{-}C^{P}$	$BCRA38L^{AP}$	BCRA38L-C-cap	cap-C-BCRE52L
υ (cm^3/g)[a]	0.73055	0.72908	0.734	0.73055	0.73780
MW_{calc}	4731.38	4573.19	4773.46	4733.27	4930.61
WinNONLIN					
MW_{obs}	10042	10395	25608[c]	26031[c]	14093.97
Stoichiometry[b]	2.12	2.27	5.36	5.50	2.86
SEDPHAT					
MW_{obs}	9995	10245	28029	25698.2	14111
Stoichiometry[c]	2.11	2.24	5.87	5.43	2.86

[a]Partial specific volumes were calculated as previously described (*71*). Partial specific volumes of disulfide bonded mixtures are averages of the components.
[b]Stoichiometry is calculated as (MW_{obs}/[MW_{calc} for a single helix]).
[c]Fit with non-random residuals

FIGURES

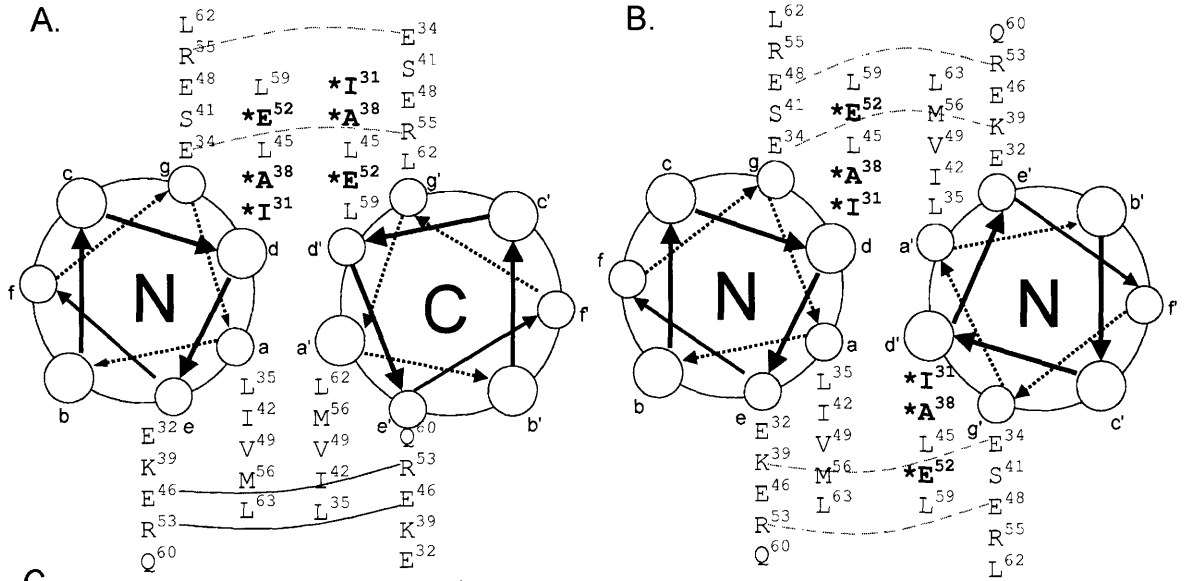


Figure 1: Helical-wheel diagram of Bcr in an antiparallel (A) and parallel (B) orientation. (A) Solid lines represent interhelical salt bridges that always form in the crystal structure. Dashed lines represent potential interhelical salt bridges that form in half of the copies in the asymmetric unit. (B) Dashed lines represent potential **g** to **e**' salt bridges that could form in a parallel dimer. (C) Peptides used in this study. C-BCR and BCR-C have the same sequence for residues 30-65 as was used in the crystal structure.

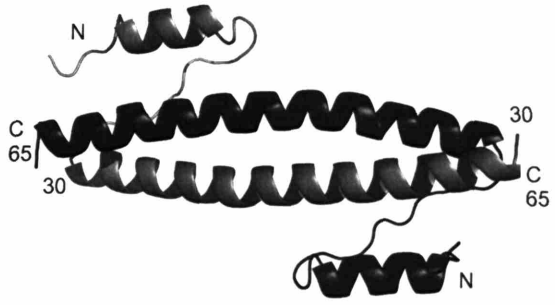
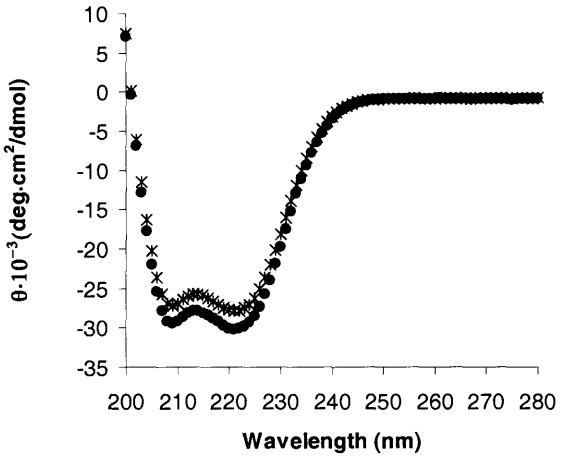


Figure 2: The Bcr oligomerization domain dimer, PDB ID 1K1F (*43*). In the crystal structure, two such dimers associate to form a tetramer.

A.



B.

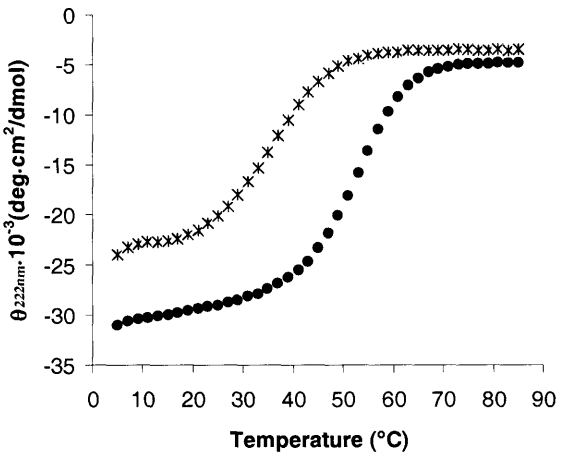


Figure 3: (A) Circular dichroism spectra of BCR-C^{P} (parallel, *) and BCR^{AP} (antiparallel, •) (25 μM peptide concentration, 50 mM sodium phosphate, 150 mM NaCl, pH 7.2) at 25°C. (B) Thermal denaturation of BCR^{AP} (antiparallel, *) and BCR^{P} (parallel, •) monitored at 222 nm in 33 mM sodium phosphate, 100 mM NaCl, 2 M GdnHCl, pH 7.2. Melting temperatures indicate that the antiparallel conformation is more stable than the parallel by ~16°C.

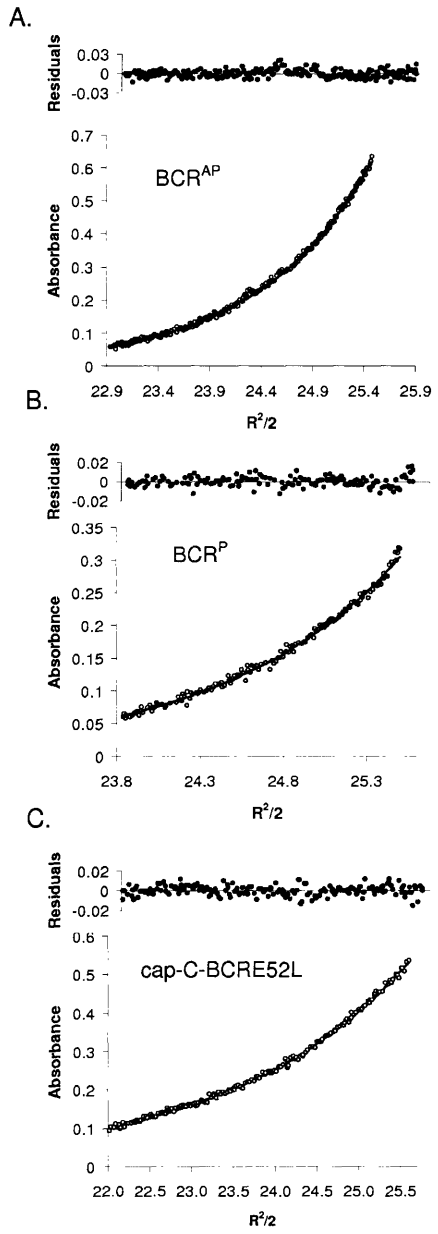


Figure 4: Analytical ultracentrifugation data for Bcr coiled coils. Global fits to data collected at three speeds and three concentrations are shown with representative experimental traces and residuals to the fit. Data shown were collected at 17,000 rpm with 50 μM total monomer concentration. (A) BCR^{AP} and (B) BCR^{P} are both two-stranded coiled coils at 25 μM, in 50 mM sodium phosphate, 150 mM NaCl, pH 7.2; (C) cap-C-BCRE52L is a single-species three-stranded coiled coil at 50 μM in 50 mM sodium phosphate, 150 mM NaCl, pH 7.2. All data were fit with WinNonLin.

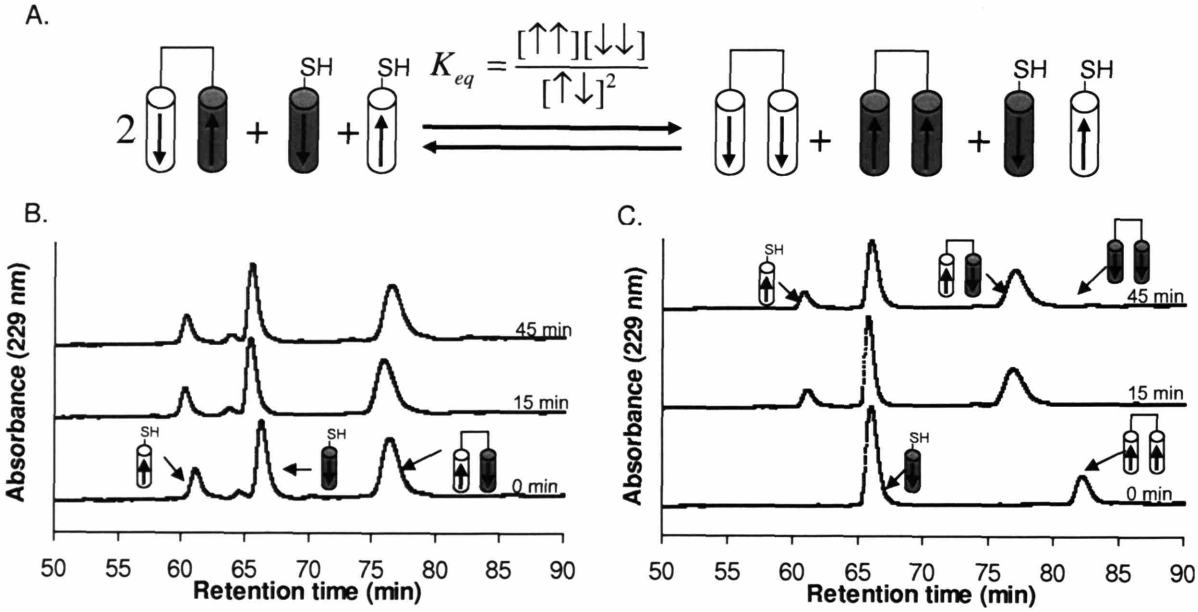


Figure 5: Disulfide-equilibrium exchange experiment used to determine the helix orientation of the Bcr coiled coil. (A) N-terminal cysteine peptides (C-BCR) are in gray, and C-terminal cysteine peptides (BCR-C) are in white. Combinations of disulfide bonded and reduced peptides were mixed together and allowed to equilibrate at 25°C, then the reaction was quenched with acetic acid and run on reverse phase HPLC. (B) & (C) Disulfide exchange reactions were initiated from two conditions. The concentration of different species was monitored as a function of time by HPLC. Both experiments gave similar K_{eq} values, 1.3×10^{-3} in (B) and 1.6×10^{-3} in (C), confirming that the reaction had reached an equilibrium strongly favoring the antiparallel species.

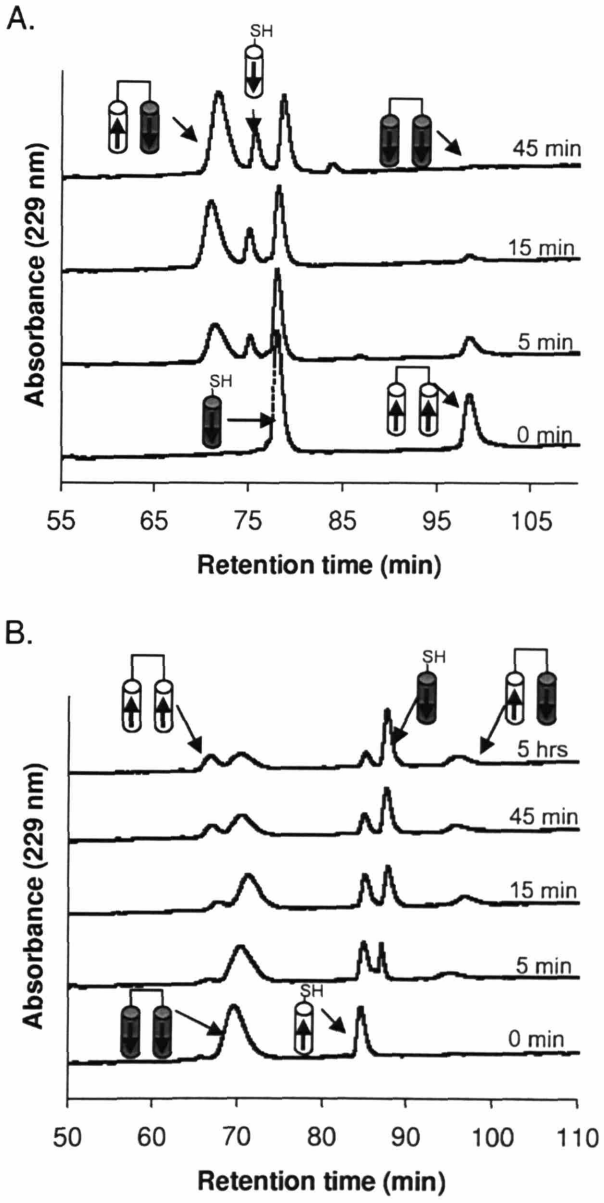
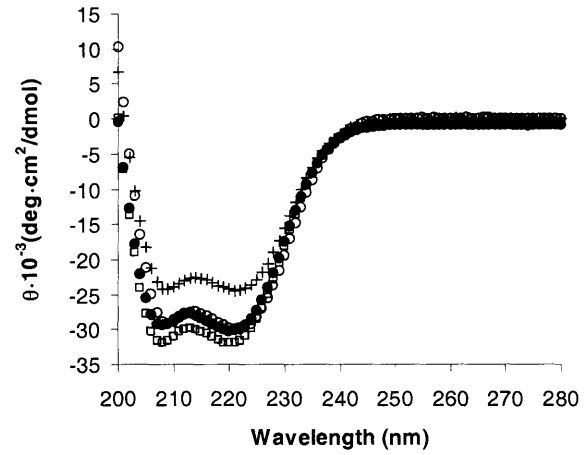


Figure 6: Helix orientation of Bcr mutants determined as in Figure 5. Representative disulfide exchange data in 50 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA pH 7.2, 25°C. (A) BCRA38L-C^P (12.5 μM) and C-BCRA38L (25 μM); $K_{eq} = 7.8x10^{-5}\pm3.2x10^{-5}$. (B) C-BCRE52$L^P$ (12.5 μM) and BCRE52L-C (25 μM); K_{eq}=2.2.

A.



B.

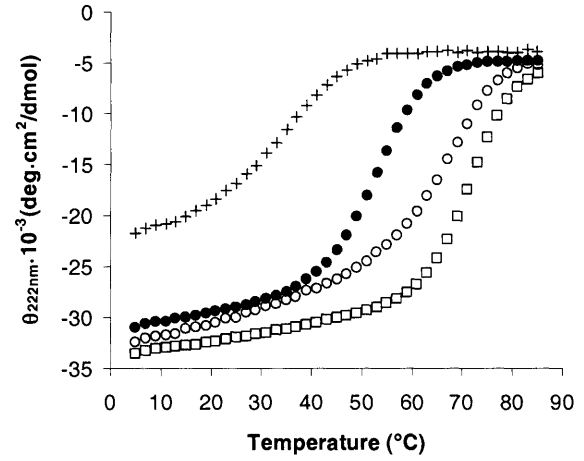


Figure 7: Circular dichroism spectra and thermal melt of Bcr mutants. (A) Circular dichroism spectra of BCRA38L^{AP} (□), BCRA38L-C-cap (+), cap-C-BCRE52L (○), compared to BCR^{AP} (•). The peptide concentration was 50 μM for alkylated peptides and 25 μM for disulfide bonded peptides in 50 mM sodium phosphate, 150 mM NaCl, pH 7.2 at 25 °C. (B) Thermal melt of BCRA38L^{AP} (□), BCRA38L-C-cap (+), cap-C-BCRE52L (○), compared to BCR^{AP} (•) monitored at 222 nm. BCR^{AP} was in 33 mM sodium phosphate, 100 mM NaCl, 2 M GdnHCl, pH 7.2. All other peptides were in 50 mM sodium phosphate, 150 mM NaCl, pH 7.2 at 25 °C.

CHAPTER FOUR

Computational Design of a Bcr Inhibitor

Collaborators: Devdoot Majumdar helped with the initial stages of design and helped develop the BCR-C and C-BCR expression and purification protocols, and Nora Zizlsperger developed purification strategies for Bcr with the swap domain and the Bcr inhibitor peptides. Nora and Jeremy Fisher developed and performed the pull-down assay.

ABSTRACT

A defining feature of 95% of Chronic Myeloid Leukemias (CML) and 20% of Acute Lymphoblastic Leukemias (ALL) is a reciprocal translocation event between chromosome 9 and chromosome 22. The reciprocal translocation joins the *c-abl* and *bcr* genes that code for a fusion protein, Bcr-Abl. Homo-oligomerization of Bcr through the coiled-coil region is the critical event that causes Abl kinase to be constitutively active, triggering numerous cytoplasmic and nuclear signal transduction pathways that lead to the eventual onset of leukemia. Although existing drugs used successfully to treat leukemia, such as Gleevec, target the Abl kinase protein, the Bcr oligomerization domain may be a good alternative or complementary target when people become resistant to current drugs. Using the Bcr oligomerization domain crystal structure, four antiparallel coiled-coil inhibitors were designed using computational methods. The goal was to design a dominant negative inhibitor to the Bcr oligomerization domain. Peptides were expressed in *E. coli*, purified, and tested using a pull-down assay. The inhibitors failed to bind to the Bcr oligomerization domain, but did bind to the Bcr coiled-coil region in a less stringent assay. Possible reasons as to why the designed inhibitors failed to bind in the stringent assay are discussed, and improvements for future designs are suggested.

INTRODUCTION

Approximately 34,810 new cases of leukemia will occur in the United States this year, according to the American Cancer Society. Of these, an estimated 4,600 new cases of Chronic Myeloid Leukemia (CML) and 3,970 new cases of Acute Lymphoblastic Leukemia (ALL) will be diagnosed (*1*). A defining feature of 95% of CML (*2*) and 25% of ALL in adults (*3*) is a reciprocal translocation event between *c-abl* on chromosome 9 and *bcr* on chromosome 22, shown in Figure 1A. The shortened chromosome 22, known as the Philadelphia Chromosome (Ph^+), codes for the Bcr-Abl fusion protein (*2*), which is localized to the cytoplasm (*4*).

The role of Abl in wild-type cells is established, but Bcr's role in the cell is not entirely clear; however, both play significant roles in causing leukemia when fused together. In wild-type Abl, tyrosine kinase domains carry signals between the cytoplasm and nucleus, playing a key apoptotic-regulation role in response to DNA damage (*5*). Abl kinase is activated by phosphorylating Tyr1294 and Tyr1127 (*6*). Found in both the cytoplasm and nucleus, Bcr (breakpoint cluster region) has many different domains, including oligomerization, serine and threonine, SH2 binding, guanine exchange factor (GEF), and guanosine triphosphatase-activating protein (GAP) domains (*4*). Bcr self-association is mediated by a coiled-coil region within its oligomerization domain and is the critical event that causes Abl kinase to be constitutively active (Figure 1C) (*6, 7*). Constitutive activity of Abl kinase triggers numerous cytoplasmic and nuclear signal transduction pathways that disrupt cell-cycle regulation and eventually lead to leukemia (*2*). However, in the absence of the Bcr coiled-coil region, kinase activity can be restored with point mutations (P1013 or P1124) to the Abl kinase SH3 domain (*6*).

The x-ray crystal structures of the Bcr oligomerization domain (*8*) and different domains of Abl (*9, 10*) have been solved. The crystal structure of the Bcr oligomerization domain (PDB

accession no. 1K1F, Figure 2A & B) reveals an antiparallel coiled coil. The monomer consists of a long α-helix that participates in the coiled coil, preceded by a loop and another short helix. The short helix associates with the long α-helix of the other monomer in the homodimer, forming a swap domain (*11*). Without the swap domains present, the long antiparallel coiled coil maintains an antiparallel orientation (*12*). The x-ray crystal structure of Abl kinase is typical for kinases, consisting of two different lobes, both containing of a mixture of β-sheets and α-helices, with an ATP binding site located in between the two lobes (*13*). Upon binding to ATP, Abl kinase undergoes a conformational change in a loop region, transforming the structure from an inactive to active state (*9*).

Existing drugs used to treat leukemia, such as Gleevec (*9, 14*), target the Abl kinase ATP binding region. Gleevec, an extremely successful drug made by Novartis, binds an inactive conformation of Abl kinase, shown in Figure 1B. When Gleevec binds the Abl ATP-binding site, intracellular signaling is abrogated, preventing proliferation and inducing apoptosis in Ph^+ cell lines. As a result of Gleevec therapy, 94% of patients have normal blood counts and 69% show no sign of the Ph^+ chromosome in their bone marrow (*4*). However, upon progression to advanced CML stages, only 15-34% of patients taking Gleevec have white blood cell counts within the normal range, and only 16-24% of patients have less than 35% of cells containing the Ph^+ chromosome (*15*). The effect of the drug lasts from 3 to 6 months for patients with advanced CML (*15*), and some people are resistant or become resistant (*14, 15*) to Gleevec due to mutations that affect the ATP binding site or shift the equilibrium to the active conformation where Gleevec is unable to bind (*15*). Therefore, it is important to explore other potential drugs and targets. In much the same way HIV is treated with combination therapies, a potential therapy combining Gleevec with a drug that targets another region may be effective (*14*).

Developing other leukemia therapies that could be used with Gleevec or on their own is an active area of research. For instance, two molecules made by Parke-Davis, PD173955 (*9*) and PD166326 (*16*), can bind to both the active and inactive conformations of Abl. Other potential combination therapies include: targeting Bcr-Abl RNA with antisense oligonucleotides, utilizing molecules designed to inhibit other signaling proteins, or trapping Bcr-Abl in the nucleus by suppressing nuclear export (*4*). Gleevec could also be combined with therapies proven to work clinically to treat CML, such as interferon-α therapy (*14*).

Because the coiled-coil region of the Bcr oligomerization domain is essential in Bcr-Abl transformation (*17-19*), a drug targeting this domain may offer a viable approach for combination therapy with Gleevec. Interestingly, the Bcr oligomerization domain, which associates as a tetramer in solution and in the crystal structure, can be replaced with a parallel dimeric coiled-coil domain, GCN4, and still cause Bcr-Abl transformation (*17*). Because GCN4 has the same transformation effect as the Bcr coiled-coil region, dimerization, not tetramerization, is required for Abl kinase constitutive activity. Adding a stoichiometric excess of the isolated Bcr coiled-coil region suppresses the Bcr-Abl transformed phenotype (*20-22*) and has been found to increase sensitivity to Gleevec (*22*). In addition, targeting Bcr may result in fewer side effects. Within the first few weeks of life, mice homozygous for *c-abl* disruption become runted and die (*23*), whereas *bcr*-null mutant mice are normal except for an increase in neutrophil reactive oxygen metabolism (*24*).

Although most drugs are small molecules, inhibiting protein-protein interactions with peptides has precedence in the literature and in the clinic. Peptide vaccines, targeting proteins other than Bcr, are currently being considered to treat CML (*14*), and some current CML treatments, such as IFN-α, are administered intravenously (*14, 25*). Miniproteins can serve as

effective inhibitor scaffolds onto which critical binding residues can be grafted. For example, the avian pancreatic polypeptide (aPP) has been used as a scaffold to design several inhibitor peptides to bind a variety of molecular targets (*26-32*). Highly specific inhibitors for Bcl-2 and Bcl-X_L, two critical proteins involved in apoptosis, were designed with this method and have nanomolar dissociation constants (*27, 30*). Utilizing a coiled-coil scaffold, several successful inhibitors that target coiled-coil structures have been made and tested. Using standard coiled-coil pairing rules, the Alber group designed a coiled-coil inhibitor that would heterodimerize with the APC tumor suppressor coiled-coil domain with greater affinity than the self-association of APC tumor suppressor or the inhibitor (*33*).

Many peptide inhibitors for HIV viral fusion, also a coiled-coil mediated process, have been designed by structural examination and modeling. One particularly exciting example is Fuzeon, made by Trimeris, which is currently on the market to treat HIV (*34*). This drug is a 36-amino acid peptide that must be taken intravenously. Many more peptide inhibitors for HIV viral fusion have been developed using protein engineering (*35-37*), and some inhibit HIV entry into cells at nanomolar concentrations *in vitro* (*38*). Solution and structural characterization of HIV glycoprotein peptide inhibitors have led to small molecule inhibitors that target viral membrane fusion in the same manner as the inhibitory peptides (*35, 39, 40*). Further, small molecule inhibitors that block other protein-protein interactions have been found (*41, 42*).

Many exciting advances in protein design have been made by the pharmaceutical industry and academic groups. In the pharmaceutical industry, computational protein design techniques have been very successful at stabilizing molecules, such as granulocyte-colony stimulating factor (G-CSF) (*34*) and human growth hormone (*43*). Enzymes with specificity for an entirely different molecule (*44*) and proteins with completely new folds (*45*) have been designed. In

addition, we have used protein design to create a heterotetramer from a homotetramer on a minimal scaffold (*46*), and very similar methodology can potentially be applied to design inhibitors for protein-protein interactions implicated in disease.

Despite successes in protein design, there are limitations to modeling large complex molecules, and a balance between computational speed and accuracy must be attained through the use of approximations in the design process. Modeling electrostatics and solvation effects is difficult due to the current inability to model solvent explicitly. Computational protein design uses many different approximations for electrostatics and solvation energy. Further, side chains are modeled in discrete conformations (rotamers), so side-chain conformations with the lowest energy for a particular site may not be present in the rotamer library and a less than optimal solution may be used. Although computational techniques are very effective, an iterative approach that combines computational design with experimental testing of the designs is needed, due to the many approximations that must be made.

The goal of this project was to computationally design α-helical inhibitor peptides to block Bcr oligomerization by forming a coiled-coil heterodimer with native Bcr. The inhibitors were designed to be dominant negative inhibitors of Bcr oligomerization. Although the Bcr oligomerization domain forms a tetramer (*11*), dimerization of the domain is the critical event that leads to the dysregulation of Abl kinase in the Bcr-Abl oncoprotein (*6, 17*). Therefore, if Bcr dimerization could be blocked by an inhibitor protein, the efficacy of a leukemia therapy based on such an inhibitor could be explored, and specific interactions between the peptide inhibitor and Bcr could aid in the design of future small molecule therapeutics. To overcome difficulties with computational protein design, several different computational techniques were

utilized to design α-helical inhibitors that potentially disrupt Bcr oligomerization. The inhibitors were expressed in *E. coli*, purified, and experimentally characterized.

METHODS

Computational design overview

For all design calculations, chains A and B or G and H from the Bcr oligomerization domain crystal structure (*8*) (1K1F) were used. The backbone of Bcr bound to its inhibitor was created using the backbone from the Bcr oligomerization domain (1K1F) structure. Chains G and H from the crystal structure were used for the stability calculation, as those chains had the lowest B values when compared to all other dimers in the crystallographic asymmetric unit. Chain G was used to model native Bcr monomer, whereas Asp 30 to Glu 66 from chain H (renumbered Asp 1 to Glu 37 for the calculations) were used to model the inhibitor.

A flow chart of the design process is shown in Figure 3. The design problem was divided into four different regions (Figure 3, Step A & Figure 4) to aid in faster computation and to make picking target sites and deconvoluting results easier. First, the analysis was broken down into swap domain vs. coiled coil. The following sites that interact with the swap domain, or have potential to do so, were designed: Q18, N21, Q22, E23, R24, R25, I28, Y29, L30, T32, L33, K36, E37. A designed site is a position where multiple amino acids are considered during Monte Carlo sampling. The coiled-coil region was divided into three sections, shown in Figure 4. Coiled coils contain a heptad repeat (**a**,**b**,**c**,**d**,**e**,**f**,**g**), where **a** and **d** are typically hydrophobic and **e** and **g** are hydrophilic residues. The following positions (denoted by position in Bcr followed by heptad position) at the Bcr-inhibitor interface were designed in separate sections: Region 1 – I2(**d**), E3(**e**), G5(**g**), L6(**a**), R8(**c**), A9(**d**), K10(**e**), S12(**g**), I13(**a**), R15(**c**), L16(**d**), and E17(**e**);

Region 2 – K10(**e**), S12(**g**), I13(**a**), R15(**c**), L16(**d**), E17(**e**), E19(**g**), V20(**a**), N21(**b**), E23(**d**), R24(**e**), and R26(**g**); Region 3: E19(**g**), V20(**a**), N21(**b**), E23(**d**), R24(**e**), R26(**g**), M27(**a**), L30(**d**), Q31(**e**), L33(**a**), L34(**a**), and K37(**d**). For each region, the sites were designed simultaneously in a large-scale selection, and 17 (Cys, Gly, Pro were excluded) and 14 (Cys, Gly, Pro, Phe, Tyr, and Trp were excluded) out of the twenty amino acids were allowed at each **e**,**g** and **a**,**d** design position, respectively. The basic energy function, discussed in the next section, was used to calculate energies of the sequences (Figure 3, Step B). Sequences were optimized and ranked for stability (Figure 5A).

Basic energy function

Energies were calculated using the following equation:

$$E_{folded} = \sum_{i=1}^{n}(E^{i}_{self,VDW} + E^{i}_{self,torsion} + E^{i}_{self,desolvation} + E^{i}_{self,electrostatics}) + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(E^{i,j}_{pair,VDW} + E^{i,j}_{pair,desolvation} + E^{i,j}_{pair,electrostatics})$$

$$E_{unfolded} = \sum_{i=1}^{n}(E^{i}_{self,VDW} + E^{i}_{self,torsion} + E^{i}_{self,desolvation} + E^{i}_{self,electrostatics})$$

$$E_{total} = E_{unfolded} - E_{folded}$$

The total energy used in the design calculation was defined as the sum of energy over self energies added to the sum over all residue-residue energies, or pair-wise interactions. The self energies of a residue are defined as interactions within a particular side chain, with the backbone, and with positions not included in the design calculation, whereas the pair-wise energies include inter side-chain interactions. Using a 12-6 Leonard-Jones potential, van der Waals interactions were calculated using van der Waals radii from CHARMM param19 scaled to 90% (*47*) for both the self and pair-wise terms. The van der Waals radii and total energy were scaled to 90% to overcome some of the small steric clashes that result from discrete rotamer approximation. Torsion energies for the self-energy terms were also computed using CHARMM param19. A distance-dependant dielectric function, which is essentially a Coulombic term with epsilon varied

as a function of atomic distance, was used to describe the electrostatics of both the self-energy and pair-wise energy terms. A dielectric constant of 4r was used for the self-energy calculations, and pair-wise electrostatic energy was calculated with $\varepsilon = 4r$ for polar-polar interactions, 8r for polar-charged interactions, and 16r for charged-charged interactions. The distance-dependant dielectric function was parameterized to agree with Poisson-Boltzman calculations for coiled coils by Nick Levenson and Jiangang Chen. The Effective Energy Function (EEF1) (*48*) from CHARMM was used to describe solvation energies for both the self and the pair-wise energy terms.

A pentapeptide model, in which each residue was modeled as the middle residue (X) in a Gly-Gly-X-Gly-Gly sequence, was used to approximate the unfolded state. The local x-ray crystal structure backbone of X and two flanking residues on either side was used. An ensemble average of all possible rotamer states was used to describe X, and the energy function mentioned above was used to describe the unfolded state.

Sequence search and side-chain repacking

Sequence space was sampled using a Metropolis Monte Carlo search algorithm with 20 cycles of 1500 steps using a linear temperature gradient from 300 to 200K or 280K. If there was very little improvement in energy, the calculation was stopped at 280K. The designs involving the swap domain were done with the Dunbrack 1999 backbone-dependent rotamer library (*49*). The Dunbrack 2002 backbone-dependent rotamer library (*50*) was implemented in the design algorithm following the swap domain study and was used to the design the coiled-coil regions. The 2002 library was thought to be more accurate and have better statistics because it contained more rotamers from a larger and more up-to-date PDB. Based on the basic energy function energies, side chains at and around design positions were placed in optimal configurations, using

the dead-end elimination algorithm (DEE) followed by the A* algorithm (*51, 52*) (Figure 3, Step B).

ICT1 design

ICT1 used computation as a guide to manually design an inhibitor. Visual evaluation of the inhibitor structures was used to assess the quality of the model and make final design decisions. Several positions found to be promising were investigated further by independently selecting residues at each site (Figure 3, Step H). Because the basic energy function used 90% van der Waals radii to decrease small steric clashes caused by rotamer approximation, each promising site was enumerated (Figure 3, Step I) and the top 5 solutions at each single-site position were minimized with 100% van der Waals radii using 500 cycles of steepest decent, followed by 500 cycles of Adapted Basis Newton Raphson (ABNR). The backbone was fixed during minimization. The van der Waals energies were recalculated with 100% van der Waals radii (Figure 3, Step J). Following minimization, residues with the most van der Waals improvement compared to wildtype and best packing, based on visual inspection, were combined to design ICT1.

Better energy functions (Design of ICT2, ICT3, ICT4)

All solutions from the large-scale selections were minimized with 100% van der Waals radii and 10 rounds of steepest decent followed by 10 rounds of ABNR (Figure 3, Step C). Following minimization, three energy function models (GB, EEF/GB, and Combo models) that calculate solvation differently were used to evaluate the structures (Figure 3, Step D). These more realistic energies were calculated using CHARMM param19 van der Waals energy (100% radii), Coulombic energy with a dielectric constant of 4, and a surface tension term computed as

the solvent accessible surface area (ASA) multiplied by 7 cal/mol•Å^2 (*53*). For the GB model, the polarization energy for the transfer of a protein from environment dielectric of $\varepsilon_{in}=4/\varepsilon_{out}=4$ to $\varepsilon_{in}=4/\varepsilon_{out}=80$ was computed (*54*). PEP was used to calculate Born radii (*55*). For the EEF/GB model, GB reaction field energies were substituted with EEF desolvation energies.

ICT2 was designed using the GB model. ICT3 was designed using the EEF/GB energy function model (Figure 3, Step D). In both designs, the top 10 most energetically favorable solutions were ranked according to helix propensity, and the solution with the largest helix propensity was chosen. Helix propensity was chosen as a method of evaluation because it was not directly taken into account in our energy function model. Helix propensity of a structure was calculated by summing the helix propensities of each residue in the protein (*56*). The sequences from each region were merged together to create one top solution (Figure 3, Step E).

The intersection of top-ranked solutions in GB and EEF/GB models (the Combo Model) was used to design ICT4 (Figure 3, Step D). The GB calculation used in the Combo model did not contain the energy of the backbone with itself or the energy of the side chains with themselves for the polarization term. Energetic cut-offs for both the EEF/GB and GB Model energies were adjusted such that only the top ~100 solutions that scored well with both models were considered. The solution with the top helix propensity was chosen from ~100 solutions for each region, and the regional sequences were merged into one solution (Figure 3, Step E).

