The Design and Structural Characterization of Oligomeric Beta Beta Alpha Mini-Proteins

by

Mayssam H. Ali

A.B. Chemistry
Harvard College, 1998

SUBMITTED TO THE DEPARTMENT OF CHEMISTRY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2004

Signature of Author:_____
Department of Chemistry
May 6, 2004

Certified by:_____
Barbara Imperiali
Ellen Swallow Richards Professor of Chemistry and Professor of Biology
Thesis Supervisor

Accepted by:_____
Robert W. Field
Haslam and Dewey Professor of Chemistry
Chairman, Departmental Committee on Graduate Student

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Sarah E. O'Connor
Assistant Professor of Chemistry
Massachusetts Institute of Technology

Barbara Imperiali
Ellen Swallow Richards Professor of Chemistry and Professor of Biology
Massachusetts Institute of Technology
Thesis Supervisor

Karen N. Allen
Professor of Physiology and Biophysics
Boston University School of Medicine

Amy E. Keating
Assistant Professor of Biology
Massachusetts Institute of Technology

The Design and Structural Characterization of Oligomeric Beta Beta Alpha Mini-Proteins

by

Mayssam H. Ali

Submitted to the Department of Chemistry
on May 6, 2004 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Chemistry

ABSTRACT

Oligomeric mini-proteins, short peptides with protein-like features, constitute valuable minimal models for the study of oligomeric proteins. Oligomerization is a common feature of cellular proteins that may confer structural and functional advantages. Oligomerization is proposed to have arisen by several evolutionary pathways.

The structural characterization of peptide **1**, a stable oligomeric mini-protein previously developed in the Imperiali group, was undertaken by X-ray crystallography, as knowledge of the structure would enable the rational design of subsequent generations of BBA oligomers varying in packing and stoichiometry. The structure of peptide **1** could not be solved by direct methods, by molecular replacement with search models derived from the monomeric precursor, or by the introduction of heavy atoms.

Two selenomethionine mutants having solution-phase properties comparable to the native were identified. The structures of these two peptides were independently solved via MAD phasing experiments, and the refined structures employed as search models for a molecular replacement solution of **1**. The structures of the three peptides are homologous, and constitute the first reported structures of a mixed α/β oligomeric mini-protein.

The X-ray crystal structures reveal that the oligomeric BBA motif has a domain-swapped architecture that supports a protein-like and water-exclusive core. The structures elucidate the unique role of unnatural amino acids in conferring native secondary structure in a short peptide sequence (21 residues per monomer). Furthermore, the crystal structures reveal that the stoichiometry of the oligomer is tetrameric, rather than trimeric, as originally proposed. A tetrameric solution-phase stoichiometry for this mini-protein family was confirmed by rigorous analytical ultracentrifugation experiments.

Heterooligomeric BBA peptides have been designed and characterized in collaboration with Christina M. Taylor and Professor Amy E. Keating of the MIT Biology Department. Acidic and basic residues were substituted along the inter-monomer interface and specific steric interactions were designed in order to disfavor homoassociation and favor heteroassociation. Heterotetramers comparable to peptide **1** in terms of structure and stoichiometry, and approaching the native homotetramer in terms of stability, have been characterized by a variety of biophysical techniques.

Thesis Supervisor: Barbara Imperiali
Title: Ellen Swallow Richards Professor of Chemistry and Professor of Biology

Acknowledgements

I would like to thank my advisor, Professor Barbara Imperiali for the excellent scientific training that I have received, and for all of her help and support over the years. Barbara has been a great source of inspiration to me. I continue to be amazed by the depth of her knowledge and the diversity of her scientific interests. I also thank her for giving me such an interesting project and the freedom to take it in new directions.

My colleagues in the Imperiali group are friendly, knowledgeable, creative and generous, and I am grateful to them for good company, helpful discussions, and experimental assistance. I thank Professor Sarah E. O'Connor for serving as my committee chair. Adam Mezo and Kevin McDonnell initiated the oligomeric BBA project that has been the focus of my graduate career. I have learned a great deal from working with them. I thank Carlos Bosques and Jen Ottesen for good company during late nights in the lab, as well as Dierdre Pearce and Vincent Tai for helpful suggestions during my early days in the group. I am grateful to Jebrell Glover for driving me and my densitometry sample to the University of Connecticut (twice!), and for many helpful discussions about analytical ultracentrifugation. Beth Vogel and Eranthie Weerapana have come to my rescue on innumerable cranky HPLC days. Seungjib Choi has a great sense of humor and an admirable work ethic. I thank Melissa Shults and Elvedin Lukovic for watching over my plants when I was away. I am grateful to Melissa Shults and Eugenio Vázquez for computer assistance. I thank Debbie Rothman, baker extraordinaire, for sharing her creations with the group. I thank Mary O'Reilly for stimulating literary and artistic conversations. Kathy Franz provided many helpful suggestions. I thank Mark Nitz for assistance with the 9050. Soonsil Hyun's quiet courage has been an example to us all. Christina Carrigan and Rob Dempski have inspired me with their athleticism. During my last few weeks of experiments Guofeng Zhang generously shared his AUC time with me, and Debby Pheasant helped me to set up a great many AUC cells. I regularly

come across Maria Ufret and Christian Hackenberger on the T, and their company invariably makes my day better. I wish the first-year graduate students in the group, Elvedin Lukovic, Galen Loving, and Langdon Martin, the best of luck in their graduate careers and beyond. I will miss you all.

Professor Karen N. Allen has been a second advisor to me, and an admired role model. I thank her for welcoming me into her group. Karen taught me crystallography and made this work possible. Karen's unfailingly encouragement and enthusiasm during the most scientifically frustrating periods of this project are gratefully acknowledged, as are her scientific insights.

I am fortunate to have had a second group of fun and talented colleagues at Boston University. I thank all of my colleagues in the Allen group for their friendship and for fruitful scientific discussions. Sushmita Lahiri and Andy Mazurkie taught me the basics of UNIX and of crystallographic data collection. Ezra Peisach has helped me with data collection, data analysis, theoretical discussions, and almost all aspects of my project. He has been an exceptional teacher and mentor. In addition to many helpful suggestions, Tracy Arakaki has been a regular companion on cookie excursions. I thank Nick Silvaggi for patiently instructing me in the fine arts of SHELXL and XFIT. Ali Malay, an adjunct group member like myself, was my companion during many late nights in the lab. Joseph Ho has lent me his ethanol wash bottle more times than I can count. I wish Manashi Sherawat and Lee Tremblay the best of luck in their qualifying exams next month (as well as in their post-quals research and in life beyond). I wish Bill Lu well with his many refinements. I will miss you all.

I thank Professor Amy E. Keating for undertaking the heterooligomeric BBA design project, and for many insightful comments and suggestions. It has been a privilege to work with her, and I have learned a great deal from her sharp mind and encyclopedic knowledge. I thank Christy Taylor for a very fruitful and enjoyable collaboration. Christy undertook all the computational aspects of the heterooligomeric BBA design, as well as the repacking of the

homooligomer. I am also grateful to Christy for experimental assistance, particularly the use of the Keating Lab autosampler.

I thank Professor Stephen L. Buchwald for his guidance and support during my first two years at MIT, and for giving me a solid foundation in organic chemistry. I also acknowledge all of my former colleagues from the Buchwald group. I thank Professor Timothy M. Swager for helpful advice and encouragement over the years. I am grateful to Professor Catherine L. Drennan and Dr. Geoffrey F. Stamper for their help in the early stages of this project.

I am indebted to a number of people for assistance with the planning and analysis of analytical ultracentrifugation experiments: Jeffrey W. Lary and Professor James L. Cole of the University of Connecticut National Analytical Ultracentrifugation Facility, Dr. Peter Schuck of the National Institutes of Health, Dr. Walter F. Stafford of the Boston Biomedical Research Institute, and Dr. James D. Lear of the University of Pennsylvania all provided many helpful suggestions. Rick Cole has answered many instrumentation questions and offered helpful advice. I thank Professor George M. Sheldrick of the University of Göttingen for his assistance with substructure searches using SHELXD and during the refinement of peptide **1**. I thank Edgar Ortiz for assistance with the R-Axis II and R-Axis IV.

I am grateful to a number of friends who have shared and enlivened my graduate experience. I thank Cyril Yee, Debbie Perlstein, Michelle Chang, Chris Chang, Jennie Ge, Kathryn Simmons, and Amanda Yarnell for their friendship and camaraderie from the first days of graduate school. My friends Shana Sturla, Maureen Fagan, Jacquin Niles, Dana Buske, and Mike Schneider have offered me encouragement and good advice throughout graduate school. I thank all the staff and volunteers at the MIT Sailing Pavilion, particularly Dwight Brown. I also thank Elizabeth Fong for years of great sailing. My friends Davide Marini, Chris Carr, Neda Vukmirovic, and Amr Farahat have shared sailing and other adventures. Sulafa Bashir and Zeina Talje make everything more fun. I thank Alexandra Tully, my oldest friend, and Siobhan Cunningham for keeping me in touch with life outside the ivory tower. I acknowledge my

childhood friends Ramzi Lebbos, Rafat Faris, Zayd Faris, and Sammy Haddad. I thank Vaso Tsakraklides, Katerina Linos, Alexi Zervos, Ari Yannopoulos, Steve Margetis, Dan Bernard, and Lise Adams for over a decade of friendship. I hope there will be many more to come. Vaso has gone through all of the ups and downs of graduate school with me, and has been a source of strength throughout. Years later, we still live parallel lives. I am very fortunate to have had Katerina, an all-round wonderful person and a true friend, here with me in Cambridge for so long. I thank Waleed Farahat for his love and support, for encouraging and challenging me, and for his silly sense of humor. I also thank him for looking after me while during the writing of this thesis, and for deciphering the intricacies of Microsoft Office.

My cousins Wael Hibri and Luz Lenis have been my friends, role models, cheerleaders, and advisers. I thank Tante Safa and Tante Anas for their love and constant encouragement. My brothers, Jamil and Ziyad, are two of my best friends. They are clever, funny, charming, thoughtful, and interesting. Most of all, I thank my parents, Hani and Raja Ali for their unconditional love, support, and generosity. My parents have encouraged me and inspired me throughout my life. I dedicate this thesis to them.

Table of Contents

List of Figures

Chapter 4

Chapter 5

Appendix

List of Tables

Appendix

List of Abbreviations

Standard one- and three-letter abbreviations were used for the 20 common amino acids. Amino acids with D chirality are denoted by the prefix D- before the three letter code, or with the single letter code in lower case, e.g. D-alanine is D-Ala or a.

2D	two-dimensional
Abu	L-α-aminobutyric acid
Abz	*o*-amino benzoic acid, anthranilic acid
Ac	acetyl
ANS	1-anilino-8-naphthalene sulfonate
AUC	analytical ultracentrifugation
ASU	asymmetric unit
BBA, ββα	beta beta alpha
Boc	*tert*-butoxycarbonyl
CD	circular dichroism
DapAbz, Daz	*o*-aminobenzoylated L-α,β-diaminopropionic acid
DapBz, Dbz	benzoylated L-α,β-diaminopropionic acid
DCM	dichloromethane
DTT	1,4-dithiothreitol, 1,4-dithio-DL-2,3-butanediol
DIPEA	*N,N*-diisopropylethylamine
DMF	*N,N*-dimethylformamide
Fmoc	fluoren-9-ylmethoxycarbonyl
FMS oil	poly-3,3,3-trifluoropropylmethylsiloxane
HBtU	2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOBt	1-hydroxybenzotriazole
HPLC	high performance liquid chromatography
MAD	multiwavelength anomalous diffraction
MIR	multiple isomorophous replacement
MIRAS	multiple isomorphous replacement with anomalous scattering
MLF	maximum likelihood target with amplitudes
MLHL	maximum likelihood target with amplitudes and Hendrickson-Lattman coefficients
MMT	p-methoxytrityl
MR	molecular replacement
NCS	non-crystallographic symmetry
NMR	nuclear magnetic resonance
NO_2-Y, NO_2-Tyr	3-nitrotyrosine
NOE	nuclear Overhauser effect
PEG	polyethylene glycol
SeMet	selenomethionine
SeMet[O]	methionine selenoxide
SIRAS	single isomorphous replacement with anomalous scattering
TFA	trifluoroacetic acid
TRIS	tris(hydroxymethyl)aminomethane

XAFS X-ray absorption fine-structure spectroscopy

Chapter 1 : Protein and Mini-Protein Oligomerization

Introduction

Oligomeric proteins, proteins composed of two or more associating polypeptide chains, comprise a sizable fraction of cellular proteins. The broad category of oligomeric proteins can be divided by subunit type, strength of subunit association, and duration of subunit association. Oligomerization may be an advantageous feature from the perspective of protein evolution for a number of reasons, including new opportunities for functional control, e.g. allosteric regulation. Many early, primitive, proteins may have been homooliogmeric or heterooligomeric in order to better support function; a study of the nature of protein oligomerization may elucidate features of protein evolution. Inter-subunit interfaces share common features with both hydrophobic protein cores and polar protein surfaces. Both hydrophobic and polar interactions are important in defining oligomeric interfaces. Protein oligomerization has probably evolved by a variety of mechanisms. Some possible mechanisms include the direct evolutionary selection for a protein oligomer, the development of a new interaction site, an exchange of association partners in a multi-domain protein, and the fusion of a monomeric protein with an existing oligomerization domain. Oligomeric mini-proteins, short peptides with discrete protein-like structures, serve as valuable models for the study of oligomeric proteins.

Characteristics of Oligomeric Proteins

Oligomeric proteins are proteins composed of multiple subunits (polypeptide chains). These subunits may be the same (a homooligomeric protein) or different (a heterooligomeric protein). The association between subunits can vary in strength and duration.

Oligomeric proteins abound in nature. It has been suggested that the average oligomeric state of cellular proteins is roughly tetrameric.[1] A recent survey suggests that 35% or more of the

proteins in a cell are oligomeric.[2] The proportion of oligomeric proteins deposited in the Protein Data Bank is significantly lower,[3] which may simply reflect experimental constraints favoring the structure determination of small monomeric proteins. Most oligomeric proteins are homooligomers.[2]

The majority of oligomeric proteins are dimeric or tetrameric. Higher-order oligomers are less prevalent[2,3] and a relatively small fraction of oligomeric structures have odd-numbered stoichiometries. Most oligomeric proteins, and essentially all homooligomeric proteins, are symmetrical. This symmetry is most frequently cyclic, dihedral, or cubic.[2]

There are several functional advantages that may be conferred by oligomerization.[4-7] (1) More complex scaffolds may better support function, e.g. by the introduction of a new active site at the interface between subunits. It has been estimated that roughly one sixth of oligomeric enzymes have an active site located at the inter-subunit interface. (2) Oligomeric proteins can be allosterically regulated, introducing an additional level of control. (3) It is more genomically efficient to produce a large protein by encoding a shorter, oligomerizing, protein. (4) There is a greater likelihood of an error-free transcript in a shorter protein sequence. A large protein composed of multiple, short, subunits, is more likely to be synthesized without errors than a single-chain protein of comparable size. (5) Where the monomer and oligomer differ in activity, additional regulatory flexibility may be achieved by regulating the conditions of oligomerization. (6) Oligomeric proteins may be subject to amplified evolutionary pressures, as deleterious mutations may be more pronounced and thus removed sooner from the gene pool. Conversely, the advantages of beneficial mutations may also be made evident sooner. (7) Larger proteins are more resistant to degradation and denaturation. Indeed, an increase in oligomerization state is one of the protein stabilization strategies adopted by thermophilic organisms.[8]

In certain cases oligomerization may be incidental to protein activity, and neither selected for, nor against, by evolutionary pressures.[4-6] Moreover, there are cases where oligomerization would be disadvantageous. A protein with a large, hence slowly-diffusing, substrate, may be

subject to evolutionary pressures to be small, hence rapidly-diffusing, so as to accelerate rates of collision with the cognate substrate and thus catalysis.[6]

A protein may homooligomerize with another molecule of the same protein, or heterooligomerize with a different protein. Some proteins are found only, or primarily, in their oligomeric state. These proteins generally have dissociation constants in the nM range. Others have a weak tendency to associate, with oligomerization dependent on environmental conditions such as concentration, temperature, and pH. Such proteins often have Kd values in the μM or even mM range. Still other proteins oligomerize dynamically in response to a stimulus such as a change in nucleotide binding, nucleotide hydrolysis or phosphorylation state. Such a change can have a dramatic effect on the affinity of the subunits for one another, often by orders of magnitude.[5,6]

Monod and coworkers have characterized homooligomeric proteins by their mode of interaction (Figure 1-1).[9] A homodimeric protein can only have an isologous inter-subunit interaction between the same surface on each monomer, giving rise to a dimer with twofold symmetry. A heterologous interaction, with two complementary sites, is possible for higher oligomers. Such an interaction can give rise to a discrete oligomer, if the interaction is circularly symmetrical, or can lead to indefinite self-association if it is not.[9] These classifications continue to provide a valuable framework for thinking about protein interfaces, but may not adequately encompass oligomeric interactions such as domain-swapping.[2]

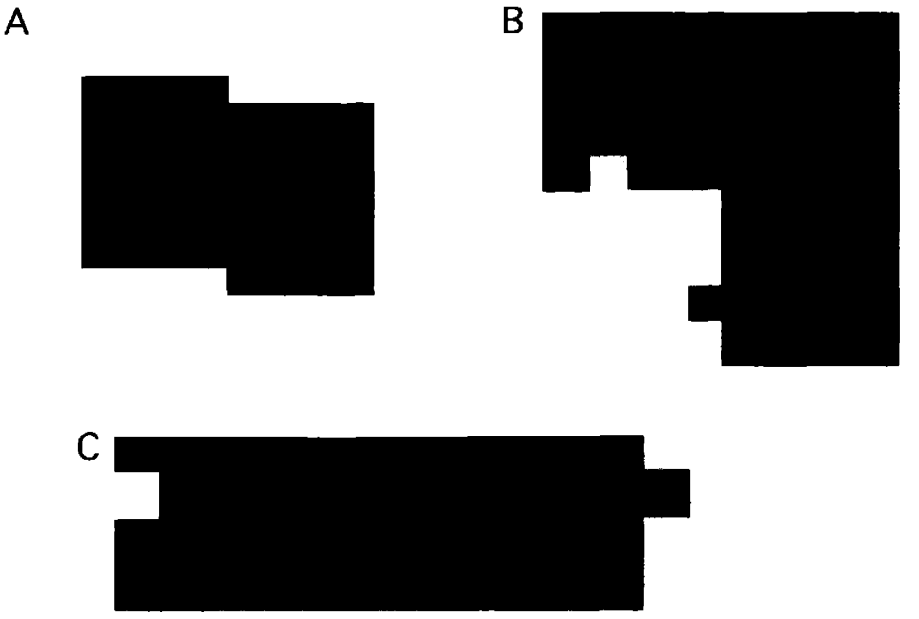


Figure 1-1. A. Isologous dimer. B. Heterologous tetramer. C. Heterologous polymer. Figure adapted from reference 9.