Refining inhibitor sequences

The final ICT2, ICT3, and ICT4 inhibitor solutions were repacked, and side-chains were minimized with 10 rounds of steepest decent followed by 10 rounds of ABNR while holding the backbone fixed (Figure 3, Step F). The minimized solutions were evaluated with the better energy functions. To ensure all the mutations were contributing to a Bcr-inhibitor structure more

stable than Bcr, wild-type residues were individually substituted at each of the mutated positions (Figure 3, Step G). The residues were placed in their optimal conformation by repacking a combination of residues at the Bcr interface. The solution was minimized, and the energy was evaluated with the better energy functions. If the energy improved upon substituting a wild-type residue for a designed residue, the wild-type residue was retained in the design. Solubility of the inhibitors was also considered; sequences chosen do not have large numbers of hydrophobic residues that might lead to insolubility. Each designed inhibitor was run through Paircoil2 (*57*) to ensure the designed inhibitor retained a high coiled-coil propensity.

Specificity Calculation

ICT1-4 were computationally checked to ensure that they would have specificity for Bcr in preference to homodimerization (Figure 3, Step K & Figure 5B). To perform the specificity calculation, the desired state (Bcr-inhibitor) and undesired states (inhibitor-inhibitor and Bcr-Bcr) had to be modeled. Ideally, the same set of chains from the crystal structure would be used to model desired and undesired states in stability and specificity calculations. Copies of Bcr in the asymmetric unit varied in the number of residues resolved at the N- and C-termini. To maximize the number of residues modeled in desired and undesired states, the A and B chains were used in the specificity calculation, as these chains had the most residues resolved. However, the A and B chains had higher B factors than the G and H chains used in the stability design. Residues from Asp 30 to Lys 68 (renumbered Asp 1 to Lys 39) were used to model the inhibitor in both the inhibitor-inhibitor and Bcr-inhibitor complexes. Bcr, in the Bcr-Bcr and Bcr-inhibitor complexes, was modeled using residues Met 1-Lys 67. To mimic backbone flexibility, both the A/B and G/H backbones were used to check designed inhibitors for rotamers that had a low probability of being found in the PDB, based on statistics in the rotamer library.

If a low-probability rotamer was found on both backbones, the rotamer conformation was considered unfavorable. However, if a low-probability rotamer occurred on one of the two backbones, the rotamer conformation was considered favorable because the rotamer could pack in a high-probability rotamer with a small amount of backbone flexibility.

Purification of Bcr and inhibitor

To construct BCR-C, residues 30-65 were removed from pMAL/BCR_{1-74} (*11*) via PCR. For BCR-C, GSKGG was added on the N-terminus, and GGC was added on the C-terminus. C-BCRFL, which contained residues 1-74, was produced by PCR from pMAL/BCR_{1-74} (*11*). Using DNA Works (*58*) to design oligos, the BCR inhibitors, shown in Figure 6, were constructed with an N-terminal GSCGG using gene synthesis techniques. The genes were cloned into pSV282 using BamHI and XhoI restriction sites. The pSV282 plasmid contains a six-residue histidine tag, *E. coli* maltose binding protein (MBP), and a tobacco etch virus (TEV) protease cleavage site upstream of the BCR gene.

BCR-C, C-BCRFL, and BCR inhibitors were grown at 37 °C and induced with 0.1 mM IPTG at 0.4 O.D. for 2 hours. Cells were pelleted, then resuspended in 50 mM Tris pH 8, 500 μM EDTA, 100 mM NaCl, with two Roche Complete protease inhibitors. The cells were sonicated, pelleted again, then the supernatant was added to an appropriate amount of amylose resin. The column was washed with 5 column volumes of 50 mM Tris pH 8, 500 μM EDTA, 100 mM NaCl in batch. About 1 mg of TEV protease per liter of culture was added to the beads, and the reaction was allowed to proceed for 2 hours at room temperature. TEV-cleaved Bcr was HPLC purified before use.

BL21-RIL cells containing pRK793 were obtained from Science Reagents to make the TEV protease. Cells were grown at 37 °C in 100 μg/mL ampicillin and 35 μg chloramphenicol.

Upon induction of expression at 0.5 O.D. with 1 mM IPTG, the temperature was reduced to 30 °C. Cells were pelleted and resuspended in 20 mM Tris (pH 8), 0.5 M NaCl. Cells were sonicated, then polytheleneimine (pH 7.9) was added to a final concentration of 0.1%. Cells were pelleted, and the supernatant was run on a NiNTA column using 20 mM Tris, 0.5 M NaCl to wash the column. The sample was eluted with 4 column volumes of 20 mM Tris (pH 8), 0.5 M NaCl, 120 mM imidazole, followed by 20 mM Tris (pH 8), 0.5 M NaCl, 300 mM imidazole. Before running on an S-200 size-exclusion column, 1 mM EDTA and 1 mM DTT were added to all fractions. The centricon system was used to concentrate fractions containing TEV.

Pull-down assay

Concentrations of proteins were measured using the Edelhoch method (*59*). C-BCRFL was biotinylated by rocking 1 mg/mL C-BCRFL, 1 mM maleimide-PEO2-biotin in PBS, pH 7.4 (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) at room temperature. Following biotinylation, C-BCRFL was HPLC purified again. Appropriate amounts of monomeric avidin beads (assuming binding capacity was 1 mg/mL) were mixed with biotinylated protein and rinsed with 4 column volumes of PBS. Monomeric avidin beads were prepared by standard protocol from Pierce (Cat #20228). Stoichiometric amounts of inhibitor BCR and biotinylated C-BCRFL were added and allowed to equilibrate for 1 hour, then the mixture was added to the avidin beads. The samples were spun down, and the supernatant was removed. The samples were eluted with 4 volumes of 2 mM biotin in PBS and run on a tris-tricine gel.

RESULTS

Four α-helical inhibitors were computationally designed to bind the Bcr oligomerization domain and adopt an antiparallel coiled-coiled structure, similar to the antiparallel coiled-coil structure in the native protein. First, the results section compares the performance of the energy function in placing side chains in conformations similar to the x-ray crystal structure. Second, the selected inhibitor sequences and important interactions that are predicted to provide stability and specificity for the Bcr-inhibitor complex are presented.

Comparison of repacked Bcr with the crystal structure

Before designing the inhibitor, the side chains on the Bcr oligomerization domain crystal structure (chain G and H) were repacked using the Dunbrack 2002 backbone-dependent rotamer library to determine how accurately the rotamer library approximated the native structure. The swap domain on the H chain was truncated to simulate the Bcr inhibitor. A rotamer within +/- 40 degrees from the native rotamer was considered correct. On the H chain, Arg 24 was found in a conformation not represented in the rotamer library, most likely due to its role in tetrameric packing. The conformation of Asn 50, on the G chain, was found to be in a low-probability rotamer (probability 0.0072). Leu 16 on the H chain was also in an unfavorable conformation in the crystal structure. Overall, the limitations of the rotamer library did not affect other residues.

Design of a potential inhibitor swap domain region

The swap domain has potential to yield additional stability and specificity to the inhibitor we design. From the large-scale selection, selected residues at Gln 18 and Gln 22 deviated from the native sequence, with some consensus among the top solutions. These positions were evaluated further by enumerating (testing all combinations of) 17 amino acids at each position

individually. However, neither position yielded a large amount of stabilization, and packing was not improved based on structural examination. Therefore, using residues in the swap domain to further stabilize and add specificity to the designed inhibitor was not considered a viable strategy.

Using the basic energy function / visual design (ICT1)

Our first approach used computation to guide the manual design of an inhibitor. Obtained from top solutions in regions 1-3, the following positions were promising because a consensus of residues at these sites deviated from the native sequence: L6, A9, L16, E17, V20, E23, and M27. These positions are far enough apart in structure space that their individual energy contributions should be unaffected by other designed positions and thus be additive. Therefore, the sites were considered independently in separate calculations. The energies of the top 5 solutions at each site are shown in Table 1a.

The residues of ICT1, L6V, L16K, E17L, V20I, (Figure 7) were chosen based on introducing stabilizing van der Waals interactions at individual sites (Table 1b) and visual inspection of packing in the model structure. Mutations to A9 and E23 did not improve van der Waals interactions and were not included in ICT1. Based on visual inspection of the solution, L6V(**a**) and V20I(**a**) enhance hydrophobic packing in the core (Figure 7B and C). L16K forms a **d-g'** salt bridge with Glu 48 on Bcr, and the aliphatic portion contributes to core packing (Figure 7D). Lys 16 also acts as a negative design element against the inhibitor-inhibitor complex due to charge repulsion, as the two lysines are in close proximity on adjacent layers (Figure 7E). Further, the close proximity also leads to poor core packing in the inhibitor-inhibitor structure, leaving a large void space in the core. As a result, Ile 20 is placed in a poor rotamer conformation to fill the void space. Leu 17(**e**) shows significant van der Waals stabilization

(Figure 7F). The stabilization energy of the ICT1 sequence over wildtype (Table 2a) was calculated with the better energy functions after having been minimization. Despite an improvement in van der Waals energy at individual sites, van der Waals energies improved much less than was anticipated when all the mutations were considered together. However, given improvements seen at individual sites, this inhibitor was still made and tested.

Effect of minimization

Minimization was found to be a critical element after running a small-scale test set where positions 16, 17, and 18 were designed with four possible amino acids enumerated at each site. Structural stability before and after minimization was vastly different, yielding a very large difference in solution ranking (Figure 8). The energies before and after minimization were evaluated with the better energy function models, allowing the energies to be broken into energy terms (van der Waals, Coulumbic + solvation, and total energy). After minimization, solutions that ranked poorly before minimization had some of the lowest overall energies in the data set. Based on the small selection, the re-ranking after minimization can be attributed to relieving small steric van der Waals clashes.

Better energy functions (ICT2, ICT3, ICT4)

The initial selection, which used the basic energy function, yielded 11,208 solutions for Region 1, 7231 solutions for Region 2, and 5093 solutions for Region 3. Model structures were constructed for each solution by placing side chains in their lowest energy rotamer conformation within the structure. The top solutions in each region were combined to create one or two inhibitor sequences. The top sequences were modeled, minimized, and evaluated with the better

energy functions. The sequences scored well with Paircoil2, indicating they were likely to form coiled coils.

ICT2 (Figure 9) was designed using the GB energy function model (GB model). An interesting trend, seen in many of the top GB model solutions, involved mutating most of the **g** positions to arginine. Rather than choosing the top design, this trend was isolated and tested as ICT2. At several **g** positions, ICT2 contained E5R, S12R, R26I, and L33R mutations (Figure 9B-E). E5R introduces an intrahelical salt bridge with Asp 1, and S12R forms a salt bridge with Glu 48. Further, the arginines at **g** also provide a negative design element against the inhibitor-inhibitor complex (Figure 9F). In addition, Met 27 was changed to leucine to avoid oxidation problems. Other than Met 27, leucine had the lowest energy of all hydrophobic residues at this site. In this particular inhibitor, Leu 27 assumes a poor rotamer conformation in the inhibitor-inhibitor complex, disfavoring the complex (Figure 9G). The stability of ICT2 relative to the native Bcr sequence is shown in Table 2.

ICT3 was designed using the EEF/GB energy function model. The top solution for ICT3 is shown in Figure 10. Again, an arginine motif at the inhibitor **g** position occurred as in ICT2, providing the same favorable interactions in the Bcr-inhibitor complex and a negative design element against the inhibitor-inhibitor complex. The calculations suggested a M27E mutation in the core. The glutamate packs in much the same way the methionine did. Unlike methionine, the glutamate carboxyl groups are able to form an intrahelical salt bridge with Arg 31 (Figure 10B), another selected residue. M27E also has potential to form an interhelical salt bridge with Lys 39. Glu 21 (**b**) makes an intrahelical salt bridge with Arg 24, and it may interact with the swap domain (Figure 10C). Position 33 (**g**), at the end of the helix, is very solvent exposed, and replacing Leu 33 with Arg adds some hydrophobic packing, but allows the charged group to be

solvent exposed (Figure 10D). L33R also interacts with the hydroxyl group on Tyr 29. R34 forms an interhelical salt bridge with E32. The **g**-position arginines can then provide some specificity for the Bcr-inhibitor complex, as in ICT2. However, some salt-bridge interactions form in the inhibitor-inhibitor complex, which are somewhat troubling. ICT3 was designed before a bug was found in the better energy function. The problem involved incorrect assignment of Born radii on different chains. The calculations were re-run after the bug was fixed, and ICT3 was no longer the top solution. However, the same general trends were found.

To design ICT4, the intersection of the top-ranking solutions from both the EEF/GB Model and GB Model was used (Figure 11). The **g**-position arginine motif was also seen in this inhibitor. Because this motif was already tested in ICT2, these residues were not mutated in the ICT4 inhibitor design. Instead, we decided to focus on some interesting core mutations that arose in this calculation. Most likely, the V20L mutation causes the inhibitor to gain more stability due to better core packing (Figure 11B). The Q31E mutation yields a possible interhelical salt bridge with Lys 39 (Figure 11C), but this interaction does not form in the repacked structure due to a competing intrahelical salt bridge. I30L packs better in the pocket formed by Bcr and the swap domain (Figure 11D). The L16A mutation could increase specificity for the Bcr-inhibitor complex. The two residues at position 16, located in the middle of the helix, are on adjacent layers in the inhibitor-inhibitor complex (Figure 11E). The proximity of the alanines yields a large void space in the inhibitor-inhibitor complex, disfavoring the complex (*46*). Leu, in the M27L mutation, (Figure 11F) assumes a low-probability rotamer in the inhibitor-inhibitor complex, also disfavoring the complex.

Results of pull-down assay

Both the negative and positive controls worked properly in the assay. For the negative controls, BCR-C or the inhibitors were added to monomeric avidin beads to determine if they bound in the absence of C-BCRFL. A band was not seen in this control (Figure 12A). For the positive control, BCR-C was added to biotinylated C-BCRFL, equilibrated, then immobilized on the streptavidin beads. Eluted BCR-C is circled in Figure 12A. ICT1-ICT4 did not bind in the pull-down assay (Figure 12A-D), but preliminary results indicate ICT2 and ICT3 bound to BCR-C in a less stringent assay.

DISCUSSION

In this study, we used computational approaches and manual intervention to design several α-helical inhibitors to the Bcr oligomerization domain, part of the Bcr-Abl oncoprotein. Although the Bcr oligomerization domain tetramerizes, the inhibitors were designed to block dimerization of the Bcr oligomerization domain, as Bcr dimerization is the critical event that leads to uncontrolled Abl kinase signaling (*6, 17*). The Bcr-inhibitor complex was designed to adopt the native Bcr coiled-coil architecture. Inhibitor sequences were computationally designed. Regions of the α-helical inhibitor interacting with the swap domain were kept native, as computational studies showed little improvement in stability upon substitution with other amino acids. In this chapter, the inhibitor designs focus on stabilizing interactions in the long coiled-coil region. In most cases, the final sequences were tweaked by manual intervention to obtain soluble sequences that tested overall trends uncovered in the computational analysis. The inhibitors were made and experimentally tested, but failed to bind BCR in pull-down assays with C-BCRFL. ICT2 and ICT3 bound in less stringent assays with BCR-C. Challenges with

computational design and potential reasons why the designed inhibitors failed to bind with high affinity are discussed.

Challenges with computational protein design

Despite successes in protein design, necessary assumptions needed to make protein design tractable create limitations on the accuracy of methods and the size of the design problem. Some difficulties include modeling aqueous protein solvation, the rotamer approximation, approximate models of the unfolded state, and lack of backbone flexibility. Techniques to overcome some of these limitations were used to design the Bcr inhibitor peptides.

Challenges modeling solvation

Modeling solvation in proteins is a major problem in computational protein design. Solvation energies have many components, including van der Waals forces, hydrogen bonding, electrostatic interactions, and many solvent molecule configurations (*60*). Solvent and electrostatic models are tightly coupled, as water competes with other side chains for interactions with polar residues. Explicit solvent models are intractable for computational protein design (*60*), but continuum solvation approaches are used frequently (*61, 62*). The finite difference Poisson Boltzman (PB) continuum solvation model is considered among the most accurate, but many other continuum solvation models used in protein design approximate PB well and use fewer computational resources (*63*).

We used three continuum electrostatics models, the Effective Energy Function (EEF), the Generalized Born solvation model (GB), and a modified GB solvation model. In the basic energy function, we used a combination of a distance-dependant dielectric function and EEF to model electrostatics and solvation. The distance-dependant dielectric function, parameterized for

coiled coils, contained a dielectric constant based on the type of interaction (charge-charge, polar-polar, etc). EEF is based on reference states taken from vacuum-to-water solvation energies for atom groups in small molecules. The solvation energy is calculated by subtracting the volume-based loss of solvent surrounding an atom from the reference-state solvation energy (*48, 62*). The EEF model, itself, did not account for charge screening due to solvent; however, by adding in the distance-dielectric model, we crudely accounted for screening.

The better energy functions, which model solvation more accurately than the basic energy function, were used to evaluate and choose ICT2-4 designs. Electrostatic screening, a vital part of solvation energy, is particularly important for designing proteins that have high specificity, as polar interactions are affected by solvent screening and play a significant role in providing specificity. A solvated charge induces ordering of polar solvent molecules around it, forming a field opposite to its own. Interaction of a charge with its own induced field is called the "reaction field" energy. The effect on interactions with other charges is called screening and this decreases the effect of the charge on other charges nearby. In both of our better energy functions, we used Generalized Born (GB) solvation energy (*53, 61, 64*) to model solvation effects. The solvation energy contains electrostatic terms and a solvent accessibility term that captures cavitation and non-electrostatic interactions with water. The solvent accessibility term helps model the hydrophobic effect, whereas the polarization term models the effects of solvent screening on side-chain interactions and contains a reaction field term (*53, 64*). The solvent screening term can be broken into the interaction of a side chain with itself, with another side chain, and with the backbone, as well as backbone with itself (*53, 63, 64*). The reaction field term is also part of the polarization energy.

One of our better energy functions, the GB Model, used all of the GB terms mentioned. In another model, elements of GB and EEF were combined. EEF models the loss of interaction with waters around an atom as a function of volume of solvent effectively displaced by neighboring atoms. As such, EEF is very similar, in theory, to the reaction field term, so the GB reaction field was substituted with EEF in the EEF/GB better energy function model. The GB reaction field is very sensitive to the unfolded-state model, which assumes that the side chain is completely solvated and not making any interactions with other side chains in the unfolded state. We know this assumption is untrue, but it is impossible to know exactly what to model for the unfolded state. As a result, the Born radii of the side-chain atoms in the unfolded state are much smaller than if inter side-chain interactions were present, as the Born radii are directly dependent on the volume proximity of surrounding atoms. EEF energies are smaller than GB reaction field energies because they use an aqueous reference state and may give less error. EEF may be better because its effects can be broken down into interaction terms, whereas GB reaction field is a single term that is not decomposable. Because both better energy functions have caveats, both energy functions were used alone and together to design three other inhibitors. In all subsequent inhibitor designs, the solutions were minimized and one of the better energy functions was used to evaluate the structure.

Rotamer approximation

Known as the rotamer approximation problem, modeling side chains with discrete conformations may have led to less optimal solutions for ICT1. The rotamer library contains a discrete number of rotamers for each residue, and rotamers are unlikely to fit exactly into the designed protein structure. Most rotamers will be close to, but not in their optimal conformation within the protein structure.

Several methods to overcome the limitations of the rotamer approximation have been employed by other groups. For instance, the van der Waals term, modeled using a 12-6 Lennard-Jones potential, has a small attractive term and a large repulsive term. The repulsive term may be softened by scaling the van der Waals radii by 5 to 10% (*46, 65-67*) when discrete rotamers are used. However, the scaling adversely effects the van der Waals energy attractive component and can lead to overpacking (*68*). Secondly, more rotamers can be added to the rotamer library to reduce the rotamer approximation problem (*69, 70*). However, extremely large rotamer libraries increase conformational search space dramatically and are mainly used for small proteins (*61*). Flexible or multiple backbone models have been used to overcome the rotamer approximation (*45, 71-73*). To accommodate discrete rotamers without a large steric clash, we use 90% van der Waals radii and 90% of the total van der Waals energy term for the initial energy calculation and repacking. Final energies are calculated after minimization with 100% van der Waals radii to prevent overpacked cores.

Minimization is another way the rotamer approximation problem can be overcome. Although minimization can help selected rotamers reach their minima within the protein structure, minimization should be done with caveats in mind. For instance, over-minimization with an unrealistic energy function or solvation model can lead to structural distortion, so a small number of minimization rounds were used. Further, minimization was done using a very crude electrostatics model present in CHARMM, which could minimize side chains, particularly charged side chains, into unrealistic conformations. Given the relatively small number of solutions sampled in this study, minimization was feasible. For a very large number of structures, minimization would be infeasible, due to the computational time needed to minimize each structure.

Based on the ICT1 inhibitor and other solutions attained with the basic energy function, we decided to minimize all structures slightly, then recalculate the energy with better energy functions. Some good solutions might appear bad due to small steric clashes, causing their energies to be high when evaluated with the basic energy function. The small clashes could be relieved with a small amount of minimization. Minimization caused many of the solutions to re-rank significantly. Interestingly, the basic energy function calculation, utilized 90% van der Waals radii to reduce steric clashes caused by the rotamer approximation. However, the signifant re-ranking of solutions showed scaling the van der Waals radii and term was insufficient in eliminating small steric clashes. In the design of ICT2, ICT3, and ICT4, a small amount of minimization removed steric clashes caused by discrete rotamers.

Toward overcoming challenges modeling the unfolded state

Using helix propensity for final solution evaluation helps compensate for our poor unfolded-state model. For each amino acid, the unfolded-state model consisted of the local pentapeptide backbone from the x-ray crystal structure, containing two glycine residues on either side. Helix propensity captures several elements, dependant on amino acid, that are not present in our unfolded state model, including loss of conformational entropy. The top solutions for ICT2-4 from each selection were further evaluated using helix propensity as a guide.

Challenges designing an inhibitor

One of the main challenges in designing a dominant negative inhibitor is the trade-off between using positive and negative design. There are many technical difficulties in including negative design, and many successful protein designs have not included negative design, only considering positive design in calculations (see Chapter 1). In this study, stability was

optimized, and specificity was checked after the sequence was designed. Using negative design in calculations often results in reduced stability (see Chapter 1), and compromising stability was not considered viable for the first round of the Bcr inhibitor design. If inhibitor-inhibitor complexes form, specificity could be optimized in a second-round design after binding to the target protein is established. Although the concepts are similar, optimizing for Bcr-inhibitor specificity is much more difficult than designing a completely *de novo* heterooligomer, as in Chapter 2. For example, when designing both partners of a heterodimer, negative design elements that disfavor both homooligomeric states can be introduced. However, negative design elements cannot be introduced into the target protein to prevent homodimerization when designing an inhibitor. The inhibitor and target protein must heterodimerize with higher affinity than the target self-associates. Further, specificity must be considered to prevent inhibitor-inhibitor complexes. In addition, the Bcr inhibitor may form a higher-order oligomer or a completely different structure. Modeling all undesired states is impossible because all undesired states cannot be predicted. Further, modeling many undesired states dramatically increases computation time, so only the undesired states most likely to form (Bcr-Bcr and inhibitor-inhibitor antiparallel dimers) were modeled.

Performance of Inhibitors

Protein design is an iterative process and insights from both positive and negative results can be incorporated into future designs. Selecting a high frequency of lysine residues and large hydrophobics on the surface are common problems with computational protein design and various remedies have been tested by other groups (*62, 66, 68*). Designed using the basic energy function, ICT1 manifests some common problems seen with computational protein design. Some selected salt-bridge interactions were not used, due to the large number of lysine residues

frequently seen in solutions. It is unclear whether this preference for lysines is realistic or a function of the energy calculations. Lysine was selected very frequently in **e** and **g** positions and with much more frequency than other charged residues in core positions. Although selected lysine and arginine residues at **a** and **d** are not completely buried in the dimer core and can make important interactions with **e'** and **g'** positions, the heptad may be disrupted and the structure destabilized if charge is placed too frequently in the core by the design calculations. The large number of rotamers that lysine and arginine have in the rotamer library may enable the hydrophobic regions of these residues to pack well in many places. On the other hand, the desolvation penalty may be modeled poorly, allowing these charged residues to be selected for core coiled-coil positions. Large hydrophobic residues were selected to be on the surface at a higher frequency than expected based on natural proteins. The basic energy function does not have a penalty for putting hydrophobic residues on the surface. For the remainder of the inhibitors, we used side-chain minimization and better energy functions to overcome problems seen in the ICT1 design.

The ICT1 design focused on hydrophobic packing, selecting residues based on individual-site top solutions, visual inspection of the model structure, and improvement in van der Waals energy over native residues at a design position. Following individual-site selections at promising positions, inhibitor sequences were modeled on the Bcr structure and evaluated. Based on visual inspection, some inhibitor structures appeared to pack better than their energy reflected, as side chains assumed good rotamer conformations and placed more hydrophobic bulk in the core.

ICT1 did not bind in the pull-down assay for several possible reasons. The Lys 16 introduced in a core **d** position may adversely affect interhelical interactions and hydrophobic

bulk in the core. Lys 16(**d**) is one heptad from Glu 23(**d**), another charged core residue. Further, E17L at **e** removed an interhelical salt-bridge interaction present in the crystal structure, potentially interfering with interhelical association. The **a**-position core mutations for this structure, which were purely hydrophobic, could be tested without Lys 16(**d**) and Glu 17(**e**) to determine if an inhibitor with **a** position mutations will bind Bcr. Further, the overall van der Waals stabilization was not nearly what was predicted based on contributions from individual sites. Some of the sites may have been too close to assume they were independent, whereas the difference may be due to small differences in side-chain repacking at sites further away from the mutated site.

Despite two potential **g-g'** interactions not forming in the crystal structure, these interactions may be important interactions in solution. ICT2 was designed using the GB model, and many solutions showed an Arg motif at the **g** position on the surface. Given the inhibitor did not bind C-BCRFL in the pull-down assay, the native **g-g'** interactions may be important, as the two potential salt bridges may form in solution or the **g** position interactions may provide additional stability to the molecule. However, preliminary results indicate ICT2 bound BCR-C in a less stringent assay. The inhibitor replaced the two potential interhelical salt bridges with one intra- and one interhelical salt bridge.

Given ICT2 did not bind C-BCRFL in the pull-down assay, it is not surprising that ICT3 did not bind C-BCRFL either. However, preliminary results indicate ICT3 bound BCR-C in a less stringent assay. ICT3 was designed with the EEF/GB model and contained the same Arg motif at **g,** as was used in ICT2. As in ICT2, two potential interhelical salt bridges from the crystal structure were replaced with one intra- and one interhelical salt bridge. Further, two additional interhelical salt bridges involving residues E27/K39 and R34/E32 have potential to

form between the long helices. Although M27E mutation yielded a potential interhelical salt bridge, having a charged residue in the core probably destabilized the interaction. Unfortunately, a bug in the better energy function was discovered after ICT3 was designed. The bug affected how Born radii were calculated between different chains. Although ICT3 was no longer the top solution, some of the same trends were seen in the solutions once the bug was fixed.

ICT4 consisted of predominantly core mutations and may have changed the oligomerization state of the inhibitors. Because both of the better energy functions are approximations, ICT4 sought to find the intersection of low-energy solutions calculated with the EEF/GB and GB energy function models. In Chapter 3 of this thesis, mutating one core residue led to a change in oligomerization state; ICT4 had four amino acids mutated in the core. Two of the core mutations were thought to increase stability. Two other mutations were thought to add negative design elements through increasing cavity size in the core or rotamer strain in models of the inhibitor-inhibitor dimer. The two mutations thought to increase stability may have caused a higher-order oligomerization state to form.

Mutating M27 to leucine may have destabilized ICT1, ICT2, and ICT4. M27 was changed to leucine to avoid methionine oxidation in subsequent biophysical tests. However, the calculations predicted leucine to be slightly destabilizing relative to methionine. Before methionine is mutated in another design round, the mutation should be tested in isolation to determine any deleterious effects.

Biophysical characterization of the oligomerization state and structure of the inhibitor may prove useful for a subsequent design round. BCR-C cannot bind if C-BCRFL is dimerized. Determining the molecular weight of the inhibitors via AUC would be easy and potentially useful to determine if the inhibitor was forming a distinct alternative state. The alternative state

could be explicitly designed against in the next round. Evaluating the Arg motif at **g** may give insight into the possible inhibitor complex. Because the Arg motif at **g** emerged when using the three better energy functions, evaluating the motif further may prove useful. ICT2 could be used, with the caveat that M27 is mutated, to further test this motif for higher-order oligomers and as a negative design technique against the inhibitor complex. The role the **g-g'** interactions play in stabilizing and directing orientation in Bcr could be probed by doing a small amount of AUC and CD.

Although the Bcr inhibitors should be able to inhibit C-BCRFL, the pull-down assay used to test the inhibitors was perhaps too stringent an assay to test the first round of design. A less stringent assay that uses only the Bcr coiled-coil region (BCR-C) could be used to test for binding. Using SulfoLink beads from Pierce, BCR-C was covalently attached to the beads through its cysteine group. The inhibitor was added to the beads, and the mixture was heated, then cooled. Preliminary results indicate that ICT2 and ICT3 bind to BCR-C in this less stringent pull-down assay, and these inhibitors could be optimized to bind tighter.

Future Directions

In retrospect, several things could have been done better in the design process, and improvements could be added to the general computational design implementation. However, determining better energy functions for protein design, implementing backbone flexibility, and compensating for the rotamer approximation are computational protein design problems that will involve much more research by many different groups.

Future designs should try mutating only one to two amino acids to determine effects on stability and specificity. ICT1-4 mutated five to nine out of 37 sites. In ICT3, about 25% of the residues were changed. By making fewer mutations, it would be easier to deconvolute the

mutational effects, and the structure's stoichiometry and topology are less likely to deviate from native.

It is unclear if sampling more sequence space would have led to a large improvement in the stability of the Bcr-inhibitor complex. In current Bcr inhibitor designs, the initial calculation with the basic energy function could have sampled more sequence space. With the number of sites and residues considered in the initial calculations, there were 2.21×10^{14}, 2.68×10^{14}, and 1.82×10^{14} solutions possible for Region 1, 2, and 3, respectively, and only 11,208, 7231, and 5093 solutions for Region 1, 2, and 3, respectively, were considered from the calculation. The number of sequences actually sampled was about 5 billionith of a percent of possible solutions. Interestingly, the top energies had extremely incremental improvement over the wild-type. The overlapping positions within Region 1, 2, and 3 yielded nearly identical top solutions. However, oftentimes energy was improved in the final sequence by substitution of a wild-type residue, indicating undersampling.

Time spent on minimization places a cap on the number of sequences that can be searched in a reasonable amount of time. Evaluating more sequences with our better energy function would have been computationally prohibitive, as the better energy function is much more computationally intense. Building and minimizing each solution also takes a great deal of computational time and is impractical for 10^{14} solutions. Merely taking the top solutions from 10^{14} possible solutions from the basic energy function calculation is not reasonable, due to the large amount of solution re-ranking after minimizing structures. However, from the design of Bcr inhibitors, we find that minimization improves the design process and better approximate van der Waals functions may improve performance in future designs.

Many solutions from the basic energy function had large hydrophobics on the surface and had intrahelical interactions, and the better energy functions seemed to mitigate the large hydrophobics on the surface, somewhat. However, incorporating an aggregated state may be helpful to design against large hydrophobic residues on the surface. Harbury and Havranek modeled the aggregated state by placing their target structure in a medium with a lower dielectric than water (*74*). This alternative state could be easily incorporated into the current design implementation. Binary patterning, as discussed in the introduction to this thesis, could also be done to limit hydrophobic residues on the surface. Another problem seen frequently was intramolecular, rather than intermolecular salt-bridge formation. Some studies have found that increasing the intermolecular component weights yield more interhelical interactions (*75, 76*). The interhelical interactions help to favor the Bcr inhibitor complex over the inhibitor-inhibitor and Bcr-Bcr complexes.

Thousands of solutions are output from a design calculation and determining useful patterns or families of solutions from the output is very difficult. Solutions from protein design could be clustered into groups and families using clustering software, similar to that used to study DNA arrays. Difficulties in finding patterns increases with the complexity of the design. When just a few residues are selected, as in the BBA heterotetramer design in chapter 2, discerning patterns was not a problem. However, the Bcr inhibitor design involved targeting many more sites in each design. Havernek et al. utilized clustering to determine families of sequences that yielded specificity for hetero versus homodimeric coiled coils. Several families, including one not found in natural coiled coils, emerged from clustering analysis and were found to be heterospecific (*74*). Clustering could be done on many crude solutions, then the top solution from each cluster could be minimized and evaluated. Not only does clustering aid in

determining patterns, it also would make our design process more automated. Any future student doing design should investigate more bioinformatic approaches, as they may provide data analysis techniques for handling large amounts of sequence data obtained from design.

Our design implementation could be improved by employing different metrics to restrict sequence space. A fast, yet accurate method, is needed to eliminate much of sequence space. Much time was spent searching parts of sequence space that were not feasible coiled coils. For coiled-coil design, a fast sequence-based method, such as SVM (support vector machine), could be used to parse sequence space, picking the top percentage of feasible solutions. Alternatively, Paircoil2 (*57*) could be run on all sequences, and feasible coiled coils could be built, minimized, and the energy could be evaluated with a better energy function. For more general structures, a secondary structure prediction program could be used to parse sequence space before energies are calculated. Another option is to use an extremely fast statistical method, such as SCADS (*77*), to parse sequence space before applying a more exact energy function.

In future design calculations, paying more attention to low-probability rotamers earlier in the design process may be useful. Obtaining solutions with common rotamers from the PDB is important for well-folded, stable structures (*78*). Currently, we have a rotamer strain term incorporated in the energy function, but all the rotamers from the PDB are weighted equally in the design algorithm. After the design calculations are done, solutions are checked to ensure no low-probability rotamers are present. For some protein design calculations, low-probability rotamers have been omitted from the rotamer library and a term that favors high-probability rotamers has been added to the energy function (*65*). Although this term causes a deviation from a completely physics-based energy function, this approach may work well for stability design. However, low-probability rotamers are important for negative design and could be useful when

designing for specificity. An optimization between low-probability rotamers in the undesired states and high-probability rotamers in the desired states among top solutions may yield greater specificity and stability. However, for this to be a useful technique, side chains would have to show the same low-probability trends on several different backbones, as a small amount of backbone relaxation may enable a high-probability rotamer to be accommodated.

Adding backbone flexibility to the design calculation may yield more sequence diversity in the solutions. Backbone flexibility may be particularly useful for designing antiparallel coiled coils, as they have more varied packing in the core. The **a** and **d** positions in antiparallel coiled coils can range from being directly across from each other to being staggered. Movement of the backbone could make the structure more accessible to sequences that are further away from wild type. When changing the sequence a great deal, the importance of modeling other competing states may increase. With several of the mentioned improvements it will probably be possible to design an effective dominant negative Bcr inhibitor.

ACKNOWLEDGMENTS

I acknowledge the use of the MIT Computational and Systems Biology Initiative Proteomics/Structural Biology Core and High-Performance Computing Facility. I acknowledge Dr. Laura Mizoue from Vanderbilt University for the generous gift of the pSV282 plasmid and Xun Zhao and Peter Kim for the Bcr plasmid. Nora Zizlsperger had helpful comments, designed the pull-down assay, tested the inhibitors. Jeremy Fisher perfected the pull-down assay and valiantly tested the inhibitors. Devdoot Majumdar worked on the early stages of inhibitor design and helped to express and purify C-BCR, BCR-C, and C-BCRFL. Gevorg Grigoryan wrote the energy code used in this study and had many helpful suggestions and comments. Shaun Deignan helped with Beowulf cluster issues.

REFERENCES

1. (2005), American Cancer Society Website.
2. Sawyers, C. L. (1999) Chronic myeloid leukemia. *N Engl J Med 340*, 1330-40.
3. Catovsky, D. (1979) Ph1-positive acute leukaemia and chronic granulocytic leukaemia: one or two diseases? *Br J Haematol 42*, 493-8.
4. Kurzrock, R., Kantarjian, H. M., Druker, B. J., and Talpaz, M. (2003) Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. *Ann Intern Med 138*, 819-30.
5. D.A. Scheinberg, J. G. J. (2004) *Treatment of Leukemia and Lymphoma*, Vol. 51, Elsevier Academic Press, New York.
6. Smith, K. M., Yacobi, R., and Van Etten, R. A. (2003) Autoinhibition of Bcr-Abl through its SH3 domain. *Mol Cell 12*, 27-37.
7. McWhirter, J. R., and Wang, J. Y. (1997) Effect of Bcr sequences on the cellular function of the Bcr-Abl oncoprotein. *Oncogene 15*, 1625-34.
8. Zhao, X., Ghaffari, S., Lodish, H., Malashkevich, V. N., and Kim, P. S. (2002) Structure of the Bcr-Abl oncoprotein oligomerization domain. *Nat Struct Biol.*
9. Nagar, B., Bornmann, W. G., Pellicena, P., Schindler, T., Veach, D. R., Miller, W. T., Clarkson, B., and Kuriyan, J. (2002) Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res 62*, 4236-43.
10. Nagar, B., Hantschel, O., Young, M. A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., and Kuriyan, J. (2003) Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell 112*, 859-71.
11. Zhao, X., Ghaffari, S., Lodish, H., Malashkevich, V. N., and Kim, P. S. (2002) Structure of the Bcr-Abl oncoprotein oligomerization domain. *Nat Struct Biol 9*, 117-20.
12. Taylor, C. M., and Keating, A. E. (2005) unpublished.
13. Schindler, T., Bornmann, W., Pellicena, P., Miller, W. T., Clarkson, B., and Kuriyan, J. (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science 289*, 1938-42.
14. Druker, B. J., O'Brien, S. G., Cortes, J., and Radich, J. (2002) Chronic myelogenous leukemia. *Hematology (Am Soc Hematol Educ Program)*, 111-35.
15. Nardi, V., Azam, M., and Daley, G. Q. (2004) Mechanisms and implications of imatinib resistance mutations in BCR-ABL. *Curr Opin Hematol 11*, 35-43.
16. Sausville, E. A. (2003) Is another bcr-abl inhibitor needed for chronic myelogenous leukemia? *Clin Cancer Res 9*, 1233-4.
17. McWhirter, J. R., Galasso, D. L., and Wang, J. Y. (1993) A coiled-coil oligomerization domain of Bcr is essential for the transforming function of Bcr-Abl oncoproteins. *Mol Cell Biol 13*, 7587-95.
18. He, Y., Wertheim, J. A., Xu, L., Miller, J. P., Karnell, F. G., Choi, J. K., Ren, R., and Pear, W. S. (2002) The coiled-coil domain and Tyr177 of bcr are required to induce a murine chronic myelogenous leukemia-like disease by bcr/abl. *Blood 99*, 2957-68.
19. Zhang, X., Subrahmanyam, R., Wong, R., Gross, A. W., and Ren, R. (2001) The NH(2)-terminal coiled-coil domain and tyrosine 177 play important roles in induction of a myeloproliferative disease in mice by Bcr-Abl. *Mol Cell Biol 21*, 840-53.

20. Wu, Y., Ma, G., Lu, D., Lin, F., Xu, H. J., Liu, J., and Arlinghaus, R. B. (1999) Bcr: a negative regulator of the Bcr-Abl oncoprotein. *Oncogene 18*, 4416-24.
21. Guo, X. Y., Cuillerot, J. M., Wang, T., Wu, Y., Arlinghaus, R., Claxton, D., Bachier, C., Greenberger, J., Colombowala, I., and Deisseroth, A. B. (1998) Peptide containing the BCR oligomerization domain (AA 1-160) reverses the transformed phenotype of p210bcr-abl positive 32D myeloid leukemia cells. *Oncogene 17*, 825-33.
22. Beissert, T., Puccetti, E., Bianchini, A., Guller, S., Boehrer, S., Hoelzer, D., Ottmann, O. G., Nervi, C., and Ruthardt, M. (2003) Targeting of the N-terminal coiled coil oligomerization interface of BCR interferes with the transformation potential of BCR-ABL and increases sensitivity to STI571. *Blood 102*, 2985-93.
23. Tybulewicz, V. L., Crawford, C. E., Jackson, P. K., Bronson, R. T., and Mulligan, R. C. (1991) Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene. *Cell 65*, 1153-63.
24. Voncken, J. W., van Schaick, H., Kaartinen, V., Deemer, K., Coates, T., Landing, B., Pattengale, P., Dorseuil, O., Bokoch, G. M., Groffen, J., and et al. (1995) Increased neutrophil respiratory burst in bcr-null mutants. *Cell 80*, 719-28.
25. Kantarjian, H. M., Smith, T. L., O'Brien, S., Beran, M., Pierce, S., and Talpaz, M. (1995) Prolonged survival in chronic myelogenous leukemia after cytogenetic response to interferon-alpha therapy. The Leukemia Service. *Ann Intern Med 122*, 254-61.
26. Schneider, T. L., Mathew, R. S., Rice, K. P., Tamaki, K., Wood, J. L., and Schepartz, A. (2005) Increasing the kinase specificity of k252a by protein surface recognition. *Org Lett 7*, 1695-8.
27. Gemperli, A. C., Rutledge, S. E., Maranda, A., and Schepartz, A. (2005) Paralog-selective ligands for bcl-2 proteins. *J Am Chem Soc 127*, 1596-7.
28. Golemi-Kotra, D., Mahaffy, R., Footer, M. J., Holtzman, J. H., Pollard, T. D., Theriot, J. A., and Schepartz, A. (2004) High affinity, paralog-specific recognition of the Mena EVH1 domain by a miniature protein. *J Am Chem Soc 126*, 4-5.
29. Rutledge, S. E., Volkman, H. M., and Schepartz, A. (2003) Molecular recognition of protein surfaces: high affinity ligands for the CBP KIX domain. *J Am Chem Soc 125*, 14336-47.
30. Chin, J. W., and Schepartz, A. (2001) Design and Evolution of a Miniature Bcl-2 Binding Protein. *Angew Chem Int Ed Engl 40*, 3806-3809.
31. Montclare, J. K., and Schepartz, A. (2003) Miniature homeodomains: high specificity without an N-terminal arm. *J Am Chem Soc 125*, 3416-7.
32. Shimba, N., Nomura, A. M., Marnett, A. B., and Craik, C. S. (2004) Herpesvirus protease inhibition by dimer disruption. *J Virol 78*, 6657-65.
33. Sharma, V., Logan, J., King, D., White, R., and Alber, T. (1998) Sequence-based design of a peptide probe for the APC tumor suppressor protein. *Curr. Biol. 8*, p823-30.
34. Lazar, G. A., Marshall, S. A., Plecs, J. J., Mayo, S. L., and Desjarlais, J. R. (2003) Designing proteins for therapeutic applications. *Curr Opin Struct Biol 13*, 513-8.
35. Jiang, S., and Debnath, A. K. (2000) Development of HIV entry inhibitors targeted to the coiled-coil regions of gp41. *Biochem Biophys Res Commun 269*, 641-6.
36. Sia, S. K., Carr, P. A., Cochran, A. G., Malashkevich, V. N., and Kim, P. S. (2002) Short constrained peptides that inhibit HIV-1 entry. *Proc Natl Acad Sci U S A 99*, 14664-9.

37. Eckert, D. M., Malashkevich, V. N., Hong, L. H., Carr, P. A., and Kim, P. S. (1999) Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell 99*, 103-15.
38. Eckert, D. M., and Kim, P. S. (2001) Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc Natl Acad Sci U S A 98*, 11187-92.
39. Si, Z., Madani, N., Cox, J. M., Chruma, J. J., Klein, J. C., Schon, A., Phan, N., Wang, L., Biorn, A. C., Cocklin, S., Chaiken, I., Freire, E., Smith, A. B., 3rd, and Sodroski, J. G. (2004) Small-molecule inhibitors of HIV-1 entry block receptor-induced conformational changes in the viral envelope glycoproteins. *Proc Natl Acad Sci U S A 101*, 5036-41.
40. Ferrer, M., Kapoor, T. M., Strassmaier, T., Weissenhorn, W., Skehel, J. J., Oprian, D., Schreiber, S. L., Wiley, D. C., and Harrison, S. C. (1999) Selection of gp41-mediated HIV-1 cell entry inhibitors from biased combinatorial libraries of non-natural binding elements. *Nat Struct Biol 6*, 953-60.
41. Vassilev, L. T., Vu, B. T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N., and Liu, E. A. (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science 303*, 844-8.
42. Oltersdorf, T., Elmore, S. W., Shoemaker, A. R., Armstrong, R. C., Augeri, D. J., Belli, B. A., Bruncko, M., Deckwerth, T. L., Dinges, J., Hajduk, P. J., Joseph, M. K., Kitada, S., Korsmeyer, S. J., Kunzer, A. R., Letai, A., Li, C., Mitten, M. J., Nettesheim, D. G., Ng, S., Nimmer, P. M., O'Connor, J. M., Oleksijew, A., Petros, A. M., Reed, J. C., Shen, W., Tahir, S. K., Thompson, C. B., Tomaselli, K. J., Wang, B., Wendt, M. D., Zhang, H., Fesik, S. W., and Rosenberg, S. H. (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature 435*, 677-81.
43. Filikov, A. V., Hayes, R. J., Luo, P., Stark, D. M., Chan, C., Kundu, A., and Dahiyat, B. I. (2002) Computational stabilization of human growth hormone. *Protein Sci 11*, 1452-61.
44. Looger, L. L., Dwyer, M. A., Smith, J. J., and Hellinga, H. W. (2003) Computational design of receptor and sensor proteins with novel functions. *Nature 423*, 185-90.
45. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science 302*, 1364-8.
46. Ali, M. H., Taylor, C. M., Grigoryan, G., Allen, K. N., Imperiali, B., and Keating, A. E. (2005) Design of a heterospecific, tetrameric, 21-residue miniprotein with mixed alpha/beta structure. *Structure (Camb) 13*, 225-34.
47. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem. 4*, 187-217.
48. Lazaridis, T., and Karplus, M. (1999) Effective energy function for proteins in solution. *Proteins 35*, 133-52.
49. Dunbrack, R. L., Jr., and Karplus, M. (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol. 230*, 543-574.
50. Dunbrack, R. L., Jr., and Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci 6*, 1661-81.
51. Desmet, J., De Maeyer, M., Hazes, B., and Lasters, I. (1992) The dead-end elimination theorem and its use in protein side-chain positioning. *Nature 356*, 539-542.
52. Goldstein, R. F. (1994) Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys. J. 66*, 1335-1340.

53. Still, W. C., Tempczyk, A., Hawley, R.C., Hendrickson, T. (1990) Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc. 112*, 6127-6129.
54. Dominy, B. N., Brooks, C.L. III. (1999) Development of a Generalized Born Model Parameterization for Proteins and Nucleic Acids. *J. Phys. Chem. B. 103*, 3765-3773.
55. Beroza, P., and Case, D. A. (1998) Calculations of proton-binding thermodynamics in proteins. *Methods Enzymol 295*, 170-89.
56. Creighton, T. E. (1993) *Proteins*, 2nd ed., W.H. Freeman and Company, New York.
57. McDonnell, A. V., Jiang, T, Keating, A.E., and Berger, B. (2005) Paircoil2: Improved Prediction of Coiled Coils from Sequence. *Unpublished.*
58. Hoover, D. M., and Lubkowski, J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res 30*, e43.
59. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry 6*, 1948-1954.
60. Park, S., Yang, X., and Saven, J. G. (2004) Advances in computational protein design. *Curr Opin Struct Biol 14*, 487-94.
61. Pokala, N., and Handel, T. M. (2001) Review: protein design--where we were, where we are, where we're going. *J Struct Biol 134*, 269-81.
62. Jaramillo, A., and Wodak, S. J. (2005) Computational protein design is a challenge for implicit solvation models. *Biophys J 88*, 156-71.
63. Pokala, N., and Handel, T. M. (2004) Energy functions for protein design I: efficient and accurate continuum electrostatics and solvation. *Protein Sci 13*, 925-36.
64. Bashford, D., and Case, D. A. (2000) Generalized born models of macromolecular solvation effects. *Annu Rev Phys Chem 51*, 129-52.
65. Kuhlman, B., and Baker, D. (2000) Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A 97*, 10383-8.
66. Dahiyat, B. I., and Mayo, S. L. (1997) De novo protein design: fully automated sequence selection. *Science 278*, 82-7.
67. Desjarlais, J. R., and Handel, T. M. (1995) De novo design of the hydrophobic cores of proteins. *Protein Sci 4*, 2006-18.
68. Pokala, N., and Handel, T. M. (2005) Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *J Mol Biol 347*, 203-27.
69. Xiang, Z., and Honig, B. (2001) Extending the accuracy limits of prediction for side-chain conformations. *J Mol Biol 311*, 421-30.
70. Peterson, R. W., Dutton, P. L., and Wand, A. J. (2004) Improved side-chain prediction accuracy using an ab initio potential energy function and a very large rotamer library. *Protein Sci 13*, 735-51.
71. Desjarlais, J. R., and Handel, T. M. (1999) Side-chain and backbone flexibility in protein core design. *J Mol Biol 290*, 305-18.
72. Nauli, S., Kuhlman, B., and Baker, D. (2001) Computer-based redesign of a protein folding pathway. *Nat Struct Biol 8*, 602-5.
73. Plecs, J. J., Harbury, P. B., Kim, P. S., and Alber, T. (2004) Structural test of the parameterized-backbone method for protein design. *J Mol Biol 342*, 289-97.
74. Havranek, J. J., and Harbury, P. B. (2003) Automated design of specificity in molecular recognition. *Nat Struct Biol 10*, 45-52.

75. Reina, J., Lacroix, E., Hobson, S. D., Fernandez-Ballester, G., Rybin, V., Schwab, M. S., Serrano, L., and Gonzalez, C. (2002) Computer-aided design of a PDZ domain to recognize new target sequences. *Nat Struct Biol 9*, 621-7.
76. Shifman, J. M., and Mayo, S. L. (2003) Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc Natl Acad Sci U S A 100*, 13274-9.
77. Kono, H., and Saven, J. G. (2001) Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. *J Mol Biol 306*, 607-28.
78. Lazar, G. A., Johnson, E. C., Desjarlais, J. R., and Handel, T. M. (1999) Rotamer strain as a determinant of protein structural specificity. *Protein Sci 8*, 2598-610.

TABLES

Table 1a. Unfolded minus folded energies for point mutations.

L6 (a)		A9 (d)		L16 (d)		E17 (e)		V20 (a)		E23 (d)		M27 (d)	
L(n)	40.53	A(n)	40.53	K	40.64	L	42.02	K	41.03	K	41.48	K	42.14
K	39.82	S	39.72	A	39.46	I	40.82	L	40.2	L	40.52	M(n)	40.99
M	38.92	V	39.19	L(n)	39.01	M	40.71	I	39.96	A	50.03	A	40.76
I	38.35	I	38.91	H	38.67	V	40.53	V (n)	39.02	S	39.25	L	40.71
V	38.31	K	38.62	S	38.01	E(n)	39.01	R	38.58	M	39.05	V	40.34

Point mutations at positions that were promising for ICT1 design. The native residue is listed in the top row, and five of the top point mutations for each position are listed in each column. Unfolded minus folded energies calculated with the basic energy function. A larger positive number indicates greater stabilization. Uncertainty in energy function is at least ±1 at each site.

Table 1b. ICT1 unfolded minus folded van der Waals energies.

	L6V	L16K	E17L	V20I
VDW(uf-f)	0.66	2.15	3.18	3.80

Unfolded minus folded van der Waals energies for inhibitor minus Bcr. A greater positive number indicates greater stabilization. Uncertainty in energy function is at least ±1 at each site.

Table 2a. Stabilization of designed inhibitors over wildtype BCR.

	ICT1	ICT2	ICT3	ICT4
VDW Energy	-0.02	12.49	40.16	2.17
EEF/GB Model	4.52	21.10	44.21	4.74
GB Model	5.74	10.63	24.37	5.03

The numbers above represent the difference of the folded minus unfolded energies of BCR and the inhibitor. A positive number indicates the inhibitor sequence is predicted to be more stable than the BCR sequence. Uncertainty in energy function is at least ±1 at each site.

Table 2b. Specificity of designed inhibitors.

	ICT1	ICT2	ICT3	ICT4
VDW Energy	22.07	8.63	-2.72	1.60
EEF/GB Energy	46.44	4.58	33.04	0.15
GB Energy	21.18	28.32	4.61	20.67

Specificity of designed inhibitors where $E_{specificity} = -2E_{BCR\text{-}Inhibitor}+E_{Inhibitor\text{-}Inhibitor}+E_{BCR\text{-}BCR}$. A positive number indicates the inhibitor is more specific for the BCR-inhibitor complex versus the inhibitor-inhibitor and BCR-BCR complexes. Uncertainty in energy function is at least ±1 at each site

FIGURES

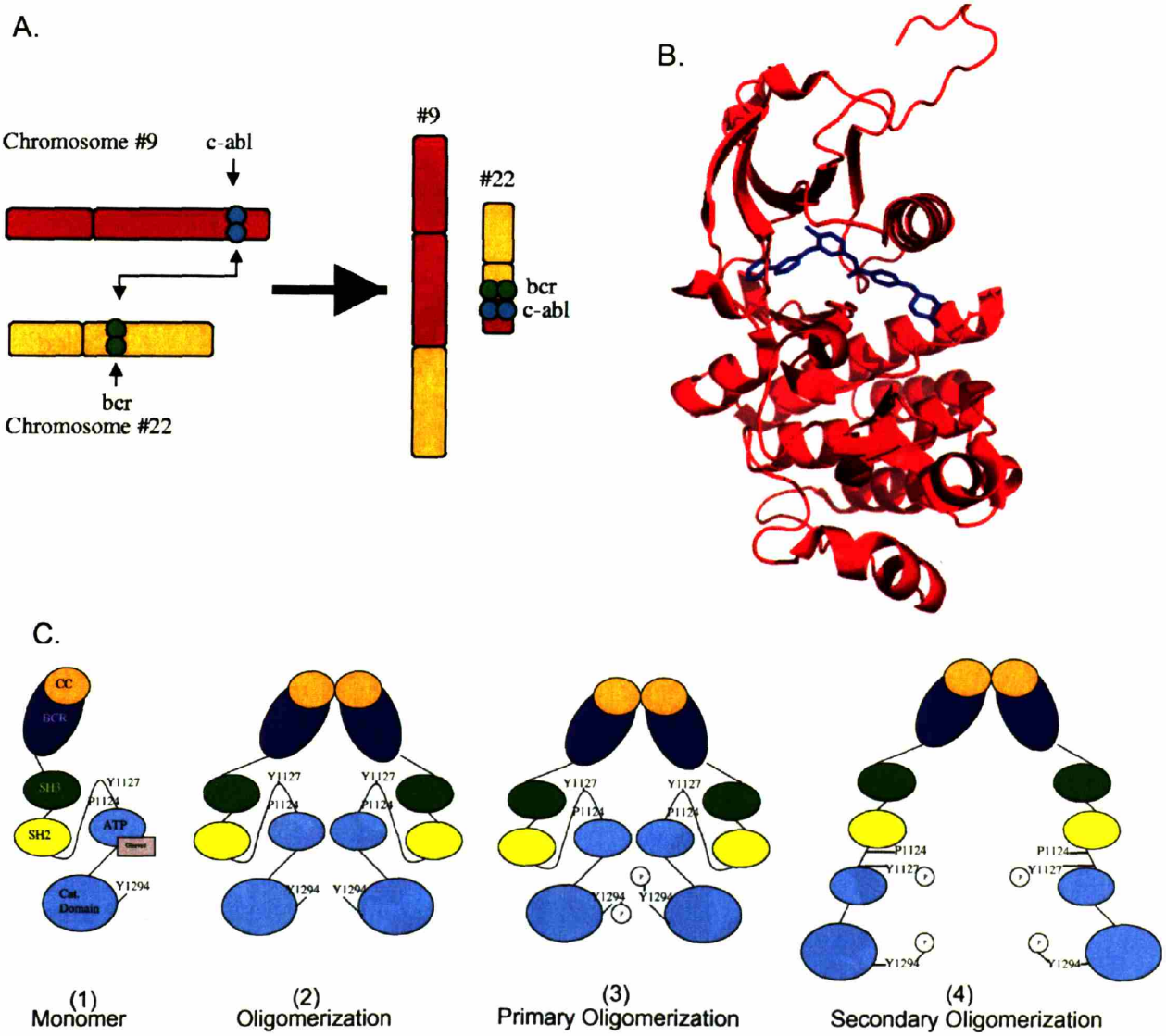


Figure 1. Mechanism of Bcr-Abl fusion and model for Bcr-Abl activation. (A) Reciprocal translocation event that leads to the Bcr-Abl fusion protein. (B) X-ray crystal structure of Abl kinase with Gleevec bound (1IEP). (C) Possible model for Bcr-Abl activation. 1) When monomeric, Bcr-Abl is unphosphorylated. This inactive conformation of Abl binds Gleevec well. 2) Oligomerization of the Bcr oligomerization domain leads to 3) primary phosphorylation at Y1294, followed by 4) secondary phosphorylation at Y1127. 4) The fully active conformation of Abl, which Gleevec cannot bind. (adapted from (*6*))

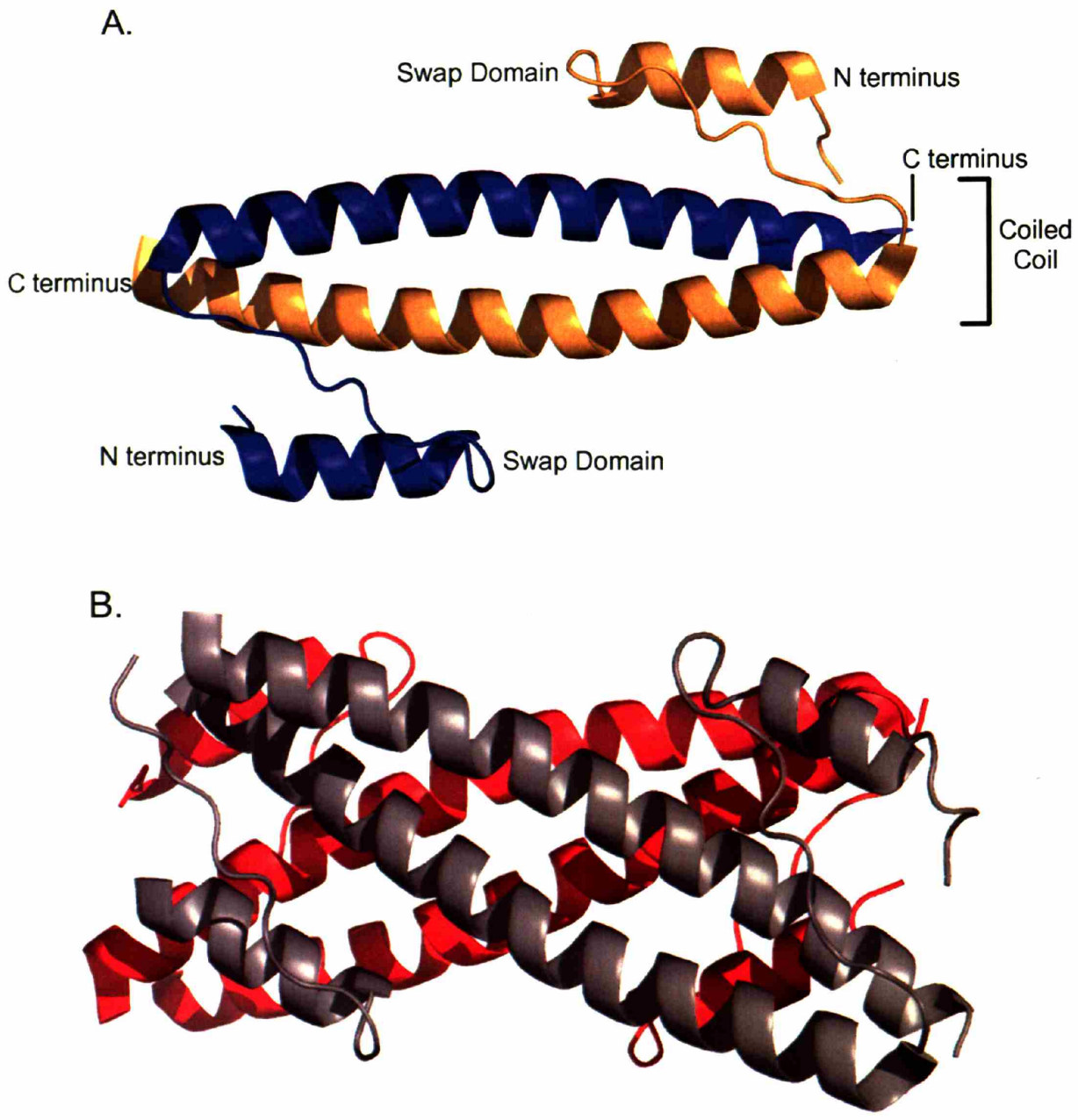


Figure 2. Bcr dimer and tetramer. (A) The BCR oligomerization domain dimer (1K1F, (*11*)). (B) Two of the dimers in (A) associate to form a tetramer in the crystal structure.

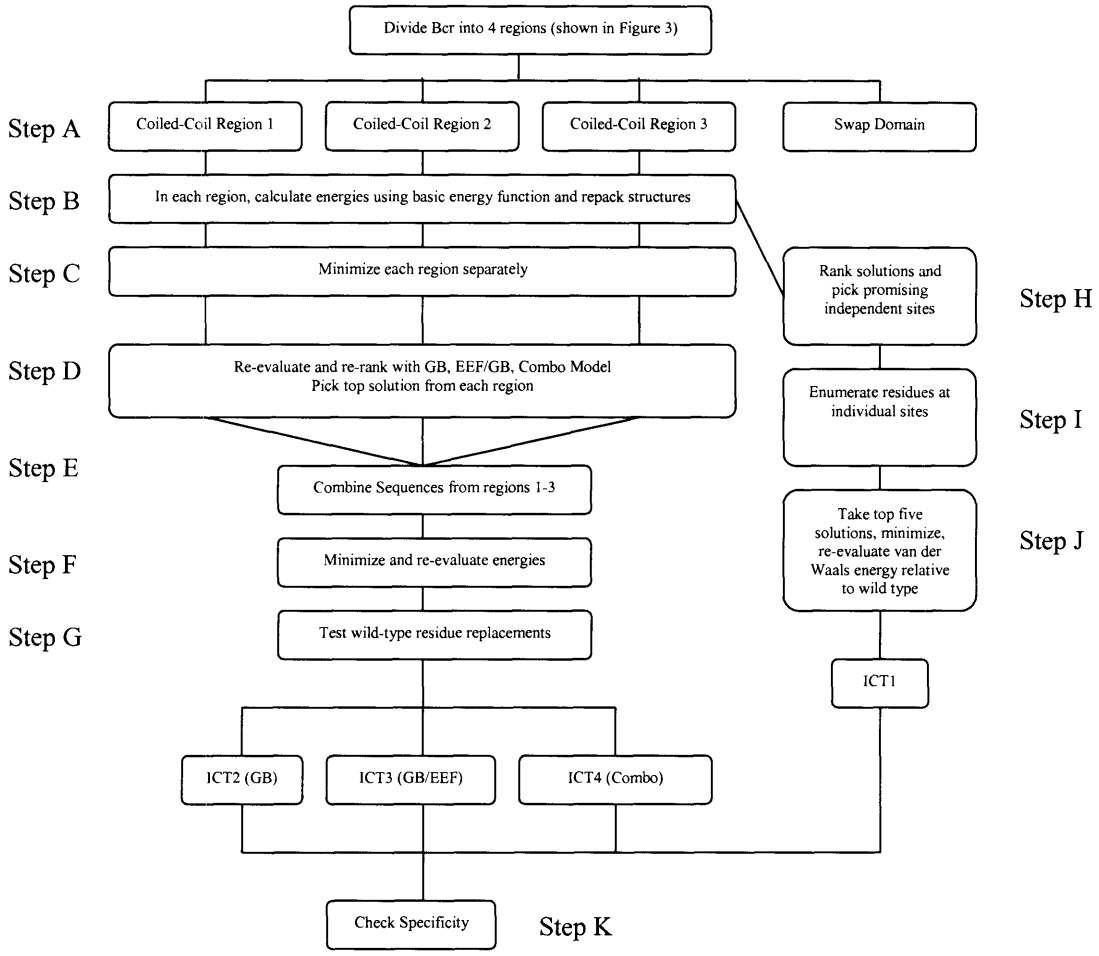


Figure 3. Flow chart of computational design.

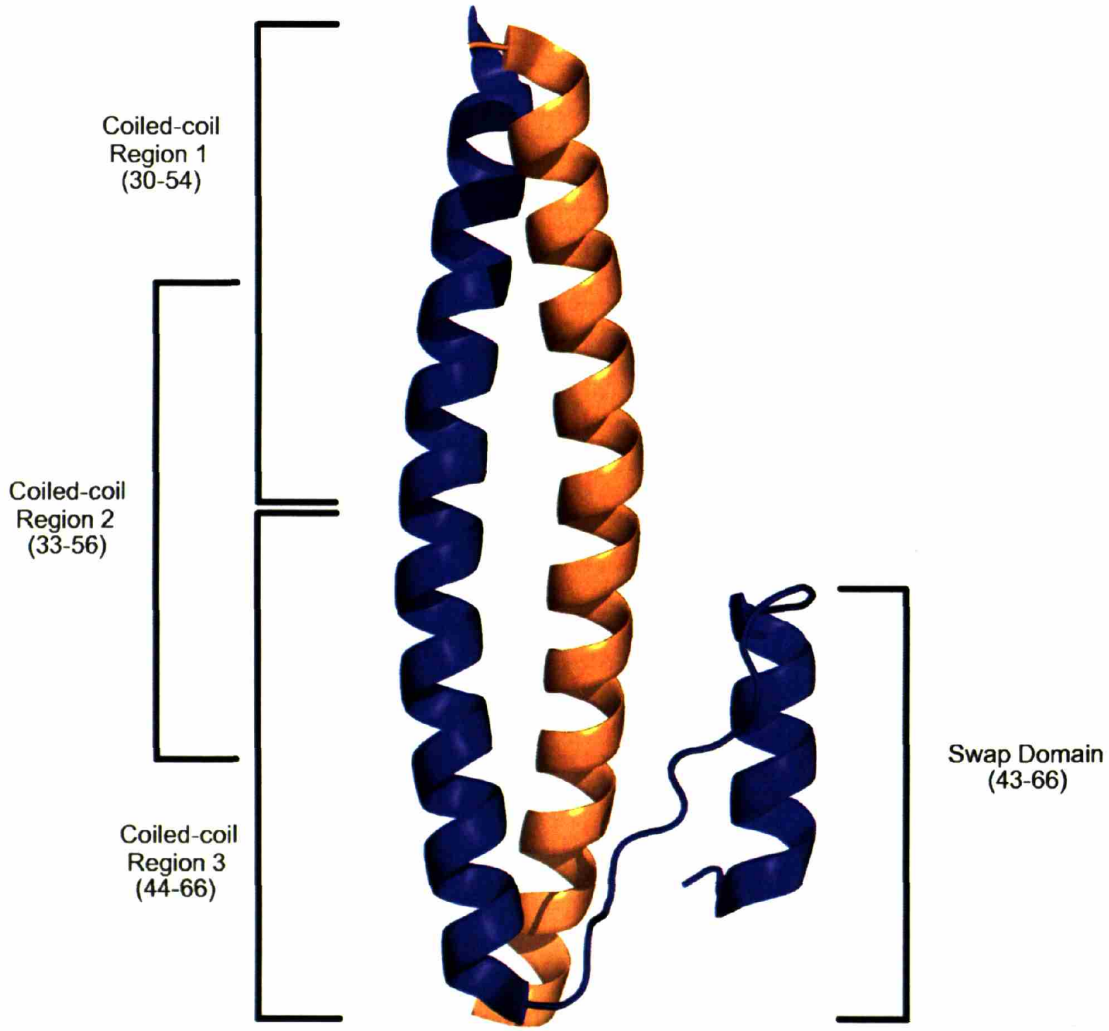


Figure 4. Bcr regions for computational design. The BCR oligomerization monomer (shown in blue) with the model inhibitor (shown in yellow) was divided into four different regions to make computation more tractable.

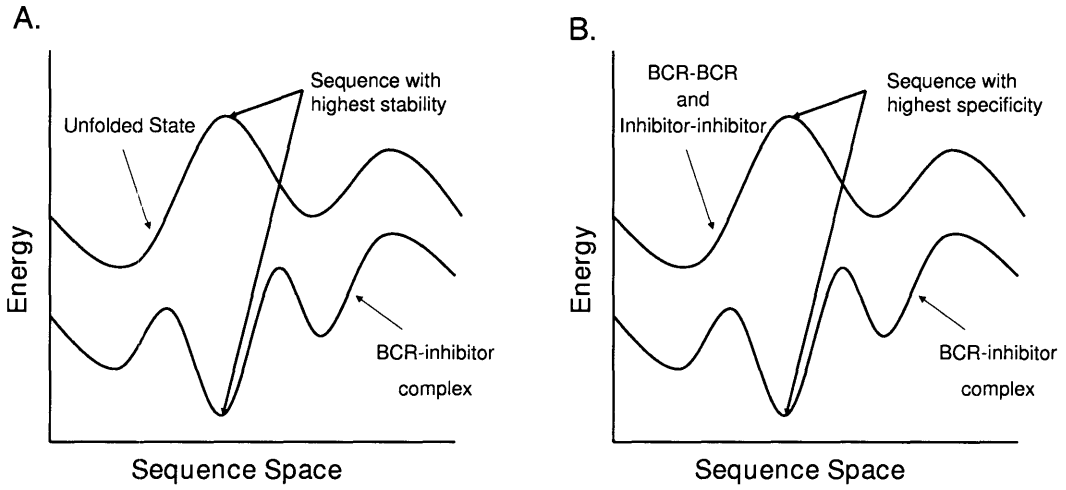


Figure 5. Stability and specificity optimization. (A) When designing for a highly stable sequence, the sequence that maximizes the energy difference between the BCR-inhibitor complex and a model for the unfolded state is chosen ($f = -E_{folded} + E_{unfolded}$). (B) When designing for specificity, the most specific sequence has the largest difference between the BCR-inhibitor versus the Bcr-Bcr and the inhibitor-inhibitor and maximizes the following equation: $f = -2E_{Bcr\_inhibitor} + E_{inhibitor\_inhibitor} + E_{Bcr\_Bcr}$

Peptide	Sequence
C-BCR	GSCGG-DIEQELERAKASIRRLEQEVNQERFRMIYLQTLLAKE-GGWK
BCR-C	GSKGG-DIEQELERAKASIRRLEQEVNQERFRMIYLQTLLAKE-GGC
ICT1	GSCGG-DIEQE**V**ERAKASIRR**KL**QE**I**NQERFR**L**IYLQTLLAKE
ICT2	GSCGG-DIEQ**R**LERAKA**R**IRRLEQEVNQERF**IL**IYLQT**R**LAKE
ICT3	GSCGG-DIEQ**R**LERAKA**R**IRRLEQEV**E**QERF**QE**IYL**R**T**RR**AK**R**
ICT4	GSCGG-DIEQELERAKASIRR**A**EQE**L**NQERFR**L**IY**IE**TLLAKE

Figure 6. BCR and BCR inhibitors that were designed and tested.