Why Study Oligomerization?

Many proteins that are monomeric in their modern or evolved state are believed to have arisen from smaller, associating, fragments, that have, via gene duplication and gene fusion, become a single encoded protein.[10] Some examples of monomeric proteins believed to have derived from homooligomeric precursors are the periplasmic binding proteins,[11,12] the eightfold β/α barrel,[13] and the ribosome anti-association factor eIF6.[14]

It is not essential that the merged genes be homologous, as in the above examples. The fusion of two different genes would result in an asymmetric protein. Each of the original associating fragments may comprise a domain of the fusion protein. Chothia found that approximately two of three prokaryotic proteins, and about 80% of eukaryotic proteins, are composed of at least two domains.[15] It is likely that many of these multi-domain proteins arose from the genetic fusion of smaller, associating single-domain proteins.

The study of oligomerization may also have relevance to the evolution of single protein domains. Russell and coworkers have proposed that many protein domains arose from antecedent domain segments (ADSs).[16] ADSs are peptide sequences, encoded by short primordial genes, that spontaneously homooligomerize in aqueous solution to adopt a fold similar to a modern domain. These non-covalent, single domain, mini-proteins would have served structural or functional roles in the early protein world. Over time, evolutionary pressures would have favored the development of a single fused gene comprised of multiple ADS repeats because of entropic and thermodynamic factors. Support for this hypothesis comes from the study of several domains exhibiting internal symmetry, with structural and sequence repeats.[16] Among these are the beta trefoil, the beta propeller, and the beta spiral.[17] Furthermore, the ability of proteolytic fragments of several independently folded domains and small proteins to reassemble into non-covalent assemblies resembling the native suggests that that may have been their evolutionary pathway.[16,18-20]

An ADS could also have formed a heterooligomer with a different ADS. Such a heterooligomer could then, following gene duplication and fusion, form a more complex single domain. Russell and coworkers propose that the protein thioredoxin (Figure 1-2) could have been one such example, formed from the combination of two $\beta\beta\alpha$ ADSs and one $\beta\alpha$ ADS.[16]

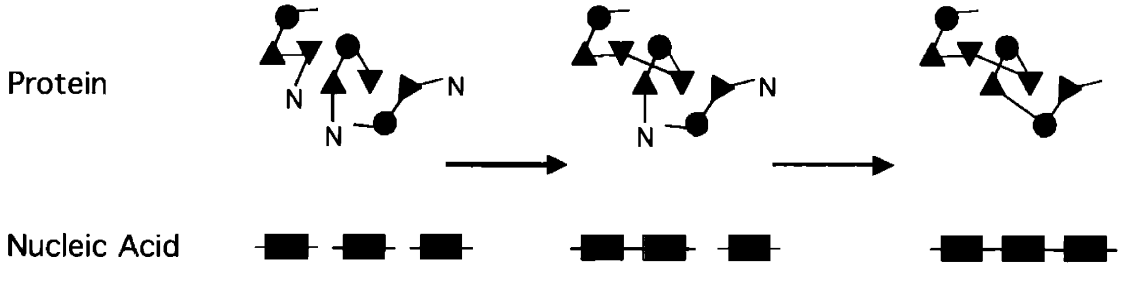


Figure 1-2. The proposed evolution of thioredoxin from three ADSs. Figure adapted from reference 16.

Knowles and coworkers have proposed a similar mechanism for the evolution of modern proteins from heterooligomeric peptides in the context of intron-early hypotheses of protein evolution.[21] In an early world, short peptide sequences, analogous to the ADSs proposed by Russell, could have been encoded by "microgenes" terminating with the amber stop codon. Some of these peptides could spontaneously associate into functionally active multi-chain protein assemblies. The terminal amber codon would serve as a mechanism for recognition of the microgenes to be spliced. It would be evolutionarily favorable for microgenes encoding a pre-protein to be spliced together, as co-expression of the peptides would improve the odds that the activity of the associating peptides would be expressed. The genes resulting from such an event would be passed on. In the modern world, these microgenes would become exons.

Although there have been some convincing arguments against the intron-early hypotheses of protein evolution,[22,23] it remains likely that whatever the genetic origins, modern proteins arose from much simpler precursors. It is also highly probable that the mini-protein precursors associated in solution to better support functionality, and that the genes encoding associating proteins, over time, were selected to be fused or coexpressed.

Studies of designed mini-proteins have revealed that it is difficult to introduce functionality into a single peptide, such as an ADS, even if the structures are well-folded.[24,25] Apparently a larger scaffold is preferable from the vantage of substrate binding and catalysis, allowing for more residues to be devoted to function rather than structure. By contrast, there has been considerable success in introducing functionality into oligomeric mini-proteins.[25-29] We could thus expect significant evolutionary pressure on an ADS or microgene product to homo- or hetero-oligomerize.

How Did Oligomerization Arise?

The study of the evolution of protein oligomerization is an active field of scientific discourse. There are several mechanisms whereby oligomerization could have arisen. It is likely

that there is not one general mechanism, but several mechanisms, and that oligomerization in different proteins has evolved by different routes. Furthermore, it is not necessarily the case that protein evolution proceeds invariably in the direction of oligomerization

Genetically, oligomeric species could be created from monomeric proteins through some combination of a few genetic events.[5,15,30] Substitutions, insertions, and deletions comprise the basis set of mutations that, over time, will result in sequence drifting. Such mutations could affect the electrostatic or geometrical properties of a surface patch, or the length or geometry of an inter-domain linker, resulting in a change in the oligomeric state of a once monomeric protein. Recombination could result in the fusion of an oligomerization domain to a previously monomeric protein. Some of these mutations would be beneficial and selected for, some detrimental and selected against, and some neutral.

Mutations can lead to the introduction, expansion, or exposure of a hydrophobic patch on the surface of a protein, thus creating a new interface for oligomerization (Figure 1-3). As all residues do not contribute equally to the oligomeric interface, it is reasonable to suppose that mutations at a few key residues necessary for mediation of an interaction could have such an effect. It has been proposed that relatively few mutations would be necessary for such a step, particularly in the case of dimeric proteins that fold via a three-step pathway.[31,32] Moreover, the close resemblance of subunit interfaces to surface residues may suggest a derivation from surface residues.[32]

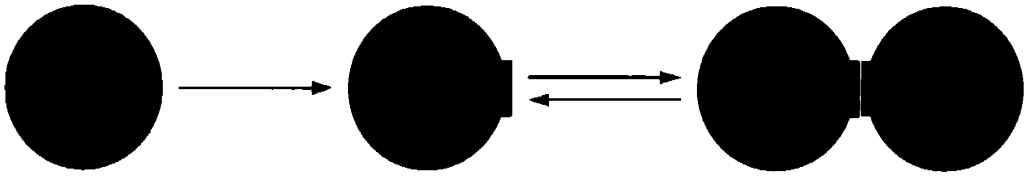


Figure 1-3. The evolution of a hydrophobic patch on the surface of a monomeric protein results in a protein capable of oligomerization.

It is also possible that oligomeric proteins that fold via a two-step pathway, i.e. wherein the individual monomers are unfolded in the absence of oligomerization, may have evolved specifically toward the dimeric state, without an intermediate stage of an independently folded monomer.[31] In particular, Xu et al.[31] proposed a direct evolutionary pathway for dimers exhibiting two-state kinetics and either having a large and flat interface or having two short intertwined chains. Some examples of proteins in this class might include dimeric coiled coils such as GCN4, which are unfolded in the absence of an oligomerization interaction.

Domain swapping is an elegant theory that proposes one mechanism whereby multi-domain proteins could evolve to a different oligomeric state. Certain proteins are known to coexist in both monomeric and oligomeric forms, wherein the interdomain interactions can occur both within a single monomer and between monomer units (Figure 1-4).[7] Roughly forty such domain-swapped proteins are currently known. These include the bacteriophage λ cro repressor, barnase, and diphtheria toxin.[7] A chance mutation to such a protein that favored a higher oligomeric state over the monomeric form could be selected for, resulting in a sustained oligomer.[7,33]

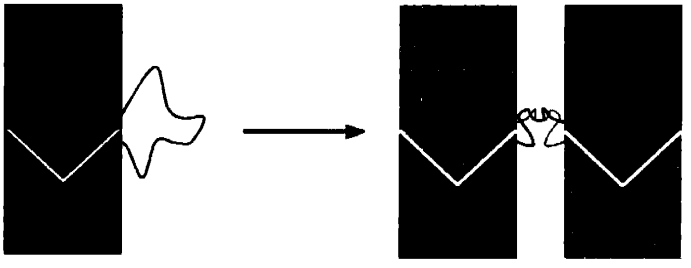


Figure 1-4. Domain swapping. A monomeric protein with two interacting domains may coexist with oligomeric state in which the domains have the same interaction between monomer units.

Moreover, it may not be necessary to proceed through the intermediacy of a true domain-swapping protein with coexisting oligomeric states. Should a mutation affect the linker region

such that the protein can no longer fold on itself, it may have the option of burying the hydrophobic surfaces in a domain-swapping interaction with another monomer. This typically occurs when the linker between the domains is shortened, precluding an intermonomer association.[34,35] Varying the linker length can result in an increase or decrease in oligomeric state.[36] Several oligomeric proteins and mini-proteins have been engineered from the corresponding monomers by shortening the linker to form a domain-swapped oligomer.[34,35,37,38]

Finally, oligomerization can arise via fusion of a gene encoding a dimerization or oligomerization domain, such as a coiled coil domain, onto a previously monomeric protein (Figure 1-5).[33] The newly oligomeric protein will then associate through this oligomerization domain, and potentially form new contacts in regions brought into proximity by the association. Mutations favoring such an interaction, if beneficial, may be selected for, thus producing a more robust oligomer.

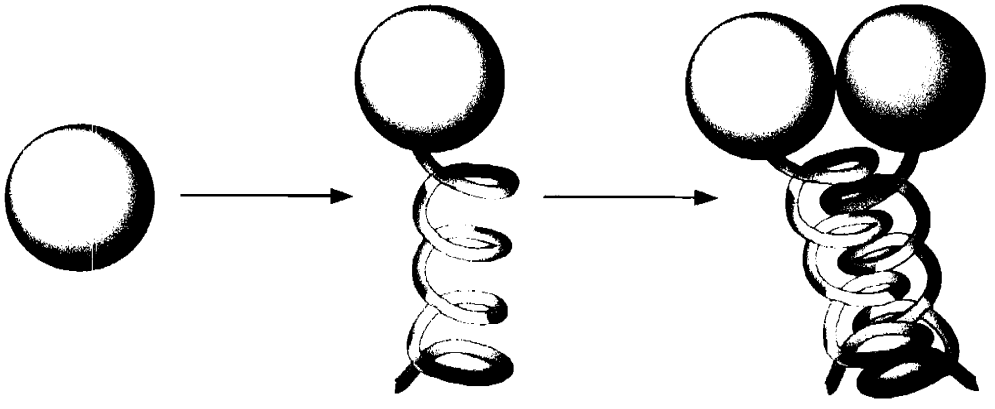


Figure 1-5. Fusion of an oligomerization domain to a monomeric protein results in an oligomeric fusion protein.

It has been proposed that the basic helix-loop-helix (bHLH) family of oligomeric transcription factors may have arisen via modular evolution as described above.[39] Over time, the primordial bHLH oligomerization domain was linked to various other proteins through gene

duplication events and insertion into other genes, resulting in the large and diverse array of modern proteins bearing bHLH domains.[39] It is likely that other oligomerization modules may also have been propagated in this fashion.

Characteristics of Oligomeric Interfaces

The specific parameters characterizing an oligomeric interface depend on the nature of that interface and on the duration of the interaction defining it. However, some general features may be described. The interaction surface tends to be circular in shape[40,41] and the interface residues tend to protrude from the surface of the protein.[3]

Protein-protein interfaces are relatively planar.[42] This is particularly pronounced for heterooligomer associations.[3,43] By contrast, many obligate homodimers and heterodimers have intertwined monomer units, and thus less planar interfaces.[3,43]

Large oligomeric interfaces are often associated with strong interactions, as might be predicted. However, a small oligomeric interface can be present in both weak and strong associations.[6] The buried surface area in obligate homodimeric proteins is usually greater than 1400 $Å^2$.[44,45] In non-obligate complexes the interface buried surface area is usually less than 2500 $Å^2$, whereas for weak and transient associations the buried surface area of the interface is less than 1000 $Å^2$.[6]

In general, the residues involved at the oligomeric interface may be slightly more conserved than other surface residues, although the extent and significance of this preference has been questioned.[6,46-48] It has been found that certain conserved residues, or "hot spots," generally at the center of an interface, are responsible for most of the binding energy of an oligomeric interaction. These hot spots are often composed of polar residues that can engage in van der Waals contacts and hydrogen bonding.[49-51]

Hydrophobic interactions play an important role in defining homooligomeric interfaces. About two thirds of the residues in subunit interfaces are nonpolar,[45,52] a much larger fraction than

in non-interacting surface residues. However, inter-subunit interfaces are less nonpolar, and have a greater proportion of hydrophilic and polar residues, than a typical protein hydrophobic core.[32]

The interaction surfaces of transient or on/off heterooligomers and weakly associating homooligomers are less nonpolar than those of tightly bound oligomers.[3,6,43] In this case the greater surface polarity may help to solubilize them in their freestanding existence.

Oligomeric interfaces are about one fifth polar, a greater proportion than is found in buried hydrophobic cores.[52] Hydrogen bonds and salt bridges are important for the stabilization of oligomeric interfaces, as suggested by the prevalence of polar hot spot residues. Early studies showed that there is about one hydrogen bond per 200 $Å^2$ of subunit interface.[45] Charged hydrogen bonds can often be found at subunit interfaces, but such interactions are less common than in the core.[52]

Oligomeric interfaces are generally tightly-packed. Moreover, the interaction surfaces often have significant electrostatic and geometrical complementarity that gives rise to the specificity of the interaction.[41,53,54] Nonetheless, most interaction surfaces are not entirely complementary. Hubbard reported that the vast majority of oligomeric proteins have one or more cavities at the inter-subunit interface; most of these cavities are filled with water.[55]

Why Study Oligomeric Mini-Proteins?

Mini-proteins are short, synthetic polypeptides that adopt a discrete, structured fold in aqueous solution. They are typically derived by paring a natural fold to a more compact size. There is a wealth of oligomeric mini-proteins in the literature, including all-alpha coiled coils and helical bundles,[27,56,57] mixed alpha/beta motifs,[37,58] and all-beta motifs.[59-61] Mini-proteins have been extensively used as minimal models for the study of key features of natural proteins, including the thermodynamic determinants of protein stability[62] and the introduction of catalytic or other functionality.[24,26,28,29] Mini-proteins are ideal for the study of oligomeric proteins for several reasons. (1) Mini-proteins are simple systems. They have far fewer variables than do natural

proteins, an attribute that is experimentally advantageous as it enables the directed study of a single feature of interest. (2) Mini-proteins, because of their small size, are highly amenable to computational study. By contrast, the natural proteins, particularly larger oligomeric proteins, are more complex and computationally intensive to model. (3) It is straightforward to design and synthesize a series of mutants with which to rigorously study a feature of interest. Moreover, nonstandard amino acids can be easily introduced with which to probe subtle features not discernible through the standard twenty amino acids.