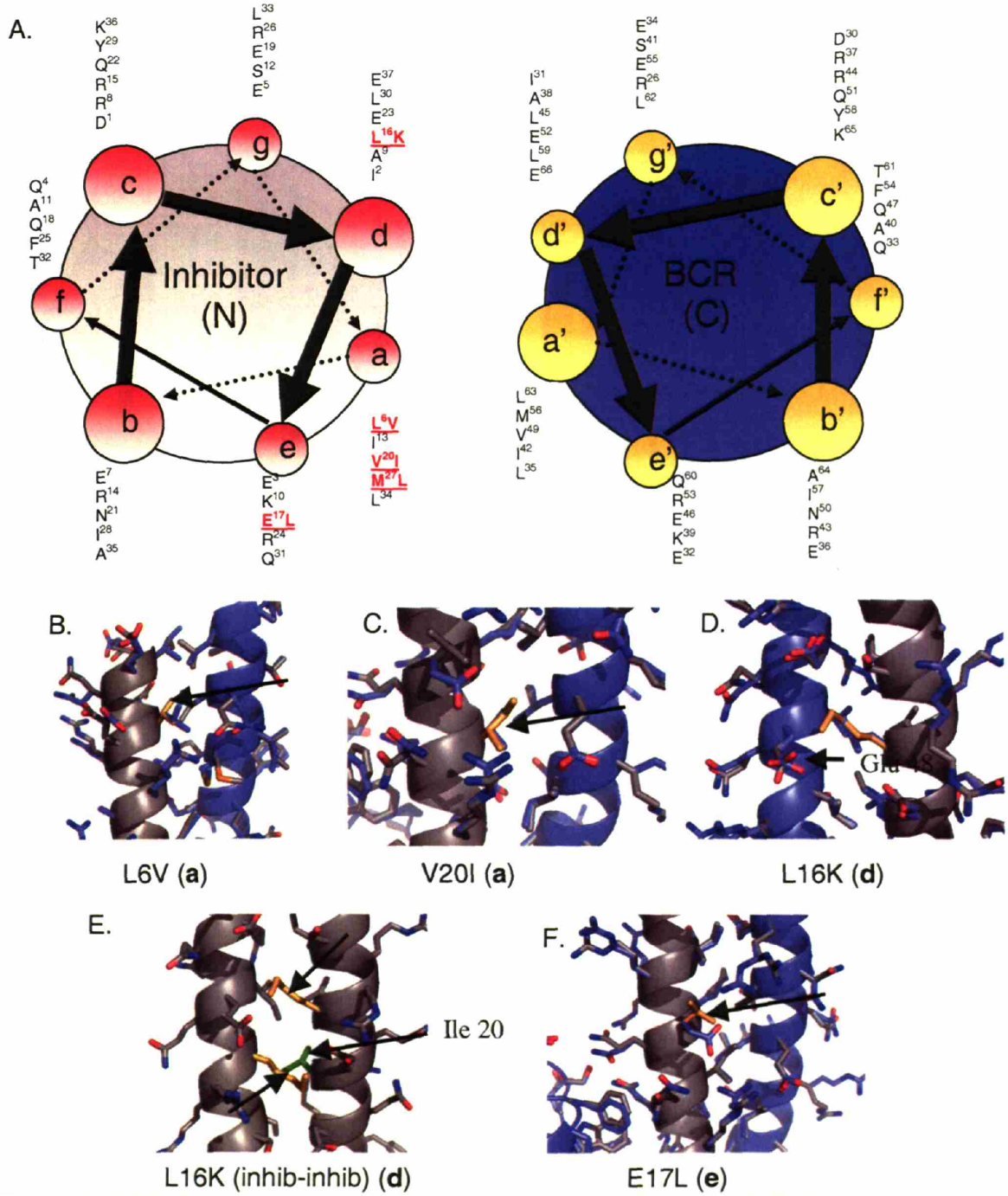


Figure 7. ICT1 designed interactions. (A) Helical wheel diagram. (B-F) Interactions were designed using the basic energy function. The Bcr homodimer x-ray crystal structure is shown in blue superimposed on the designed inhibitor and repacked Bcr structure in gray. The designed residue is shown in orange and marked with an arrow, and surrounding residues with poor rotamer conformations are shown in green.

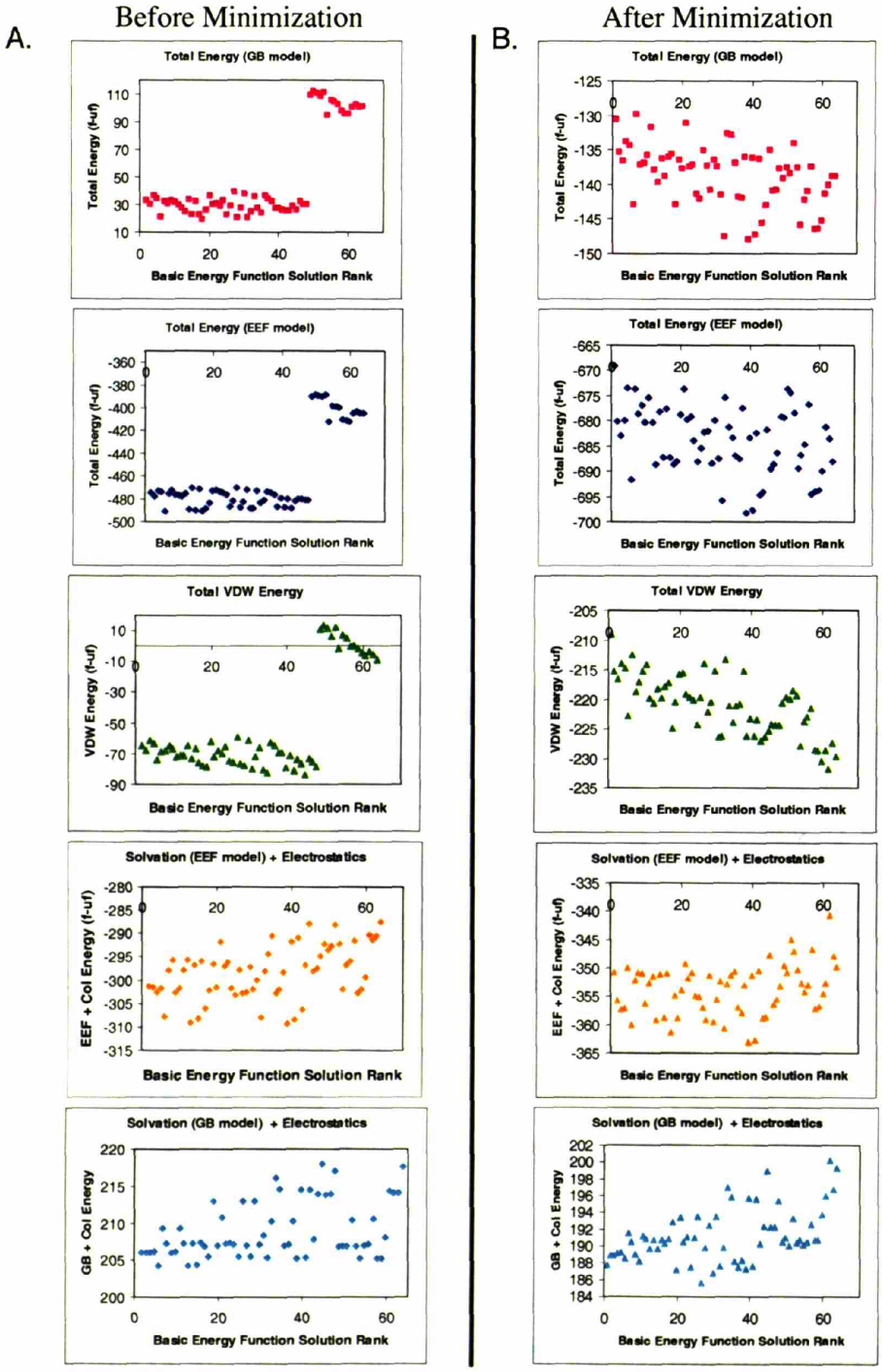


Figure 8. Effect of minimization on energy. Solutions from a very small-scale selection before and after minimization evaluated by van der Waals, solvation and electrostatics, and total energy. (A) Solutions before minimization. (B) Solutions after minimization.

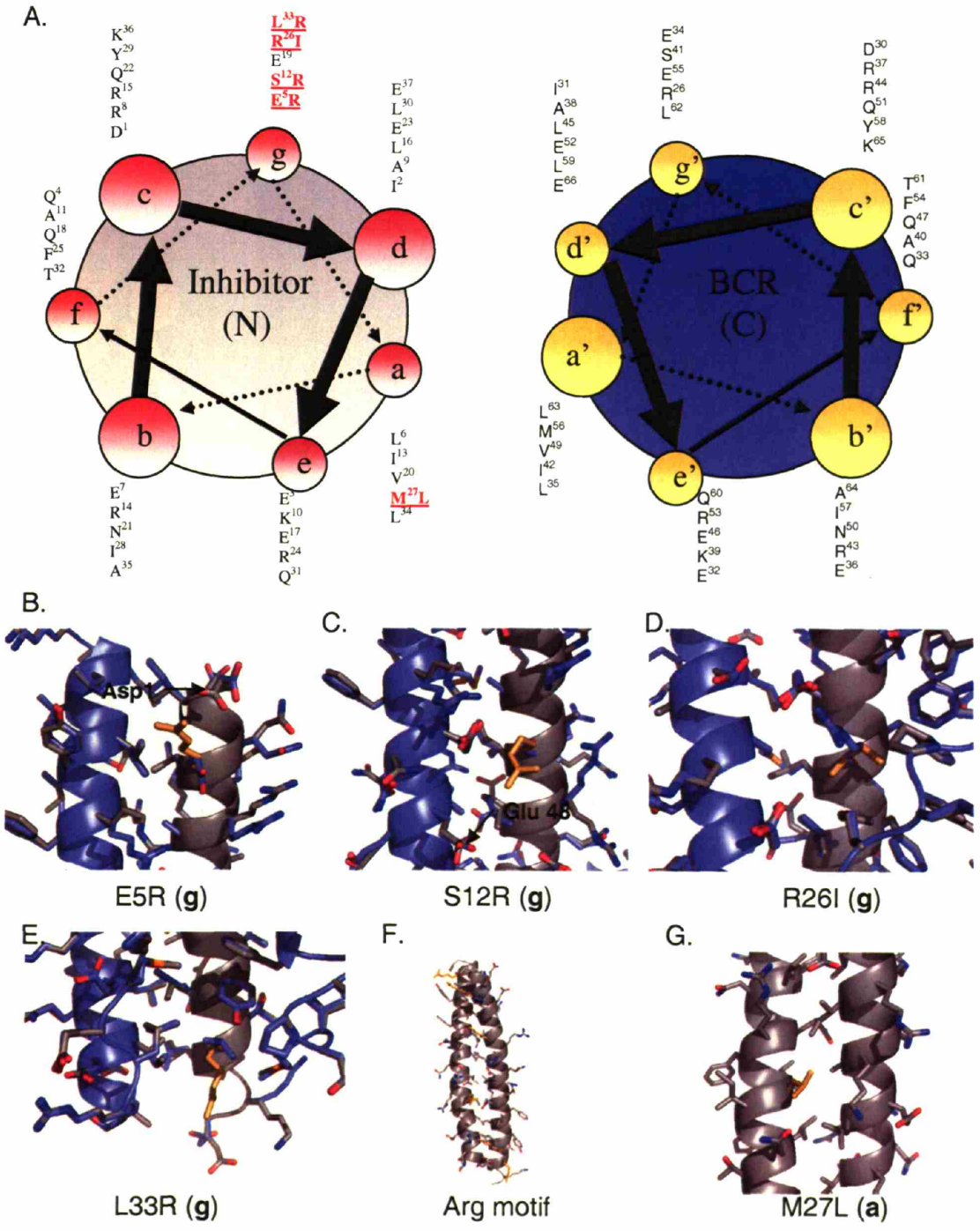


Figure 9. ICT2 designed interactions. (A) Helical wheel diagram. (B-G) Interactions were designed using GB to calculate the solvation term (GB Model). The Bcr homodimer x-ray crystal structure is shown in blue, and the designed inhibitor and repacked Bcr structure are shown in gray. The designed residue is shown in orange and marked with an arrow.

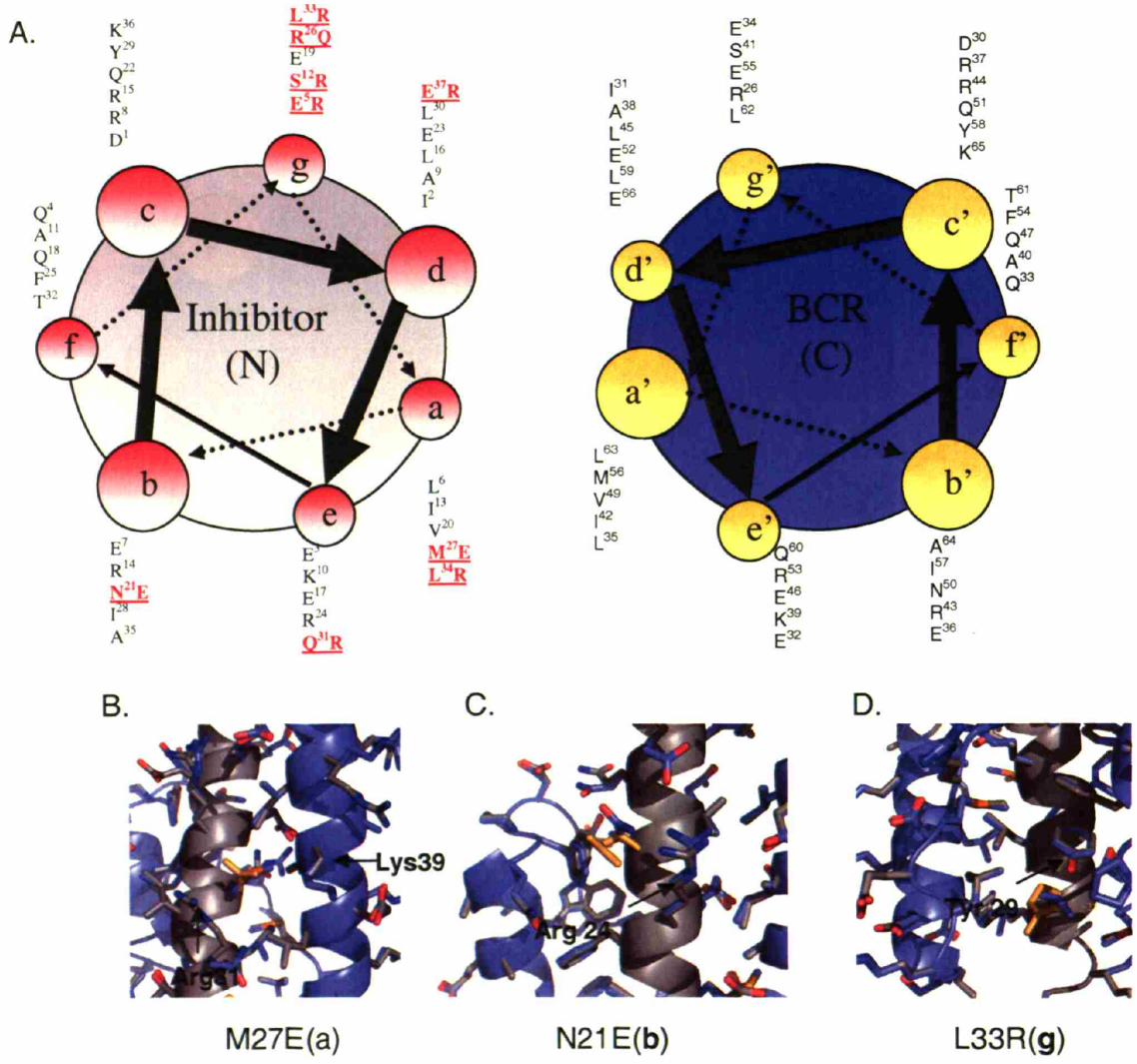


Figure 10. ICT3 designed interactions. (A) Helical wheel diagram. (B-D) ICT3 interactions designed with the EEF/GB Model. The Bcr homodimer x-ray crystal structure is shown in blue, and the designed inhibitor and repacked Bcr structure are shown in gray. The designed residue is shown in orange and marked with an arrow.

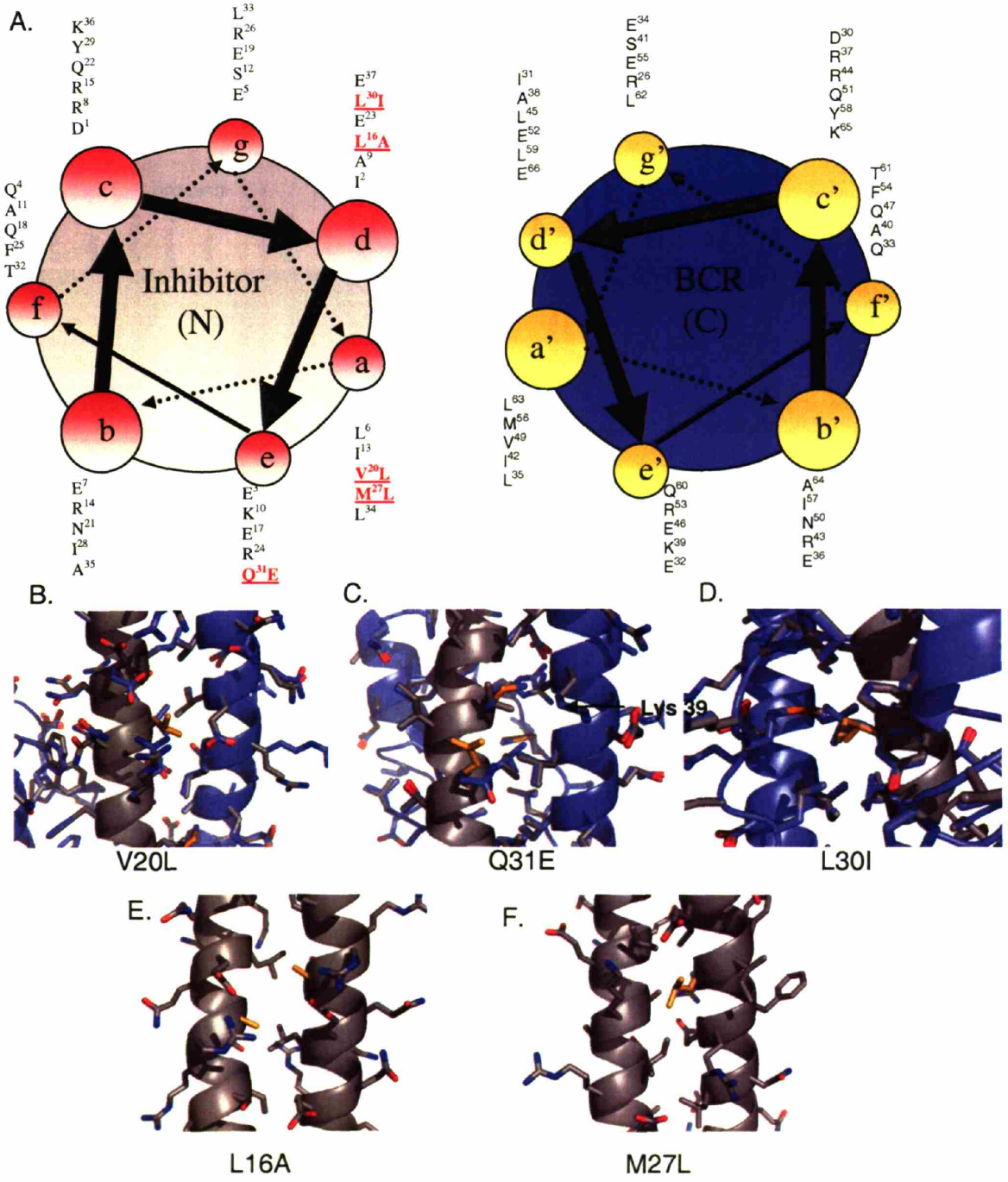


Figure 11. ICT4 designed interactions. (A) Helical wheel diagram. (B-F) Interactions from the ICT4 inhibitor, designed using both GB and EEF/GB energy functions ("Combo Model"). The Bcr homodimer x-ray crystal structure is shown in blue, and the designed inhibitor and repacked Bcr structure are shown in gray. The designed residue is shown in orange and marked with an arrow.

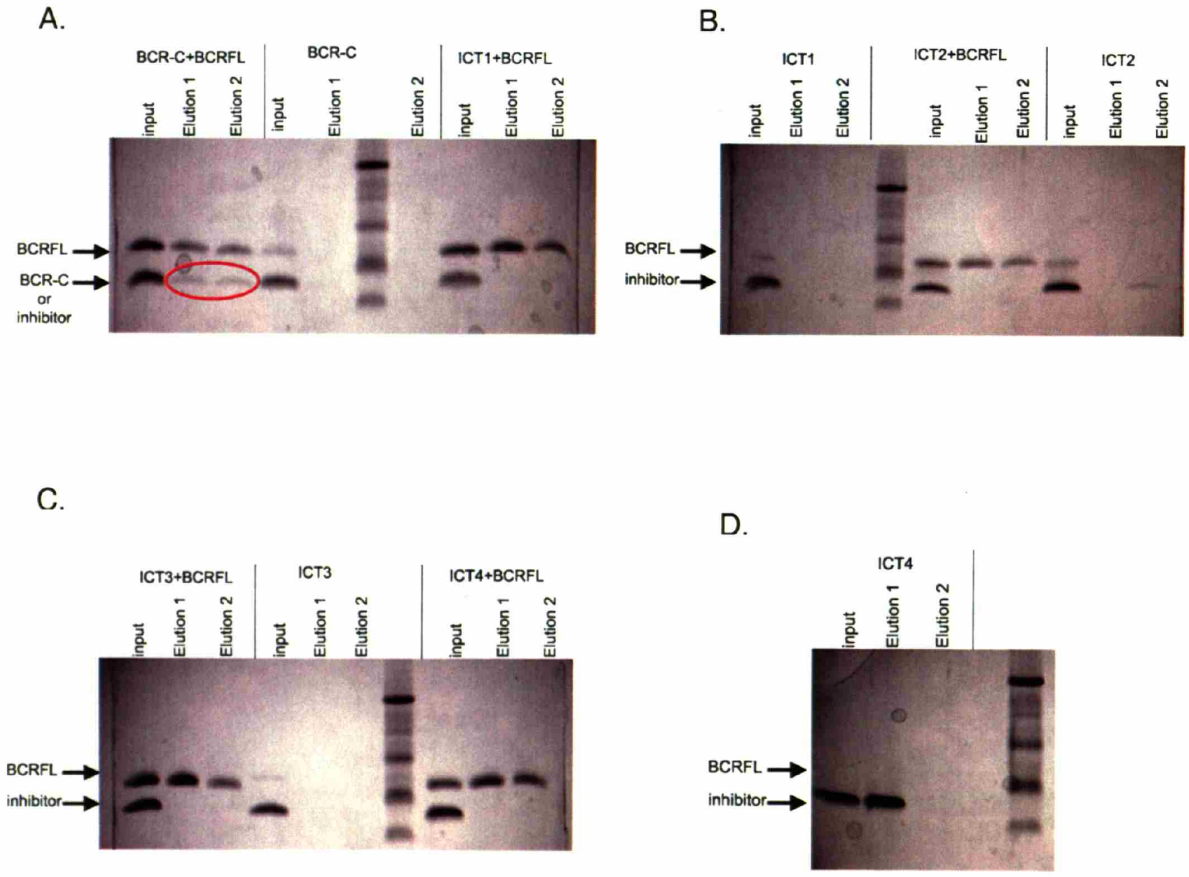


Figure 12. Pull-down assay. (A) Positive and negative control for BCR-C. Pull-down for ICT1. (B) Negative control for ICT1. Pull-down and negative control for ICT2. (C) Pull-down and negative control for ICT3. Pull-down for ICT4. (D) Negative control for ICT4. Experiment run by J. Fisher.

APPENDIX A

Insights into the Computational Design of a Heterospecific ββα Protein

Both successful and unsuccessful protein designs can yield important information which can lead to improvements in the design process. In the first part of this appendix, attempts to stabilize the BBA homotetramer (**BBAT2**) and additional details about design of the BBA heterotetramers (**BBAhetT1** and **BBAhetT2**) are discussed. In the second part, an analysis of the **BBAhetT1** crystal structure and designed structure is presented to suggest ways the design process could be improved to obtain more accurate protein designs. A discussion of a third computational design round with experimental verification follows. Finally, a fourth round of computational design was done, and mutations are suggested for experimentation.

Attempts to Stabilize the BBA Homotetramer (BBAT2)

Early work on this project attempted to stabilize **BBAT2** using computational design techniques. **BBAT2** contains five layers (Figure 1), three of which are unique. Layers A and A' (containing residues Ile 3 and benzoylated L-α,β-diaminopropionic acid (DapBz) 20), layers B and B' (containing residues Phe 8 and Leu 16), and layer C (containing Leu 12) all contribute to packing in the core. Each layer was computationally redesigned separately to increase stability. Optimizing stability using layers B, B', and C simultaneously resulted in minuscule energetic improvement.

For layers A and A', problems modeling the DapBz molecule in CHARMM and constructing a good rotamer library for DapBz, led to higher energies for solutions containing DapBz than solutions containing nearly every other amino acid at position 20. Problems modeling DapBz potentially stemmed from introducing constraints in CHARMM that dictate planar geometry in the DapBz ring structure and carbonyl group during minimization. Optimal DapBz rotamers were found by evaluating energies about chi angles in DapBz using CHARMM. Chi angles that resulted in minima were used to create the DapBz rotamer library. Greater

sampling around the low-energy chi angles may have also reduced problems with DapBz in the design calculations. When all residues were considered at layers A and A', Lys/Tyr and Lys/Phe combinations did well. When charged residues were excluded from the selection, Ile/Met and Leu/Tyr scored well. These residues did not fill the core as well as DapBz, however. None of the stabilization designs seemed promising based on structural inspection and energy score.

Additional Details Regarding the Design of the BBA Heterotetramers

Details regarding the design of the BBA heterotetramers (**BBAhetT1** and **BBAhetT2**) can be found in Chapter 2. The purpose of this section is to discuss the thought process and computational details behind design rounds 1 and 2. Because both core and surface regions do not interact strongly, rounds 1 and 2 of the design process treated the core and surface regions separately, making computation faster and results easier to interpret. Experimental data were also used to guide the design process. Sites modeled in alternate conformations or those mutated to SeMet in the **BBT2** crystal structures were considered to be excellent design candidates because they were unlikely to affect the basic BBA structure. To mimic backbone flexibility, most of the computational selections were done on both available experimental SeMet-mutated **BBAT2** structures: 1SNE (denoted bb1) and 1SNA (denoted bb2).

Design of the core

In the core, layers A/A', B/B', and C were individually selected for heterospecificity. Similar steric matching patterns emerged from the designs of B/B' and C layers. Redesigning the layer A/A' was problematic due to modeling DapBz.

Designing the B/B' layer for heterospecificity was not promising, as many core residue pairings that emerged were polar and not paired with an appropriate hydrogen-bond donor or

acceptor. The promising sequences and heterospecific energy scores are shown in Table 1a. If hydrogen bonds had been satisfied in the core, this would have been a fine method for introducing specificity, as has been seen in several previous coiled-coil studies (*1, 2*). However, unpaired hydrogen bond donors and acceptors in the core are likely to destabilize the structure significantly. In our calculation, hydrogen bonding was calculated as a combination of electrostatic interactions, modeled using a distance-dependant dielectric and van der Waals interactions. Only hydrogens on polar atoms were explicitly represented. The EEF solvation term should have prevented unpaired hydrogen bonds in the core, as desolvating a polar atom is unfavorable in the absence of a compensating interaction.

Obtaining an unstable result is not surprising however, given specificity was optimized, not stability. When optimizing specificity, the energy between the stabilities of the heterotetramer and two homotetramers was maximized. Large clashes found in homotetramer states, combined with good van der Waals packing in the heterotetramer state, caused the energy difference between the homo- and heterotetramer states to be large, regardless of the whether core residues were hydrophobic, charged, or polar. The heterotetramer stability was not optimized in the specificity calculation; however, a stability calculation would have disfavored unsatisfied hydrogen bonds in the core.

Many of the best solutions for layer B/B' contained low-probability rotamers, which are considered undesirable in a potential design. Rotamers seen infrequently in the PDB often cause strain, and strained rotamer conformations can adversely affect protein stability and conformational specificity (*3, 4*). Using probabilities from the Dunbrack rotamer library, the rotamer conformations for the top-scoring layer B and B' solutions were checked. In nearly all promising solutions, at least one rotamer conformation had low probability, shown in white in

Table 1b. The structure of BBA may accommodate a limited number of residues in the B layers; layers A and A' reach into the core, restricting where residues in the B and B' layers can pack. In addition, DapBz was not repacked, further restricting the accessible space. Because the best solutions contained unfulfilled hydrogen-bonding partners in the core and low-rotamer probabilities, introducing heterospecificity into this layer was not explored further.

The middle BBA layer (layer C), containing residue 12, was also redesigned for heterospecificity. Although more promising, some of the same problems seen in the B/B' layers existed in layer C. Some solutions were solely hydrophobic, eliciting no worry about destabilization due to unsatisfied hydrogen bonds, shown in Table 2a. Some residues selected, like Tyr and Trp, left possible hydrogen-bond donors and acceptors unsatisfied, as was seen in layers B and B'.

Layer C was much more promising, as many solutions used only rotamer conformations seen often in the PDB, shown in Table 2b. One high-scoring solution for heterospecificity was also among the top solutions when heterotetramer stability was optimized. The stability optimization results for both backbones ranked Phe/Ala as 5^{th} when all combinations of amino acids (except for Cys, Gly, and Pro) were enumerated at position 12.

The specificity energies would have been more realistic had minimization been incorporated into the selection process to relieve the large steric clashes in the homotetrameric state. In the design of the B/B' and C layers, looking at both the specificity and stability energies of individual components (i.e. the two homotetramer and heterotetramer states) can be helpful, as specificity energy is sometimes very high due to bad steric clashes on one homotetramer model. For example, in Tables 1a and 2a, Trp and Phe are predicted to make good heterotetramers based on specificity. Trp and Phe have steric clashes in the homotetramer state which creates a large

van der Waals term and unrealistically high heterospecificity energies. However, steric clash is relieved in the heterotetramer. A small amount of minimization could reduce the steric clash in the homotetramer state, making the specificity score more realistic.

To introduce a small degree of backbone flexibility to relieve steric clashes, the design was done on bb1 (1SNE) and bb2 (1SNA) backbones, which each deviate by ~0.7 Å from the **BBAhetT1** crystal structure. Designs based on the two SeMet-mutated **BBAT2** x-ray crystal structures showed large differences in component stabilities (AAAA, BBBB, and ABAB) and heterospecificity energies. Further, solutions were ranked differently between the two crystallographic backbones. This makes a compelling argument for adding a small amount of backbone flexibility to design calculations, as one solution may give a large heterooligomeric specificity score on one backbone and not another. The backbones giving the lowest energy structure for each hetero- and homotetramer would be considered in the specificity score calculation.

Surface Selection

The surface of **BBAT2** has two distinct faces that were treated separately in our designs. Residues on the surface that do not interact in **BBAT2** but have potential to do so were targeted, including position 13 on one face and positions 11 and 18 on the other. Position 13 is solvent exposed and located near the middle of **BBAT2** (See Figure 6 in Chapter 2). When position 13 was mutated to SeMet to solve the **BBAT2** structure, the structure remained unperturbed, indicating position 13 could be altered without disrupting the overall BBA fold. Position 13 residues are directly across from each other on two **BBAT2** faces, making position 13 a potential site for introducing a salt bridge.

Based on the predicted specificity energies and structures from the calculations, Glu and Lys were clearly the best solutions at position 13, shown in Table 3. Predicted structures showed Glu and Lys at an ideal distance, making an interhelical salt bridge, whereas Asp and Lys were further apart and not in positions as conducive to salt bridge formation. Asp and Arg can interact, but structurally Glu and Lys form a salt bridge with closer contacts. Glu and Arg are long enough to interact, but their large size precludes good salt-bridge interaction geometry at this position. Glu and Lys were chosen for position 13 on the designed heterotetramer because of the results of the computational selection results, evaluation of the predicted structure, and the ACID/BASE peptide model which had been shown to yield heterospecificity in a coiled coil (*5*).

In a separate selection for specificity, positions 11 and 18, both surface exposed residues on either end of the helix, were redesigned. In **BBAT2**, monomers associate in an antiparallel fashion, giving position A11/B18 and A18/B11 potential to form interhelical interactions. Using two different backbones (bb1 and bb2) and a 17-residue amino-acid library (Gly, Cys, and Pro were excluded), a selection was done using a Monte Carlo algorithm to sample sequence space. Using a temperature gradient from 300 to 200 K, Monte Carlo selection was run for 128 rounds, with 1500 steps per round. The top results are in Table 4. This selection yielded charged residues at most of the 11 and 18 positions; however, in most cases at least one other type of amino acid was selected. Isoleucine came up a great deal in one of the positions in many of the top solutions.

Based on knowledge that the current energy function does not penalize surface exposed hydrophobics and with concerns over solubility, a second selection was limited to only charged residues. The sequences were enumerated, due to the smaller size of the selection, and evaluated in terms of specificity. Table 5 shows the results from the smaller selection. Lys and Glu

formed salt bridges in some predicted structures. In some cases, the rotamer that has potential to make an interhelical salt bridge was not selected, causing the interaction to not look as favorable in the selected structure. In the top solutions selected for the ACID peptide, Glu and Asp yielded a slightly more specific peptide than Glu and Glu at positions 11 and 18, respectively. Based on models, Arg and Glu are too large to make a good salt bridge between positions 11 and 18. Further, Asp interacts with Lys on the other helix with greater ease than Glu at position 18 due to its length.

Therefore, the selections predict an ACID/BASE chemistry on the surface to introduce heterospecificity. This design also makes sense based on previous heterooligomer design literature (21). Lys residues were placed at the 11, 13, and 18 positions of one helix (BASE), and with the exception of the 18 positions, Glu was placed at the 11 and 13 positions on the other helix (ACID). Asp was chosen for position 18 on the ACID peptide based on the predicted structures.

A final design calculation incorporated the surface residues fixed in the ACID/BASE motif and a very limited library consisting of Phe, Ala, Leu, and Trp at position 12. The ACID motif was enforced on the A peptide, and the BASE motif on the B peptide. The results are shown in Table 6. Again, Trp ranked near the top for specificity, but ranked low for stability. In this calculation, the low stability can most likely be attributed to the unsatisfied hydrogen-bond acceptor in Trp, leaving the polar NH buried in the hydrophobic core. The heterotetramer specificity score was significantly less for the heterotetramer containing Phe/Ala than Trp/Ala. However, the stabililty score for Phe/Ala was much greater than that for Trp/Ala. Thus, Trp was eliminated as a possibility for position 12. Phe packed well in position 12 and gave a substantial amount of heterotetramer specificity. Phe/Ala at position 12 and the ACID/BASE surface

patterning (denoted **A-Ala** and **B-Phe**) were chosen for testing in the first round of protein design.

Experimentally, **A-Ala** and **B-Phe** formed heterotetramers, based on fluorescence experiments, and **A-Ala** and **B-Phe** were monomeric at 25 μM concentrations. However, **B-Phe** also self-associated at concentrations of 50 μM and above. This heterotetramer will be referred to as **BBAhetT1**. In addition, the crystal structure of **BBAhetT1** was solved.

Steric matching in the core and charge complementarity on the surface were also evaluated independently. A peptide containing Leu/Leu at position 12 with ACID/BASE surface patterning (**A-Leu/B-Leu**) tested the heterospecificity arising from charged complementarity on the surface. Calculations predicted these peptides would be heterospecific, as shown in Table 7, but not to the extent of other solutions. The predicted specificity energy arising from the **A-Leu/B-Leu** is likely to be closer to realistic specificity energies, because core steric clashes that make van der Waals energies extremely high were not present. Experimentally, **A-Leu** and **B-Leu** both self-associated in solution at low concentrations, but form a heterotetramer when mixed together.

Optimization – Round 2

The experimental results from round 1 indicated that **A-Ala/B-Phe** form a heterotetramer, but the thermal stability of this complex was much lower than for **BBAT2**. The predicted packing for Phe/Ala in position 12 was not ideal, as there was empty space in the core that could potentially be filled to stabilize the complex. All 17 amino acids (excluding Cys, Pro, and Gly) were evaluated in design calculations at this position for their ability to stabilize the heterotetramer and preserve specificity.

Although polar residues in the core are typically not favorable, the Phe/Ser structures packed well. The hydroxyl group filled the core much better and left a large hole in the Ser/Ser homotetramer state. Rather than using Ser, a hydrophobic non-natural amino acid, L-α-aminobutyric acid (Abu), was chosen to stabilize the heterotetramer. Met looked reasonable, but was too large in the heterotetramer structure and had potential to be too favorable in the homotetramer state. According to round 1 calculations, Phe/Met was heterospecific, but did not have a heterospecificity score nearly as large as other amino-acid combinations on both backbones. Rather than trying Met, a slightly shorter non-natural amino acid, 2-aminohexanoic acid (norleucine or Nle), was added to the calculation. A comparison of structures is shown in Figure 2.

Another calculation checked the heterospecificity of Phe/Abu and Phe/Nle compared to peptides tested in round 1. The selection results are shown in Table 8. On one of the backbones, Phe/Abu was calculated to be more stable than Phe/Ala. Phe/Abu packed well in the heterotetramer state, leaving a large hole to destabilize the homotetramer state. When the structure was minimized and only van der Waals and surface area-based solvation energy contributions considered, Abu was predicted to stabilize the structure by about ~4 kcals/mol on both model backbones. Depending on the backbone used, very little, if any, specificity was predicted to be lost by changing Ala to Abu. Phe/Nle was more stable than Phe/Abu and also packed well, but some heterospecificity was compromised by greater stabilization from the extra methyl group in the Nle homotetramer. Based on these results, only **A-Abu** was made and tested. The experimental results from round 2 indicated that **A-Abu/B-Phe** on the surface was more stable and still retained heterospecificity.