Mini-proteins have been used to systematically examine general features of oligomeric proteins. For example, the roles of buried polar residues[63-66] in oligomeric interfaces and of inter-subunit salt bridges[67-70] have been extensively researched using coiled coil and helical bundle mini-protein models.

Determinants of a unique oligomeric state have been explored through the design of oligomeric and monomeric mini-proteins. Substitutions of core polar residues in coiled-coil peptides have been found to result in the loss of a specific oligomeric state.[71] Shortening an inter-domain linker has resulted in a domain-swapped oligomeric mini-protein.[37] Negative design features preventing self-association have been elucidated through the iterative design of all-beta mini-proteins.[72]

The folding kinetics of alpha,[73,74] beta,[75] and alpha/beta[76] mini-proteins have been computationally and experimentally investigated in order to derive insights abut the folding processes of natural proteins. The folding pathways of a smaller number of oligomeric coiled coils mini-proteins[77,78] have also been studied. The relative scarcity reflects the limited number of oligomeric mini-proteins in the literature.

As noted above, designed mini-proteins, and in particular designed oligomeric mini-proteins, bear a close resemblance to proposed primordial proteins. It may be possible to deduce principles about the early protein environment, and about the evolutionary selection processes

they faced, through the exercise of introducing function and elaborating structure in mini-proteins.

Conclusions

Oligomeric proteins are prevalent in nature, comprising roughly one third of cellular proteins.[2] A number of functional advantages favor the evolution of oligomers from monomeric precursors. The study of protein oligomerization can provide insights into the early protein environment and the evolution of modern proteins. The specific composition of an inter-subunit interface is dependent on the nature and affinity of the interactions comprising that interface, but both hydrophobic and polar interactions play key roles in most interfaces. There are several mechanisms whereby protein oligomerization could have arisen, including the evolution of a new site of interaction, the direct evolution of a protein for the oligomeric state, a mutation in the hinge region leading to the a domain-swapped oligomer, and the fusion of a preexisting oligomerization domain. Oligomeric mini-proteins constitute simple and tractable model systems for the study of oligomeric proteins.

References

(1) Goodsell, D. S. *Trends Biochem. Sci.* **1991**, *16*, 203-206. Inside A Living Cell.

(2) Goodsell, D. S.; Olson, A. J. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 105-153. Structural Symmetry and Protein Function.

(3) Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13-20. Principles of Protein-Protein Interactions.

(4) Goodsell, D. S.; Olson, A. J. *Trends Biochem. Sci.* **1993**, *18*, 65-68. Soluble Proteins: Size, Shape and Function.

(5) Nooren, I. M. A.; Thornton, J. M. *EMBO J.* **2003**, *22*, 3486-3492. Diversity of Protein-Protein Interactions.

(6) Nooren, I. M. A.; Thornton, J. M. *J. Mol. Biol.* **2003**, *325*, 991-1018. Structural Characterisation and Functional Significance of Transient Protein-Protein Interactions.

(7) Liu, Y.; Eisenberg, D. *Protein Sci.* **2002**, *11*, 1285-1299. 3D Domain Swapping: As Domains Continue to Swap.

(8) Walden, H.; Bell, G. S.; Russell, R. J. M.; Siebers, B.; Hensel, R.; Taylor, G. L. *J. Mol. Biol.* **2001**, *306*, 745-757. Tiny TIM: A Small, Tetrameric, Hyperthermostable Triosephosphate Isomerase.

(9) Monod, J.; Olsen, K. W.; Sabesan, M. N.; Buehner, M.; Ford, G. C.; Rossman, M. G. *J. Mol. Biol.* **1965**, *12*, 88-118. On the Nature of Allosteric Transitions: A Plausible Model.

(10) Taylor, W. R.; May, A. C. W.; Brown, N. P.; Aszódi, A. *Rep. Prog. Phys.* **2001**, *64*, 517-590. Protein Structure: Geometry, Topology and Classification.

(11) Dassa, E.; Saurin, W.; Hofnung, M.; Kuan, G.; Saier, M. H., Jr. *Res. Microbiol.* **1995**, *146*, 271-278. Phylogenetic Analyses of the ATP-Binding Constituents of Bacterial Extracytoplasmic Receptor-Dependent ABC-Type Nutrient Uptake Permeases.

(12) Saurin, W.; Dassa, E. *Protein Sci.* **1994**, *3*, 325-344. Sequence Relationships Between Integral Inner Membrane Proteins of Binding Protein-Dependent Transport Systems: Evolution by Recurrent Gene Duplications.

(13) Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M. *Science* **2000**, *289*, 1546-1550. Structural Evidence for Evolution of the β/α Barrel Scaffold by Gene Duplication and Fusion.

(14) Groft, C. M.; Beckmann, R.; Sali, A.; Burley, S. K. *Nat. Struct. Biol.* **2000**, *7*, 1156-1164. Crystal Structures of Ribosome Anti-Association Factor IF6.

(15) Chothia, C.; Gough, J.; Vogel, C.; Teichmann, S. A. *Science* **2003**, *30*, 1701-1703. Evolution of the Protein Repertoire.

(16) Lupas, A. N.; Ponting, C. P.; Russell, R. B. *J. Struct. Biol.* **2001**, *134*, 191-203. On the Evolution of Protein Folds: Are Similar Motifs in Different Protein Folds the Result of Convergence, Insertion, or Relics of an Ancient Peptide World?

(17) Ponting, C. P.; Russell, R. R. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 45-71. The Natural History of Protein Domains.

(18) Galaktos, N. G.; Walsh, C. T. *Biochemistry* **1987**, *26*, 8475-8480. Specific Proteolysis of Native Alanine Racemases From *Salmonella typhimurium*: Identification of the Cleavage Site and Characterization of the Clipped Two-Domain Proteins.

(19) Girons, I. S.; Gilles, A.-M.; Margarita, D.; Michelson, S.; Monnot, M.; Fermandjian, S.; Danchin, A.; Bârzu, O. *J. Biol. Chem.* **1987**, *262*, 622-629. Structural and Catalytic Characteristics of *Escherichia coli* Adenylate Kinase.

(20) Pelletier, J. N.; Campbell-Valois, F.-X.; Michnick, S. W. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12141-12146. Oligomerization Domain-Directed Reassembly of Active Dihydrofolate Reductase From Rationally Designed Fragments.

(21) Bertolaet, B. L.; Seidel, H. M.; Knowles, J. R. *Science* **1995**, *268*, 1367-9. Introns and the Origin of Protein-Coding Genes.

(22) Patthy, L. *Matrix Biol.* **1996**, *15*, 301-310. Exon Shuffling and Other Ways of Module Exchange.

(23) Stoltzfus, A.; Spencer, D. F.; Zuker, M.; John M. Logsdon, J.; Doolittle, W. F. *Science* **1994**, *265*, 202-207. Testing the Exon Theory of Genes: The Evidence from Protein Structure.

(24) Shogren-Knaak, M. A.; Imperiali, B. *Bioorg. Med. Chem.* **1999**, *7*, 1993-2002. Modulating Pyridoxamine-Mediated Transamination Through a Beta Beta Alpha Motif Peptide Scaffold.

(25) McDonnell, K. A. Ph.D. Thesis, Massachusetts Institute of Technology, 2001. Towards Incorporation of Catalytic Function Into Small Folded Peptide Scaffolds.

(26) Robertson, D. E.; Farid, R. S.; Moser, C. C.; Urbauer, J. L.; Mulholland, S. E.; Pidikiti, R.; Lear, J. D.; Wand, A. J.; DeGrado, W. F.; Dutton, P. L. *Nature* **1994**, *368*, 425-431. Design and Synthesis of Multi-Haem Proteins.

(27) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochem.* **1999**, *68*, 779-819. *De Novo* Design and Structural Characterization of Proteins and Metalloproteins.

(28) Broo, K. S.; Nilsson, H.; Nilsson, J.; Floodberg, A.; Baltzer, L. *J. Am. Chem. Soc.* **1998**, *120*, 4063-4068. Cooperative Nucleophilic and General-Acid Catalysis by the $HisH^+$-His Pair and Arginine Transition State Bindign in Catalysis of Ester Hydrolysis Reactions by Helix-Loop-Helix Motifs.

(29) Allert, M.; Kjellstrand, M.; Broo, K.; Nilsson, Å.; Baltzer, L. *J. Chem. Soc., Perkin Trans. 2* **1998**, 2271-2274. A Designed Folded Polypeptide Model System That Catalyses the Decarboxylation of Oxaloacetate.

(30) Long, M.; Betrán, E.; Thornton, K.; Wang, W. *Nat. Rev. Genet.* **2003**, *4*. The Origin of New Genes: Glimpses From the Young and Old.

(31) Xu, D.; Tsai, C.-J.; Nussinov, R. *Protein Sci.* **1998**, *1998*, 533-544. Mechanism and Evolution of Protein Dimerization.

(32) D'Alessio, G. *Prog. Biophys. Mol. Biol.* **1999**, *72*, 271-98. The Evolutionary Transition from Monomeric to Oligomeric Proteins: Tools, The Environment, Hypotheses.

(33) *Proc. Natl. Acad. Sci. USA* **1994**, *1*, 411-412. The Evolution of Oligomerization.

(34) Pei, X. Y.; Holliger, P.; Murzin, A. G.; Williams, R. L. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 9637-9642. The 2.0-Å Resolution Crystal Structure of a Trimeric Antibody Fragment with Noncognate V_H-V_L Domain Pairs Shows a Rearrangement of V_H CDR3.

(35) Albright, R. A.; Mossing, M. C.; Matthews, B. W. *Biochemistry* **1996**, *35*, 735-742. High-Resolution Structure of an Engineered Cro Monomer Shows Changes in Conformation Relative to the Native Dimer.

(36) MacBeath, G.; Kast, P.; Hilvert, D. *Protein Sci.* **1998**, *7*, 1757-1767. Probing Enzyme Quaternary Structure by Combinatorial Mutagenesis and Selection.

(37) Mezo, A. R.; Ottesen, J. J.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 1002-1003. Discovery and Characterization of a Discretely Folded Homotrimeric Peptide.

(38) Kuhlman, B.; O'Neill, J. W.; Kim, D. E.; Zhang, K. Y. J.; Baker, D. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10687-10691. Conversion of Monomeric Protein L to an Obligate Dimer By Computational Protein Design.

(39) Morgenstern, B.; Atchley, W. R. *Mol. Biol. Evol.* **1999**, *16*, 1654-1663. Evolution of bHLH Transcription Factors: Modular Evolution by Domain Shuffling?

(40) Chothia, C.; Janin, J. *Nature* **1975**, *256*, 705-708. Principles of Protein-Protein Recognition.

(41) Jones, S.; Thornton, J. M. *Prog. Biophys. Mol. Biol.* **1995**, *63*, 31-65. Protein-Protein Interactions: A Review of Protein Dimer Structures.

(42) Argos, P. *Protein Eng.* **1988**, *2*, 101-113. An Investigation of Protein Subunit and Domain Interfaces.

(43) Jones, S.; Thornton, J. M. *J. Mol. Biol.* **1997**, *272*, 121-132. Analysis of Protein-Protein Interaction Sites Using Surface Patches.

(44) Miller, S.; Lesk, A. M.; Janin, J.; Chothia, C. *Nature* **1987**, *328*, 834-6. The Accessible Surface Area and Stability of Oligomeric Proteins.

(45) Janin, J.; Miller, S.; Chothia, C. *J. Mol. Biol.* **1988**, *204*, 155-64. Surface, Subunit Interfaces and Interior of Oligomeric Proteins.

(46) Caffrey, D. R.; Somaroo, S.; Hughes, J. D.; Mintseris, J.; Huang, E. S. *Protein Sci.* **2004**, *13*, 190-202. Are Protein-Protein Interfaces More Conserved in Sequence Than the Rest of the Protein Surface?

(47) Valdar, W. S. J.; Thornton, J. M. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 108-124. Protein-Protein Interfaces: Analysis of Amino Acid Conservation in Homodimers.

(48) Grishin, N. V.; Phillips, M. A. *Protein Sci.* **1994**, *3*, 2455-2458. The Subunit Interfaces of Oligomeric Enzymes Are Conserved to a Similar Extent to the Overall Protein Sequences.

(49) Clackson, T.; Wells, J. A. *Science* **1995**, *267*, 383-386. A Hot Spot of Binding Energy in a Hormone-Receptor Interface.

(50) Bogan, A. A.; Thorn, K. S. *J. Mol. Biol.* **1998**, *280*, 1-9. Anatomy of Hot Spots in Protein Interfaces.
(51) Hu, Z.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 331-342. Conservation of Polar Residues as Hot Spots at Protein Interfaces.
(52) Miller, S. *Protein Eng.* **1989**, *3*, 77-83. The Structure of Interfaces Between Subunits of Dimeric and Tetrameric Proteins.
(53) McCoy, A. J.; Epa, V. C.; Colman, P. M. *J. Mol. Biol.* **1997**, *268*, 570-584. Electrostatic Complementarity at Protein/Protein Interfaces.
(54) Lawrence, M. C.; Colman, P. M. *J. Mol. Biol.* **1993**, *234*, 946-950. Shape Complementarity at Protein/Protein Interfaces.
(55) Hubbard, S. J.; Argos, P. *Protein Sci.* **1994**, *3*, 2194-2206. Cavities and Packing at Protein Interfaces.
(56) Hill, R. B.; Raleigh, D. P.; Lombardi, A.; DeGrado, W. F. *Acc. Chem. Res.* **2000**, *33*, 745-754. *De Novo* Design of Helical Bundles as Models for Understanding Protein Folding and Function.
(57) Kohn, W. D.; Hodges, R. S. *Trends Biotechnol.* **1998**, *16*, 379-389. *De Novo* Design of α-Helical Coiled Coils and Bundles: Models for the Development of Protein-Design Principles.
(58) Das, C.; Shankaramma, S. C.; Balaram, P. *Chem. Eur. J.* **2001**, *7*, 840-847. Molecular Carpentry: Piecing Together Helices and Hairpins.
(59) Quinn, T. P.; Tweedy, N. B.; Williams, R. W.; Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 8747-8751. Betadoublet: *De Novo* Design, Synthesis, and Characterization of a β-Sandwich Protein.
(60) Ilyina, E.; Roontga, V.; Mayo, K. H. *Biochemistry* **1997**, *36*, 5245-5250. NMR Structure of a *De Novo* Designed, Peptide 33mer with Two Distinct, Compact β-Sheet Folds.
(61) Venkatraman, J.; Gowda, G. A. N.; Balaram, P. *J. Am. Chem. Soc.* **2002**, *124*, 4987-4994. Design and Construction of an Open Multistrand β-Sheet Polypeptide Stabilized by a Disulfide Bridge.
(62) Zhou, N. E.; Zhu, B.-Y.; Kay, C. M.; Hodges, R. S. *Biopolymers* **1992**, *32*, 419-426. The Two-Stranded α-Helical Coiled-Coil is an Ideal Model for Studying Protein Stability and Subunit Interactions.
(63) Akey, D. L.; Malashkevich, V. N.; Kim, P. S. *Biochemistry* **2001**, *40*, 6352-6360. Buried Polar Residues in Coiled-Coil Interfaces.
(64) Lumb, K. J.; Kim, P. S. *Biochemistry* **1995**, *34*, 8642-8648. A Buried Polar Interaction Imparts Structural Uniqueness in a Designed Heterodimeric Coiled Coil.
(65) Eckert, D. M.; Malashkevich, V. N.; Kim, P. S. *J. Mol. Biol.* **1998**, *284*, 859-865. Crystal Structure of GCN4-pI_QI , a Trimeric Coiled Coil with Buried Polar Residues.
(66) Oakley, M. G.; Kim, P. S. *Biochemistry* **1998**, *37*, 12603-12610. A Buried Polar Interaction Can Direct the Relative Orientation of Helices in a Coiled Coil.
(67) McClain, D. L.; Binfet, J. P.; Oakley, M. G. *J. Mol. Biol* **2001**, *313*, 371-383. Evaluation of the Energetic Contribution of Interhelical Coulombic Interactions for Coiled Coil Helix Orientation Specificity.
(68) Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Mol. Biol* **1994**, *237*, 500-512. The Role of Interhelical Ionic Interactions in Controlling Protein Folding and Stability. *De Novo* Designed Synthetic Two-Stranded α-Helical Coiled Coils.
(69) Burkhard, P.; Ivaninski, S.; Lustig, A. *J. Mol. Biol.* **2002**, *2002*, 901-910. Improving Coiled-Coil Stability by Optimizing Ionic Interactions.
(70) Spek, E. J.; Bui, A. H.; Lu, M.; Kallenbach, N. R. *Protein Sci.* **1998**, *7*, 2431-2437. Surface Salt Bridges Stabilize the GCN4 Leucine Zipper.

(71) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401-1407. A Switch Between Two-, Three-, and Four-Stranded Coiled Coils in GCN4 Leucine Zipper Mutants.

(72) Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2754-2759. Natural β-Sheet Proteins Use Negative Design to Avoid Edge-to-Edge Aggregation.

(73) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548-14549. The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations.

(74) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952-12953. Smaller and Faster: The 20-Residue Trp-Cage Folds in 4 μs.