Evaluation of the Design Process with the Crystal Structure

The crystal structure of the **BBAhetT1** was solved, allowing further evaluation of the design process and energy function performance. A brief evaluation of the designed structure versus the crystal structure was done. Possible improvements to the design process, such as use of wild-type rotamers and different energy functions used for repacking, were evaluated. Unless stated otherwise, energies were re-evaluated with 100% van der Waals radii with the complete Generalized Born (GB) solvation model.

Overall agreement of designed and experimental structure

The x-ray crystal structure and designed structure of **BBAhetT1** superimpose well. For the purpose of this section, the structure designed using the **BBAT2** backbone (bb1) was used for comparison, unless otherwise noted. The core residues superimposed extremely well. The backbone/side-chain RMSD was 1.76 Å, and the backbone RMSD was 0.71 Å. Most of the amino acids in boundary positions were designed to give interhelical salt bridges. With the exception of Leu 15, Tyr 6, and the designed interhelical salt bridges, residues that were repacked agreed well with the crystal structure.

Assessing computational approximations using side-chain repacking

There are differences between repacking and design calculations. When repacking, the amino-acid sequence and backbone are known and the energy function dictates side-chain placement. With design, both the amino-acid sequence and side-chain placement are important. When repacking, discrete rotamers are the main approximation (beside those in the energy function); however, both discrete rotamers and backbone are being approximated in design. Thus, repacking accuracy does not necessarily mean better performance in design. Both

approximations can decrease the accuracy in obtaining exact inter side-chain and side-chain / backbone interactions, and combining the approximations in design calculations can further diminish accuracy. Nonetheless, side chain repacking was an effective way to evaluate the accuracy of the design methods.

Given the backbone of the designed and crystal structures were very close, the question remains as to whether small differences in backbone can account for discrepancies in side-chain placement in the two structures. Designing on the heterotetramer crystal structure backbone versus the homotetramer backbone yielded a 7.4% improvement in the correct placement of rotamers at chi 1; however, only 70.4% of chi 1 angles were correct. A dramatic improvement in side-chain placement was seen when energies used for repacking were calculated using a pairwise decomposable GB solvation model (*6, 7*) and the heterotetramer crystal structure backbone. With the EEF (effective energy function) model (*8*) used in the initial design calculations, 70.4% of chi 1 positions were correct, whereas 83.3% were correct with the GB solvation model. Such drastic differences between EEF and GB solvation models were not seen with the designed structure backbone, as the performance only increases from 63.0% (EEF) to 68.5% (GB). The energy functions, particularly the GB solvation model, seem to be very sensitive to the backbone used. Further, there are significant differences between repacking accuracy on the exact backbone versus an approximate backbone.

In predicted structures, the position of Tyr 6, located toward the end of the BBA helix on the surface, was consistently wrong. Tyr 6 packs against the helices in the **BBAhetT1** crystal structure, but is much more solvent exposed in the **BBAhetT1** designed structure. The hydroxyl group on Tyr 6 can interact with positions 11 and 18. In the x-ray crystal structure, two Tyr 6 residues interact with Glu and Lys at positions 11 and 18, but the other two Tyr 6 residues do not

interact with Asp and Lys because they are involved in crystal contacts. In the crystal structure, Tyr 6 is close to one of the rotamers in the rotamer library, but this rotamer does not allow it to be packed ideally against the BBA structure.

To obtain better repacking results for Tyr 6, wild-type rotamers were added to the rotamer library. Wild-type rotamers refers to rotamers taken from the structure on which the design was done. For example, if the **BBAhetT1** crystal structure was used in the design, the rotamers from the **BBAhetT1** crystal structure were used. Using the **BBAhetT1** crystal structure, and expanded rotamer library, two of the four Tyr 6 residues were placed correctly. Upon using a GB solvation model (with solvent accessible surface area parameter scaled by 7 cal/mol•Å^2), rather than EEF, all four Tyr 6 residues were placed in their correct positions when the expanded library was used. With the designed backbone, only 2 of the 4 were placed correctly with the GB solvation model. Experience in the lab following this analysis has found expanding the aromatic rotamers +/- 10° gives improved repacking performance. This expanded library might have helped in BBA design.

The performance of GB versus EEF at this position indicates that GB may model the hydrophobic effect better than EEF. Regardless of whether the 11/18 salt bridge forms, Tyr 6 packs in the same manner. Further, Tyr 6 packs in the same manner in **BBAT2**. Therefore, the problem is most likely due to hydrophobic interactions and not interaction with positions 11 and 18. Solvating the largely hydrophobic Tyr 6 probably incurred a much larger penalty with GB than it did with EEF. This may have caused Tyr 6 to pack against BBA in structures repacked using GB. Positions 11 and 18 in **BBAT2** do not interact with Tyr 6, so Tyr 6 in **BBAT2** would not be biased by those positions if it were repacked. Repacking the homotetramer structure

should be done before conclusions are drawn as to whether GB modeled the hydrophobic effect or charged interactions better than EEF.

Overall, adding wild-type rotamers to design calculations yielded slight improvements in side-chain placement in repacking tests. Improvements were seen regardless of backbone or energy function used. In a few cases, there was a slight decrease in the percentage of correctly placed rotamers when wild-type rotamers were added, but overall there was some improvement. Therefore, incorporating wild-type rotamers into design might be useful without adding a great deal of computational time. In addition, the GB solvation energy function was better at modeling Tyr 6.

The most striking difference between the designed and x-ray crystal structures was the formation of salt bridges. In the **BBAhetT1** crystal structure, nearly all designed surface residues made some sort of crystal contact, with only two out of six designed intratetramer salt bridges forming. Modeling the crystal lattice and accounting for crystal contacts in design calculations was impossible because the space group in which the protein crystallizes cannot be predicted. In fact, **BBAhetT1** crystallized in a different space group than all three of the **BBAT2** native and SeMet crystal structures. In the case of the two position 11 to 18 designed salt bridges that formed, as predicted, in the crystal structure, there were no competing crystal contacts nearby.

Interesting differences in salt-bridge formation were seen between very similar backbone structures. When residues were repacked on the **BBAhetT1** crystal structure backbone, fewer designed salt bridges formed than on the designed backbone. Although not very important for repacking core hydrophobics, the very similar backbones made a larger difference in the placement of residues on the surface.

As expected, modeling solvation with GB versus EEF allowed formation of different salt bridges. Overall, GB yielded fewer designed salt bridges at selected positions (11, 13, and 18), but about the same number or more total salt bridges. Adding wild-type rotamers to the calculation caused more salt bridges to form, when GB versus EEF was used. In predicted structures, salt bridges tended to be further apart when GB versus EEF solvation was used. Any information taken from this analysis should consider the caveat that crystal contacts, which influenced salt-bridge formation in **BBAhetT1**, were not taken into account in this analysis. Both energy functions should be tested using the crystal lattice of the **BBAhetT1** structure to determine which energy function performs better. Using the lattice, performance in predicting salt bridges can be more accurately determined. Solution structures of **BBAhetT1** would be necessary to draw definitive conclusions about the effects of backbone, incorporation of wild-type rotamers, and energy function performance, in regard to salt-bridge formation.

Mitigating Homo-oligomerization – Round 3 Design

For an ideal, highly specific heterooligomer, individual components would not self-associate. However, in our study, analytical ultracentrifugation experiments indicated the **B-Phe** monomer self-associated at concentrations 50 μM and above. To reduce **B-Phe** monomer self-association, another round of design was done.

Both x-ray SeMet **BBAT2** crystal structures (bb1 and bb2) were used for round 3 calculations, and a "core" and "surface" strategy, similar to the strategy used in round 1, was used for round 3. Round 1 calculations on layer C were checked again for any other natural amino-acid combinations that might yield greater heterospecificity, but none seemed promising. Position 12 was reselected to try to improve heterospecificity.

The "core" strategy focused on methyl-substituted Phe derivatives at position 12, which could add stability to the heterotetramer state and also make the **B-Phe** homotetramer state less favorable by introducing further steric clashes. The following phenylalanine derivatives, shown in Figure 3, were added to a restricted library consisting of Leu, Ala, Phe, Abu, and Nle for round 3 selections: 2-methyl-L-phenylalanine (Otl), 4-methyl-L-phenylalanine (Tol), and 3-methyl-L-phenylalanine (Mtl). Phe rotamer conformations were used for all Phe derivatives. The results are shown in Table 10a and b. As before, specificity was optimized by maximizing the following equation: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. The stability of the homotetramer and heterotetramer structures were also compared. The stabilities of BBBB and ABAB were approximated by using the folded energy without subtracting the unfolded state. Because the BBBB structures were nearly identical, only differing by one or two methyl groups, this was a reasonable approximation. Heterotetramer structures ABAB could also be compared in this manner. The stability of ABAB was also calculated as unfolded minus folded energy.

Based on the results of the "core" selection, Otl/Abu and Otl/Ala are predicted to increase both the stability and the specificity. Mtl was predicted to increase specificity by a great deal, but to destabilize the structure. The current calculations do not model backbone flexibility; however stability may not be compromised with Mtl in the core, due to backbone relaxation in solution. Tol was not predicted to contribute much more to specificity or stability. Destabilizing the homotetrameric state was the main goal of the third selection round, so efforts were focused on characterizing structures with Mtl/Abu.

The "surface" strategy involved varying residues at positions 11, 13, and 18 on **B-Phe**. The first approach to mitigate self-association of **B-Phe** involved trying various lysine derivatives. Lysine residues on the surface of **B-Phe** have many degrees of freedom, possibly

making them less effective in preventing self-association by charge repulsion; the lysine residues may be able to make alternative interactions, preventing charge repulsion from being an effective negative design strategy. Removing methylene groups eliminates many degrees of freedom, causing the side chain to make a more defined interaction. In addition, the helices are closer on either end of the helix than in the middle. In some of the predicted structures, the Lys/Glu and Lys/Asp pairs would have been able to make better interactions had a methylene been removed from one of the amino acids. Two lysine derivates, L-ornithine (Orn) and 2,4-diaminobutyric acid (Dab), (shown in Figure 4) differ by one and two methyl groups from lysine, respectively, and were computationally tested to determine if the mutations would mitigate self-association of **B-Phe** and increase heterospecificity of **BBAhetT2**.

Positions 11 and 18 were selected in one calculation, and position 13 was selected in another. In the position 13 calculation, all combinations of the following amino acids were computed: Ala, Orn, Lys, Arg, Glu, Asp, Dab. The results shown in Table 11 favor retaining Glu on **A-Abu** and Lys on **B-Phe**. Therefore, Glu and Lys were kept at position 13.

In the position 11 and 18 selection, all combinations of the following amino acids were computed: Glu, Arg, Orn, Lys, Asp, Dab. The results are shown in Table 12. Based on the calculations, two Lys residues at 11 and 18 on **B-Phe** still appear to provide the most specificity. In the bb2 model, the Orn/Orn and Orn/Dab pairs are more stable on **B-Phe** than the Lys/Lys pair, whereas the Lys/Lys pair is more stable on **B-Phe** than Orn/Orn and Orn/Dab in the bb1 model calculation. Therefore, the stability results are ambiguous when both backbones are considered. All three pairs, Lys/Lys, Orn/Orn, and Orn/Dab, are capable of making salt bridges with residues on **A-Abu** in the heterotetramer based on modeling studies.

The second approach to mitigate self-association of **B-Phe** involved changing positions 11 and 18 to arginine. With a pKa of 10.4, some of the lysines on **B-Phe** may deprotonate upon association, causing less charge repulsion (and perhaps even allowing hydrogen bonding) in the homotetrameric state. Arginine has a pKa of 12, which is much higher than Lys, making Arg harder to deprotonate. Our calculations used EEF for solvation energy, which does not include a model for neutral Lys; thus, this could not have been computationally tested using our methods. Based on modeling in the heterotetramer state, arginine did not appear optimal at position 11 and 18 due to its longer side chain. However, Arg might eliminate the homotetrameric state by providing more repulsion in the homotetrameric state.

A final calculation was conducted to determine the heterospecificity and stability of simultaneous core and surface mutations, shown in Table 13a and b, before deciding on peptides to test. The final peptides tested are listed in Table 14. Briefly, **B3** changed positions 11 and 18 to arginine to address the pKa argument. **B4** and **B5** changed positions 11 and 13 to shorter lysine derivatives, decreasing the degrees of freedom the side chain can sample. **B6-B8** added more steric clashes to the core to decrease self-association. The peptides were synthesized and biophysically characterized by Mayssam Ali. All BASE peptides showed very little structure via CD at 50 μM, but all show some amount of self-association by AUC when concentrations from 50 μM to 1 mM are fit to a stoichiometric model (shown in Table 15). In experiments, arginine was more effective than Lys at keeping homotetramers from forming, most likely due to the pKa argument. Based on AUC experiments, Arg on the surface was more effective in preventing homotetramerization than designing additional steric clashes in **B-Phe**. The design calculations do not take pKa into consideration, so this element of potential specificity was introduced rationally. The Mtl substitution also reduced some homotetramerization in experiments. If Mtl

in the core and arginine at positions 11 and 18 were combined in a designed peptide, the homotetramerization at high concentration may be substantially reduced.

Future Directions

Heterospecificity is an important design concept. A molecule that can form heterospecific interactions with no self-association would be useful for biomaterials or targeting applications, such as drug delivery or protein purification. In additional attempts to prevent **B-Phe** from self-associating, a subsequent round of design was done. Leu 15, located at the helical interface, was used as an additional target site. Leu 15 residues on adjacent helices face one another. Promising solutions emerged from this round of design, but were not experimentally tested. Steric and charge complementarity both emerged at position 15 as viable means for making heterospecific peptides. For instance, the following pairs of residues (listed as Leu 15 on helix A / Leu 15 on helix B) were found to be highly heterospecific and also stabilizing: Ala/Tyr, Leu/Tyr, Glu/Arg, Asp/Arg. This round of design offers an additional method to prevent self-association of **B-Phe**.

Simultaneously optimizing for stability and specificity

Future research in this area should focus on methods for optimizing both specificity and stability simultaneously during the design process. In the BBA heterotetramer design, specificity and stability were optimized separately. The solutions that ranked extremely high in the specificity optimization did not rank at the top of the stability optimization. This makes sense, as stability was often sacrificed for specificity in previous studies (*1, 2, 9*). However, optimizing for specificity and stability simultaneously may yield better solutions. Havernak and Harbury successfully simultaneously optimized for stability and specificity by including the unfolded,

aggregated, and undesired coiled-coil state in their optimization for homo- or heterospecific coiled coils (*10*). We know that if only stability is taken into account, the solutions are often not completely heterospecific (*9*).

Both specificity and stability could be incorporated into one design calculation by defining an objective function that would effectively weight both stability and specificity into a score, as done by Havernak and Harbury. However, this was tried in a couple BBA selections and was found to be difficult due to large scores that often result from steric clashes in specificity calculations, placing stability and specificity energies on different scales. Thus, a weighting system may not be effective due to large specificity scores, and scale factors would most likely need to be experimentally determined. An alternative is a score that incorporates both elements without use of raw energies, perhaps as a function of rank (*11*). Evaluating based on rank would cause the stability and specificity score to be of a similar magnitude. Essentially, this is what was done manually in the BBA heterotetramer design. In addition, steric clashes that cause the large specificity scores could be relieved with a small amount of minimization or backbone flexibility, making the specificity and stability scores more comparable.

Realistic negative design

Incorporating minimization, a larger rotamer library, or backbone flexibility into the design process may also be useful in quantifying specificity, as large rotamer clashes will be reduced, and energy will be reduced as a result (as shown in Figure 9a and b). Alternatively, the repulsive term in the Leonard Jones potential could be modified such that repulsive energies stay within reasonable limits. Ideally, the specificity score that results from a steric clash will be the same order of magnitude as specificity resulting from an electrostatic clash. However, the energies from the steric clashes in this design far exceeded the score from electrostatic clashes.

Part of the success of the BBA design was due to the separation of the "core" and "surface", separating the steric and the electrostatic clashes, allowing them to be evaluated on their relative scales. Given the performance of the various peptides, it would be interesting to determine how the peptides score if minimization and / or backbone flexibility are incorporated into design. In this study, minimization was only used before the structures were examined and was not an element in the design process. Further, testing the plasticity limits of the BBA backbone by introducing more steric bulk into the core would also be interesting to test experimentally. Given the crystal structure of both the homo- and heterostructures, this system could give additional insights into the design software with a little more computation and experimentation. Because steric matching is an important design technique, it is important to make attempts to improve this in the present design algorithm.

Other Potential Improvements

The computational design program used in the BBA heterotetramer design eliminates high-energy rotamers based on an energy cutoff. Ideally, the undesired state will pack poorly and have many steric clashes, and the desired state will pack well and have few steric clashes. To calculate specificity, the energies of both states must be calculated, however this may not be possible if the undesired state is extremely poor. In an extremely poor undesired state, all rotamers may be eliminated at a position, preventing the specificity score from being calculated. Unfortunately, these sequences are of the most interest for introducing heterospecificity because the energy of the undesired state should not be favorable. The extent to which this happens has not been explored.

Experimentally, the **A-Ala** and **A-Abu** showed no self-association, even at high concentrations. This was surprising, as the majority of specificity in the calculations came from

steric clashes on **B-Phe**, not **A-Ala** and **A-Abu**. Therefore, the current energy function does not penalize a sparsely packed core nearly to the extent that steric clashes are penalized. However, based on the results, the sparsely packed core may be a better negative design technique. Perhaps, exaggerating the attractive component of the Leonard-Jones potential would cause the energies to appear to favor packing in the core versus a cavity. Checking potential solutions with a program, such as PROVE (*12*), that checks for the quality of packing in the core may be useful for both desired and undesired structures.

REFERENCES

1. Lumb, K. J., and Kim, P. S. (1995) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry 34*, 8642-8648.
2. Oakley, M. G., and Kim, P. S. (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry 37*, 12603-10.
3. Johnson, E. C., Lazar, G. A., Desjarlais, J. R., and Handel, T. M. (1999) Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Structure Fold Des 7*, 967-76.
4. Lazar, G. A., Johnson, E. C., Desjarlais, J. R., and Handel, T. M. (1999) Rotamer strain as a determinant of protein structural specificity. *Protein Sci 8*, 2598-610.
5. O'Shea, E. K., Lumb, K. J., and Kim, P. S. (1993) Peptide 'Velcro*': design of a heterodimeric coiled coil. *Curr. Biol. 3*, 658-667.
6. Dominy, B. N., and Brooks, C. L. (1999) Development of a generalized born model parametrization for proteins and nucleic acids. *Journal of Physical Chemistry B 103*, 3765-3773.
7. Grigoryan, G., and Keating, A. E. (2005) unpublished.
8. Lazaridis, T., and Karplus, M. (1999) Effective energy function for proteins in solution. *Proteins 35*, 133-52.
9. Bolon, D. N. (2005).
10. Havranek, J. J., and Harbury, P. B. (2003) Automated design of specificity in molecular recognition. *Nat Struct Biol 10*, 45-52.
11. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) High-resolution protein design with backbone freedom. *Science 282*, 1462-1467.
12. Pontius, J., Richelle, J., and Wodak, S. J. (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol 264*, 121-36.
13. Ali, M. H. (2004) in *Chemistry* pp 203, Massachusetts Institute of Technology, Cambridge.

TABLES

Phe8/Leu16 (Layer B & B')								
A/B	**AAAA-bb1**	**AAAA-bb2**	**BBBB-bb1**	**BBBB-bb2**	**ABAB-bb1**	**ABAB-bb2**	**bb1 Spec. Score (Rank)**	**bb2 Spec. Score (Rank)**
PHE/ARG TRP/TYR	-180.54	-182.03	592.67	321.25	-134.79	-121.28	681.71(1)	381.76(59)
PHE/GLU TRP/TYR	**-161.15**	**-177.82**	**592.67**	**321.25**	**-125.09**	**-114.38**	**681.68 (2)**	**378.19 (188)**
TRP/TYR PHE/ILE	**592.67**	**321.25**	**-52.37**	**-58.14**	**-70.29**	**-59.31**	**680.87(5)**	**381.74(61)**
HIS/ARG TRP/TYR	-200.77	-198.89	592.67	321.25	-142.71	-131.18	677.32(32)	384.73(24)
HIS/THR TRP/TYR	-132.05	-135.35	592.67	321.25	-108.08	-98.66	676.78(35)	383.22(33)
ASP/GLU TRP/TYR	-174.62	-181.54	592.67	321.25	-129.22	-121.79	676.49(39)	383.28(31)
HIS/SER TRP/TYR	-145.12	-148.36	592.67	321.25	-113.34	-103.96	674.22(61)	380.81(95)
TRP/TYR ILE/VAL	**592.67**	**321.25**	**27.76**	**-2.02**	**-26.48**	**-30.58**	**673.39(85)**	**380.39(115)**
TRP/TYR MET/ARG	592.67	321.25	-163.83	-161.12	-122.08	-111.26	672.98(114)	382.64(42)
ASN/ARG TRP/TYR	-197.95	-192.68	592.67	321.25	-138.81	-126.84	672.34(157)	382.25(51)
TYR/TYR PHE/ILE	**377.01**	**187.59**	**-52.37**	**-69.44**	**-66.969**	**-68.63**	**458.57(445)**	**225.41(1886)**
HIS/TYR HIS/VAL	**319.47**	**187.59**	**-67.74**	**-69.44**	**-79.46**	**-68.63**	**410.65 (6234)**	**255.41 (1351)**
HIS/TYR SER/PHE	319.49	-------	-26.29	--------	131.8	----------	29.57(24211)	----------
ILE/ALA HIS/TRP	**-29.07**	**-60.79**	**-10.92**	**59.65**	**-80.24**	**-85.81**	**120.49(11153)**	**170.48(4067)**
TYR/TRP ILE/ALA	**-4.34**	**59.55**	**-29.07**	**-60.79**	**-73.91**	**-84.05**	**114.41(13041)**	**166.85(4865)**
PHE/ILE HIS/PHE	**-52.37**	**-58.14**	**-31.26**	**-30.85**	**-98.6**	**-102.43**	**113.58(13233)**	**115.88(10706)**
VAL/TYR TRP/TRP	391.03	106.03	160.81	227.9	223.49	-39.27	104.87(14391)	412.47(1)
THR/TYR TRP/TRP	359.67	75.36	160.81	227.9	208.28	-49.53	103.93(14544)	402.32(13)
ILE/TYR TRP/TRP	418.12	109.11	160.81	227.9	237.58	-32.96	103.78(14558)	402.93(11)

GLU/TYR TRP/TRP	326.93	43.95	160.81	227.9	192.25	-68.64	103.25(14629)	409.14(3)
ALA/TYR TRP/TRP	372.18	85.58	160.81	227.9	215.11	-39.95	102.77(14689)	393.38(17)
MET/TYR TRP/TRP	359.38	91.18	160.81	227.9	210.57	-44.5	99.04 (14979)	408.09(5)
GLN/TYR TRP/TRP	327.19	49.09	160.81	227.9	194.55	-65.42	98.91 (14983)	407.83(7)
ASN/TYR TRP/TRP	326.59	51.78	160.81	227.9	194.76	-57.79	97.88 (15068)	395.25(15)
LYS/TYR TRP/TRP	322.62	56.35	160.81	227.9	193	-60.24	97.43 (15100)	404.72(9)
HIS/ILE TRP/TYR	------	-74.77	-------	321.25	--------	-69.05	----------	384.57(26)

Table 1a: Energies for layer B. Energies of undesired homotetramer states, desired heterotetramer states, and specificity scores for both backbones in layer B are shown. In the far left hand column, the residues are represented as A8/A16 over B8/B16. Therefore, A8 and B16 would form a layer and A16 and B8 would form a layer, due to the antiparallel orientation of the helices. The rows in bold were ones considered the most promising, as no charged residues were present in the solution. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. Rank refers to the solution's overall rank relative to other solutions within the individual specificity selection. A and B do not refer to ACID and BASE, at this point in the calculation.

Rotamers for bb1 Phe8/Leu16 (Layer B and B')									
Rank (specificity)	E_{ABAB}	8A	16A	8A	16A	8B	16B	8B	16B
2	-125.087	PHE.02	GLU.26	PHE.02	GLU.23	TRP.03	TYR.04	TRP.05	TYR.05
5	-70.2853	TRP.03	TYR.04	TRP.05	TYR.05	PHE.02	ILE.01	PHE.02	ILE.01
85	-26.4832	TRP.03	TYR.04	TRP.05	TYR.05	ILE.03	VAL.02	ILE.03	VAL.02
445	-66.9631	TYR.03	TYR.04	TYR.02	TYR.05	PHE.02	ILE.01	PHE.02	ILE.01
6432	-79.46	HIS.03	TYR.04	HIS.02	TYR.05	HIS.03	VAL.02	HIS.02	VAL.02
11153	-80.2421	ILE.04	ALA.00	ILE.04	ALA.00	HIS.03	TRP.08	HIS.02	TRP.08
13041	-73.9105	TYR.03	TRP.08	TYR.02	TRP.08	ILE.04	ALA.00	ILE.04	ALA.00
13233	-98.6025	PHE.02	ILE.01	PHE.02	ILE.01	HIS.03	PHE.05	HIS.02	PHE.05
Rotamers for bb2 Phe8/Leu16 (Layer B and B')									
Rank (specificity)	E_{ABAB}	8A	16A	8A	16A	8B	16B	8B	16B
61	-59.3136	TRP.05	TYR.04	TRP.03	TYR.04	PHE.02	ILE.01	PHE.02	ILE.01
115	-30.5791	TRP.05	TYR.04	TRP.03	TYR.04	ILE.04	VAL.02	ILE.04	VAL.02
188	-114.38	PHE.02	GLU.23	PHE.02	GLU.26	TRP.05	TYR.04	TRP.03	TYR.04
1351	-68.6267	HIS.03	TYR.04	HIS.03	TYR.04	HIS.02	VAL.02	HIS.02	VAL.02
1886	-60.30	TYR.02	TYR.04	TYR.02	TYR.04	PHE.02	ILE.01	PHE.02	ILE.01
4067	-85.8107	ILE.04	ALA.00	ILE.04	ALA.00	HIS.03	TRP.07	HIS.02	TRP.08
4865	-84.0482	TYR.02	TRP.07	TYR.02	TRP.08	ILE.04	ALA.00	ILE.04	ALA.00
10706	-102.431	PHE.02	ILE.01	PHE.02	ILE.01	HIS.03	PHE.05	HIS.02	PHE.05

Table 1b. Rotamer conformations for the selection of layer B and B'. Based on probabilities in the Dunbrack backbone-dependent rotamer library, rotamers that occur frequently in the PDB are shown in color. If there were 10 or less rotamers in the library for an amino acid, rotamers with probabilities below 0.1 were considered low-probability rotamers. For amino acids with greater than 10 rotamers, rotamers with probabilities within the top quarter to half of rotamers for a particular amino acid were considered high-probability. Rank refers to the solution's overall rank relative to other solutions within the individual specificity selection. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. E_{ABAB} refers to the folded-state energy, so a more negative number indicates greater stability. To calculate stability accurately, the unfolded-state energy would need to be subtracted, but this method provides a rough estimate. A and B do not refer to ACID and BASE, at this point in the calculation.

Leu12 Layer (Layer C)								
A/B	AAAA-bb1	AAAA-bb2	BBBB-bb1	BBBB-bb2	ABAB-bb1	ABAB-bb2	bb1 Spec. Score (Rank)	bb2 Spec. Score (Rank)
PHE/ALA	233.08	287.35	-89.33	-97.82	-105.49	-94.52	354.74 (5)	397.93 (47)
TYR/SER	230.60	299.66	-108.90	-97.82	-115.23	-108.97	352.18(9)	352.18 (9)
PHE/ILE	233.08	287.35	-60.36	-49.11	-81.41	-79.15	333.57 (19)	396.52 (52)
TRP/ALA	265.92	568.02	-89.33	-78.47	-73.24	-65.97	323.07 (23)	661.48 (1)
TRP/SER	265.92	568.02	-108.90	-97.82	-81.97	-95.55	320.97(25)	661.17 (3)
PHE/MET	233.08	287.35	-106.07	-94.52	-94.58	-100.06	316.19 (29)	392.95 (59)
TRP/THR	265.92	568.02	-105.30	-94.40	-65.97	-88.72	292.56 (41)	651.07 (5)
TRP/ILE	230.59	568.02	-60.36	-49.11	-22.67	-57.60	250.82 (66)	634.12 (7)
PHE/LEU	233.08	287.35	-93.93	-92.34	-52.36	-82.20	243.89 (69)	359.42 (77)

Table 2a. Energies for layer C. Energies of undesired homotetramer states, desired heterotetramer states, and specificity scores for both backbones are shown. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. Rank refers to the solution's overall rank relative to other solutions within the individual specificity selection. A and B do not refer to ACID and BASE, at this point in the calculation.

Rotamers for bb2 Leu12/Leu12 (Layer C)				
Het E	A	A	B	B
-94.5293	**PHE.02**	**PHE.02**	**ALA.00**	**ALA.00**
-79.1424	PHE.02	PHE.02	ILE.03	ILE.08
-88.6506	TYR.02	TYR.02	VAL.01	VAL.01
-108.983	TYR.02	TYR.02	SER.02	SER.02

Rotamers for bb1 Leu12/Leu12 (Layer C)				
Het E	A	A	B	B
-105.491	**PHE.02**	**PHE.02**	**ALA.00**	**ALA.00**
-80.4182	PHE.02	PHE.02	ILE.08	ILE.08
-57.2019	TYR.02	TYR.02	VAL.00	VAL.01
-115.24	TYR.02	TYR.02	SER.00	SER.01

Table 2b. Rotamer analysis of some of the top solutions from the initial selection for layer C. Favorable rotamer conformations are shown in color. Based on probabilities in the Dunbrack backbone-dependent rotamer library, rotamers that occur frequently in the PDB are shown in color. If there were 10 or less rotamers in the library for an amino acid, rotamers with probabilities below 0.1 were considered low-probability rotamers. For amino acids with greater than 10 rotamers, rotamers with probabilities within the top quarter to half of rotamers for a particular amino acid were considered high-probability. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. E_{ABAB} refers to the folded-state energy, so a more negative number indicates greater stability. To calculate stability accurately, the unfolded-state energy would need to be subtracted, but this method provides a rough estimate. A and B do not refer to ACID and BASE, at this point in the calculation.

Pos 13 A/B	AAAA-bb1	AAAA-bb2	BBBB-bb1	BBBB-bb2	ABAB-bb1	ABAB-bb2	Spec bb1	Spec bb2
Glu/Lys	**-281.01**	**-293.63**	**-295.0054**	**-305.03**	**-290.45**	**-302.0694**	**4.89**	**5.47**
Lys/Glu	**-295.005**	**-305.03**	**-281.0173**	**-293.63**	**-290.45**	**-302.0679**	**4.89**	**5.47**
Asp/Lys	-279.46	-291.01	-295.0054	-305.03	-288.68	-299.38	2.89	2.72
Lys/Asp	-295.01	-305.03	-279.4677	-291.01	-288.68	-299.38	2.89	2.72
Arg/Asp	-336.25	-336.25	-279.46	-291.01	-304.01	-314.66	2.14	2.06
Asp/Arg	-299.46	-291.01	-326.411	-336.25	-304.01	-314.66	2.15	2.06
Glu/Arg	-281.01	-293.63	-326.411	-336.25	-304.73	-316.65	2.04	3.41
Arg/Glu	-326.4107	-336.25	-281.01	-293.63	-304.73	-316.65	2.04	3.41

Table 3. Design of position 13. Energies of undesired homotetramer states, desired heterotetramer states, and specificity scores for both backbones are shown. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. Rank refers to the solution's overall rank relative to other solutions within the individual specificity selection. A and B do not refer to ACID and BASE, at this point in the calculation.

A11/A18	B11/B18	AAAA-bb1	AAAA-bb2	BBBB-bb1	BBBB-bb2	ABAB-bb1	ABAB-bb2	Specificity bb1 (rank)	Specificity bb2 (rank)
Ile/Leu	Trp/His	-156.05	-199.68	-295.04	-316.83	-248.49	-269.10	45.87 (1)	39.85 (385)
Ile/Asp	Arg/Arg	-204.09	-230.39	-369.34	-380.29	-308.33	-327.758	43.22(23)	44.84(6)
Ile/Asp	Lys/Arg	-204.09	-230.39	-343.66	-354.081	-295.271	-314.61	42.78(25)	44.81(7)
Ile/Lys	Glu/Glu	-295.04	-228.51	-285.52	-334.76	-297.27	-304.65	40.88 (4216)	45.977(1)
Phe/Trp	Ile/Ser	-268.71	------	-181.78	--------	-244.546	--------	38.60 (178)	--------
Arg/Ser	Ile/Glu	-321.06	-331.89	-208.57	-235.97	-283.96	-303.91	38.28(203)	39.98(359)
Lys/Val	Ile/Asp	-261.49	-273.09	-204.09	-230.39	-251.63	-272.12	37.68(264)	40.76(200)
Asn/Trp	Ile/Gln	-306.18	-317.534	-209.06	-233.38	-275.46	-294.89	35.68(819)	38.87(768)
Glu/Asp	Lys/Lys	-316.08	-330.23	-315.8	-325.9	-324.91	-337.1	17.93(3917)	18.07(3641)
Glu/Glu	Lys/Lys	-320.45	-334.76	-315.8	-325.8	-326.08	-338.83	15.9(4051)	17.007(3701)
Asp/Asp	Lys/Lys	-304.71	-320.57	-315.8	-325.8	-317.75	-330.65	14.96(4142)	14.85(4004)

Table 4. Design of position 11 and 18. Energies of undesired homotetramer states, desired heterotetramer states, and specificity scores for both backbones are shown. Salt-bridging pairs are A11/B18 and B11/A18. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. Rank refers to the solution's overall rank relative to other solutions within the individual specificity selection. A and B do not refer to ACID and BASE, at this point in the calculation.

A11/B18	B11/A18	AAAA-bb1	AAAA-bb2	BBBB-bb1	BBBB-bb2	ABAB-bb1	ABAB-bb2	Spec bb1	Spec bb2
Lys/Arg	Glu/Asp	-343.66	-362	-316.08	-330.23	-339.95	-350.83	20.17	17.36
Glu/Glu	Lys/Arg	-320.45	-334.76	-343.66	-354.08	-341.12	-352.62	18.14	16.39
Lys/Arg	Asp/Asp	-343.66	-320.57	-304.71	-334.76	-333.23	-345.51	18.09	16.36
Glu/Asp	Lys/Lys	-316.08	-330.23	-315.8	-325.9	-324.91	-324.91	17.93	18.07
Arg/Arg	Glu/Glu	-369.33	-380.29	-320.46	-354.08	-353.31	-364.97	16.83	14.89
Glu/Glu	Lys/Lys	-320.45	-334.76	-325.8	-325.8	-326.08	-326.08	15.9	17.01
Asp/Asp	Lys/Lys	-315.8	-320.57	-325.8	-325.8	-317.75	-317.75	14.96	14.85

Table 5. Energies for position 11 and 18 selection. Energies of undesired homotetramer states, desired heterotetramer states, and specificity scores for both backbones are shown. Positions 11 and 18 were reselected with a restricted library. Salt-bridging pairs are A11/B18 and B11/A18. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. A and B do not refer to ACID and BASE, at this point in the calculation.

	AAAA-bb1	AAAA-bb2	BBBB-bb1	BBBB-bb2	ABAB-bb1	ABAB-bb2	Spec-bb1 (rank)	Spec-bb1 (rank)	Stability bb1
Phe/Ala	**-132.63**	**-82.23**	**-460.49**	**-461.64**	**-492.69**	**-485.24**	**392.25 (1)**	**426.62(5)**	**139.88**
Trp/Ala	-124.14	197.78	-460.49	-461.64	-466.74	-476.59	348.86 (3)	689.31(1)	102.55
Leu/Phe	-454.83	-475.70	-138.67	-82.23	-437.87	-473.29	282.24 (5)	388.54(8)	87.43

Table 6. Results of final round 1 selection. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized for the specificity calculation: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity energy. The following equation is maximized for the stability calculation: $f = -E_{folded} + E_{unfolded}$, so a large, positive stability energy indicates a stable molecule. A and B refer to the ACID and BASE peptide.

bb1					
	AAAA	**BBBB**	**ABAB**	**Spec-bb1 (rank)**	**Stability**
Phe/Ala	**-138.67**	**-454.79**	**-492.41**	**391.38 (2)**	**139.61**
Phe/Leu	-138.67	-454.83	-437.88	282.26 (5)	87.44
Leu/Leu	-361.98	-364.25	-376.63	27.03	134.82
Native	----------	----------	----------	----------	91.84
bb2					
	AAAA	**BBBB**	**ABAB**	**Spec-bb2 (rank)**	**Stability**
Phe/Ala	**-96.29**	**-447.8**	**-485.25**	**426.40 (6)**	**133.97**
Phe/Leu	-96.29	-461.75	-473.28	388.53 (8)	124.42
Leu/Leu	-372.99	-377.39	-389.14	27.91	142.22
Native	----------	----------	----------	----------	90.83

Table 7. Final energies for round 1. Energies of undesired homotetramer states, desired tetramer states, specificity, stability. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. The following equation is maximized for the stability calculation: $f = -E_{folded} + E_{unfolded}$, so a large, positive stability energy indicates a stable molecule. A and B refer to the ACID and BASE peptide.

Leu12	AAAA-bb1	BBBB-bb1	ABAB-bb1	Specificity-bb1	Stability-bb1
Phe/Ala	**-132.64**	**-435.84**	**-474.78**	**381.08(6)**	**129.68**
Phe/Abu	**-132.64**	**-438.57**	**-476.68**	**382.15(2)**	**132.47**
Phe/Nle	-132.64	-443.64	-478.93	381.58(4)	135.1
Phe/Leu	-132.64	-440.54	-422.06	270.95(8)	79.22

Leu12	AAAA-bb2	BBBB-bb2	ABAB-bb2	Specificity-bb2	Stability-bb2
Phe/Ala	**-96.29**	**-447.8**	**-485.25**	**426.40 (4)**	**133.97**
Phe/Abu	**-96.29**	**-449.02**	**-483.63**	**421.95 (7)**	**133.29**
Phe/Nle	-96.29	-456.41	-486.73	420.75 (11)	136.79
Phe/Leu	-96.29	-461.75	-473.29	388.53 (348)	124.42

Table 8. The results of round 2 selection. Only the Base/Acid combination is shown in this table. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. The following equation is maximized for the stability calculation: $f = -E_{folded} + E_{unfolded}$, so a large, positive stability energy indicates a stable molecule. A and B refer to the ACID and BASE peptide.

Backbone 1

A/B	AAAA	BBBB	ABAB	Specificity
A-Ala/B-Phe	-454.7	-138.7	-492.4	391.4
A-Leu/B-Leu	-364.3	-362.0	-376.6	26.9

Backbone 2

A/B	AAAA	BBBB	ABAB	Specificity
A-Ala/B-Phe	-447.8	-96.3	-485.3	426.5
A-Leu/B-Leu	-377.4	-373.0	-389.1	27.8

Table 9a. Energies of different complexes according to the energy function employed for sequence selection. AAAA, BBBB, and ABAB are homotetramers and heterotetramers, respectively, of the peptides indicated in the first column. The energies reported are E_{fold}-E_{unfold}. Specificity is defined as (E_{AAAA}+E_{BBBB}-2E_{ABAB}) and is positive when the heterotetramer is preferred. The **B-Phe** homotetramer is severely disfavored by steric clashes. The **A-Ala** homotetramer is predicted to be more stable than the A-Leu homotetramer because the geometry of Leu packing is non-ideal in these structures.

Backbone 1

A/B	AAAA	BBBB	ABAB	Specificity
A-Ala/B-Phe	-439.1	-421.0	-485.3	109.8
A-Leu/B-Leu	-467.5	-442.1	-474.8	40.0

Backbone 2

A/B	AAAA	BBBB	ABAB	Specificity
A-Ala/B-Phe	-395.1	-366.6	-428.7	95.7
A-Leu/B-Leu	-426.8	-423.0	-439.2	28.6

Table 9b. Refined energies of minimized complexes. Structures from the design calculations were relaxed using ten steps of steepest decent minimization to relieve steric clashes of the side chains. The backbone was fixed during minimization. Energies were evaluated included the following terms for folded and unfolded states: 100% van der Waals energy with 100% radii from CHARMM param19, 4/4 Coulomb interaction, 4/4 4/80 screening estimated by a Generalized Born model, and EEF1 desolvation. Side-chain relaxation makes the energies of all complexes more comparable, and the A-Leu homotetramer is now correctly predicted to be more stable than the **A-Ala** homotetramer.

12/12	AAAA-bb1	BBBB-bb1	ABAB-bb1	Specificity-bb1	Stability-bb1
MTL/ABU	**252.28**	**-355.63**	**-358.36**	**613.37**	**97.23**
ABU/MTL	**-355.63**	**252.28**	**-358.34**	**613.31**	**97.21**
MTL/ALA	**252.28**	**-352.94**	**-355.92**	**611.18**	**93.89**
ALA/MTL	**-352.94**	**252.28**	**-355.90**	**611.14**	**93.87**
MTL/NLE	252.28	-360.67	-359.37	610.35	98.65
NLE/MTL	-360.67	252.28	-359.35	610.31	98.63
TOL/ALA	40.51	-352.94	-363.80	415.17	100.37
ALA/TOL	-352.94	40.51	-363.78	415.14	100.35
OTL/ABU	-13.70	-355.63	-370.69	372.05	113.26
ABU/OTL	-355.63	-13.70	-370.69	372.04	113.25
OTL/NLE	-13.70	-360.67	-372.00	371.63	115.97
NLE/OTL	-360.67	-13.70	-372.99	371.62	115.96
OTL/ALA	-13.70	-352.94	-368.79	370.95	110.45
ALA/OTL	-352.94	-13.70	-368.79	370.95	110.45
PHE/ABU	***-30.50***	***-355.63***	***-370.97***	***355.80***	***109.01***
ABU/PHE	*-355.63*	*-30.50*	*-370.96*	*355.79*	*109.00*
PHE/NLE	-30.50	-360.67	-373.16	355.15	111.60
NLE/PHE	-360.67	-30.50	-373.15	355.14	111.60
PHE/ALA	-30.50	-352.94	-369.09	354.74	106.22
ALA/PHE	-352.94	-30.50	-369.08	354.73	106.22
TOL/ABU	40.51	-355.63	-325.15	335.17	62.62
ABU/TOL	-355.63	40.51	-325.14	335.16	62.62
TOL/NLE	40.51	-360.67	-325.85	331.54	63.73
NLE/TOL	-360.67	40.51	-325.84	331.53	63.72

Table 10a. Round 3 results for position 12 with one backbone, bb1. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. The following equation is maximized for the stability calculation: $f = -E_{folded} + E_{unfolded}$, so a large, positive stability energy indicates a stable molecule. A and B refer to the ACID and BASE peptide. The round 2 peptide is shown in italics, and the peptides tested in round 3 are in bold.

12/12	AAAA-bb2	BBBB-bb2	ABAB-bb2	Specificity-bb2 (rank)	Stability-bb2
MTL/ALA	**1093.97**	**-414.16**	**-419.30**	**1518.40(1)**	**108.46**
ALA/MTL	**-414.16**	**1093.97**	**-419.22**	**1518.24(2)**	**108.37**
MTL/ABU	**1093.97**	**-415.30**	**-418.14**	**1514.94(3)**	**108.23**
ABU/MTL	**-415.30**	**1093.97**	**-418.05**	**1514.76(4)**	**108.14**
MTL/NLE	1093.97	-422.63	-418.73	1508.79(5)	109.23
NLE/MTL	-422.63	1093.97	-418.64	1508.62(6)	109.14
TOL/ALA	21.39	-414.16	-431.03	469.28(9)	118.78
ALA/TOL	-414.16	21.39	-431.02	469.27(10)	118.77
TOL/ABU	21.39	-415.30	-418.13	442.35(11)	106.82
ABU/TOL	-415.30	21.39	-418.11	442.32(12)	106.80
TOL/NLE	21.39	-422.63	-417.19	433.13(13)	106.28
NLE/TOL	-422.63	21.39	-417.17	433.11(14)	106.27
ALA/OTL	-414.16	-36.05	-430.19	410.17(15)	123.05
OTL/ALA	-36.05	-414.16	-430.19	410.17(16)	123.05
OTL/ABU	-36.05	-415.30	-428.49	405.63(17)	122.28
ABU/OTL	-415.30	-36.05	-428.48	405.62(18)	122.28
OTL/NLE	-36.05	-422.63	-431.51	404.33(19)	125.71
NLE/OTL	-422.63	-36.05	-431.50	404.32(20)	125.70
PHE/ALA	-48.24	-414.16	-430.17	397.93(21)	118.49
ALA/PHE	-414.16	-48.24	-430.16	397.93(22)	118.48
PHE/ABU	***-48.24***	***-415.30***	***-428.50***	***393.47(23)***	***117.75***
ABU/PHE	*-415.30*	*-48.24*	*-428.49*	*393.45(24)*	*117.74*
PHE/NLE	-48.24	-422.63	-431.55	392.22(25)	121.21
NLE/PHE	-422.63	-48.24	-431.54	392.20(26)	121.20

Table 10b. Round 3 results for position 12 with one backbone, bb2. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. The following equation is maximized for the stability calculation: $f = -E_{folded} + E_{unfolded}$, so a large, positive stability energy indicates a stable molecule. A and B refer to the ACID and BASE peptide. The round 2 peptide is shown in italics, and the peptides tested in round 3 are in bold.

13/13	AAAA-bb1	BBBB-bb1	ABAB-bb1	Specificity-bb1 (rank)	Stability-bb1
GLU/LYS	**-281.02**	**-295.01**	**-290.46**	**4.89(2)**	**94.75**
LYS/GLU	**-295.01**	**-281.02**	**-290.46**	**4.89(1)**	**94.75**
GLU/ORN	-281.02	-258.71	-271.11	2.49(6)	93.40
ORN/GLU	-258.71	-281.02	-271.11	2.49(5)	93.40
GLU/DAB	-281.02	-255.25	-269.30	2.33(12)	92.52
DAB/GLU	-255.25	-281.02	-269.30	2.33(11)	92.52
ARG/GLU	-326.41	-281.02	-304.73	2.04 (16)	93.18
GLU/ARG	-281.02	-326.41	-304.74	2.04 (15)	93.18

13/13	AAAA-bb2	BBBB-bb2	ABAB-bb2	Specificity-bb2	Stability-bb2
LYS/GLU	**-305.04**	**-293.64**	**-302.07**	**5.47(1)**	**107.26**
GLU/LYS	**-293.64**	**-305.03**	**-302.07**	**5.47(2)**	**107.26**
ARG/GLU	-336.25	-293.64	-314.66	3.41(3)	105.96
GLU/ARG	-293.64	-336.25	-314.66	3.41(4)	105.96
ORN/GLU	-268.06	-293.64	-299.39	2.41(9)	105.23
GLU/ORN	-293.64	-268.06	-299.38	2.41(10)	105.23
GLU/DAB	-293.64	-265.14	-279.23	2.28(14)	104.61
DAB/GLU	-265.14	-293.64	-279.23	2.28(13)	104.61

Table 11. Round 3 results for position 13 with both backbones. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. The following equation is maximized for the stability calculation: $f = -E_{folded} + E_{unfolded}$, so a large, positive stability energy indicates a stable molecule. A and B refer to the ACID and BASE peptide.

11/18	11/18	AAAA-bb1	BBBB-bb1	ABAB-bb1	Specificity-bb1	Stability-bb1
GLU/ASP	**LYS/LYS**	**-316.08**	**-315.80**	**-324.91**	**17.93(11)**	**100.04**
LYS/LYS	**GLU/ASP**	**-315.80**	**-316.08**	**-324.91**	**17.93(12)**	**100.04**
GLU/ASP	**ARG/ARG**	**-316.08**	**-369.34**	**-351.36**	**17.31(13)**	**94.72**
ARG/ARG	**GLU/ASP**	**-369.34**	**-316.08**	**-351.36**	**17.31(14)**	**94.72**
ORN/DAB	GLU/ASP	-235.00	-316.08	-283.39	15.69(35)	95.56
GLU/ASP	ORN/DAB	-316.08	-235.00	-283.38	15.68(36)	95.56
GLU/ASP	DAB/DAB	-316.08	-229.79	-280.53	15.19(45)	93.47
DAB/DAB	GLU/ASP	-229.79	-316.08	-280.52	15.17(46)	93.47
ORN/ORN	GLU/ASP	-238.83	-316.08	-285.01	15.11(51)	96.14
GLU/ASP	ORN/ORN	-316.08	-238.83	-284.00	15.09(52)	96.14

11/18	11/18	AAAA-bb2	BBBB-bb2	ABAB-bb2	Specificity-bb2	Stability-bb2
GLU/ASP	**LYS/LYS**	**-330.23**	**-325.91**	**-337.11**	**18.07(3)**	**114.48**
LYS/LYS	**GLU/ASP**	**-325.91**	**-330.23**	**-337.10**	**18.05(4)**	**107.27**
GLU/ASP	**ORN/DAB**	**-330.23**	**-244.49**	**-295.76**	**16.79(11)**	**110.10**
ORN/DAB	**GLU/ASP**	**-244.49**	**-330.23**	**-295.75**	**16.79(12)**	**110.10**
ARG/ARG	**GLU/ASP**	**-380.29**	**-330.23**	**-363.09**	**15.66(39)**	**108.72**
GLU/ASP	ORN/ORN	-330.23	-248.75	-297.21	15.44(41)	110.58
ORN/ORN	GLU/ASP	-248.75	-330.23	-297.21	15.43(42)	110.58
GLU/ASP	DAB/DAB	-330.23	-240.65	-292.66	14.43(68)	107.89
DAB/DAB	GLU/ASP	-240.65	-330.23	-292.66	14.43(69)	107.89
GLU/ASP	ARG/ARG	-330.23	-380.29	-359.59	13.20(95)	108.72

Table 12. Round 3 results for positions 11 and 18 with both backbones. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. The following equation is maximized for the stability calculation: $f = -E_{folded} + E_{unfolded}$, so a large, positive stability energy indicates a stable molecule. A and B refer to the ACID and BASE peptide.

Pos 11, 13, 18	Pos 11, 13, 18	Pos 12	AAAA-bb1	BBBB-bb1	ABAB-bb1	Specificity-bb1	Stability-bb1
LYS/LYS/LYS	**GLU/GLU/ASP**	**MTL/ABU**	**149.48**	**-438.58**	**-464.53**	**639.96(4)**	**109.30**
LYS/LYS/LYS	**GLU/GLU/ASP**	**MTL/ALA**	**149.48**	**-435.84**	**-462.06**	**637.76(14)**	**105.93**
LYS/LYS/LYS	GLU/GLU/ASP	TOL/ALA	-61.76	-435.84	-469.59	441.57(112)	112.06
LYS/LYS/LYS	GLU/GLU/ASP	OTL/ABU	-117.12	-438.58	-477.19	398.68(149)	125.65
LYS/LYS/LYS	GLU/GLU/ASP	OTL/ALA	-117.12	-435.84	-475.27	397.57(161)	122.83
LYS/LYS/LYS	***GLU/GLU/ASP***	***PHE/ABU***	***-132.64***	***-438.58***	***-476.69***	***382.16(230)***	***120.63***
LYS/LYS/LYS	*GLU/GLU/ASP*	*PHE/ALA*	*-132.64*	*-435.84*	*-474.78*	*381.08(243)*	*117.81*
LYS/LYS/LYS	GLU/GLU/ASP	TOL/ABU	-61.76	-438.58	-431.18	362.03(349)	74.56
ORN/LYS/ORN	**GLU/GLU/ASP**	**MTL/ABU**	**226.24**	**-438.58**	**-424.64**	**636.94(22)**	**105.41**
ORN/LYS/ORN	**GLU/GLU/ASP**	**MTL/ALA**	**226.24**	**-435.84**	**-422.17**	**634.75(48)**	**102.05**
ORN/LYS/ORN	GLU/GLU/ASP	TOL/ALA	15.00	-435.84	-429.70	438.56(120)	108.17
ORN/LYS/ORN	GLU/GLU/ASP	OTL/ABU	-40.36	-438.58	-437.30	395.67(175)	121.77
ORN/LYS/ORN	GLU/GLU/ASP	OTL/ALA	-40.36	-435.84	-435.38	394.56(190)	118.94
ORN/LYS/ORN	**GLU/GLU/ASP**	**PHE/ABU**	**-55.88**	**-438.58**	**-436.80**	**379.14(262)**	**116.74**
ORN/LYS/ORN	**GLU/GLU/ASP**	**PHE/ALA**	**-55.88**	**-435.84**	**-434.90**	**378.06(286)**	**113.93**
ORN/LYS/ORN	GLU/GLU/ASP	TOL/ABU	15.00	-438.58	-391.30	359.02(371)	70.67
ORN/LYS/DAB	**GLU/GLU/ASP**	**MTL/ABU**	**230.00**	**-438.58**	**-422.96**	**637.34(17)**	**104.78**
ORN/LYS/DAB	**GLU/GLU/ASP**	**MTL/ALA**	**230.00**	**-435.84**	**-420.50**	**635.15(45)**	**101.41**
ORN/LYS/DAB	GLU/GLU/ASP	TOL/ALA	18.76	-435.84	-428.02	438.96(118)	107.53
ORN/LYS/DAB	GLU/GLU/ASP	OTL/ABU	-36.60	-438.58	-435.62	396.07(172)	121.13
ORN/LYS/DAB	GLU/GLU/ASP	OTL/ALA	-36.60	-435.84	-433.70	394.96(184)	118.30
ORN/LYS/DAB	**GLU/GLU/ASP**	**PHE/ABU**	**-52.12**	**-438.58**	**-435.12**	**379.54(258)**	**116.10**
ORN/LYS/DAB	**GLU/GLU/ASP**	**PHE/ALA**	**-52.12**	**-435.84**	**-433.22**	**378.47(275)**	**113.29**
ORN/LYS/DAB	GLU/GLU/ASP	TOL/ABU	18.76	-438.58	-389.61	359.42(369)	70.03

Table 13a. Complete round 3 selection with positions 11, 12, 13, and 18 with bb1. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. The following equation is maximized for the stability calculation: $f = -E_{folded} + E_{unfolded}$, so a large, positive stability energy indicates a stable molecule. A and B refer to the ACID and BASE peptide. The round 2 peptide is shown in italics, and the peptides tested in round 3 are in bold.

Pos 11, 13, 18	Pos 11, 13, 18	Pos 12	AAAA-bb2	BBBB-bb2	ABAB-bb2	Specificity-bb2	Stabilities-bb2
LYS/LYS/LYS	**GLU/GLU/ASP**	**MTL/ALA**	**690.6352**	**-447.8011**	**-474.6799**	**1192.19392(1)**	**120.55314**
LYS/LYS/LYS	**GLU/GLU/ASP**	**MTL/ABU**	**690.6352**	**-449.0266**	**-473.5395**	**1188.68754(15)**	**120.3411**
LYS/LYS/LYS	GLU/GLU/ASP	TOL/ALA	-22.9095	-447.8011	-486.1679	501.62526(109)	130.65541
LYS/LYS/LYS	GLU/GLU/ASP	TOL/ABU	-22.9095	-449.0266	-473.3273	474.71856(136)	118.7432
LYS/LYS/LYS	GLU/GLU/ASP	OTL/ALA	-81.4939	-447.8011	-486.0323	442.76969(190)	135.57483
LYS/LYS/LYS	GLU/GLU/ASP	OTL/ABU	-81.4939	-449.0266	-484.3858	438.25111(210)	134.85669
LYS/LYS/LYS	GLU/GLU/ASP	PHE/ALA	-96.2967	-447.8011	-485.2504	426.40309(271)	130.16659
LYS/LYS/LYS	***GLU/GLU/ASP***	***PHE/ABU***	***-96.2967***	***-449.0266***	***-483.6387***	***421.95417(290)***	***129.48328***
ORN/LYS/ORN	**GLU/GLU/ASP**	**MTL/ALA**	**768.0355**	**-447.8011**	**-434.9324**	**1190.09921(6)**	**116.79741**
ORN/LYS/ORN	**GLU/GLU/ASP**	**MTL/ABU**	**768.0355**	**-449.0266**	**-433.7902**	**1186.58938(29)**	**116.58364**
ORN/LYS/ORN	GLU/GLU/ASP	TOL/ALA	54.5479	-447.8011	-446.4202	499.58727(114)	126.89955
ORN/LYS/ORN	GLU/GLU/ASP	TOL/ABU	54.5479	-449.0266	-433.5781	472.67755(141)	114.98583
ORN/LYS/ORN	GLU/GLU/ASP	OTL/ALA	-4.0414	-447.8011	-446.2847	440.72698(195)	131.81903
ORN/LYS/ORN	GLU/GLU/ASP	OTL/ABU	-4.0414	-449.0266	-444.6369	436.20587(225)	131.09962
ORN/LYS/ORN	**GLU/GLU/ASP**	**PHE/ALA**	**-19.1217**	**-447.8011**	**-445.532**	**424.14135(276)**	**126.43999**
ORN/LYS/ORN	**GLU/GLU/ASP**	**PHE/ABU**	**-19.1217**	**-449.0266**	**-443.9191**	**419.6899(306)**	**125.75541**
ORN/LYS/DAB	**GLU/GLU/ASP**	**MTL/ALA**	**772.2246**	**-447.8011**	**-433.4433**	**1191.31014(4)**	**116.28679**
ORN/LYS/DAB	**GLU/GLU/ASP**	**MTL/ABU**	**772.2246**	**-449.0266**	**-432.3**	**1187.79789(23)**	**116.07181**
ORN/LYS/DAB	GLU/GLU/ASP	TOL/ALA	58.7488	-447.8011	-444.9317	500.8111(112)	126.38945
ORN/LYS/DAB	GLU/GLU/ASP	TOL/ABU	58.7488	-449.0266	-432.0887	473.89956(139)	114.47483
ORN/LYS/DAB	GLU/GLU/ASP	OTL/ALA	0.1596	-447.8011	-444.7957	441.94998(193)	131.30846
ORN/LYS/DAB	GLU/GLU/ASP	OTL/ABU	0.1596	-449.0266	-443.1472	437.42737(216)	130.58831
ORN/LYS/DAB	**GLU/GLU/ASP**	**PHE/ALA**	**-14.9669**	**-447.8011**	**-444.0782**	**425.38841(274)**	**125.96461**
ORN/LYS/DAB	**GLU/GLU/ASP**	**PHE/ABU**	**-14.9669**	**-449.0266**	**-442.4645**	**420.93547(297)**	**125.27928**

Table 13b. Complete round 3 selection with positions 11, 12, 13, and 18 with bb2. The energy function consisted of VDW, EEF1 solvation, and distance-dependent dielectric energy terms. The following equation was maximized: $f = -2E_{ABAB} + E_{AAAA} + E_{BBBB}$. E_{ABAB}, E_{AAAA}, and E_{BBBB} are E_{folded} energies. A sequence predicted to be very specific for the heterotetramer has a large positive specificity score. The following equation is maximized for the stability calculation: $f = -E_{folded} + E_{unfolded}$, so a large, positive stability energy indicates a stable molecule. A and B refer to the ACID and BASE peptide. The round 2 peptide is shown in italics, and the peptides tested in round 3 are in bold.

		hairpin									*helix*											
Peptide		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
B3	Ac	Y	R	I	p	S	Y	D	F	a	D	R	F	K	K	L	L	R	R	A	Z	G	NH_2
B4	Ac	Y	R	I	p	S	Y	D	F	a	D	Orn	F	K	K	L	L	R	Orn	A	Z	G	NH_2
B5	Ac	Y	R	I	p	S	Y	D	F	a	D	Orn	F	K	K	L	L	R	Dab	A	Z	G	NH_2
B6	Ac	Y	R	I	p	S	Y	D	F	a	D	K	Mtl	K	K	L	L	R	K	A	Z	G	NH_2
B7	Ac	Y	R	I	p	S	Y	D	F	a	D	Orn	Mtl	K	K	L	L	R	Orn	A	Z	G	NH_2
B8	Ac	Y	R	I	p	S	Y	D	F	a	D	Orn	Mtl	K	K	L	L	R	Dab	A	Z	G	NH_2

Table 14. Sequences of **B3**, **B4**, **B5**, **B6**, **B7**, and **B8**. Abbreviations: p=D-proline, a = D-alanine, Z= benzoylated L-α,β-diaminopropionic acid, Mtl= *m*-methylphenylalanine, Dab= L-α,β-diaminobutyric acid (adapted from (*13*)).

Peptide	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**
All Data						
NONLIN	2.9	3.2	3.3	3.3	3.6	3.7
SEDPHAT	3.0	3.4	3.5	3.5	3.7	3.8
Lowest Concentration Data						
NONLIN	2.0	3.4	3.2	2.8	4.0	4.8
SEDPHAT	1.7	3.4	3.1	2.7	3.0	4.5

Table 15. Results of AUC experiments with peptides from round 3. Stoichiometries reported correspond to the weight-average molecular weight found by NONLIN or SEDPHAT.
1. Peptide concentrations are **B3**: 100, 300, 600 μM, **B4**: 200, 600, 1000 μM, **B5**: 100, 300, 600 μM, **B6**: 50, 300, 700, **B7**: 50, 500, 1000 μM, **B8**: 50, 300, 600 μM (adapted from (*13*)). This experiment was run by Mayssam Ali.

FIGURES

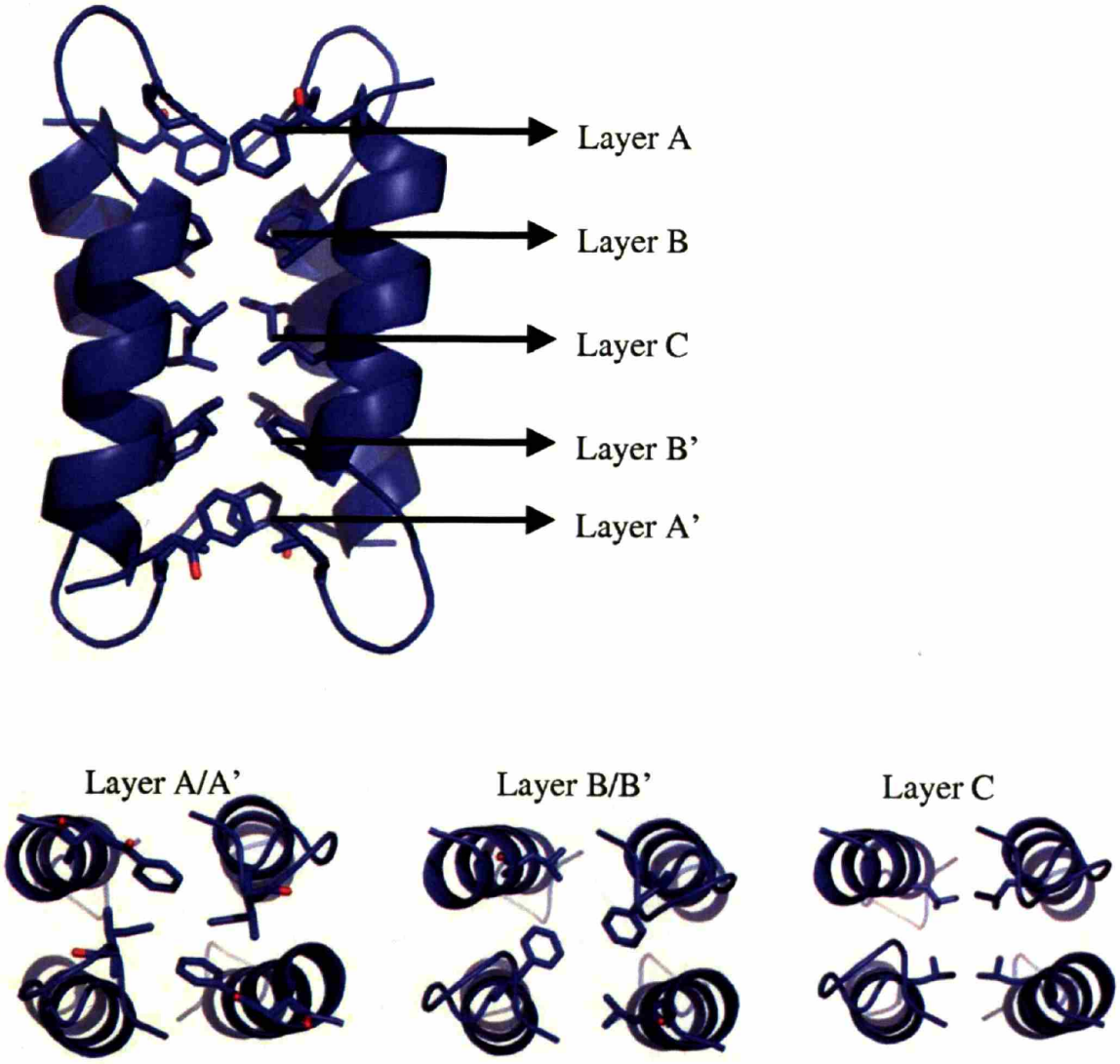
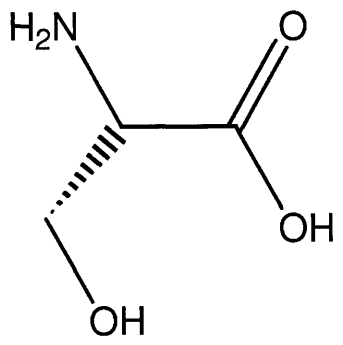
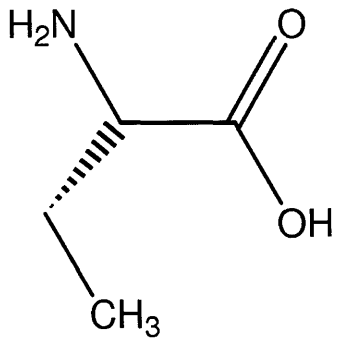


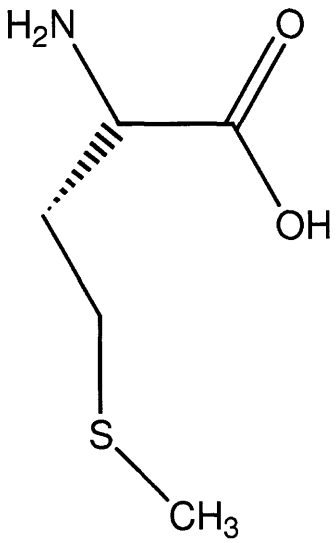
Figure 1. BBA homotetramer crystal structure. The BBA core consists of five palindromic layers (A, B, C, B', and A'). The layers are shown above: A and A' (DapBz/Ile), B and B' (Phe/Leu), and C (Leu/Leu).



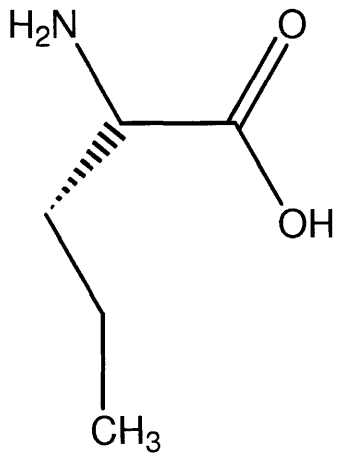
Serine



L-α-aminobutyric acid (Abu)

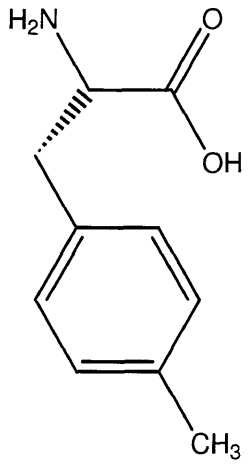


Methionine

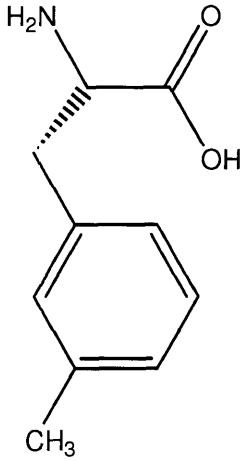


2-aminohexanoic acid (Nle)

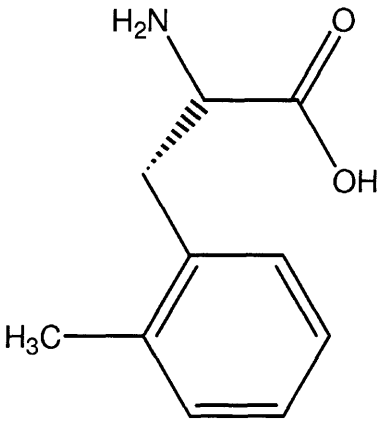
Figure 2. Structures of amino acids analyzed in round 2 for their potential to stabilize the heterotetramer.



4-methyl-L-phenylalanine (Tol)

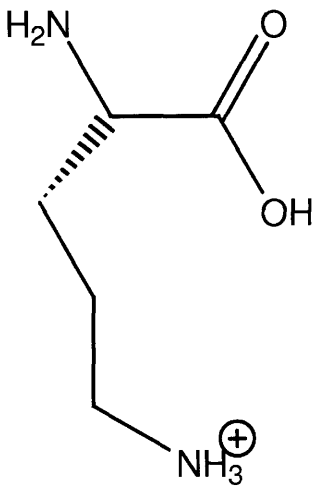


3-methyl-L-phenylalanine (Mtl)

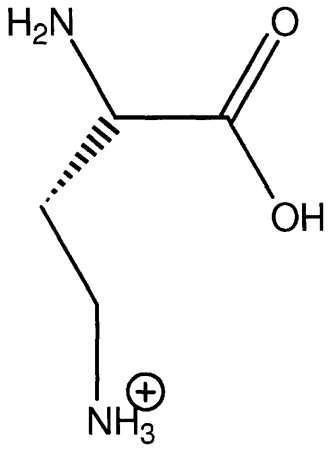


2-methyl-L-phenylalanine (Otl)

Figure 3. Structures of amino acids evaluated in round 3 for their potential to add additional specificity in the core.



L-ornithine (Orn)



2,4-diaminobutyric acid (Dab)

Figure 4. Structures of amino acids evaluated in round 3 for their potential to add additional specificity on the surface.

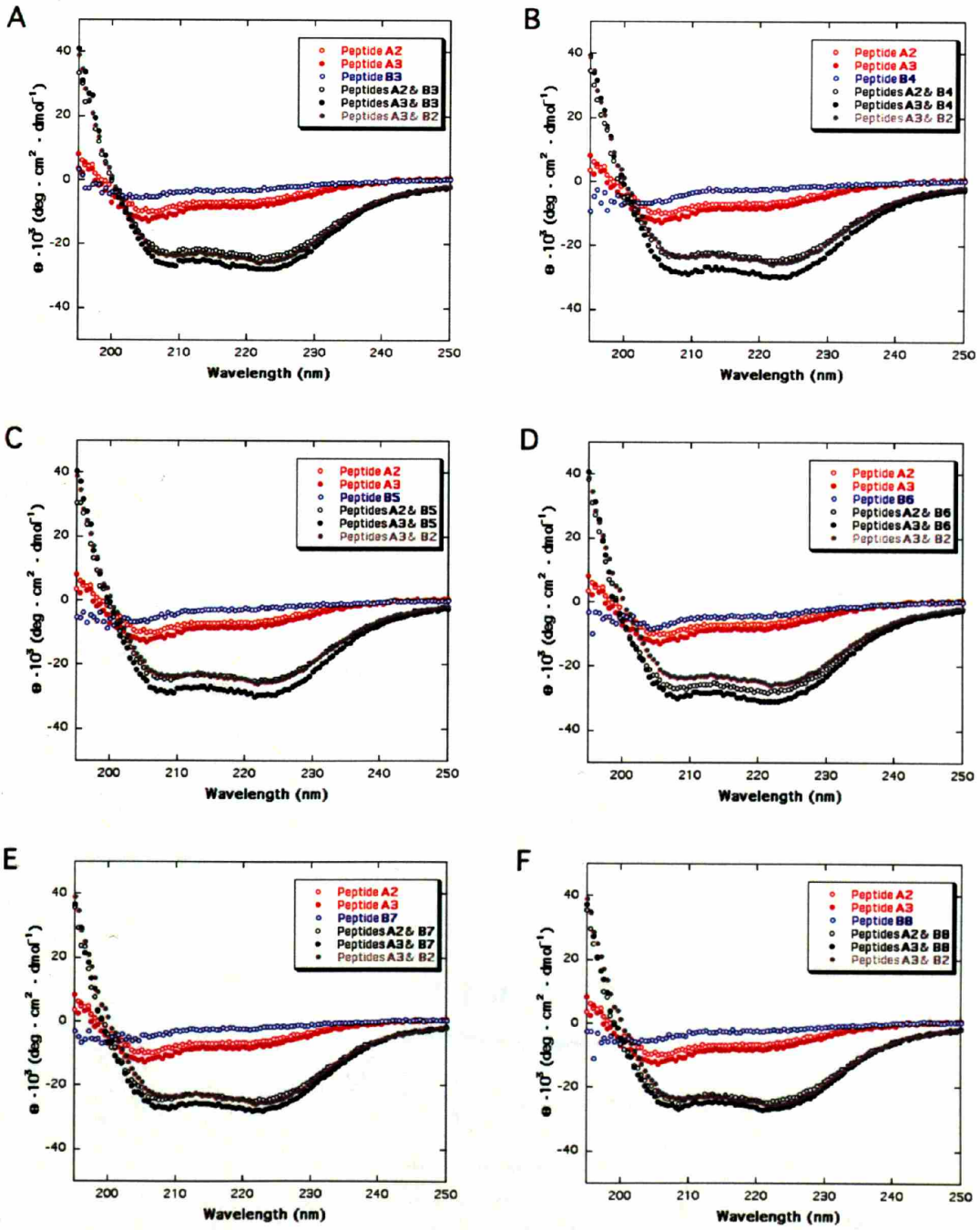


Figure 5. CD spectra of peptides designed in round 3 alone and in combination with **A-Ala (A2)** and **A-Abu (A3)**. All spectra are compared to **A3-B2 (A-Abu/B-Phe)**. A: Peptide **B3**. B: Peptide **B4**. C: Peptide **B5**. D: Peptide **B6**. E: Peptide **B7**. F: Peptide **B8**. (Taken from (13)). These experiments were run by Mayssam Ali.