(75) Jager, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2001**, *311*, 373-393. The Folding Mechanism of a Beta-Sheet: The WW Domain.

(76) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102-106. Absolute Comparison of Simulated and Experimental Protein-Folding Dynamics.

(77) Wendt, H.; Leder, L.; Härmä, H.; Jelesarov, I.; Baici, A.; Bosshard, H. R. *Biochemistry* **1997**, *36*, 204-213. Very Rapid, Ionic Strength-Dependent Association and Folding of a Heterodimeric Leucine Zipper.

(78) Ibarra-Molero, B.; Makhatadze, G. I.; Matthews, C. R. *Biochemistry* **2001**, *40*, 719-731. Mapping the Energy Surface for the Folding Reaction of the Coiled-Coil Peptide GCN4-p1.

Chapter 2: Towards a $(BBA)_n$ Structure

Introduction: An Oligomeric BBA

Quaternary structure is a fundamental feature of natural proteins. It has been estimated that approximately one third of cellular proteins are oligomeric.[1] Oligomerization may confer certain structural and functional advantages relative to monomeric proteins, as discussed in the preceding chapter. Oligomeric mini-proteins constitute minimal model systems for quaternary structure formation. Moreover, they constitute platforms for probing whether more complex oligomeric structures might ultimately support function.

An oligomeric ββα (BBA) mini-protein was recently designed by Dr. Adam R. Mezo from the prototypic BBA motif (Figure 2-1) developed previously by Dr. Mary D. Struthers.[2,3] The BBA motif was chosen as the building block as it represents a structurally well-defined and discretely folded mini-protein motif incorporating domains of both alpha and beta structure. The final generation BBA peptide, **BBA5**, was utilized for these studies.[3]

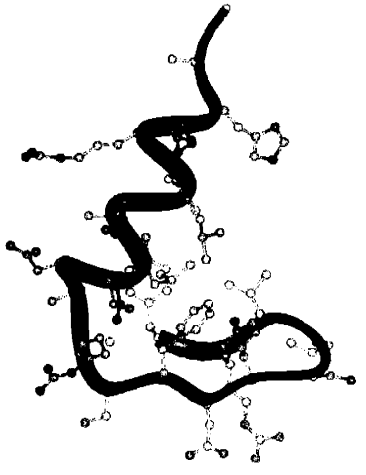


Figure 2-1. NMR structure of the monomeric mini-protein **BBA5**.

The oligomerization strategy employed by Dr. Mezo – "domain swapping," or the exchange of association partners – mimics an important evolutionary mechanism for protein oligomerization.[4-7] As it has been reported that shortening the linker between domains favors domain-swapped oligomers,[8-10] a bias towards oligomerization was introduced by varying the length of the three-amino acid hinge region between the alpha helix and the beta hairpin of **BBA5**.

A small library of ββα peptides was created. Each member of the library was synthesized with a fluorophore incorporated into the BBA helix and additionally with a quencher incorporated into the BBA hairpin. An oligomeric BBA peptide, **BBAT1**, was discovered by means of a fluorescence quenching assay for homoassociation (Figure 2-2).[11,12] As the fluorophore and quencher must be in close proximity to result in fluorescent quenching, the observation of fluorescence quenching was suggestive of a domain-swapping interaction between alpha and beta secondary structural elements. This hypothesis was supported by the existence of NOE interactions between helix and hairpin residues (*vide infra*).

BBAT1 associates in a discrete oligomeric state and exhibits many of the hallmarks of larger proteins, including cooperative folding, lack of ANS binding (indicative of a well-packed hydrophobic core), and limited deuterium exchange rates. The oligomeric state of **BBAT1** was initially assigned as trimeric on the basis of sedimentation equilibrium analytical ultracentrifugation experiments.

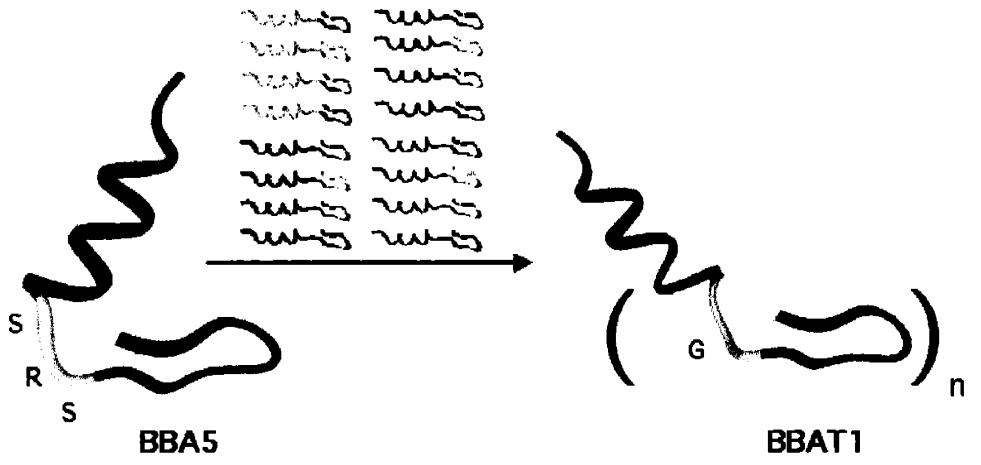


Figure 2-2. An oligomeric ββα peptide was obtained from a small library of BBA derivatives. The winning peptide features a shortened linker between the alpha and beta domains, resulting in a domain-swapping interaction that buries the hydrophobic face of the peptide in inter-monomer associations.

The discovery of **BBAT1** immediately raised questions regarding the nature of the inter-monomer interactions, the forces responsible for tight and cooperative association, and the role of the hinge region in determining the specific oligomeric state. Biophysical studies and detailed mutational studies provided incomplete glimpses into some of these issues.[11-13] A complete structural elucidation would explain the biophysical observations and detailed knowledge of the structure would enable us to rationally design subsequent generations of BBA oligomers varying in packing and oligomerization. It would also aid in the introduction of function in this mini-protein scaffold.

NMR experiments conducted prior to the studies presented in this chapter revealed that each monomer in the assembly is quite similar to the parent **BBA5**, and that each monomer exists in a symmetrical environment.[11] NOEs were found between helix and hairpin residues.[11] As the shortened linker does not allow a monomer to fold upon itself, these NOE interactions were taken as evidence of a domain-swapped topology, whereby the helix of one monomer interacts with the hairpin of another BBA unit. Unfortunately, the degeneracy of the NMR signals, proof of a high

degree of symmetry in **BBAT1**, prohibited further structure determination. The circular dichroism spectrum of the oligomeric species, in particular the ratio of ellipticities at 222 nm and 208 nm,[14,15] suggests that there are interactions between individual helices, a pattern most frequently found in a coiled coil or helical bundle.

The conformation of the key hinge residue was not known. Mutational studies suggested that the flexible glycine hinge adopts dihedral angles favored by D-amino acids.[16] Specifically, in peptide **1**, the incorporation of a D-alanine at the hinge position (residue 9), in place of the original glycine in **BBAT1** (Table 2-1), imparts a greater thermodynamic stability due to a more limited sampling of conformational space. While both peptides demonstrate cooperative folding, the transition point for thermal denaturation (T_m) of peptide **1** was considerably higher than that of **BBAT1** (64 °C *vs.* 40 °C).

		hairpin									*helix*												
Peptide		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
BBA5	Ac	Y	R	V	p	S	Y	D	F	S R S	D	E	L	A	K	L	L	R	Q	H	A	G	NH_2
BBAT1	Ac	Y	R	I	p	S	Y	D	F	G	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2
1	Ac	Y	R	I	p	S	Y	D	F	a	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2

Table 2-1. Sequences of peptides **BBA5**, **BBAT1** and **1**. Abbreviations: p=D-proline, a=D-alanine, Z=benzoylated L-α,β-diaminopropionic acid.

Dr. Kevin A. McDonnell proposed that peptides **BBAT1** and **1** exist in solution as a trimeric coiled coil-core encased by beta sheets (Figure 2-3).[16] This structure could account for the observed interhelical associations, domain swapping, and symmetry. However, it requires that the inter-domain linker be long and flexible enough to connect the two domains, which may require some loss of structure at the end of the hairpin and/or the beginning of the helix, and therefore be inconsistent with the overall resemblance to **BBA5** observed by NMR..[11]

Figure 2-3. A trimeric BBA structure proposed by Dr. Kevin A. McDonnell.

X-Ray Crystallography

We chose to pursue the structural determination of peptide **1** by X-ray crystallography. There are several steps in the X-ray structural determination of a new protein or mini-protein between purification of the protein of interest and publication of a final model (Figure 2-4). Optimal conditions must first be determined for the growth of single, robust, crystals with good diffraction properties. After a data set is collected, it is necessary to translate the diffraction data into an interpretable electron density map. This is known as the phase problem. There are three primary methods for obtaining crystallographic phases: direct methods, molecular replacement, and heavy atom-based methods such as multiple isomorphous replacement (MIR) and multi-wavelength anomalous diffraction (MAD). Once starting phases have been found, an initial model is generated, and subjected to several cycles of refinement until it is an accurate depiction of the protein structure in the crystalline state.

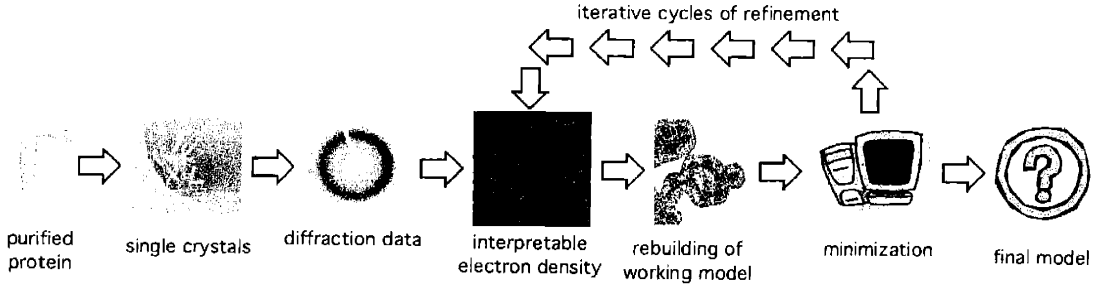


Figure 2-4. The process of obtaining an X-ray crystal structure.

This chapter will focus on early work towards obtaining diffraction-quality crystals, as well as initial attempts to obtain phase information through direct methods, molecular replacement, MIR, and MAD. The phasing section will be divided into four parts: direct methods, molecular replacement, heavy atoms incorporated by crystal soaks, and heavy atoms covalently-incorporated prior to crystallization. The discussion of heavy atoms will thus be divided by introduction of the heavy atom, rather than the data collection technique. Where possible, we made use of both isomorphous replacement and anomalous scattering with heavy atom derivatives.

Crystallization of Peptide 1

We selected peptide **1** for crystallization trials because of the higher transition point for thermal denaturation (T_m, 64 °C *vs.* 40 °C), and the greater solubility relative to **BBAT1**. Biophysically, the two peptides are comparable in all other regards.

Peptides were crystallized using the hanging drop method (Figure 2-5): a drop of peptide stock mixed with precipitant is incubated in a sealed container over a reservoir solution of the precipitant.[17] As the two solutions approach equilibrium by vapor diffusion, the concentration of peptide increases and its solubility limit is reached. If conditions are favorable, crystallization occurs. Other possible outcomes are precipitation of the peptide or the production of an oil. The conditions leading to these results can sometimes be optimized to induce crystallization.

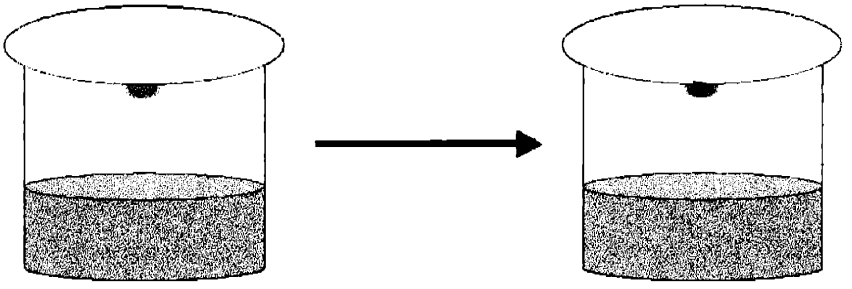


Figure 2-5. A hanging drop crystallization.

Initial crystallization conditions were obtained by assaying several commercial sparse matrix crystallization[18] kits. Of several initial hits, two crystallization conditions were found to be particularly useful (**a**: 25% *t*-butanol in 0.1M TRIS buffer, pH 8.5, **b**: 100 mM Na HEPES buffer pH 7.5, 10% *v/v* *i*-propanol, 20% *w/v* PEG 4000). All crystals used for diffraction studies were grown from either of these two conditions: most of the earlier experiments described in this chapter were performed using crystals grown from condition **a**, and the later work from crystals grown from condition **b**.

When using condition **a**, crystallization occurs in approximately two weeks at room temperature, but is occasionally unreliable. Crystals grow as rods and needles (Figure 2-6). The growth and size of crystals is highly dependent on concentration of the stock solution of peptide **1**; better crystals are obtained from more concentrated stocks (≥ 10 mg/ml). These crystals are somewhat difficult to manipulate because the high proportion of *t*-butanol results in very low solution viscosity and a high tendency toward evaporation. Thus opening a well to remove a crystal often results in the loss of all the remaining crystals in that well. Extensive efforts to improve upon these crystallization conditions, e.g. by varying the organic co-solvent or co-solvent content, pH, or temperature, did not result in any marked improvement in the crystals, and thus the commercial formulations of these conditions were used for room temperature growth of crystals.

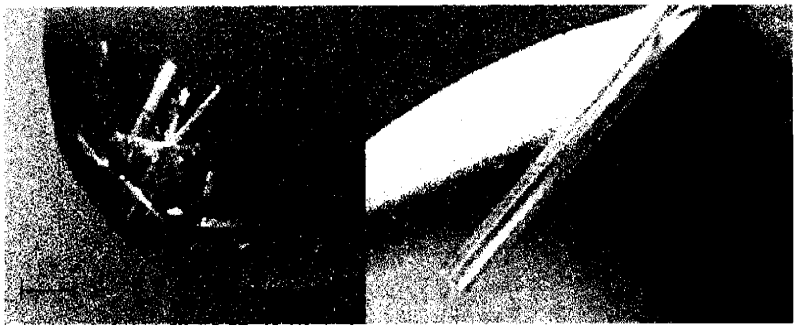


Figure 2-6. Crystals of peptide **1** grown from condition **a**. Bar indicates 0.1 mm.

The alternative condition, **b**, produces more robust crystals that can be more easily manipulated due to the lower organic co-solvent content. Condition **b** is also preferable with regard to ease and reliability of crystallization: crystals grow overnight at room temperature as rods and needles, and are relatively insensitive to environmental conditions such as temperature and humidity. Crystals are considerably easier to manipulate than those grown from condition **a** as the lower organic content results in drops that are less prone to evaporation (allowing greater time for experimental manipulations, and for the utilization of more of the crystals contained within a drop) and that are more viscous (making it easier to pick up crystals from the drop). Growth at 4 °C or 18 °C does not affect the size or morphology of the crystals grown relative to those grown at room temperature.

Crystals of **1** grown from either condition have the same morphology, space group and unit cell. We scaled together two data sets collected from crystals grown from conditions **a** and **b**, and obtained a low R_{merge} value of 13.3% (to 2.3 Å), further indicating that the two sets of crystallization conditions are equivalent.

Data Collection

Peptide **1** crystallizes in the orthorhombic crystal system with unit cell a = 38.94 Å, b = 56.17 Å, c = 32.32 Å, $\alpha = \beta = \gamma = 90°$. Examination of the systematic absences of the axial

reflections did not allow for an unambiguous determination of the space group. A 1.2 Å data set was collected at the 14-BM-C beam line at the BIOCARS facility at Argonne National Laboratory by Dr. Ezra Peisach. Crystals were frozen in a stream of N_2 gas cooled to -180 °C using Paratone-N (Hampton Research) as a cryoprotectant. Data were collected in two sweeps of varied exposure time to ensure high completeness of both high resolution and low resolution data. The DENZO and SCALEPACK programs[19] were used for data indexing, reduction, and scaling (Table 2-2).

Unit cell (Å,°)	a = 38.94 b = 56.17 c = 32.32 α = 90 β = 90 γ = 90
Space group	$P2_12_12$
X-ray source	BNL X12C
Wavelength (Å)	0.900
Resolution (Å)	∞ – 1.2
Total/ unique reflections	78,662 21,067
Completeness[1] (%)	92.5 (54.7)
I/σ(I)	26.8 (3.7)
R_{merge}[2] (%)	3.4 (22.9)
Solvent content (%)	28

Table 2-2. Data collection statistics for peptide **6**. [1]Values for the outermost shell (1.55-1.50 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.