APPENDIX B

Allosteric Inhibition of Zinc-Finger Binding in the Major Groove of DNA by Minor-Groove Binding Ligands

Reprinted from:
Doan H. Nguyen-Hackley, Elizabeth Ramm, Christina M. Taylor, J. Keith Joung, Peter B. Dervan, and Carl O. Pabo. Biochemistry (2004), 43(13), 3880-3890.

C.M.T. did the computer modeling section, some of the protein purification, and helped with the gel shift assays.

ABSTRACT

In recent years, two methods have been developed that may eventually allow the targeted regulation of a broad repertoire of genes. The engineered protein strategy involves selecting Cys_2His_2 zinc finger proteins that will recognize specific sites in the major groove of DNA. The small molecule approach utilizes pairing rules for pyrrole-imidazole polyamides that target specific sites in the minor groove. To understand how these two methods might complement each other, we have begun exploring how polyamides and zinc fingers interact when they bind the same site on opposite grooves of DNA. Although structural comparisons show no obvious source of van der Waals collisions, we have found a significant "negative cooperativity" when the two classes of compounds are directed to the overlapping sites. Examining available crystal structures suggests that this may reflect differences in the precise DNA conformation, especially with regard to width and depth of the grooves, that is preferred for binding. These results may give new insights into the structural requirements for zinc finger and polyamide binding and may eventually lead to the development of even more powerful and flexible schemes for regulating gene expression.

INTRODUCTION

Recent developments in the design and selection of Cys_2His_2 zinc fingers *(1, 2)* and in the design of novel polyamides *(3, 4)* suggest that each method can be used to target specific sites in the genome. Cys_2His_2 zinc finger proteins, in which each domain contains 28-30 amino acid residues, are the most common DNA-binding motif in higher eukaryotes. Each zinc finger has a conserved $\beta\beta\alpha$ motif, and amino acids near the N-terminus of the α-helix contact bases in the major groove of B-DNA *(2, 5-7)*. The X-ray crystal structure of the three-finger Zif268 protein (Figure 1A, refs *6, 7)* illustrates the basic features of recognition and has provided the basis for most of the subsequent work in this field. The Cys_2His_2 zinc finger framework appears to be very adaptable, and specific variants have been selected that bind to many different desired target sites in duplex DNA *(8-11)*.

Another strategy for recognition and regulation has been developed that uses polyamides consisting of N-methylpyrrole (Py)[1] and N-methylimidazole (Im) rings, which can be linked together to recognize a predetermined DNA sequence *(3, 4)*. Sequence-specific polyamide-DNA recognition depends on binding in the minor groove with side-by-side amino acid pairings. Simple rules were developed for designing polyamides that can target desired DNA sequences *(4)*. These rules have been carefully validated through characterization of synthesized polyamides via DNase I footprinting, affinity cleavage, two-dimensional nuclear magnetic resonance (NMR) *(12)*, and X-ray crystallographic methods *(13-15)*. To illustrate the docking arrangement, Figure 1B shows the X-ray crystal structure of a polyamide *(15)* bound in the minor groove of B-DNA.

In this work, we have begun studies to investigate how polyamides and zinc fingers may interact as they bind to overlapping sites on double-stranded DNA. Earlier studies have explored

how polyamides interfere with the binding of several classes of transcription factors *(16-23)*. Polyamides have inhibited minor-groove binding proteins such as TATA-binding protein (TBP), as well as minor groove contacting proteins like the lymphoid enhancer factor (LEF-1) *(17)*. In the case of the purely major-groove binding protein GCN4, polyamides were clearly shown to co-occupy in the minor groove of the GCN4 binding site *(21)*. An extended Arg-Pro-Arg tripeptide attached to a polyamide did inhibit GCN4 binding by neutralizing a phosphate contact made by the protein with the DNA backbone *(22)*. Polyamides have been shown to inhibit the binding of a zinc finger protein, TFIIIA *(16, 20)*. However, inhibition of this zinc finger protein was the result of the polyamide being targeted to and displacing the minor-groove-spanning fourth finger of the nine-finger protein. Therefore, studies to date on polyamide/protein interactions would suggest that an unmodified polyamide should co-occupy the minor groove face of a purely major groove zinc finger protein binding site.

In this study, the interaction of polyamides and zinc fingers that bind to opposing grooves of the same DNA target site was explored. The zinc finger proteins chosen for this study included Zif268 and a set of Zif268 variants that had been selected to recognize rather different DNA sequences *(9)*. These three other variants had been selected to recognize the TATA box, the nuclear receptor element, and the p53 binding sites, and they are referred to (respectively) as $TATA_{ZF}$, NRE_{ZF}, and $p53_{ZF}$. All these proteins bind their sites with nanomolar dissociation constants, recognizing sites in the DNA major groove and discriminating effectively against nonspecific DNA *(9, 24)*. For this project, polyamides **1-4** were designed to specifically target the very same TATA, NRE, p53, and Zif268 binding sites. These hairpin polyamides bind their sites with at least nanomolar dissociation constants, as determined by quantitative DNase I footprinting experiments *(25-28)*. Biochemical studies, using gel mobility shift experiments and

DNase I analysis, give information about how these compounds interact as they bind. Computer modeling based on previous structural studies of zinc finger-DNA and polyamide-DNA complexes help us to interpret these results.

MATERIALS AND METHODS

Protein Production and Purification

The Zif268 zinc finger region (residues 333-421) was subcloned into a pET-21d expression vector (Novagen) and was transformed into the Escherichia coli strain BL21(DE3) containing the pLysE plasmid. Cultures were grown and induced as described (Novagen). After the cells were harvested, they were lysed and sonicated as recommended (Novagen). The peptides were denatured and reduced in 6.4 M guanidine·HCl, 150 mM DTT, and 50 mM Tris-HCl, pH 8.0, at 75 °C for 30 min and acidified to pH ~ 2.0 by addition of 10% trifluoroacetic acid (TFA). The peptides were purified by reverse-phase batch extraction on Sep-pack C-18 cartridges (Waters) as described *(29)*, followed by purification on a C4 reversed-phase (Vydac) high-performance liquid chromatography column using a gradient of 22-35% acetonitrile (ACN) containing 0.1% TFA. The purified peptide fractions were then refolded anaerobically in the binding buffer that we used for gel shifts, after supplementing it with a 0.5 M excess of $ZnCl_2$. Refolded peptides were stored at -80 °C in 10 μL aliquots; each aliquot was used once for binding studies and then discarded. The active concentrations of peptides were determined in stoichiometric competition experiments.

$TATA_{ZF}$, NRE_{ZF}, and $p53_{ZF}$ zinc finger peptides, selected from residues 333-421 of Zif268 *(24)*, used in these studies were purified by Scot A. Wolfe, Robert Grant, and Sandra Fay-

Richard, respectively. Since they were preparing these samples for crystallographic studies, they used several additional purification steps, essentially as described *(30)*.

Hairpin Polyamide Syntheses and Characterization

The pyrrole-imidazole hairpin polyamides **1-4** were prepared by manual and machine-assisted solid-phase methods *(31)* (Figures 2 and 3). Purity and identity of each compound were verified by a combination of analytical HPLC, 1H NMR, and matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF). Polyamides **1-3** have been described previously *(32, 33)* and polyamide **4**, PyIm-β-Im-$(R)^{H}2^{N}$γ-PyImPyPy-β-Dp, is described here: UV (H_2O) λ_{max}304 (60 800); MALDI-TOF MS 1230.65 (1230.59 calcd for [M + H] $C_{55}H_{72}N_{23}O_{11}^{+}$). Lyophilized samples of polyamides were stored in at -80 °C. Polyamide concentrations were determined spectrophotometrically with extinction coefficients estimated based on the number of aromatic rings using the relation 8690 M^{-1} cm^{-1} per aromatic ring for the absorption maximum at 290-315 nM *(28)*.

Gel Mobility Shift Assay

Double-stranded oligonucleotides used for gel mobility shift assays were essentially identical to DNA sites used *(9)* when selecting the $TATA_{ZF}$, NRE_{ZF}, and $p53_{ZF}$ zinc finger proteins. A few nucleotides in the sequence flanking the zinc finger binding sites also were changed so that the zinc fingers and polyamides would have overlapping binding sites. The 27 bp DNA duplexes that were used for gel mobility shift assays are boxed in Figure 4, and the nine base pair zinc finger target sites near the center of each duplex are underlined. (Note that in the form shown in Figure 4, 27 bp duplexes are embedded within larger DNA segments that were later used for footprinting studies.) The oligonucleotide binding site for wt Zif268 was slightly

different from that used by Greisman et al. *(9)* containing GCGGGGGCG rather than GCGTGGGCG. However, this construct was advantageous because it readily provided an overlapping polyamide binding site and should have no other effect. (Zif268 binds extremely well to either site and there is no direct contact with this base in the crystal structure *(6)*). Each 27 bp strand of these binding sites was synthesized using standard phosphoramidite chemistry on an Applied Biosystems model 392 DNA synthesizer. Following deprotection, oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis. Duplexes were annealed and quantified, and then 0.3 pmol of each DNA fragment were end-labeled with [α-^{32}P]-ATP (New England Biolabs) using Klenow exo- to fill-in the overhangs. Unincorporated nucleotides were removed by use of Sephadex G-25 quick spin columns (Boehringer Mannheim), and the DNA was resuspended in 1 mL of the buffer used for all gel shift experiments, which contained 5 mM HEPES, pH 7.8, 50 mM KCl, 50 mM KGlu, 50 μM KoAc, 5 mM $MgCl_2$, 5% glycerol, 0.1% NP-40, 20 μM $ZnSO_4$, 100 μg/mL BSA, 1 mM $CaCl_2$, and 1 mM DTT. The nucleotides were stored at -20 °C.

Using the buffer above, we performed binding reactions and equilibrium binding studies for zinc finger proteins in both the presence and absence of polyamides. Zinc finger dissociation constants were determined as previously described *(9)*, except that 0.5 pM of labeled duplex DNA was used and equilibration was done at room temperature for 16-18 h for all gel shift experiments. The reaction mixtures were then subjected to 12% native polyacrylamide gel electrophoresis for about 2 h with 0.5× TBE running buffer. Radioactive signals were quantitated by PhosphorImage analysis (Molecular Dynamics).

Equilibrium dissociation constants (K_D's) were determined by linear regression using the Scatchard equation:

$$\Theta/[P] = 1/K_D - \Theta/K_D \qquad (1)$$

in which Θ equals the fraction of DNA bound ([PD]/([PD] + [D])), and [P] equals the free protein concentration (after applying corrections to account for the percent of the protein that was active for DNA binding).

DNase I Footprinting

Oligonucleotides used for footprinting experiments (Figure 4) were designed so that they included the sequences of the 27 bp duplexes that had been used for gel mobility shift assays, and additional six to seven base pair "spacers" were added so that the protein-binding sites would be slightly further apart. Oligonucleotide 1 contains binding sites for $TATA_{ZF}$, NRE_{ZF}, and $p53_{ZF}$; oligonucleotide 2 contains binding sites for Zif268, $TATA_{ZF}$, and $p53_{ZF}$. The synthetic duplexes were designed such that the ends were ready for cloning into a pBluescript II SK(+) plasmid (Stratagene) that had been cut with EcoRI and BamHI. After growth in E. coli, DNA probes for footprinting were prepared by digestion of the appropriate plasmids with EcoRI and NotI restriction enzymes, giving a 134 bp fragment for oligonucleotide 1 and a 143 bp fragment for oligonucleotide 2. After digestion, these DNA fragments were radioactively labeled by using Klenow enzyme (New England Biolabs) and α-^{32}P labeled nucleotides to fill in the overhanging ends. The labeled oligonucleotides were purified using 5% native PAGE, were precipitated with EtOH, and were counted for specific activity.

Footprinting experiments were done essentially as previously described *(25-28)*. Quantitative footprinting for polyamides **1-4** was performed using the appropriate restriction fragment (oligonucleotide 1 or 2) in triplicate under both conventional TKMC buffer (10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$) and the same binding buffer used in the gel shift experiments. Polyamide/DNA solutions were allowed to equilibrate at 22

°C for 16-18 h. Footprinting reactions were initiated by the addition of the appropriate amount of DNase I to give ~50% intact DNA and allowed to proceed for 7 min at 22 °C. The reactions were stopped by addition of 50 μL of a solution containing 1.25 M NaCl, 100 mM EDTA, 0.2 mg/mL glycogen, and 28 μM base-pair calf thymus DNA and ethanol precipitated. The reactions were resuspended in 1 μL of TBE/80% formamide loading buffer, denatured by heating at 80 °C for 10 min, and placed on ice. The reaction products were separated by electrophoresis. Radioactive signals were visualized with a Molecular Dynamics Typhoon phosphorimager followed by quantitation using ImageQuant software (Molecular Dynamics). Equilibrium association constants for the polyamides were calculated as previously described *(29)* and found to be identical in both TKMC buffer and the gel shift buffering conditions.

Computer Modeling Experiments

Structures of polyamide-DNA complexes were taken from the Protein Data Bank *(13-15)* and were compared with known zinc finger-DNA complexes for $TATA_{ZF}$ *(34)*, $p53_{ZF}$ *(35)*, and Zif268 *(6)*. The structures were aligned (using the PROTEUS Program *(36)*) via phosphorus atoms, C1' atoms, or common atoms in the set of superimposed base pairs. Several different alignment strategies were tested, including schemes in which (1) the entire zinc finger DNA binding site was aligned with the polyamide DNA, (2) the polyamide binding sites in the known polyamide-DNA structures were aligned with the expected polyamide binding sites on the zinc finger DNA, (3) the GC or AT base pair in the polyamide-DNA structure was aligned with the corresponding base pair in the DNA-zinc finger structure, or (4) all the common atoms in a base pair were aligned. The RMS deviation was noted for each of the alignments, and a visual inspection of the alignment was made using Insight *(37)*.

As another strategy for comparison of the DNA structures, Curves *(38)* was run on all the polyamide-DNA complexes and all the zinc finger-DNA complexes. In addition, Curves also was run on B-DNA *(39)*. Comparisons of the output from Curves focused on the width and depth of the major and minor grooves and on base pair parameters such as buckle, shear, propeller twist, opening, and X and Y displacement. For major and minor groove comparisons, the groove dimensions were taken from (1) the region of the polyamide-DNA complex closest to the polyamide binding site and (2) the region on the zinc finger-DNA complex that was closest to the expected polyamide binding site present in our experiments. The average of the buckle, shear, propeller twist, and opening parameters were taken from the global base-base parameter output of Curves. Parameters for individual base pairs were also compared separately.

RESULTS

Gel mobility shift experiments and DNase I footprinting experiments were used to analyze binding of the zinc finger proteins, binding of polyamides, and interactions between the zinc fingers and the match polyamides that recognized overlapping sites. In competition studies, control experiments also were performed with mismatch polyamides to help ensure the specificity of the observed effects. To facilitate comparison of different studies, all experiments were carried out with identical buffer conditions, temperature, equilibration times of the binding reactions, and order of the addition of the components into the reaction mixtures.

Polyamide Equilibrium Dissociation Constants

A set of pyrrole-imidazole polyamides was designed, synthesized, and purified for our competitive binding studies. Structures of these polyamides and their expected modes of binding with the DNA minor groove, as predicted on the basis of "pairing rules" *(4)*, are shown in Figures

2 and 3. Polyamides **1-4** are referred to as PA_{p53}, PA_{TATA}, PA_{NRE}, and PA_{Zif268} (where our nomenclature indicates which binding site each polyamide has been designed to recognize).

We used quantitative DNase I footprint titration analysis to determine the apparent dissociation constants for the polyamides. Oligonucleotide 1 was used for PA_{p53} and PA_{NRE}; oligonucleotide 2 was used for PA_{TATA} and PA_{Zif268} (Figure 4). The oligonucleotides were labeled from the 3'-end. DNase I footprinting experiments (Figure 5), done at a series of different polyamide concentrations, reveal the location of the binding site for each polyamide and allow determination of the binding constants. These footprinting experiments (Figure 5) show that PA_{TATA} and PA_{Zif268} polyamides also bind specifically to other match and single base pair mismatch sites (as defined by the pairing rules) found on the restriction fragment at the same or higher concentration, respectively. Quantitative analysis of the degree of protection observed at each polyamide concentration allowed us to present the data of the DNase I footprinting experiments as binding isotherms (Figure 6). The apparent K_D, equal to $1/K_A$, was determined from the binding isotherms by fitting the data points with a modified Hill equation, as previously described *(28)*. These binding isotherms and apparent dissociation constants for each polyamide represent the average of three independent experiments, and all these polyamides have at least nanomolar affinity for their expected target sites.

Determination of Zinc Finger Dissociation Constants

This study used Zif268 and three variants, specific for the p53, NRE, and TATA binding sites, that had been selected via phage display. (Note that the sequential selection protocol *(9)* used to obtain these variants allowed extensive changes in the recognition site, since six amino acids in each finger had been randomized.) Quantitative gel shift analysis was used to determine the fraction of the DNA fragment bound at a series of protein concentrations (Figure 7), and KD

values were calculated from the slope of Scatchard plots. Corrections were made for the active concentration of each protein, which was estimated using stoichiometric competition experiments.

The sequences of the 27 bp oligonucleotides used in gel mobility shift experiments are boxed in Figure 4. (Note that in this figure the oligonucleotides are shown embedded within larger sites that were later used for DNase I footprinting.) The dissociation constants for the zinc fingers (Table 1) were determined from the average of three independent experiments, and the values are comparable to those previously reported *(9)*, even though slightly different buffers were used.

Dissociation Constant of Zinc Finger Protein in the Presence of Specific Polyamides

Binding of the polyamides could not be directly observed in our gel shift experiments, but competition studies clearly revealed that the polyamides affected formation, stability or both of the zinc finger-DNA complexes. Two types of competition experiments were done to analyze the effects of polyamides on zinc finger binding.

The first set of experiments used a fixed polyamide concentration that would have given ~90% occupancy of the free DNA site, and gel shift experiments were used to determine the apparent binding constant of the zinc finger proteins under these conditions. These gel shift experiments were essentially identical to those done with the protein alone. The long equilibration time, done for all gel shift experiments, ensured that the apparent binding did not depend on the order of addition of the components. Figure 7 shows one set of gel shift experiments, where the apparent binding constant of Zif268 is determined in the presence and absence of the polyamide that recognizes an overlapping site. Binding isotherms for this reaction are shown in Figure 7C, and similar results were obtained for the other polyamide/zinc

finger competitive binding titrations. In every case, the presence of the corresponding polyamide significantly decreases (by 13-35-fold) the apparent affinity of the zinc finger protein for the binding site (Table 1).

Interactions between the polyamides and zinc fingers were also studied in experiments that used fixed zinc finger protein concentrations with variable amounts of polyamide. In each of these experiments, the zinc finger proteins were present at active concentrations equal to 10× their respective dissociation constants, and thus about 90% of the DNA was initially shifted by the proteins. Similar binding reactions with increasing amounts of polyamide were conducted, and after allowing for full equilibration, we monitored the effect of the polyamides on formation of the zinc finger-DNA complexes. Gel shift results obtained with the PA_{Zif268} polyamide and the Zif268 protein (Figure 8A) show that the polyamide interferes with formation or stability of the zinc finger-DNA complex. Similar results were obtained using other zinc finger proteins and polyamides that compete for the same binding sites. Protein bound was calculated as a function of polyamide concentration, each data point representing the average of three independent experiments. Data were fit with the modified Hill equation *(28)*, enabling us to estimate "inhibition constants". These represent the polyamide concentrations that give 50% inhibition of formation of the respective protein-DNA complex, and we find that these "inhibition constants" are in the subnanomolar range (Table 2).

Testing for Specificity of Zinc Finger/Polyamide Interference Effects

Several types of experiments were done to test the specificity of the observed interference effects between polyamides and zinc fingers. For example, Figure 8 shows gel shift data obtained with the Zif268 protein-DNA complex. This figure compares the interference effects observed with PA_{Zif268} (the match polyamide with specific binding for this site) and interference

effects observed with PA_{NRE} (which is specific for the NRE site and a mismatch polyamide with no binding site on the 27 bp Zif268 oligonucleotide). There is marked interference by PA_{Zif268}, as indicated by the gel shift results in Figure 8A and the sigmodal plot shown in Figure 8C. However, similar concentrations of PA_{NRE} have no measurable effect at this site, as there is no reduction in Zif268 binding even at high polyamide concentrations (Figure 8B,C). Similar data were obtained using other zinc finger/polyamide combinations, and these results show that interference is dependent on having a polyamide that can compete for the same binding site.

Competitive DNase I Footprinting Experiments

Since binding of polyamides could not be directly observed in any of our gel shift experiments, we used DNase I footprinting to further explore the mechanism of polyamide/zinc finger interference effects. In principle, these experiments should allow us to directly monitor polyamide-DNA interactions, zinc finger-DNA interactions, and possible formation of the ternary complex.

Figure 9 shows the results of DNase I footprinting experiments that explored interference effects between the Zif268 protein and the PA_{Zif268} polyamide. These experiments used oligonucleotide 2 (Figure 4), labeled from the 5' end, to provide a Zif268 protein and PA_{ZF} polyamide binding site. The PA_{NRE} polyamide, which does not have a binding site on this oligonucleotide, was used as a "mismatch" control. Six sets of experiments were performed, using the PA_{Zif268} polyamide at concentrations of 1, 5, and 10 nM (panel A) and then using the PA_{NRE} polyamide at corresponding concentrations (panel B). When the Zif268 protein was present, it was used at a concentration that would have been expected to give approximately 90% occupancy of the free DNA site. As indicated at the top of the respective lanes (Figure 9), four conditions were tested within each set of reactions: (1) the first lane within each set has no

protein and no polyamide; (2) the second lane has Zif268 protein but no polyamide; (3) the third lane has Zif268 protein and one of the polyamides; (4) the fourth lane has the same polyamide but contains no protein.

We used Gilbert-Maxam G+A sequencing reactions to determine the precise base positions where cleavage occurred in these footprinting experiments and to locate the binding sites for the Zif268 protein and for the PA_{Zif268} polyamide, which are marked on the left side of the gels. (Note that experiments in panel B used the PA_{NRE} polyamide, which does not have a binding site on this oligonucleotide.) As expected, there is significant overlap between the region where Zif268 footprints and the region where PA_{Zif268} footprints: the PA_{Zif268} binding site includes bases 32-39, while the Zif268 protein binding site includes bases 35-43 (when counting from the 5' end). However, it appears that band 43, at the far end of the Zif268 binding site, can only be effectively protected by the protein.

Results obtained with the mismatch polyamide (panel B) are entirely straightforward: Zif268 occupies its binding site in lanes 2 and 3 of each experiment, and the PA_{NRE} polyamide (which is present in lanes 3 and 4 but has no binding site in this oligonucleotide) has no effect on the footprinting patterns. The results with the PA_{Zif268} polyamide (panel A) are somewhat more complicated, but they are the real crux of the experiment. In this panel, lane 2 of each set (protein alone) shows the expected footprint for the Zif268 protein. Lane 4 of each set (polyamide alone) shows the footprints obtained when the PA_{Zif268} polyamide is present at concentrations of 1, 5, and 10 nM. Effects are localized to the expected binding site at low concentration (1 nM), but additional protection due to specific binding of a single base pair mismatch site present in the restriction fragment (near the top of the gel) is seen at higher concentrations.

Obviously, the most interesting results in each set of reactions from panel A involve lane 3, in which both the Zif268 protein and the PA_{Zif268} polyamide were present. At the higher polyamide concentrations (5 and 10 nM), the footprinting patterns obtained with the polyamide/zinc finger combinations (lane 3) appear surprisingly similar to those obtained with the polyamide alone (lane 4). The simplest interpretation of these patterns is that binding of the polyamide displaces the zinc fingers, making the patterns look very similar in lanes 3 and 4. It is somewhat more problematic to explain the pattern obtained in lane 3 of the first set (at a 1 nM concentration of PA_{Zif268}), but this may indicate a mixture of species with the Zif268 protein bound to some fraction of the DNA sites and the PA_{Zif268} polyamide bound to most of the remaining DNA.

Computer Modeling of Polyamide and Zinc Finger Structures

In a further attempt to understand the mechanisms of interference, we examined the structure of various zinc finger-DNA and polyamide-DNA complexes. X-ray crystal structures are now available for the $TATA_{ZF}$ *(34)*, $p53_{ZF}$ *(35)*, and Zif268 *(6)* zinc finger-DNA complexes. X-ray crystal structures also are available for several polyamide-DNA complexes *(12, 13, 15)*, and we thought that these structures would be useful for initial modeling, even though the exact sequences of the polyamides and of the binding sites are somewhat different than those used in our current experiments. Our first major conclusion is that we do not see any obvious basis for a steric collision between the zinc fingers and the polyamides. All the polyamide-DNA complexes indicate that these compounds bind exclusively in the minor groove, while the zinc finger-DNA complexes show that these proteins bind in the major groove. [The only exception here involves one lysine (Lys189) in the TATA structure with a high-temperature factor that might reach into the minor groove, but the polyamide binds at the opposite end of this site, far from Lys189.]

Various zinc finger-DNA and polyamide-DNA complexes were superimposed, and these were compared in an attempt to understand the basis for the "negative cooperativity" that we had observed in biochemical studies. Alignments done with the program PROTEUS *(36)* typically used 3-4 sets of phosphates or 3-4 sets of C1' atoms to superimpose the complexes. The rms deviations in these alignments varied from about 0.6 to about 2 Å, depending on (1) which structures were aligned, (2) which atoms were used for alignment, and (3) how many base pairs were superimposed. However, there were striking and consistent differences in the groove width, and these were large enough that they could readily explain how the polyamides interfere with zinc finger binding. In nearly every alignment that we carried out, the polyamide-DNA complex had a wider and deeper minor groove than that observed in the zinc finger-DNA complexes. There also were a number of cases where superimposing the complexes indicated shifts in the precise arrangement of the base pairs, and these could affect the position of hydrogen bond donors and acceptors that are critical for recognition.

These initial qualitative observations of the structures were confirmed by using the Curves program *(38)* to calculate dimensions of the major and minor grooves. We find that the major grooves in zinc finger-DNA complexes are consistently wider and deeper than the major grooves in the polyamide-DNA complexes. In this set of structures, the Zif268 complex has the deepest major groove (6-8 Å), while the $TATA_{ZF}$ and $p53_{ZF}$ complexes have the widest major grooves (13-14.5 Å). The major groove width of the polyamide-DNA complexes ranged from 9 to 10.5 Å, while the major groove depth ranged from 3 to 5.5 Å. Differences in minor groove width are not quite as consistent when comparing the two classes of complexes, but there is a general tendency for the polyamide-DNA complexes to have wider minor grooves than the zinc finger-DNA complexes.

The initial superimpositions indicated that the orientations of the bases were slightly different in the two types of complexes, and these differences were confirmed and analyzed using the program Curves. In general, we find the following: (1) The two base pairs in the middle of the binding region have negative opening angles in the polyamide-DNA complexes (-12.22° ± 5.71°). No such region is observed in the zinc finger-DNA structures, which on average have more positive opening angles (1.86° ± 4.23° averaged over the entire binding region of all of the complexes). (2) The zinc finger complexes have a larger negative X-displacement of the base pairs than do the polyamide complexes. Averaged over the respective binding sites for each set of complexes, the X-displacement values are -1.54 ± 0.28 Å for the zinc finger-DNA complexes and -0.13 ± 0.35 Å for the polyamide-DNA complexes. (3) X-displacement parameters for the polyamide-DNA complexes are generally closer to those of B-DNA, whereas opening angles for zinc finger-DNA complexes are generally closer to those of B-DNA. [For a typical B-DNA structure *(39)* averaged over the entire sequence, the values are -0.52 ± 0.25 Å and -0.31° ± 5.27°, respectively.] No significant differences in shear or buckle were noted when comparing the zinc finger and polyamide complexes with those of B-DNA.

DISCUSSION

Compounds that can target specific sites on double-stranded DNA may provide tools for the regulation of gene expression. Excellent progress has been reported in *(1)* the design and selection of Cys_2His_2 zinc fingers for these purposes *(1, 2)* and 2) the design of pyrrole-imidazole polyamides that will bind according to well-defined "recognition rules" *(4)*. These two broad classes of compounds may provide new reagents for molecular medicine and gene therapy, and recent studies with zinc fingers specifically designed to turn on the VegF gene have shown that such designer transcription factors can work effectively in animal models *(40)*.

Our central goal in this paper was to explore how zinc fingers and polyamides might interact when targeted to overlapping recognition sites. We thought this would be interesting from a structural and physical/chemical perspective, and we thought this work also might provide a basis for the eventual development of new regulatory schemes (perhaps where polyamides were used to help regulate the binding of zinc finger proteins or visa versa). Our strategy in these studies was to (1) pick a set of zinc finger proteins that had been carefully characterized, (2) design polyamides that would target sequences that overlap the binding sites of these proteins, and (3) carefully characterize the DNA-binding affinity for each set of compounds under identical buffer conditions. We then used a set of biochemical studies (involving gel mobility shifts and DNase I footprinting experiments) to see what happened when zinc fingers and polyamides that recognized the same DNA site were mixed together.

Our central conclusion from these biochemical studies is that polyamides interfere with the binding of zinc finger proteins when the two compounds recognize overlapping, or partially overlapping, binding sites on the minor and major groove sides of the DNA, respectively. This is clearly demonstrated by interference experiments that use the gel mobility shift assay to monitor zinc finger binding (Figures 7 and 8), and this result holds for every combination of zinc fingers and polyamides that we have tested, when they recognize overlapping or partially overlapping binding sites. It also is clear that this "negative cooperativity" requires direct interactions of the polyamide with the zinc finger binding site: polyamides directed to other DNA sites show no interference with the binding of a given zinc finger protein (as indicated, for example, in the data of Figure 8B).

In an attempt to understand the structural and energetic basis for this "negative cooperativity", we have examined and compared crystal structures that are available for a set of

zinc finger-DNA and polyamide-DNA complexes. In every case, structures of the relevant zinc finger-DNA complexes show that the proteins bind in the major groove, while structures of polyamide-DNA complexes show that these compounds bind in the minor groove. There does not appear to be any basis for a van der Waals collision (or any other direct contact) between the protein and the polyamide when they bind to overlapping sites. However, structural comparisons do show striking differences between the DNA conformations in the polyamide-DNA complexes and those in the zinc finger-DNA complexes. Differences in the groove dimensions are quite clear: (1) The minor groove tends to be wider in the polyamide-DNA complexes than in the zinc finger-DNA complexes (Figure 10), while (2) the major groove tends to be wider (and often deeper) in the zinc finger-DNA complexes than in the polyamide-DNA complexes. Quantitative comparison of the DNA structures (using output from the program Curves *(38)*) reveals many other differences between zinc finger-DNA and polyamide-DNA complexes, and many of these differences will affect the precise position and orientation of key hydrogen bond donors and acceptors within the DNA site. At this stage, it seems very plausible that differences in the preferred DNA conformations could explain the "negative cooperativity" that we have observed when zinc fingers and polyamides are targeted to the same site. We assume that (1) the free DNA is somewhat "plastic" and that each compound can induce an appropriate conformation when it binds alone but that (2) simultaneous binding is difficult because the two types of compounds prefer somewhat different DNA conformations. (Note that if the conformational preferences of a given DNA sequence were more rigid, it might be hard to design polyamides for one class of sites and might be hard to design zinc fingers for another class of sites. This does not seem to be the primary problem.) In short, our modeling suggests that negative cooperativity may involve "allosteric" effects of changes in the DNA structure, and considering differences in

the relative dimensions of the major and minor groove (Figure 10) makes it easy to picture why these changes may be so important.

We undertook this study to investigate how polyamides and zinc fingers interact when they bind to overlapping sites on double-stranded DNA and show here an example of allosteric inhibition of a major-groove binding protein by unmodified polyamides. One could envision that designed zinc finger proteins could be displaced by small molecule polyamides, thereby providing both an on and off switch for gene regulation. Furthermore, phage selection technology may allow for the generation of artificial zinc finger proteins that only bind their target DNA sites with high affinity in the presence of a bound polyamide in the minor groove of the site. The results of this study give key insights into how a zinc finger/polyamide system might be designed to regulate gene expression.

ACKNOWLEDGMENTS

$TATA_{ZF}$, NRE_{ZF}, and $p53_{ZF}$ zinc finger peptides were overexpressed, purified, and submitted from Scot A. Wolfe, Robert Grant, and Sandra Fay-Richard, respectively. The pET-21d expression vector transformed in BL21(DE3) was submitted by Bryan S. Wang. We thank Ezra Peisach for a critical reading of this manuscript, and we thank members of the Pabo lab for their support and useful discussions throughout the experimental process and writing of this manuscript. C.M.T. was supported by a NSF predoctoral fellowship, J.K.J. was supported by a HHMI Physician Postdoctoral Fellowship, E.R. and C.O.P. were supported by the Howard Hughes Medical Institute, D.H.N. was supported by a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship, and P.B.D. was supported by NIH Grant GM 51747.

REFERENCES

1. Pabo, C. O., Peisach, E., and Grant, R. (2003) Design and Selection of Novel Cys_2His_2 Zinc Finger Proteins, *Annu. Rev. Biochem. 70*, 313-340.
2. Jamieson, A. C., Miller, J. C., and Pabo, C. O. (2003) Drug Discovery with Engineered Zinc Finger Proteins, *Nat. Rev. Drug Discovery 2*, 361-368.
3. Dervan, P. B. (2001) Molecular recognition of DNA by small molecules, *Bioorg. Med. Chem. 9*, 2215-2235.
4. Dervan, P. B., and Edelson, B. S. (1999) Recognition of the DNA minor groove by pyrrole-imidazole polyamides, *Curr. Opin. Struct. Biol. 13*, 284-299.
5. Wolfe, S. A., Nekludova, L., and Pabo, C. O. (2000) DNA Recognition by Cys_2His_2 Zinc Finger Proteins, *Annu. Rev. Biophys. Biomol. Struct. 29*, 183-212.
6. Elrod-Erickson, M., Rauld, M. A., Nekludova, L., and Pabo, C. O. (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions, *Structure 4*, 1171-1180.
7. Pavletich, N. P., and Pabo, C. O. (1991) Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å, *Science 252*, 809-817.
8. Rebar, E. J., and Pabo, C. O. (1994) Zinc Finger Phage: Affinity Selection of Fingers with New DNA-Binding Specificities, *Science 263*, 671-673.
9. Greisman, H. A., and Pabo, C. O. (1997) A General Strategy for Selecting High-Affinity Zinc Finger Proteins for Diverse DNA Target Sites, *Science 275*, 657-661.
10. Choo, Y., and Klug, A. (1994) Selection of DNA Binding Sites for Zinc Fingers Using Rationally Randomized DNA Reveals Coded Interactions, *Proc. Natl. Acad. Sci. U.S.A. 91*, 11168-11172.
11. Segal, D. J., Dreier, B., Beerli, R. R., and Barbas, C. F. (1999) Toward Controlling Gene Expression at Will: Selection and Design of Zinc Finger Domains Recognizing Each of the 5'-GNN-3' DNA Target Sequences, *Proc. Natl. Acad. Sci. U.S.A. 96*, 2758-2763.
12. DeClairac, R. P. L., Geierstanger, B. N., Mrksich, M., Dervan, P. B., and Wemmer, D. E. (1997) NMR Characterization of Hairpin Polyamide Complexes with the Minor Groove of DNA, *J. Am. Chem. Soc. 119*, 7909-7916.
13. Keilkopf, C. L., Baird, E. E., Dervan, P. B., and Rees, D. C. (1998) Structure of a photoactive rhodium complex intercalated into DNA, *Nat. Struct. Biol. 5*, 104-109.
14. Keilkopf, C. L., White, S., Szewczyk, Y. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (1998) A Structural Basis for Recognition of A·T and T·A Base Pairs in the Minor Groove of B-DNA, *Science 282*, 111-115.
15. Kielkopf, C. L., Bermer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (2000) Structural effects of DNA sequence on T·A recognition by hydroxypyrrole/pyrrole pairs in the minor groove, *J. Mol. Biol. 295*, 557-567.
16. Gottesfeld, J. M., Nealy, L. Traunger, J. W., Baird, E. E., and Dervan, P. B. (1997) Regulation of gene expression by small molecules, *Nature 387*, 202-205.
17. Dickinson, L. A., Gulizia, R. J., Traunger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) Inhibition of RNA Polymerase II Transcription in Human Cells by Synthetic DNA-Binding Ligands, *Proc. Natl. Acad. Sci. U.S.A. 95*,12890-12895.

18. Sluka, J. P., Horvath, S. J., Glasgow, A. C., Simon, M. I., and Dervan, P. B. (1990) Importance of minor-groove contacts for the recognition of DNA by the binding domain of Hin recombinase, *Biochemistry 29*, 6551-6561.
19. Dickinson, L. A., Trauger, J. W., Baird, E. E., Dervan, P. B., Graves, B. J., and Gottesfeld, J. M. (1999) Inhibition of Ets-1 DNA Binding and Ternary Complex Formation between Ets-1, NF-κB, and DNA by a Designed DNA-binding Ligand, *J. Biol. Chem. 274*, 12765-12773.
20. Nealy, L. Trauger, J. W., Baird, E. E., Dervan, P. B., and Gottesfeld, J. M. (1997) Importance of minor groove binding zinc fingers within the transcription factor IIIA-DNA complex, *J. Mol. Biol. 274*, 439-445.
21. Oakley, M. G., Mrksich, M., and Dervan, P. B. (1992) Evidence that a minor groove-binding peptide and a major groove-binding protein can simultaneously occupy a common site on DNA, *Biochemistry 31*, 10969-10975.
22. Bremer R. E., Baird, E. E., and Dervan, P. B. (1998) Inhibition of major-groove-binding proteins by pyrrole-imidazole polyamides with an Arg-Pro-Arg positive patch, *Chem. Biol. 5*, 119-133.
23. Winston, R. L., Ehley, J. A., Baird, E. E., Dervan, P. B., and Gottesfeld, J. M. (2000) Asymmetric DNA Binding by A Homodimeric bHLH Protein, *Biochemistry 39*, 9092-9098.
24. Wolfe, S. A., Greisman, H. A., Ramm, E. I., and Pabo, C. O. (1999) Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code, *J. Mol. Biol. 285*, 1917-1934.
25. Brenowitz, M., Senear, D. F., Shea, M. A., and Ackers, G. K. (1986) Quantitative DNase Footprint Titration - A Method for Studying Protein-DNA Interactions, *Methods Enzymol. 130*, 132-181.
26. Brenowitz, M., Senear, D. F., Shea, M. A., and Ackers, G. K. (1986) "Footprint" Titrations Yield Valid Thermodynamic Isotherms, *Proc. Natl. Acad. Sci. U.S.A. 83*, 8462-8466.
27. Senear, D. F., Brenowitz, M., Shea, M. A., and Ackers, G. K. (1986) Energetics of cooperative protein-DNA interactions: comparison between quantitative deoxyribonuclease footprint titration and filter binding, *Biochemistry 25*, 7344-7354.
28. Trauger, J. W., and Dervan, P. B., (2001) Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes, *Methods Enzymol. 340*, 450-466.
29. Elrod-Erickson, M., and Pabo, C. O. (1999) Binding Studies with Mutants of Zif268. Contribution of Individual Side Chains to Binding Affinity and Specificity in the Zif268 Zinc Finger-DNA Complex, *J. Biol. Chem. 274* (27), 19281-19285.
30. Elrod-Erickson, M., Benson, T. E., and Pabo, C. O. (1998) High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition, *Structure 6* (4), 451-464.
31. Baird, E. E., and Dervan, P. B. (1996) Solid Phase Synthesis of Polyamides Containing Imidazole and Pyrrole Amino Acids, *J. Am. Chem. Soc. 118*, 6141-6146.
32. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996) Recognition of DNA by designed ligands at subnanomolar concentrations, *Nature 382*, 559-561.
33. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1996) Recognition of a 5'-(A,T)GGG(A,T)2-3' Sequence in the Minor Groove of DNA by an Eight-Ring Hairpin Polyamide, *J. Am. Chem. Soc. 118*, 8198-8206.

34. Wolfe, S. A., Grant, R., and Pabo, C. O. (2001) Beyond the "Recognition Code": Structures of Two Cys2His2 Zinc Finger/TATA Box Complexes, *Structure 9*, 717-723.
35. Grant, R., Richards, S. F., and Pabo, C. O. (2001) Personal communication.
36. Pabo, C. O., and Nekludova, L. (2000) Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? *J. Mol. Biol. 301*, 597-624.
37. MSI, San Diego, CA.
38. Stofer, E., and Lavery, R. (1994) Measuring the Geometry of DNA Grooves, *Biopolymers 34*, 337-346.
39. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R. E. Structure of a B-DNA Dodecamer: Conformation and Dynamics, *Proc. Natl. Acad. Sci. U.S.A. 78*, 2179-2183.
40. Rebar, E. J., Huang, Y., Hickey, R., Nath, A. K., Meoli, D., Nath, S., Chen, B., Xu, L., Liang, Y., Jamieson, A. C., Zhang, L., Spratt, S. K., Case, C. C., Wolffe, A., Giordano, F. J. (2002) Induction of angiogenesis in a mouse model using engineered transcription factors, *Nat. Med. 8*, 1427-1432.

TABLES

Table 1: Equilibrium Dissociation Constants and Free Energies of DNA-Zinc Finger Protein Binding in the Absence and in the Presence of Polyamides

polyamide/protein	K_D (nM)[a]
$p53_{ZF}$	0.240 ± 0.018
$PA_{p53}/p53_{ZF}$	3.200 ± 0.283
TATAZF	0.204 ± 0.030
$PA_{TATA}/TATA_{ZF}$	4.320 ± 0.582
NRE_{ZF}	0.101 ± 0.009
PA_{NRE}/NRE_{ZF}	3.520 ± 0.248
Zif268	0.010 ± 0.003
PA_{Zif268}/Zif268	0.182 ± 0.009

[a] The reported equilibrium dissociation constants are apparent for the proteins in the presence of polyamides. All constants are the mean values obtained from three or more gel mobility shift experiments.

Table 2: Equilibrium Inhibition Constants

polyamide/protein	app K_I (nM)[a]
$PA_{p53}/p53_{ZF}$	0.20 ± 0.07
$PA_{TATA}/TATA_{ZF}$	0.04 ± 0.02
PA_{NRE}/NRE_{ZF}	16.40 ± 3.90
PA_{Zif268}/Zif268	0.03 ± 0.02

[a] The reported equilibrium inhibition constants are the mean values obtained from three gel mobility shift experiments.

FIGURES

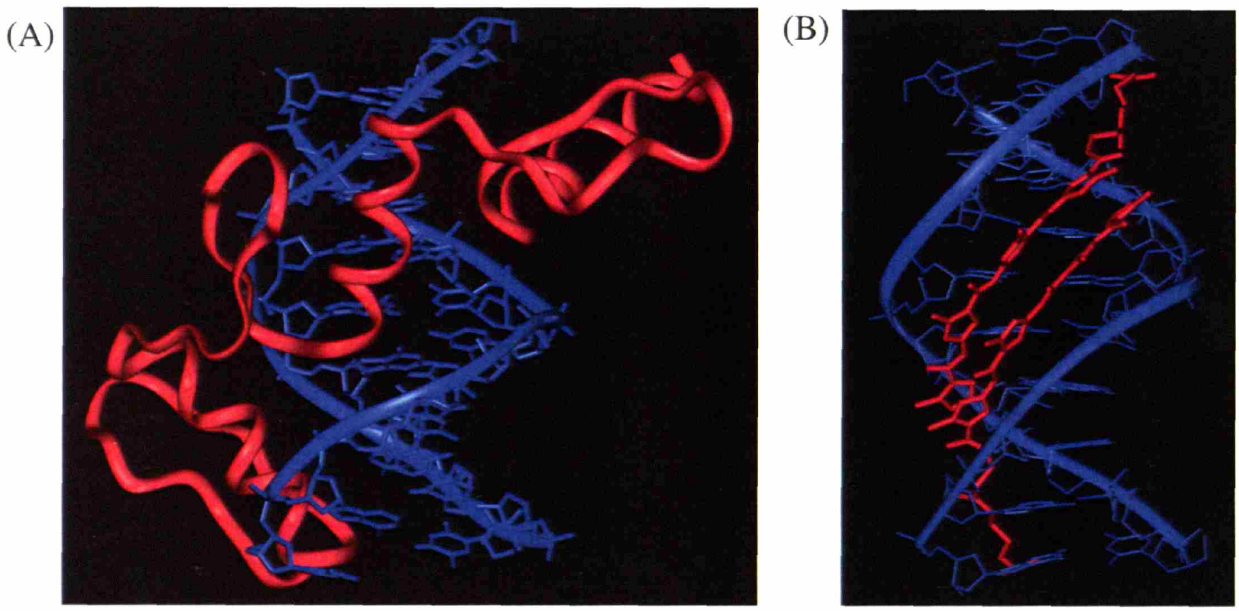


Figure 1. Crystal structure of Zif268 and polyamide bound to DNA. (A) Crystal structure of the Zif268 zinc finger protein (red, with ribbon representing protein backbone) bound in the major groove of B-DNA (blue, with ribbon connecting phosphates) [coordinates are from ref 6] and (B) crystal structure of a polyamide (red) bound in minor groove of B-DNA (blue) [coordinates are from ref *15*, 1CVY].

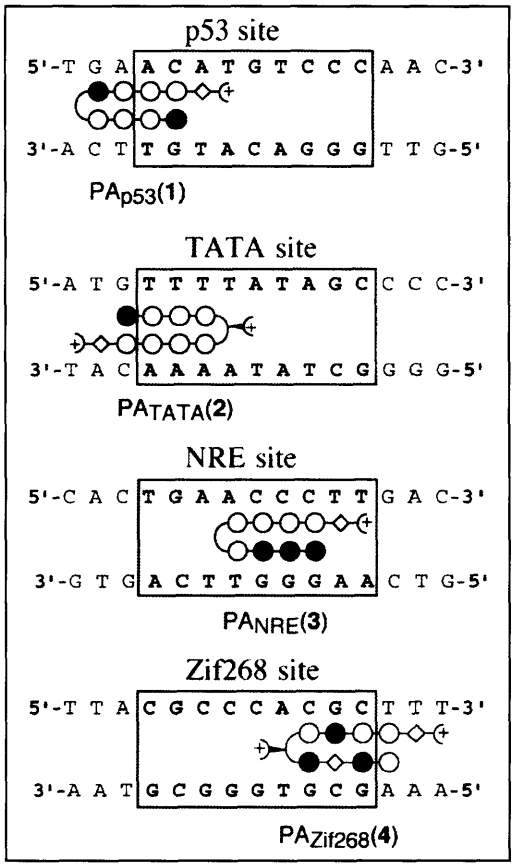
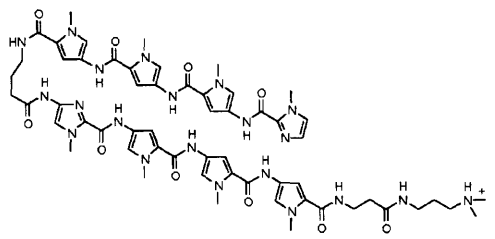
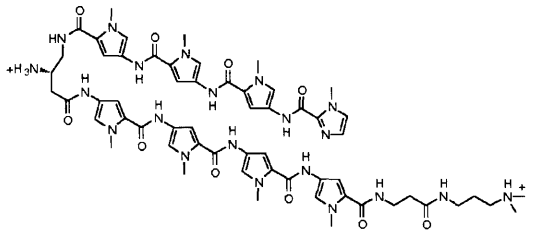


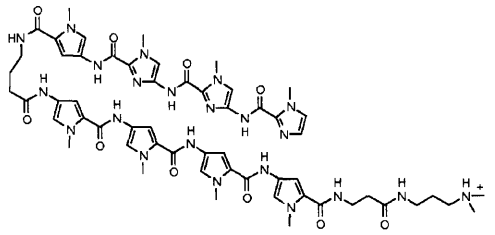
Figure 2. Minor-groove binding models expected for hairpin complexes of ImPyPyPy-γ-ImPyPyPy-β-Dp (**1**, PA_{p53}), ImPyPyPy-(*R*)$^{H}2^{N}$γ-PyPyPyPy-β-Dp (**2**, PA_{TATA}), ImImImPy-γ-PyPyPyPy-β-Dp (**3**, PA_{NRE}), and ImPyPyPy-(*R*)$^{H}2^{N}$γ-PyPyPyPy-β-Dp (**4**, PA_{Zif268}) targeted to their match sites in the minor groove opposite the $p53_{ZF}$, $TATA_{ZF}$, NRE_{ZF}, and $Zif268_{ZF}$ major-groove binding sites. Shaded and unshaded circles represent imidazole and pyrrole carboxamides, respectively, and the β-alanine residue is represented by an unshaded diamond. Boxes enclose the nine base pair sites that are recognized by the corresponding zinc finger proteins.



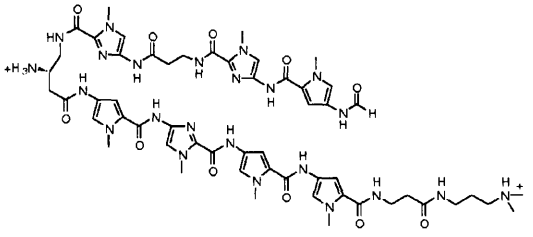
1 ImPyPyPy-γ-ImPyPyPyPy-β-Dp



2 ImPyPyPy-(R)^{H2N}γ-PyPyPyPyPy-β-Dp



3 ImImImPy-γ-PyPyPyPyPy-β-Dp



4 PyIm-β-Im-(R)^{H2N}γ-PyImPyPy-β-Dp

Figure 3. Full chemical structures of the hairpin polyamides: PA_{p53}, ImPyPyPy-γ-ImPyPyPyPy-β-Dp (**1**), PA_{TATA}, ImPyPyPy-(*R*)$^{H}2^{N}$γ-PyPyPyPyPy-β-Dp (**2**), PA_{NRE}, ImImImPy-γ-PyPyPyPyPy-β-Dp (**3**), and PA_{Zif268}, ImPyPyPy-(*R*)$^{H}2^{N}$γ-PyPyPyPyPy-β-Dp (**4**).

Figure 4. Oligonucleotide sequences used for DNase I footprinting experiments, with each of the nine base pair target sites for the zinc finger proteins underlined. Oligonucleotide 1 contains the $TATA_{ZF}$, NRE_{ZF}, and $p53_{ZF}$ binding sites; oligonucleotide 2 contains the Zif268, $TATA_{ZF}$, and $p53_{ZF}$ binding sites. Gel mobility assays used shorter oligonucleotides (27 bp segments each containing a single binding site), and the corresponding segments used in these experiments are shown as boxed regions surrounding each of the nine base pair binding sites. In this orientation, the site that has been cleaved by *Eco*RI is at the left end of the oligonucleotide; the site cleaved by *Not*I is on the right.

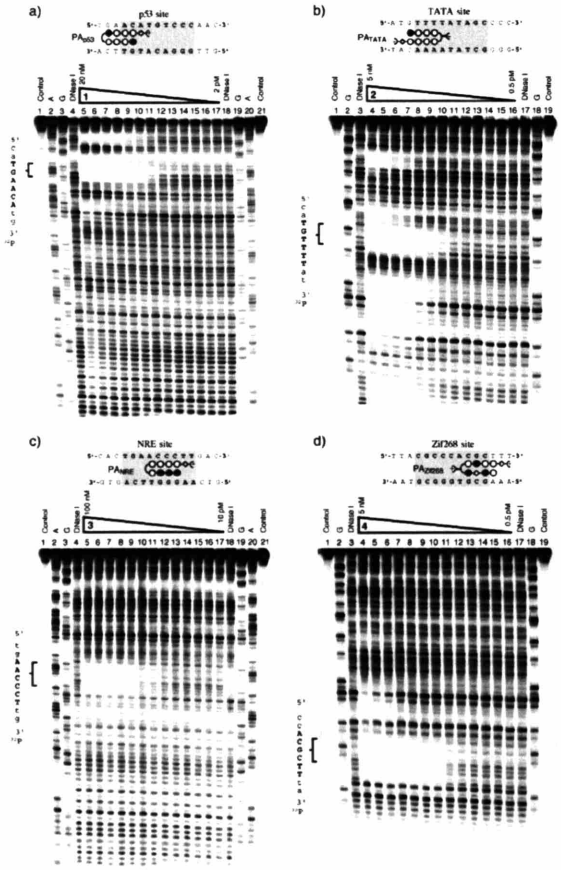


Figure 5. Quantitative DNase I footprint titration experiments used to monitor binding of polyamides to the oligonucleotides shown in Figure 4. Oligonucleotides were obtained as EcoRI/NotI restriction fragments from the pBluescript II SK+ plasmid and were 3'-^{32}P-labeled. Panel a contains PA_{p53} (1) on oligonucleotide 1: lanes 1 and 21, intact DNA; lanes 2 and 20, A-specific chemical sequencing reaction; lanes 3 and 19, G-specific chemical sequencing reaction; lanes 4 and 18, DNase I digestion products in the absence of polyamide; lanes 5-17, DNase I digestion products in the presence of 20, 10, 5, 2, 1, 0.5, 0.2, and 0.1 nM and 50, 20, 10, 5, and 2 pM polyamide, respectively. Panel b contains PA_{TATA} (2) on oligonucleotide 2: lanes 1 and 19, intact DNA; lanes 2 and 18, G-specific chemical sequencing reaction; lanes 3 and 17, DNase I digestion products in the absence of polyamide; lanes 4-16, DNase I digestion products in the presence of 5, 2, 1, 0.5, 0.2, and 0.1 nM and 50, 20, 10, 5, 2, 1, and 0.5 pM polyamide, respectively. Panel c contains PA_{NRE} (3) on oligonucleotide 1: lanes 1 and 21, intact DNA; lanes 2 and 20, A-specific chemical sequencing reaction; lanes 3 and 19, G-specific chemical sequencing reaction; lanes 4 and 18, DNase I digestion products in the absence of polyamide; lanes 5-17, DNase I digestion products in the presence of 100, 50, 20, 10, 5, 2, 1, 0.5, 0.2, and 0.1 nM and 50, 20, and 10 pM polyamide, respectively. Panel d contains PA_{Zif268} (4) on oligonucleotide 2: lanes 1 and 19, intact DNA; lanes 2 and 18, G-specific chemical sequencing reaction; lanes 3 and 17, DNase I digestion products in the absence of polyamide; lanes 4-16, DNase I digestion products in the presence of 5, 2, 1, 0.5, 0.2, and 0.1 nM and 50, 20, 10, 5, 2, 1, and 0.5 pM polyamide, respectively.

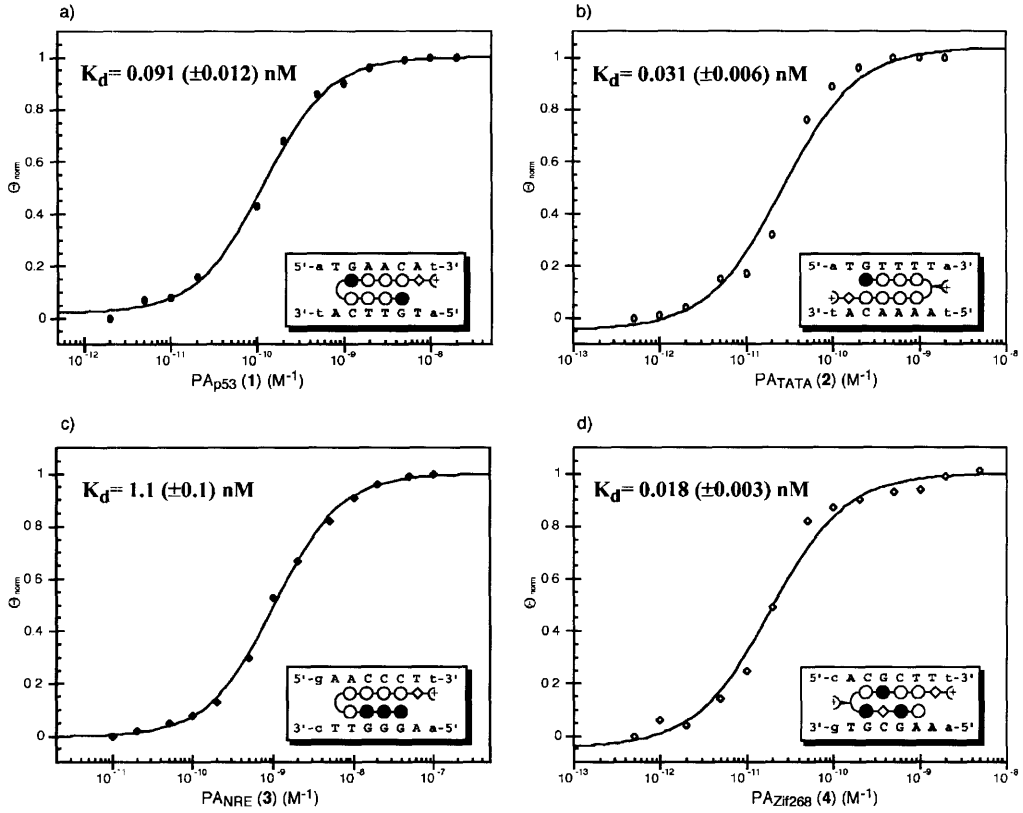


Figure 6. Binding isotherms derived from the DNase I quantitative footprint titration experiments (Figure 5) as shown for the PA_{p53} (a), PA_{TATA} (b), PA_{NRE} (c), and PA_{Zif268} (d) polyamides. In each case, isotherms represent binding at the match target site in the minor groove, which has the sequence shown in the lower right panel of the figure and which overlaps with the nine base pair site in the major groove recognized by the corresponding zinc finger protein. Θ_{norm} points were obtained using storage phosphor autoradiography and processed as previously described *(28)*. Each data point represents the average of three independent quantitative footprint titration experiments, and the solid curves are the best-fit Langmuir binding titration isotherms obtained from a nonlinear least-squares algorithm where n = 1.

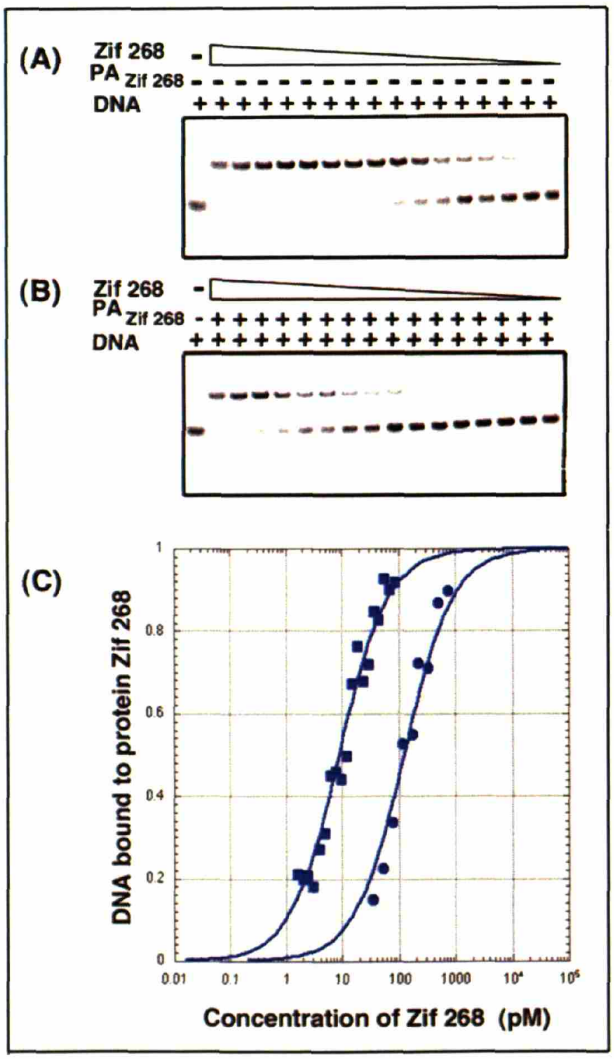


Figure 7. Gel mobility shift experiment showing that the PA_{Zif268} polyamide interferes with binding of Zif268 protein. Panel A presents binding of Zif268 protein in the absence of polyamide, where upper band on the gel represents the protein-DNA complex. Protein was diluted 1.5× between lanes, and the 27 bp DNA probe had a concentration of 0.5 pM. Panel B presents a similar binding experiment to that in panel A, but this was performed in the presence of a constant amount of polyamide. The concentration was chosen as $10\,K_D^{PA}$ and thus is expected to give about 90% saturation of the site in the absence of protein. Panel C presents the Zif268 binding isotherms in the absence (■) and in the presence (·) of PA_{Zif268}. Each data point represents the average of the three independent gel mobility shift experiments. The binding curves show that a protein concentration of 10 pM gives 50% occupancy when no polyamide is present; a protein concentration of 182 pM is required to give 50% occupancy when the polyamide is present.

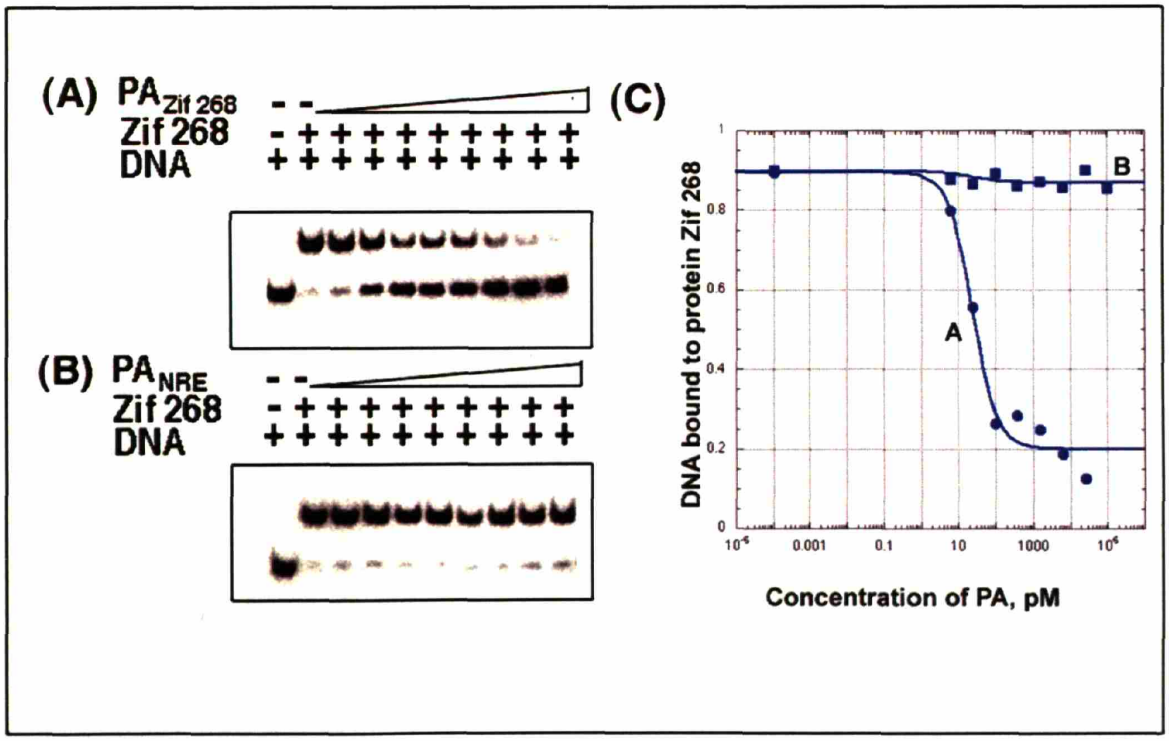


Figure 8. Specificity of polyamide interference effects. In these experiments, the active concentration of protein was constant and equal to $10\,K_D^{prot}$ (giving ~90% of saturation of DNA binding sites when no polyamide was present). The DNA probe was present at 0.5 pM, and the polyamide concentration varies 4-fold between adjacent lanes. Panel A presents representative gel mobility shift results for Zif268 protein binding to DNA probe in the presence of increasing concentrations of PA_{Zif268} polyamide. Panel B presents the corresponding experiments using the Zif268 protein, the Zif268 binding site, and the polyamide PA_{NRE} (note that this polyamide will not bind to the site and thus represents a "mismatch" control for the purposes of this experiment). Panel C presents the effects of polyamides PA_{Zif268} (·) and PA_{NRE} (■) on Zif268 protein binding. Each data point represents the average of three independent experiments, and the difference between these curves emphasizes the specificity of polyamide binding and interference.

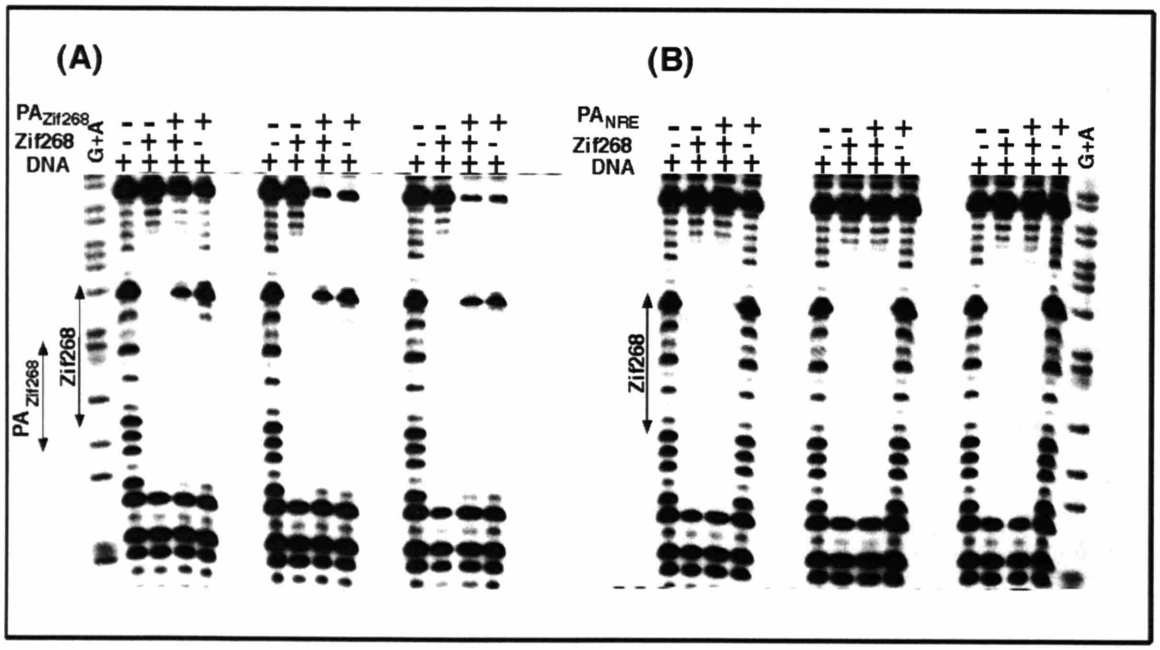


Figure 9. DNase I footprinting analysis of polyamide/zinc finger protein interference effects. Panel A presents reactions using Zif268 protein, Zif268 DNA site, and PA_{Zif268} polyamide. The first lane within each group of four has no protein and no polyamide; the second lane has protein but no polyamide; the third lane has protein and polyamide; the forth lane has polyamide alone. When present, the concentration of Zif268 protein was equal to $10K_D$; concentrations of polyamides were 1 nM in the first set of reactions (i.e., the first four lanes), 5 nM in the second set of reactions, and 10 nM in the third set of reactions. Lanes labeled G+A show markers prepared with the Gilbert-Maxam G+A reaction protocol, and binding sites for the Zif268 protein and PA_{Zif268} polyamide are marked on the left side of the gel. Panel B presents the control experiment (with a similar arrangement of lanes and choice of concentration) using the Zif268 protein, Zif268 DNA, and PA_{NRE} polyamide (note that the PA_{NRE} polyamide does not have a binding site on this oligonucleotide and thus serves as a "mismatch" control for these experiments).

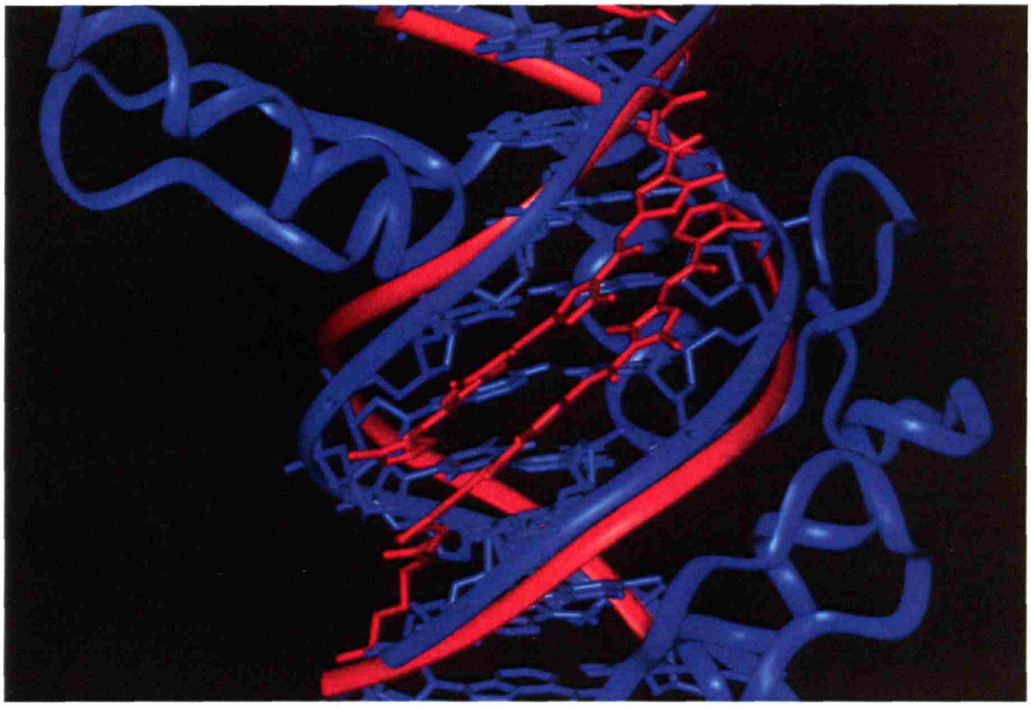


Figure 10. Superposition of zinc finger-DNA complex [coordinates are from ref *34*] (blue) and polyamide-DNA complex [coordinates are from ref *13*] (red) oriented to highlight differences in minor groove width. Such "allosteric" changes in the groove dimensions may explain the "negative cooperativity" observed in our binding studies with polyamides and zinc fingers.