Very rarely, peptide **1** also crystallizes in a hexagonal lattice with unit cell a = b = 45.63 Å, c = 32.13 Å, α = β = 90°, γ = 120°. These crystals were determined to belong to either of the two enantiomorphic space groups $P6_222$ or $P6_422$ by an examination of the symmetry axes and the systematic absences. These hexagonal crystals are identical to the eye to the more prevalent orthorhombic crystals and can grow from a well containing orthorhombic crystals. This space group has been observed only twice. A 1.8 Å data set was collected at the Boston University Core Facility for Macromolecular Crystallography. Crystals were frozen in a stream of N_2 gas cooled to -180 °C using Paratone-N (Hampton Research) as a cryoprotectant. The DENZO and

SCALEPACK programs[19] were used for data indexing, reduction, and scaling. Because of the irreproducibility of this crystal growth, and the existence of higher-resolution data, this data set was not pursued further at this stage.

Crystal Characterization

A rough estimate of the contents of the unit cell can be determined from a knowledge of the unit cell and space group using the Matthews coefficient, or the crystal volume per unit of protein molecular weight.[20] The Matthews coefficient, V_M, is defined as

$$V_M = \frac{V}{M \cdot n \cdot s}$$

where V is the volume of the unit cell, M is the molecular weight of a monomer, n is the number of molecules per asymmetric unit and s is the number of asymmetric units (ASU) per unit cell. The Matthews coefficient for proteins typically fall in a roughly normal distribution characterized by Matthews.[20] A V_M calculation for peptide **1** (Table 2-3) suggests that there are probably three monomers per asymmetric unit (the median value of V_M is 2.6[20]). Two or four monomers per asymmetric unit would also be possible, but the former would indicate a relatively high solvent content, and the latter a relatively low solvent content.

monomers per ASU	V_M	Solvent Content (%)
1	6.9	82.1
2	3.5	64.1
3	2.3	46.2
4	1.7	28.2
5	1.4	10.3

Table 2-3. Matthews coefficient determination for peptide **1**.

We searched for axes of noncrystallographic symmetry (NCS) using a self-rotation function with GLRF,[21] and noted that whereas two- and four-fold symmetry axes were found, we did not observe any indication of three-fold symmetry (Figure 2-7). We also searched for NCS using the complementary programs POLARRFN[22] and GETAX.[23] Again, we found strong peaks indicating two- and four-fold symmetry, but nothing suggestive of a three-fold axis. The two- and four-fold axes found could be due to four-fold noncrystallographic symmetry. The failure to observe a three-fold NCS axis could indicate that there was no three-fold NCS, or that the three-fold symmetry was inexact.

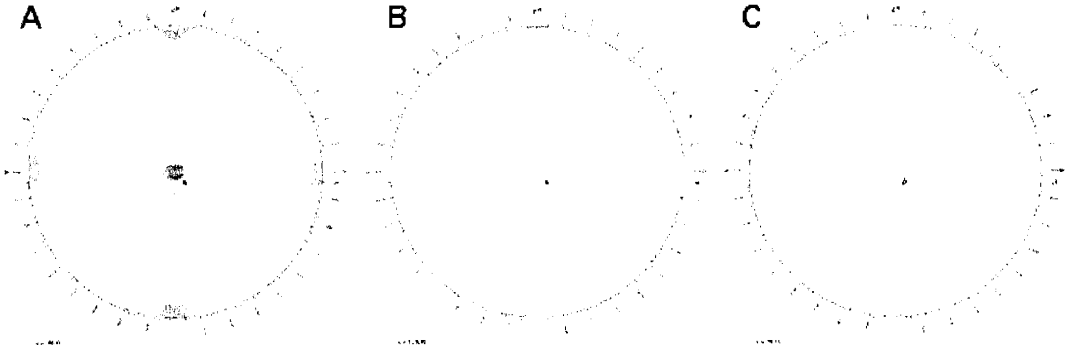


Figure 2-7. Results of the self-rotation function using GLRF from 10 to 4.0 Å. A. There are several peaks on a $\kappa = 180°$ plot indicative of two-fold symmetry. The larger peaks are likely due to crystallographic symmetry axes along a and c, and the smaller peaks along the diagonal due to noncrystallographic symmetry. B. There are no peaks on a $\kappa = 120°$ plot suggestive of three-fold symmetry. C. There is a peak along the a axis on a $\kappa = 90°$ plot that may be indicative of four-fold or pseudo four-fold noncrystallographic symmetry. Contour levels in A start at 250 and in B and C start at 200.

As the NCS findings were intriguing, but not necessarily inconsistent with the presence of three monomers in the asymmetric unit, we sought an alternative means of confirming the proposed trimeric oligomeric state in the crystal form. It is possible to directly measure the number of monomers in the asymmetric unit of a crystal by measuring the density of that crystal and then applying information about the unit cell and symmetry of the cell.

The density of a microscopic crystal is measured indirectly by determining the crystal buoyant point in a series of ficoll solutions of varying density.[24] Ficoll is a high molecular weight polymer of sucrose which makes a relatively benign medium for protein crystals. The density of ficoll solutions can be measured or calculated relatively easily.

A crystal dropped in a ficoll solution of lower density will sink; a crystal dropped in a ficoll solution of higher density will float (Figure 2-8). When the density of the crystal matches that of the solution in which it rests, it is neutrally buoyant, and can be observed to hover approximately midway in the solution column.

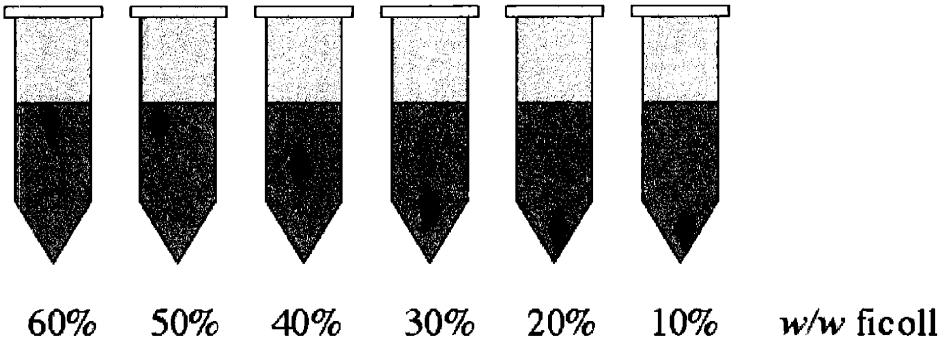


Figure 2-8. A crystal density measurement: crystals are placed in tubes containing ficoll solutions of varying densities. The tubes are inspected to determine the position of neutral buoyancy.

The protein volume fraction, Φ_p, is a function of the crystal density ρ_c, the solvent density ρ_s, and the protein partial specific volume, $\bar{\nu}_p$, as defined by equation (1).

$$(1)\ \Phi_p = \frac{(\rho_c - \rho_s)}{\left(\frac{1}{\bar{\nu}_p}\right) - \rho_s}$$

The partial specific volume $\bar{\nu}_p$ is generally calculated as a weighted average of the amino acid composition. These values are defined for a two-component, protein and water, system, but are often used as reasonable approximations in the case of proteins in dilute aqueous buffers.[25-27]

The number of monomers per unit cell depends on Φ_p, $\bar{\nu}_p$, the unit cell volume V, Avogadro's number N_A, and the protein's molecular weight M. It is given by equation (2).

$$(2)\ n_{cell} = \frac{\Phi_p N_A V}{\bar{\nu}_p M} = \left[\frac{(\rho_c - \rho_s)}{\left(\frac{1}{\bar{\nu}_P}\right) - \rho_s}\right]\frac{N_A V}{\bar{\nu}_p M}$$

The number of asymmetric units per unit cell depends on the symmetry of the space group crystal to which the crystal belongs. In equation (3), dividing n_{cell} by s, the number of symmetry operators, gives n_{ASU}, the number of monomers per asymmetric unit:

$$(3)\ n_{ASU} = \left[\frac{(\rho_c - \rho_s)}{\left(\frac{1}{\bar{\nu}_p}\right) - \rho_s}\right]\frac{N_A V}{\bar{\nu}_p M s}$$

Crystals grown from either condition **a** or **b** yielded the same results: 2.9 and 2.8 molecules per asymmetric unit, respectively, corresponding to 50% *w/w* ficoll. These values were consistent with the interpretation of prior sedimentation experiments, suggesting that the same oligomeric species that was found in solution was also found in the asymmetric unit of the crystal. However, as this value was derived using a partial specific volume appropriate for

proteins in dilute aqueous solutions, and not for solutions having a relatively high percentage of an organic cosolvent, the final outcome of these experiments was inconclusive.

The results of the crystal density experiment, in conjunction with the Matthews coefficient determination, suggested that peptide **1** exists as a homooligomer in the crystalline form. However, the stoichiometry of the peptide association could not be unambiguously determined in the absence of a crystal structure.

Direct Methods

We first attempted to solve the structure of peptide **1** by direct methods. Direct methods solutions rely on statistical relationships between the normalized structure factor amplitudes, $|E_{hkl}|$. They make use of the physical constraint that the electrons contributing to the X-ray scattering are found in discrete atoms, and that electron density is greater than or equal to zero throughout. Direct methods solutions are only possible when the phase problem is overdetermined, i.e. when there is considerably more data than there are parameters describing the model. This is true for very small molecules, and for large molecules that diffract to very high resolution, wherein considerably more data is available. For these cases, neither experimental phase information nor a search model is necessary for a structure solution.

Older direct methods approaches are limited to structures of roughly 200 atoms.[28,29] Shake-and-Bake and related dual-space algorithms that alternate between reciprocal space refinement and real-space filtering have extended the practical size limit, permitting direct methods solutions of relatively small proteins at high resolution. Larger structures can be solved when the protein in question contains heavy atoms or anomalously-scattering atoms.[29] The largest structure solved to date by direct methods that does not contain any heavy atoms is that of triclinic lysozyme, with 1001 protein atoms, including ten anomalously scattering sulfur atoms, per asymmetric unit.[30]

Direct methods solutions are generally applicable when the available data extends to atomic resolution (1.2 Å or better), and is of high quality ($I/\sigma(I) \geq 2$, $R_{merge} \leq 30\%$ in the outermost resolution shell).[28,31,32] This is commonly referred to as Sheldrick's 1.2 Å rule.

The atomic-resolution data set collected for peptide **1** satisfies these criteria: 1.20 Å resolution, $R_{merge} = 22.9\%$ (1.22-1.20 Å shell), $\langle I \rangle / \langle \sigma(I) \rangle = 3.7$ (Figure 2-9). Each monomer of peptide **1** has the molecular formula $C_{118}H_{177}O_{38}N_{31}$, and thus 187 non-hydrogen protein atoms per monomer. For comparison, the structure of triclinic lysozyme was solved with data extending to 0.85 Å.

Shell limit	Lower Angstrom	Upper	Average I	Average error	Average stat.	Norm. Chi**2	Linear R-fac	Square R-fac
	99.00	3.26	594231.4	20899.6	6498.6	0.738	0.027	0.031
	3.26	2.59	287716.0	9228.2	2966.3	1.089	0.034	0.033
	2.59	2.26	159139.7	4645.9	1490.5	1.171	0.040	0.041
	2.26	2.05	100968.7	3648.9	1429.8	1.880	0.049	0.056
	2.05	1.90	58580.5	2172.8	962.6	2.029	0.058	0.067
	1.90	1.79	34677.7	1314.5	642.7	1.897	0.059	0.061
	1.79	1.70	21672.7	864.6	477.7	1.788	0.065	0.064
	1.70	1.63	16418.9	718.2	445.7	1.701	0.072	0.075
	1.63	1.57	13063.7	647.0	443.0	1.980	0.085	0.081
	1.57	1.51	11985.8	684.0	490.0	2.674	0.112	0.103
	1.51	1.46	8827.1	623.4	498.1	2.837	0.144	0.128
	1.46	1.42	6476.5	577.6	492.7	3.238	0.198	0.163
	1.42	1.39	4704.1	494.8	453.0	2.260	0.206	0.187
	1.39	1.35	4435.3	543.8	499.5	2.765	0.254	0.222
	1.35	1.32	3015.9	499.7	478.4	2.296	0.297	0.259
	1.32	1.29	4713.8	622.5	566.5	2.854	0.250	0.185
	1.29	1.27	2193.8	572.1	557.8	2.562	0.400	0.307
	1.27	1.24	2822.1	635.3	614.3	2.907	0.391	0.298
	1.24	1.22	2688.7	670.5	647.6	2.407	0.329	0.229
	1.22	1.20	2638.8	719.5	703.9	2.520	0.317	0.229
All reflections			72256.5	2693.9	1105.3	1.966	0.041	0.034

Figure 2-9. SCALEPACK[19] statistics showing the high quality of data in the highest resolution shell.

The programs SnB,[33] incorporating the Shake-and-Bake algorithm, SHELXS,[34] and SHELXD[29,31,35] ("Half-Baked") were investigated. Professor George Sheldrick, the author of the latter two programs, was consulted regarding their implementation, and his suggestions incorporated in our efforts. Substructure searching with SnB and SHELXD was also attempted.

None of these attempts were fruitful. We concluded that at our relatively modest atomic resolution there were simply too many atoms for this approach to be successful.

Molecular Replacement

In molecular replacement solutions, a structurally-homologous search model is used to obtain phase information. The search model must first be located in space via a rotation function followed by a translation function. Starting phases are then generated from the placed search model. Molecular replacement works best when there is significant homology between the target structure and the search model (rms deviations less than 2.0 Å between C_α atoms, generally corresponding to sequence homologies better than 25%).[36]

The sequences of **BBA5** and peptide **1** are highly homologous. Moreover, 2D NMR studies indicated that the oligomeric peptide possessed secondary structure that was essentially identical to the parent peptide. The NMR-derived structure of **BBA5** was thus predicted to be a satisfactory search model.

We first searched iteratively for three monomers of **BBA5**. Those experiments were unsuccessful. We postulated that this was because the angle between alpha and beta domains of each monomer unit differed substantially between peptide **1** and **BBA5**, thus several strategies to improve the search model were attempted. First, the **BBA5** helix and hairpin segments were used iteratively as search models, singly and in succession, to avoid defining an angle between subunits. Second, two virtual libraries, having 22 and 28 members, respectively, of different hinge dihedral angles were created. The libraries were designed by selecting allowed regions from the Ramachandran plots for glycine and alanine and inverting ϕ and ψ to account for the reversal of chirality. Models were then created from the **BBA5** NMR structure by expunging the three-residue (SRS) linker was and reconnecting the helix and hairpin with a D-alanine residue. The ϕ and ψ angles of D-alanine residues were varied to get the desired values (Figure 2-10).

Each of these models was tested for its ability to produce a molecular replacement solution. Finally, structures of parallel trimeric coiled coils were located from the Protein Data Bank, trimmed to the appropriate length if necessary, and used as search models for molecular replacement.

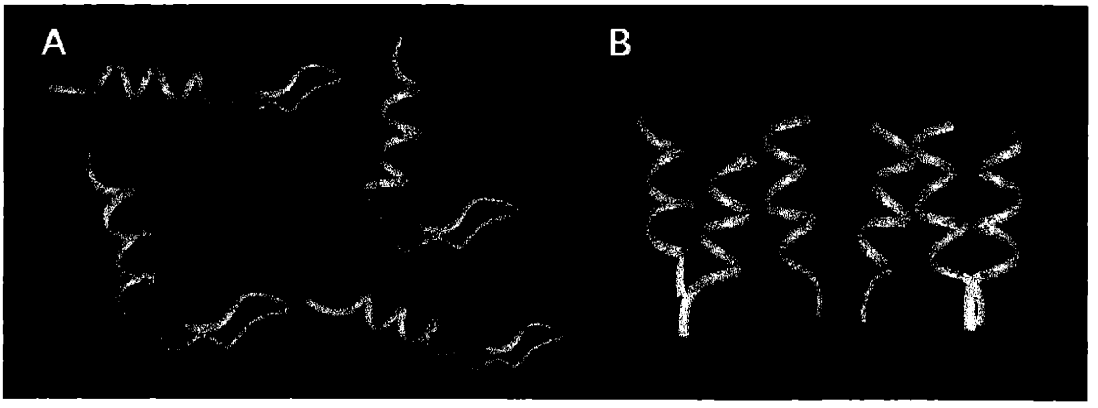


Figure 2-10. A. Representative members of the virtual hinge dihedral angle library. B. Representative members of the virtual coiled coil library.

Despite extensive experimentation with molecular replacement models, algorithms (AMoRe,[37] MOLREP,[38] EPMR[39,40]), resolution limits, and other parameters, no clear solution was found. This failure was attributed primarily to the lack of a highly homologous search model.

Several other factors may also have hindered a facile molecular replacement solution. (1) The primary search model was the monomeric **BBA5**. The oligomeric state of peptide **1** in the asymmetric unit meant that any one monomer model could contain at most one third (or one fourth) of the scattering mass in the asymmetric unit, and thus even a correct solution might have a low correlation coefficient, making it difficult to identify. A single domain would contain half as much of the scattering mass, making it even less likely that a correct solution could be identified. (2) Molecular replacement with a model derived from an NMR structure has been

reported to be problematic in a number of instances.[41,42] (3) It is noteworthy that coiled coils are generally held to be difficult structures to solve by molecular replacement.

We concluded that, in the absence of additional structural information leading to a better model, a molecular replacement approach was unlikely to yield a successful solution and that it would be preferable to focus on an approach that did not require any prior knowledge of the structure.

Heavy Atoms: Crystal Soaks

Heavy atoms introduced into a protein crystal by soaking can be used to obtain phase information by MIR or MAD or some combination of isomorphous replacement with anomalous signal (MIRAS, SIRAS). In order for this strategy to be successful the heavy atom soaked into the crystal must occupy defined sites within the lattice, and must not damage or alter the crystal lattice during the course of the soak.

MIR takes advantage of the greater X-ray scattering power of "heavy atoms" (generally metals in the lower d-block and the f-block of the periodic table) relative to the lighter atoms of carbon, nitrogen, and oxygen that are the majority of atoms in a protein. Two data sets collected in the presence and in the absence of a heavy metal can be compared, and the differences ascribed to the heavy atom. The heavy atom position can then be located using Patterson methods,[17] and the phases used as a starting approximation for the phases of the model as a whole. In principle, two derivative data sets having heavy atoms at unique positions within the unit cell are needed. In practice, often more than two derivatives are used.

MAD, or multi-wavelength anomalous diffraction, relies on the ability of heavy atoms to scatter X-rays anomalously near the absorption edge in a wavelength-dependent fashion.[17] Three wavelengths are typically collected: a peak data set, in which the anomalous signal is maximal, an inflection data set, in which the anomalous signal is half-maximal, and a remote data set in which the anomalous signal is diminished. As a single crystal containing an anomalously-scattering

atom provides two "derivative" and one native data sets, crystal non-isomorphism is eliminated as a source of error.[17]

Peptide **1** does not have any natural metal binding sites that can be exploited, nor does it contain any histidines, cysteines or methionines, the most common metal-binding amino acids for crystallography.[43] Our options included introducing one of these metal-binding residues into the BBA structure, as well as investigating other reactive peptide functional groups found in BBA. Aspartate and glutamate carboxylates react with lanthanides and other oxophilic metals, and the ε-amino group of lysine reacts with gold and platinum compounds.

One of our first attempts involved the preparation of cysteine-containing BBA derivatives so that we could prepare a mercury derivative for phasing purposes. Tucker et al. have had success with this strategy.[44] We introduced cysteine at two positions in peptide **1**. Peptide **C5**, containing a Ser5Cys substitution (Table 2-4), was very poorly soluble in phosphate buffer and in water, and could not be crystallized. We hoped that introducing a mercury compound could help to solubilize the peptide for crystallization, but **C5** was also insoluble in buffered aqueous $HgCl_2$.

Peptide **C20**, having a Dbz20Cys substitution, was designed to generate a derivative that would be approximately isosteric to DapBz after reaction with phenylmercuric chloride. It is also not very soluble. Peptide **C20** was reacted with PhHgCl, which did not increase solubility. Crystallization trials prepared with the reaction mixture (~5 mg/ml in H_2O) did not yield any crystals.

		hairpin									*helix*											
Peptide		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
1	Ac	Y	R	I	p	S	Y	D	F	a	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2
C5	Ac	Y	R	I	p	C	Y	D	F	a	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2
C20	Ac	Y	R	I	p	S	Y	D	F	a	D	E	L	A	K	L	L	R	Q	A	C	G	NH_2

Table 2-4. Cysteine derivatives of peptide **1**. Abbreviations: p=D-proline, a=D-alanine, Z=benzoylated L-α,β-diaminopropionic acid.

Introducing a heavy atom into a crystal can often result in damage to that crystal. In certain instances cross-linking of crystals with glutaraldehyde has been shown to result in increased mechanical stability and greater resistance to the perturbation of a heavy atom soak.[45] A few experiments were performed to determine whether cross-linking would be of use for crystals of peptide **1**. Crystals were cross-linked by glutaraldehyde vapor diffusion, dipped in 20% glycerol cryoprotectant, flash-frozen, and screened. It was found that cross-linking significantly reduced the resolution and increased the mosaicity of all the crystals screened. We thus chose to optimize heavy atom soaking conditions without cross-linking.

The affinity of peptide **1** for various heavy atoms, useful information for soaking experiments, was investigated by means of a gel shift assay prior to the soaking experiments.[46] As peptide **1** has no net charge, it does not migrate under native (non-denaturing) conditions during gel electrophoresis unless it is stably complexed to an ion. The peptide was thus incubated with various heavy atoms, subjected to native gel electrophoresis under standard and reverse polarity conditions, and the resultant gel examined for evidence of a complex.

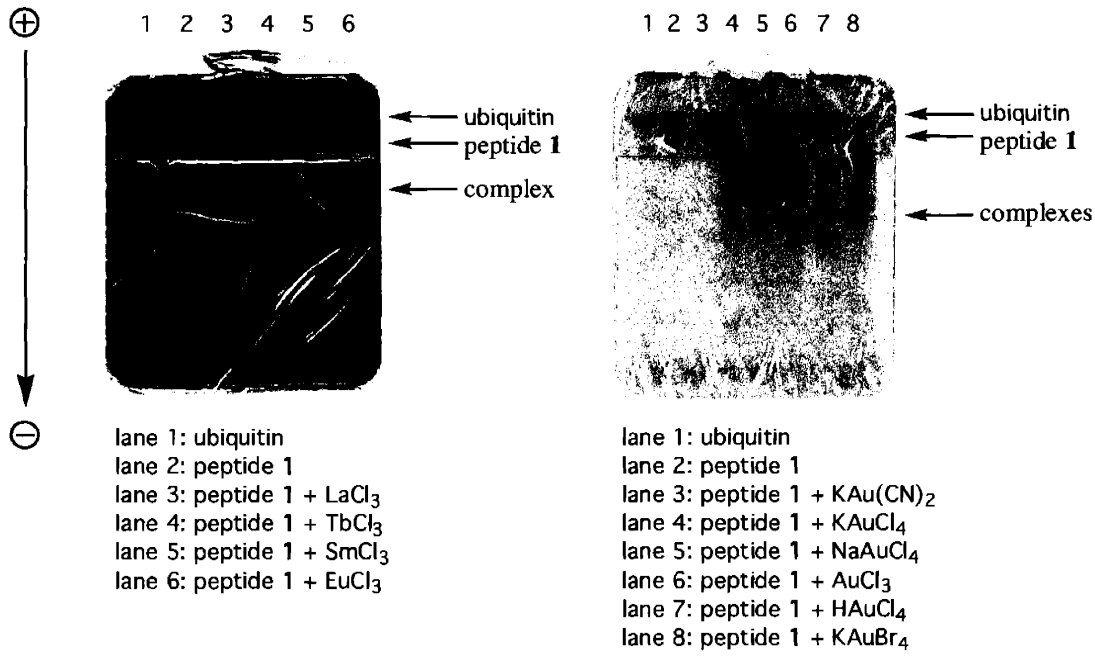


Figure 2-11. Peptide **1** interacts with a number of lanthanide (left) and gold (right) compounds to form positively charged complexes, suggesting that these compounds could make specific contacts in a crystal for heavy atom phasing.

Peptide **1** interacts to form discrete, positively-charged, complexes with the lanthanide salts $LaCl_3$, $TbCl_3$, $SmCl_3$, and $EuCl_3$ after overnight incubation, resulting in a single band on a native gel shifted towards the negative pole (Figure 2-11). Peptide **1** also forms non-specific complexes with a number of gold derivatives after overnight incubation (Figure 2-11).

The metals that were found to bind in a native gel assay were subsequently used for crystal-soaking experiments. $SmCl_3$ and $LaCl_3$ were found to result in significant crystal damage and poor diffraction when soaked into crystals of **1**, but soaks with $TbCl_3$ and $EuCl_3$ resulted in adequate diffraction. Data sets were collected from crystals of **1** soaked in $TbCl_3$, $EuCl_3$, $KAuCl_4$ and K_2PtCl_4. None of these data sets were of very high quality, and no usable phase information was obtained. $OsCl_3$ and Na_2WO_4 derivatives were also investigated (the former binds to carboxylates, and the latter could replace any tightly bound phosphates that may have crystallized

with the peptide). The data from the $OsCl_3$ and Na_2WO_4 derivatives were of poor quality and were not used further

Noble gas derivatives, particularly xenon and krypton, can also be used for obtaining phase information.[47,48] Rather than interacting with a specific functional group, the noble gas generally binds in a hydrophobic cavity of the protein, providing a way to derivatize proteins with few reactive functional groups. Noble gas derivatives are often more isomorphous than are other heavy atom derivatives.[47,48]

It was hypothesized that the hydrophobic core of peptide **1** might accommodate a noble gas. Attempts were made to introduce both krypton and xenon. Dr. Ezra Peisach collected a data set from a crystal of **1** pressurized with krypton after an XAFS scan of that crystal indicated Kr was weakly present. Anomalous and difference Patterson maps suggested that Kr could be present. However, we were unable to obtain phase information from this derivative either singly or in combination with other derivatives.

Triiodide ion has been reported to bind both in hydrophobic cavities and electrostatically on the surface.[49] Triiodide, generated *in situ* from the reaction of KI and I_2, was selected for further exploration because of the strong anomalous signal of iodine, and because of the unusual mode of binding of triiodide. As before, no usable phase information was obtained from this derivative.

A number of groups have successfully derivatized their crystals by briefly dragging them through concentrated solutions of NaI and NaBr.[50,51] We hoped that this method might be less damaging than a longer soak in a less concentrated metal solution. We screened a range of recommended concentrations of NaI and NaBr. Crystals were dipped into concentrated halide solution for *ca.* 30 seconds, flash-frozen, and irradiated. No diffraction or extreme "smeariness" was observed with 1.0 M and 0.75 M halide solutions. Two data sets were collected at 0.5 M NaI and 0.5 M NaBr, at which concentration the smeariness was considerably reduced. However, both

data sets had extremely high mosaicities. Ultimately neither data set yielded any useful information.

All efforts at heavy atom phasing of soaked crystals were hindered by the difficulty of introducing a large atom into a small and tightly-packed crystal. In many cases the crystals were visibly damaged or destroyed by soaking. In others the crystal appeared to the eye to have withstood soak-induced damage, but diffracted very poorly or not at all. Nonetheless, several usable data sets were collected (Table 2-5).

Derivative	Source	a (Å)	b (Å)	c (Å)	Wavelength(s) (Å)	Resolution (Å)
—	BNL X25	32.2	39.3	56.5	1.1000	1.5
—	APS 14-BM-C	32.3	38.9	56.2	0.9000	1.2
$EuCl_3$	BU	32.0	37.0	56.2	1.5418	3
$TbCl_3$	BNL X4A	32.1	40.8	60.0	1.65004, 1.64938, 1.62772	2.5
$KAuCl_4$	APS	31.9	41.2	56.5	1.0408	3.2
K_2PtCl_4	APS 14-BM-D	32.1	38.6	55.0	1.070	2.4
Kr	APS	32.2	39.2	57.5	0.8649	2.4
$KI+I_2$	APS 14-BM-D	32.3	39.4	57.3	1.6531	2.8
NaI	BU	32.0	38.8	55.6	1.5418	2.9
NaBr	BNL X4A	32.1	39.0	56.0	0.91980, 0.91959	1.9

Table 2-5. A comparison of the unit cells obtained from crystals of peptide **1** soaked with various metals, indicating a high degree of non-isomorphism between various derivatives. All crystals belonged to the orthorhombic space group $P2_12_12$, with $\alpha = \beta = \gamma = 90°$.

Extensive efforts were made to obtain phase information from these data sets, singly and in various combinations. Although a handful of attempts resulted in high figures of merit, as reported by SOLVE,[52] none resulted in resulted in interpretable electron density maps. The unsatisfactory outcomes of this approach were attributed to the significant non-isomorphism between native and derivative data sets, and between derivative data sets. Low occupancies of the heavy atoms may also have been a contributing factor. The non-isomorphism between derivatives was ascribed to the significant perturbation posed by the introduction of a large atom into a small unit cell.

The data collected in a diffraction experiment are the location and intensities of the measured reflections. The intensity of a reflection, I, is related to F, the magnitude of the structure factor describing that reflection, by the equation $|F|^2 = I$. A small difference in F can thus be translated to a measurable change in I.

Measurable changes in I, relative to a native data set, are the basis of the phasing power of a heavy atom in an MIR experiment. The quantity $\left(\frac{\Delta F}{|F|}\right)$ is the relative change in the magnitude of the structure factor due to the introduction of a heavy atom. It is related to the atomic number, Z_H, and to the number of heavy atoms in the asymmetric unit, N_H, by equation (4), where N_P is the number of (non-hydrogen) atoms in the protein $Z_{eff} \approx 6.7$ is an effective atomic number for a protein molecule.[46]

$$(4)\quad \left(\frac{\Delta F}{|F|}\right) = \sqrt{\frac{2N_H}{N_P}} \cdot \frac{Z_H}{Z_{eff}} = \sqrt{\frac{2N_H}{N_P}} \cdot \frac{Z_H}{6.7}$$

Any of the heavy atoms investigated, if present with complete occupancy throughout the crystal, would have resulted in measurable differences in intensity that could be used for MIR phasing, as shown in Table 2-6.

Atom	Z_H	$(\Delta F)/(\|F\|)$
I	53	47.2%
Sm	62	55.3%
Eu	63	56.1%
Gd	64	57.0%
Tb	65	57.9%
W	74	65.9%
Os	76	67.7%
Pt	78	69.5%
Au	79	70.4%

Table 2-6. Phasing power of the heavy atoms used in these experiments. Values calculated assuming three monomers (N_P=561) and one heavy atom (N_H=1) per asymmetric unit.

An MIR approach should thus be successful if the problems of incomplete occupancy and of non-isomorphism could be overcome. However, we were unable to improve data collection conditions by further screening of heavy atoms, soaking concentrations, and soak times. The intrusion of a heavy atom into the small unit cell resulted in damage to the crystals, high mosaicity, and significant non-isomorphism.

We thus elected to focus our efforts on introducing a heavy atom covalently at the peptide level, thus ensuring that N_H=3 for all crystals having three molecules per asymmetric unit. Introducing a heavy atom that anomalously scatters X-rays would obviate the need for isomorphism with another derivative.

Peptide-Level Incorporation of Heavy Atoms

Mutational studies indicated that the oligomeric BBA is highly sensitive to mutations in its hydrophobic core.[13] Conservative substitutions such as halogenated and sulfur-containing amino acids were investigated first.

In several instances a tyrosine residue has been iodinated, in a protein crystal, *via* an electrophilic aromatic substitution reaction, and the resultant *o*-iodotyrosine used for phase determination.[53-56] Our experiences with soaking heavy atoms into crystals of **1** suggested that iodine soaks might not be feasible for reasons of crystal damage and of incomplete occupancy.

Instead we chose to chemically synthesize four derivatives (Table 2-7) halogenated at tyrosine-1 and tyrosine-6, thus guaranteeing homogeneity of the peptide. Two commercially available amino acids, one an iodotyrosine derivative, and the other a bromophenylalanine derivative, were investigated (Figure 2-12).

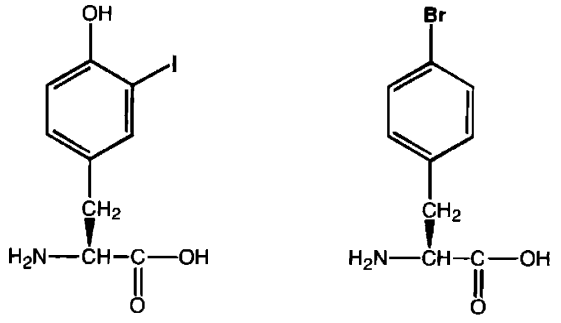


Figure 2-12. Halogenated aromatic amino acids introduced into peptide **1** for phasing purposes.

		hairpin									*helix*											
Peptide		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
1	Ac	Y	R	I	p	S	Y	D	F	a	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2
I1	Ac	IY	R	I	p	S	Y	D	F	a	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2
I6	Ac	Y	R	I	p	S	IY	D	F	a	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2
Br1	Ac	BF	R	I	p	S	Y	D	F	a	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2
Br6	Ac	Y	R	I	p	S	BF	D	F	a	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2

Table 2-7. Sequences of halogenated derivatives of peptide **1**. Abbreviations: p=D-proline, a=D-alanine, Z=benzoylated L-α,β-diaminopropionic acid, IY=*o*-iodotyrosine, BF=*p*-bromophenylalanine.

Peptide **I1** is very insoluble and the concentrations necessary for crystallization could not be achieved. Peptide **I6** was initially deemed more promising, as initial crystallizations yielded a large number of crystals having the same morphology as the parent peptide. Data were collected to 1.8 Å resolution. When these data were processed there was no evidence of anomalous scatter due to the iodine atom. Mass spectroscopic analysis of a dissolved crystal from the same well confirmed the absence of the iodine. Crystallographic lore suggests that iodinated aromatic derivatives are very light-sensitive. One explanation for these findings may be light-catalyzed homolytic cleavage of the C-I bond (BDE ~65 kcal/mol[57]), followed by H-atom abstraction from the solvent. Crystallization of **I6** from the same conditions was repeated, unsuccessfully, in a darkroom, suggesting that only the dehalogenated, native, peptide could be crystallized.

Peptide **Br1** was also poorly soluble in phosphate buffer. A solution in water (1.6 mM, 4.3 mg/ml), in which it is somewhat more soluble, was used to set up several crystallizations

using condition **a**. These yielded small, thin, branch-like crystals unsuitable for obtaining phase information. A crystal screen of peptide **Br6** yielded only precipitate.

More concentrated solutions of the halogenated amino acids could be achieved by dissolving the peptides with a small fraction of the organic cosolvent *t*-butanol. In all cases this led to precipitation of the peptides after setup.

Heavy halogens were also introduced in the benzoyl cap of DapBz (peptides **Z1**, **Z2**, **Z3**, Figure 2-13 and Table 2-8). Each of these peptides was poorly soluble in aqueous solutions, but could be solubilized in 1:1 phosphate buffer: *t*-butanol. Attempts to crystallize these peptides were unsuccessful.

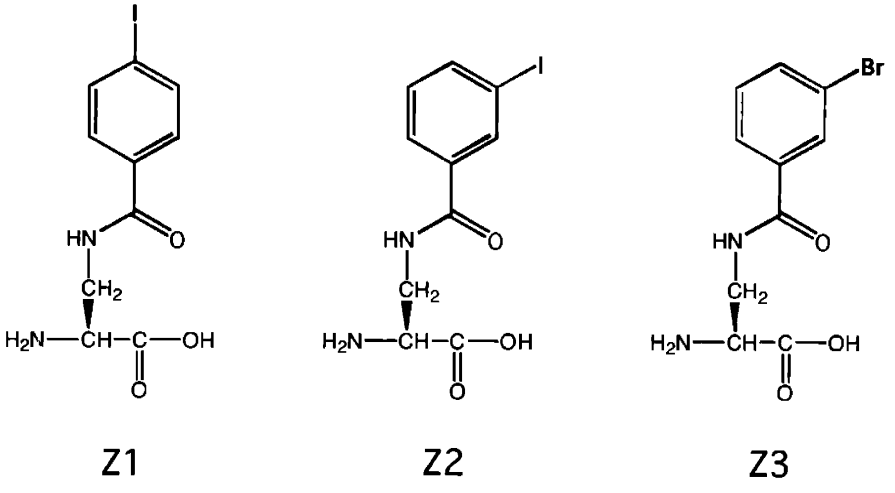


Figure 2-13. Halogenated derivatives of DapBz-20.

		hairpin									*helix*												
Peptide		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
1	Ac	Y	R	I	p	S	Y	D	F	a	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2
Dbz1	Ac	Y	R	I	p	S	Y	D	Y	a	D	E	L	A	K	L	L	R	Q	A	Z1	G	NH_2
Dbz2	Ac	Y	R	I	p	S	Y	D	Y	a	D	E	L	A	K	L	L	R	Q	A	Z2	G	NH_2
Dbz3	Ac	Y	R	I	p	S	Y	D	Y	a	D	E	L	A	K	L	L	R	Q	A	Z3	G	NH_2

Table 2-8. Derivatives of peptide **1** halogenated at DapBz-20. Abbreviations: p=D-proline, a=D-alanine, Z=benzoylated L-α,β-diaminopropionic acid. Z1, Z2, and Z3 are defined in Figure 2-13.

The poor solubility of these peptides in aqueous solutions indicated that peptide **1** is tightly packed and susceptible to structural disruption by seemingly minor changes to its hydrophobic core. These structural disruptions could result in nonspecific aggregation and poor solubility. We thus became interested in exploring heavy atom substitutions at charged, solvent-exposed residues, which we expected to be unlikely to affect critical contacts necessary for tertiary and quaternary structure.

γ-Thialysine (Figure 2-14) is an isosteric lysine analog that has been used to perturb the pK_a of an active site lysine[58] and for chemical modification rescue of a Lys → Cys mutant.[59] The substitution of γ-thialysine for lysine-14, producing peptide **γK**, was expected to be non-perturbing, and to result in a native-like derivative. Phases could be obtained for crystals of peptide **γK** from the anomalous signal of the sulfur atom.[60-62]

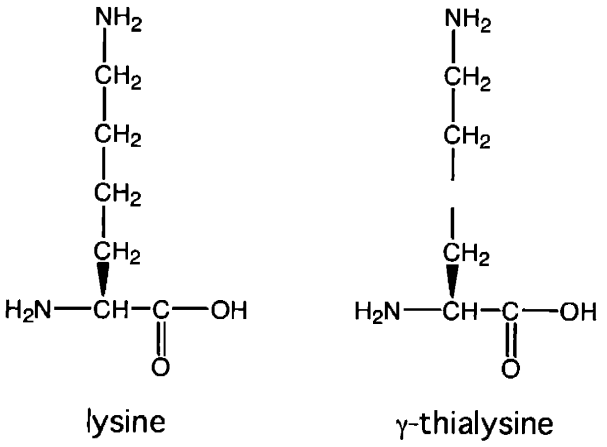


Figure 2-14. Chemical structures of lysine and γ-thialysine.

To synthesize peptide **γK**, a Lys14Cys mutant was prepared, largely with an automated peptide synthesizer, with the exception of the critical cysteine residue, which was hand-coupled using conditions known to minimize racemization.[63] The cysteine residue was orthogonally protected as Cys(MMT), selectively deprotected on-resin following completion of the synthesis, and coupled with $Boc(H)NCH_2CH_2Br$ (Figure 2-15).

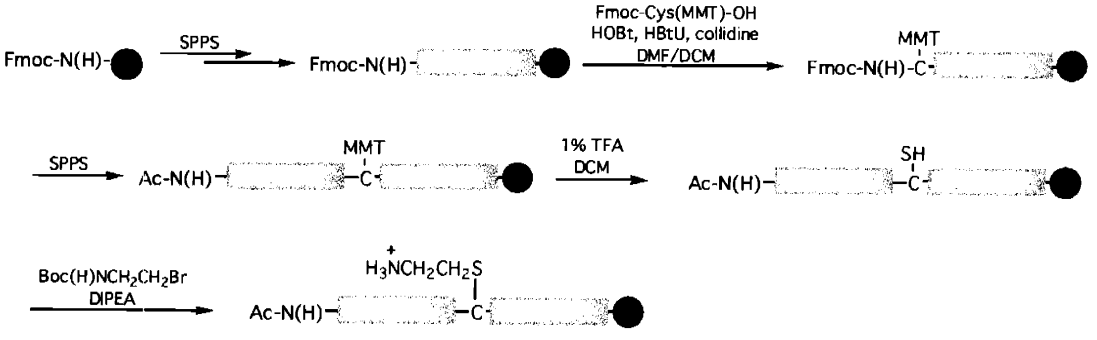
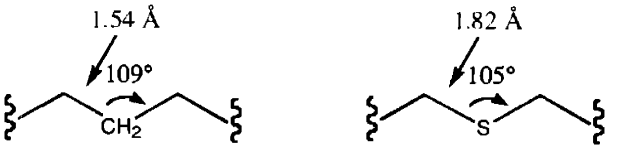


Figure 2-15. Synthetic scheme for the solid-phase synthesis of peptide **γK**.

Peptide **γK** is considerably less soluble than the parent peptide **1**. A similar effect has been reported for methionine to selenomethionine substitutions, where the difference in solubility

has been attributed to the greater hydrophobic surface area in selenomethionine.[64] This may also be the case for a methylene to thioether substitution (Figure 2-16). The limited solubility of the peptide makes it difficult to achieve the high concentrations necessary for crystallization, and makes simple manipulations, e.g. low-speed spinning to remove particulates, difficult because of the tendency of the peptide to precipitate.



electronegativity	2.55	2.58
atomic radius (Å)	0.91	1.27
covalent radius (Å)	0.77	1.02

Figure 2-16. A comparison of the methylene and thioether linkages.

The circular dichroism spectrum of **γK** has the same features as that of peptide **1**, but a much weaker ellipticity (Figure 2-17). This suggests that it associates more weakly but in the same manner as peptide **1**. Neither peptide exhibits significant pH dependence by circular dichroism, and both appear to associate over the pH range 4-8, as indicated by a ratio of $\theta_{222\ nm}/\theta_{208\ nm}$ greater than one.[14,15]

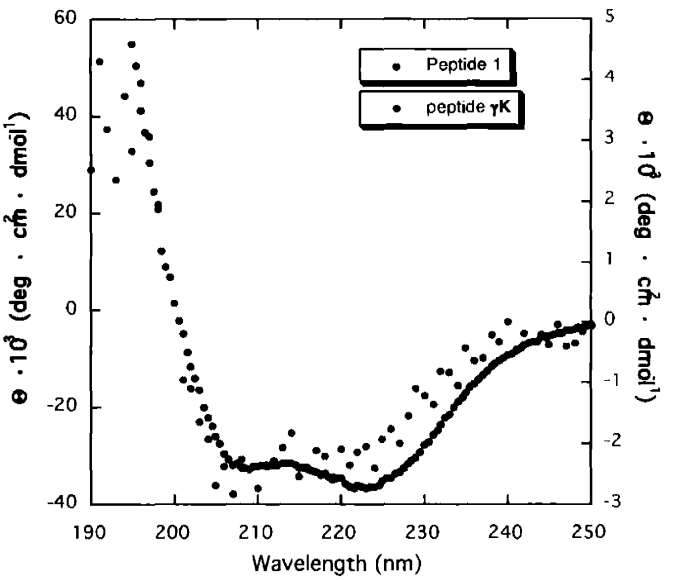


Figure 2-17. Circular dichroism spectra of 50 μM **1** and **γK** in 10 mM phosphate buffer, pH 7.

Several attempts were made to establish that peptide **γK** associates by analytical ultracentrifugation. The first experiment performed indicated a degree of association, between a dimer and a trimer, in aqueous solution. However, three subsequent experiments were unable to reproduce this finding, and indicated that the peptide behaved as an ideal monomer in solution. Peptide **γK** exhibits a tendency to precipitate upon spinning at low speed in a benchtop centrifuge, a tendency which may also have manifested itself upon several days of spinning at higher speed in an analytical ultracentrifuge. Spinning at the elevated speeds necessary for sedimentation of the peptide may have precipitated all but the monomeric species in the latter experiments.

The crystallization of **γK** was attempted numerous times. Conditions **a** and **b**, which yield high-quality crystals of **1**, did not result in crystal growth over several trials. Microseeding from a native crystal of **1** was next attempted. Three commercial additive screens (Hampton Research) were investigated in conjunction with crystallization condition **b**. Two commercial sparse matrix

crystallization screens (Hampton Research) were also evaluated. In all cases, **γK** either precipitated upon setup, precipitated shortly after setup and/or oiled out of solution.

The poor solubility of **γK** made it difficult to achieve the concentrated solutions in aqueous buffer suitable for crystallization. Several crystallization solutions were made by easing the peptide into solution with the addition of a small volume of *t*-butanol, acetonitrile, or other organic solvent. We reasoned that the organic cosolvent should not interfere with crystallization, as upon equilibration with a reservoir solution having a 200-300x greater volume, the proportion of the cosolvent in the crystallization drop would be significantly lower than in the initial peptide stock solution. Unfortunately, what was observed was precipitation, rather than crystallization, of the peptide as the proportion of the solubilizing organic cosolvent decreased.

After a number of covalent derivatives had been explored, and found unsuccessful, we chose to investigate selenomethionine substitutions. Selenomethionine, an analog of methionine with good anomalous scattering properties, is the most common covalently-incorporated heavy atom derivative.[65] Peptide **1** does not contain any methionines, precluding a simple methionine to selenomethionine substitution. As mutational studies had suggested that substitutions at Gln-18 were less disruptive than at other positions in peptide **1**,[13] a Gln18SeMet mutant was prepared (Table 2-9). An analytical ultracentrifugation experiment indicated that selenomethionine peptide was oligomeric in solution, with a stoichiometry intermediate between two and three (Table 2-10).

		hairpin									*helix*											
Peptide		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
1	Ac	Y	R	I	p	S	Y	D	F	a	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2
Gln18SeMet	Ac	Y	R	I	p	S	Y	D	Y	a	D	E	L	A	K	L	L	R	Se	A	Z	G	NH_2

Table 2-9. Sequence of Gln18SeMet derivative of peptide **1**. Abbreviations: p=D-proline, a=D-alanine, Z=benzoylated L-α,β-diaminopropionic acid.

	Concentration (μM)		
	20	100	500
oligomeric state	2.6	2.8	2.7

Table 2-10. Initial analytical ultracentrifugation data for Gln18SeMet peptide. Data were collected at 45,000 rpm, 5 °C, and analyzed with a linear plot of ln(abs) vs $\left(r^2/2\right)$.

This derivative crystallizes readily over a few hours from condition **b** to yield small, plate-like crystals. It has a smaller unit cell than does the native **1**, with dimensions a = 49.89 Å, b = 31.85 Å, c = 21.85 Å, $\alpha = \beta = \gamma = 90°$. The space group could not be unambiguously determined by examination of the systematic absences. Crystal density measurements indicate that there are 1.6 monomers per asymmetric unit.

A MAD experiment was performed by Dr. Geoffrey F. Stamper at the 14-BMD beam line at the BIOCARS facility at Argonne National Laboratory. Data were collected at two wavelengths (0.9778 Å, edge and 0.97930 Å, peak) to ~2.3 Å. Despite the presence of several promising peaks on an anomalous Patterson map, the structure could not be solved by MAD phasing. Selenium substructure searches[66] were also unsuccessful.

We attributed this inability to obtain phase information to a heterogeneous oxidation state at the selenium atom. Selenoethers are prone to oxidation to the corresponding selenoxides or selenones.[67] In a protein, the extent of selenomethionine oxidation is variable and dependent upon the local environment. Oxidation results in heterogeneity, which could be problematic from the vantage of crystallization, although crystallization is frequently unaffected.[64] More significantly, oxidation also affects the MAD experiment, as the selenium K edge is shifted to higher energy, broadened, and the white line of the peak reduced or split. A 'smeary' anomalous signal from an electronically heterogeneous sample may preclude obtaining phase information from a MAD experiment.[64,68]

Although the initial experiment was unsuccessful, it was anticipated that the structure of the native peptide **1** could be determined by a MAD experiment given (a) a selenomethionine

derivative structurally homologous to **1**, (b) a homogeneous selenium oxidation state in the selenomethionine derivative, and (c) a suitable selection of wavelengths for the MAD experiment. We therefore sought to prepare, in parallel, other selenomethionine derivatives which might prove to have more suitable properties.

Conclusions

Optimal conditions for the growth of robust and strongly-diffracting crystals of peptide **1** were determined. Extensive attempts were made to determine initial phases. Direct methods approaches failed because of the large size of the target at the moderate atomic resolution (1.2 Å) of the data. Molecular replacement, primarily with search models derived from the NMR structure of **BBA5**, was unsuccessful because of the lack of a homologous search model. It was found that large heavy atoms could not be introduced into a small unit cell by soaking of a crystal without causing crystal damage and non-isomorphism. Several mutants covalently incorporating heavy atoms were prepared, and found to differ substantially from the native **1**, indicating that peptide **1** is tightly packed and that seemingly small changes can disrupt the native structure. A Gln18SeMet derivative of peptide **1** showed considerable promise. The failure to obtain phases from a MAD experiment on crystals of Gln18SeMet was attributed to electronic heterogeneity at selenium consequent to heterogeneous oxidation.

Acknowledgements

I am grateful to Dr. Geoffrey F. Stamper and Professor Catherine L. Drennan for their assistance in the earliest stages of this project. I thank Dr. Kevin A. McDonnell for assistance and helpful discussions at the start of this project. I am indebted to Professor Karen N. Allen for welcoming me into her lab, and for teaching me crystallography. I thank Dr. Ezra Peisach for experimental help, useful suggestions, and computer assistance. All of the synchrotron data sets were collected by Professor Karen N. Allen, Dr. Ezra Peisach, and Dr. Tracy L. Arakaki, for

which I am extremely grateful. Professor George M. Sheldrick provided us with several helpful suggestions regarding the implementation of SHELXS and SHELXD.

Materials and Methods

Peptide Synthesis

Peptides were prepared by standard Fmoc-based solid-phase peptide synthesis. Unless otherwise noted, all peptide synthesis reagents were purchased from Applied Biosystems or Novabiochem. The *C*-terminal dipeptide was coupled by hand using HBtU/HOBt/DIPEA chemistry. DapBz was incorporated as Fmoc-Dap(alloc)-OH (Bachem), deprotected on the solid phase resin,[69] and side chain capped with benzoic acid or a halogenated benzoic acid (Aldrich). The synthesis was then completed using a 9050 PepSynthesizer™ or an Advanced ChemTech 396W peptide synthesizer. Fmoc-Phe(4-Br)-OH and Fmoc-Phe(*o*-I)-OH were purchased from Advanced ChemTech. Fmoc-SeMet-OH (AnaSpec) was coupled by hand with PyBOP/HOBt/DIPEA.

Cleavage and Purification

Peptides were cleaved from the resin for 2-6 hours with 2.5% triisopropylsilane, 2.5% water, and 95% trifluoroacetic acid, after which the cleavage cocktail was filtered to remove the resin and the filtrate concentrated under a gentle stream of nitrogen. When the volume was between ~1-2 ml the peptide was precipitated from ice-cold 2:1 ether:hexane, and centrifuged. The supernatant was decanted and the pellet resuspended in ice-cold 2:1 ether:hexane and washed 1-2 more times. The supernatant was then decanted, and the pellet dissolved in ~10% acetonitrile and immediately purified or lyophilized. High performance liquid chromatography (HPLC) was performed using a Waters Prep LC 4000 System equipped with a Waters 2487 absorbance detector (set at 228 and 280 nm), a C_{18} reverse-phase analytical column (Beckman Ultrasphere, 5 μm, 4.6 x 150 mm) and a C_{18} reverse-phase preparative column (YMC-pack, 250 x 20 mm). A standard gradient of 80:20

to 20:80 (water:acetonitrile, 0.1% TFA) over 25 minutes (with a flow rate of 1 ml/min on the analytical column, and 15 ml/min on the preparative column) was employed in all cases. Purity was determined by electrospray mass spectroscopy (PerSeptive Biosystems Mariner™ BioSpectrometry Workstation, using the Turbo Ion Source) and by analytical HPLC.

	Calculated			Observed		
Peptide	**M**	$(MH_2)^{2+}$	$(MH_3)^{3+}$	**M**	$(MH_2)^{2+}$	$(MH_3)^{3+}$
1	2556.3	1279.2	853.1	2558.1	1280.1	853.7
C5	2572.3	1287.2	858.4	2573.2	1287.6	858.7
C20	2469.2	1235.6	824.1	2469.4	1235.7	—
I1	2681.9	1341.9	895.0	2683.0	1342.5	895.3
I6	2681.9	1341.9	895.0	2682.8	1342.4	895.3
Br1	2619.0	1310.5	874.0	2621.0	1311.5	874.3
Br6	2619.0	1310.5	874.0	2619.8	1310.9	874.3
Dbz1	2681.9	1341.9	895.0	2683.0	1342.5	895.3
Dbz2	2681.9	1341.9	895.0	2683.0	1342.5	895.3
Dbz3	2635.0	1318.5	879.3	1635.0	1318.5	879.3
γK	2574.3	1288.1	859.1	2574.8	1288.4	859.3
SeMet	2606.0	1304.0	869.7	2607.8	1304.9	870.2

Table 2-11. Mass spectrometry data for peptide **1** and derivatives.

γK Synthesis

The *C*-terminal dipeptide was hand-coupled, followed by automated synthesis through Cys-14. Fmoc-Cys(MMT)-OH (Novabiochem) was hand-coupled under racemization-free conditions (HBtU, HOBt, collidine, 1:1 DCM:DMF),[63] and the synthesis completed using a 9050 PepSynthesizer™ as above. Cys(MMT) was deprotected using 95:1:5 DCM:TFA:TIS (3 x 15ml x 2 min) and then coupled with $Boc(H)NCH_2CH_2Br$ (Fluka, 20 eq amine + 40 eq *N,N*-diisopropylethylamine). After several rounds of coupling, but incomplete conversion to the γ-thialysine derivative, resin was cleaved with 2.5% 1,2-ethanedithiol, 2.5% water, 1% triisopropylsilane, and 94% trifluoroacetic acid.

Extinction Coefficient Determination

The extinction coefficient of peptide **1** was determined by correlating the absorbance of a concentrated stock solution at 280 nm (measured with a Shimadzu UV160U spectrophotometer) with its concentration, determined by duplicate quantitative amino acid analysis measurements. The extinction coefficient of peptide **1** at 280 nm was thus found to be 3090 cm^{-1} M^{-1}.

Circular Dichroism

Peptides were dissolved in degassed 10 mM phosphate buffer, pH 6.7. Stock solution concentrations were determined based on absorbance at 280 nm, and appropriate dilutions used to prepare 50 μM samples for analysis. Data were collected with an Aviv Circular Dichroism Spectrometer Model 202 using strain-free quartz cells having a path length of 0.1 cm. Wavelength scan experiments were performed in duplicate, from 340 to 192 nm, at 25 °C, with an averaging time of 0.5 s and a step size of 1.0 nm.

Analytical Ultracentrifugation

Concentrated stock solutions of peptides were prepared in 10 mM phosphate buffer at pH 7.2. Stock solution concentrations were determined based on absorbance at 280 nm, and dilutions made to cover a range of concentrations. Peptide solutions were loaded into standard two-sector epon centerpieces and spun, at 25 °C, in a Beckman XL-I analytical ultracentrifuge at 30,000, 40,000 and 50,000 rpm for eighteen to twenty-four hours at each speed. Data were analyzed using NONLIN[70] and plots of ln(abs) vs $\left(r^2/2\right)$.[71] Molecular weights were obtained from sigma using a partial specific volume, $\bar{\nu}$, calculated in SEDNTERP.

Native Gel Electrophoresis

Aliquots of peptide **1** (3 μl of a 1 mM stock solution in 10 mM HEPES buffer, pH 7.2) were incubated overnight with 4 μl of heavy atom solution (100 mM or greater, Hampton Research and Aldrich) and 2 μl of glycerol. Solutions were loaded onto a high density PhastGel™ (Amersham Biosciences) and subjected to electrophoresis under native conditions using a PhastSystem™ (Amersham Biosciences). All mixtures were subjected to electrophoresis with both standard and reverse electrode polarities. Gels were visualized by staining with Coomassie Blue.

Crystallization

Peptide 1. Peptide **1** was dissolved at ≥8 mg/ml in 10 mM phosphate at pH 7.2 for crystallization. Initial crystallization conditions were obtained by assaying several sparse-matrix screens (Hampton Research, Emerald BioStructures). Crystals were grown using vapor diffusion with hanging-drop geometry by mixing 1.5 μl of protein with an equal volume of reservoir solution.

Derivatives. Solutions of heavy-atom-incorporating peptides were prepared to be as concentrated as possible (in most cases well below 8 mg/ml). If necessary, the peptides were solubilized by the addition of a small quantity of an organic cosolvent (acetonitrile or *t*-butanol). Crystals were grown using vapor diffusion with hanging-drop geometry by mixing 1.5 μl of protein with an equal volume of reservoir solution. Crystallization conditions for peptide **1** (**a**: 25% *t*-butanol in 0.1M TRIS buffer, pH 8.5, **b**: 100 mM Na HEPES buffer pH 7.5, 10% *v/v i*-propanol, 20% w/v PEG 4000) were assayed before commercial sparse matrix kits.

Crystal Density Measurements

Individual solutions were prepared as described in the literature.[24] Solution densities were measured or calculated following the literature reference. One to three crystals were gently placed in clear centrifuge tubes containing ficoll solutions of varying densities. The crystals were

allowed to equilibrate overnight on the benchtop or were gently spun at low speed in a benchtop centrifuge in order to reach equilibrium. The tubes were then carefully examined under a microscope to determine the location of the crystal in the tube.

Data Collection

*Native Peptide **1***. Crystals were frozen in a stream of N_2 gas cooled to -180 °C using Paratone-N (Hampton Research) as a cryoprotectant. A 1.2 Å data set was collected at the 14-BM-C beam line at the BIOCARS facility at Argonne National Laboratory. Data were collected in two sweeps of varied exposure time to ensure high completeness of both high resolution and low resolution data. The DENZO and SCALEPACK package[19] was used for data indexing, reduction, and scaling.

*Crystal Soaks of **1***. Stock solutions of heavy atom reagents were prepared by dissolving the metal in saturated mother liquor obtained from the reservoir of a crystal-containing well. Heavy atom solution was then added to the drop for a final concentration of ~10 mM, and the crystals soaked for one hour to overnight, during which period the crystals were monitored for signs of damage. Intact-appearing soaked crystals were then briefly immersed in cryoprotectant (20% glycerol in saturated mother liquor or Paratone-N (Hampton Research)) and frozen in a stream of N_2 gas cooled to -180 °C. Data sets were collected at the Boston University Core Facility for Macromolecular Crystallography and various beam lines of the Advanced Photon Source, Argonne National Laboratory and the National Synchrotron Light Source, Brookhaven National Laboratory. The DENZO and SCALEPACK package[19] was used for data indexing, reduction, and scaling.

Phasing

SnB[33] and SHELXD[31] were used to search for a direct methods solution. The following programs were used for molecular replacement: AMoRe,[37] MOLREP,[38] and EPMR.[39,40] Heavy-atom sites were

searched for using the automated heavy atom solution program SOLVE.[52] Patterson maps were generated using XtalView.[72] Electron density maps were viewed in the graphics program O.[73]

References

(1) Goodsell, D. S.; Olson, A. J. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 105-153. Structural Symmetry and Protein Function.

(2) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *Science* **1996**, *271*, 342-345. Design of a Monomeric 23-Residue Polypeptide with Defined Tertiary Structure.

(3) Struthers, M. D.; Ottesen, J. J.; Imperiali, B. *Fold. Des.* **1998**, *3*, 95-103. Design and NMR Analyses of Compact, Independently Folded BBA Motifs.

(4) Liu, Y.; Eisenberg, D. *Protein Sci.* **2002**, *11*, 1285-1299. 3D Domain Swapping: As Domains Continue to Swap.

(5) Heringa, J.; Taylor, W. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 416-421. Three-Dimensional Domain Duplication, Swapping and Stealing.

(6) Bennett, M. J.; Schlunegger, M. P.; Eisenberg, D. *Protein Sci.* **1995**, *4*, 2455-2468. 3D Domain Swapping: A Mechanism for Oligomer Assembly.

(7) Bennett, M. J.; Choe, S.; Eisenberg, D. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3127-3131. Domain Swapping: Entangling Alliances Between Proteins.

(8) Ogihara, N. L.; Ghirlanda, G.; Bryson, J. W.; Gingery, M.; DeGrado, W. F.; Eisenberg, D. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 1404-1409. Design of Three-Dimensional Domain-Swapped Dimers and Fibrous Oligomers.

(9) Pei, X. Y.; Holliger, P.; Murzin, A. G.; Williams, R. L. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 9637-9642. The 2.0-Å Resolution Crystal Structure of a Trimeric Antibody Fragment with Noncognate V_H-V_L Domain Pairs Shows a Rearrangement of V_H CDR3.

(10) Albright, R. A.; Mossing, M. C.; Matthews, B. W. *Biochemistry* **1996**, *35*, 735-742. High-Resolution Structure of an Engineered Cro Monomer Shows Changes in Conformation Relative to the Native Dimer.

(11) Mezo, A. R.; Ottesen, J. J.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 1002-1003. Discovery and Characterization of a Discretely Folded Homotrimeric Peptide.

(12) Mezo, A. R.; Cheng, R. P.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 3885-3891. Oligomerization of Uniquely Folded Mini-Protein Motifs: Development of a Homotrimeric Beta Beta Alpha Peptide.

(13) McDonnell, K. A. Ph.D. Thesis, Massachusetts Institute of Technology, 2001. Towards Incorporation of Catalytic Function Into Small Folded Peptide Scaffolds.

(14) Zhou, N. E.; McKay, C.; Hodges, R. S. *Biochemistry* **1992**, *31*, 5739-5746. 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.

(15) Zhou, N. E.; Zhu, B.-Y.; Kay, C. M.; Hodges, R. S. *Biopolymers* **1992**, *32*, 419-426. The Two-Stranded α-Helical Coiled-Coil is an Ideal Model for Studying Protein Stability and Subunit Interactions.

(16) McDonnell, K. A.; Imperiali, B. *J. Am. Chem. Soc.* **2002**, *124*, 428-433. Oligomeric Beta Beta Alpha Miniprotein Motifs: Pivotal Role of Single Hinge Residue in Determining the Oligomeric State.

(17) Drenth, J. *Principles of Protein X-Ray Crystallography*; 2nd ed.; Springer: New York, 1994.

(18) Jancarik, J.; Kim, S. H. *J. Appl. Cryst.* **1991**, *24*, 409-411. Sparse Matrix Sampling: A Screening Method for Crystallization of Proteins.
(19) Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307-326. Processing of X-ray Diffraction Data Collected in Oscillation Mode.
(20) Matthews, B. W. *J. Mol. Biol* **1968**, *33*, 491-497. Solvent Content of Protein Crystals.
(21) Tong, L.; Rossman, M. G. *Acta Crystallogr., Sect. A* **1990**, *46*, 783-792. The Locked Rotation Function.
(22) Collaborative Computational Project, N. *Acta Crystallogr., Sect. D* **1994**, *50*, 760-763. The *CCP4* Suite: Programs for Protein Crystallography.
(23) Vonrhein, C.; Schulz, G. E. *Acta Crystallogr., Sect. D* **1999**, *55*, 225-229. Locating Proper Non-Crystallographic Symmetry in Low-Resolution Electron-Density Maps with the Program GETAX.
(24) Westbrook, E. M. *Methods Enzymol.* **1985**, *114*, 187-196. Crystal Density Measurements Using Aqueous Ficoll Solutions.
(25) Charlwood, P. A. *J. Am. Chem. Soc.* **1956**, *79*, 776-781. Partial Specific Volumes of Proteins in Relation to Composition and Environment.
(26) McMeekin, T. L.; Marshall, K. *Science* **1952**, *116*, 142-143. Specific Volumes of Proteins and the Relationship to their Amino Acid Contents.
(27) Durchschlag, H. In *Thermodynamic Data for Biochemistry and Biotechnology*; Hinz, H.-J., Ed.; Springer-Verlag: Berlin, 1986.
(28) Dauter, Z.; Lamzin, V. S.; Wilson, K. S. *Curr. Opin. Struct. Biol.* **1997**, *7*, 681-688. The Benefits of Atomic Resolution.
(29) Usón, I.; Sheldrick, G. M. *Curr. Opin. Struct. Biol.* **1999**, *9*, 643-648. Advances in Direct Methods for Protein Crystallography.
(30) Deacon, A. M.; Weeks, C. M.; Miller, R.; Ealick, S. E. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9284-9289. The Shake-and-Bake Structure Determination of Triclinic Lysozyme.
(31) Sheldrick, G. M. *Acta Crystallogr., Sect. A* **1990**, *46*, 467-473. Phase Annealing in SHELX-90 - Direct Methods for Larger Structures.
(32) Morris, R. J.; Bricogne, G. *Acta Crystallogr., Sect. D* **2003**, *59*, 615-617. Sheldrick's 1.2 Å Rule and Beyond.
(33) Miller, R.; Gallo, S. M.; Khalak, H. G.; M.Weeks, C. *J. Appl. Cryst.* **1994**, *27*, 613-621. SnB: crystal structure determination via Shake-and-Bake.
(34) Sheldrick, G. M. *Acta Crystallogr., Sect. A Suppl. S* **1984**, *40*, C440-C440. SHELX-84 - A Program System for Crystal-Structure Solution and Refinement.
(35) Schneider, T. R.; Sheldrick, G. M. *Acta Crystallogr., Sect. D* **2002**, *58*, 1772-1779. Substructure Solution with SHELXD.
(36) Taylor, G. *Acta Crystallogr., Sect. D* **2003**, *59*, 1881-1890. The Phase Problem.
(37) Navaza, J. *Acta Crystallogr., Sect. A* **1994**, *50*, 157-163. *AMoRe*: an Automated Package for Molecular Replacement.
(38) Vagin, A.; Teplyakov, A. *J. Appl. Cryst.* **1997**, *30*, 1022-1025. MOLREP: An Automated Program For Molecular Replacement.
(39) Kissinger, C. R.; Gehlhaar, D. K.; Fogel, D. B. *Acta Crystallogr., Sect. D* **1999**, *55*. Rapid Automated Molecular Replacement by Evolutionary Search.
(40) Kissinger, C. R.; Gehlhaar, D. K.; Smith, B. A.; Bouzida, D. *Acta Crystallogr., Sect. D* **2001**, *57*, 1474-1479. Molecular Replacement by Evolutionary Search.
(41) Chen, Y. W.; Dodson, E. J.; Kleywegt, G. J. *Structure* **2000**, *8*, R213-R220. Does NMR Mean "Not for Molecular Replacement"? Using NMR-Based Search Models to Solve Protein Crystal Structures.
(42) Chen, Y. W. *Acta Crystallogr., Sect. D* **2001**, *57*, 1457-1461. Solution Solution: Using NMR Models for Molecular Replacement.

(43) Petsko, G. A. *Methods Enzymol.* **1985**, *114*, 147-167. Preparation of Isomorphous Heavy-Atom Derivatives.

(44) Tucker, A. D.; Baty, D.; Parker, M. W.; Pattus, F.; Lazdunski, C.; Tsernoglou, D. *Protein Eng.* **1989**, *2*, 399-405. Crystallographic Phases Through Genetic Engineering: Experiences With Colicin A.

(45) Lusty, C. J. *J. Appl. Cryst.* **1999**, *32*, 106-112. A Gentle Vapor-Diffusion Technique for Cross-Linking of Protein Crystals For Cryocrystallography.

(46) Boggon, T. J.; Shapiro, L. *Structure* **2000**, *8*, R143-R149. Screening for Phasing Atoms in Protein Crystallography.

(47) Cohen, A.; Ellis, P.; Kresge, N.; Soltis, S. M. *Acta Crystallogr., Sect. D* **2001**, *57*, 233-238. MAD Phasing with Krypton.

(48) Prangé, T.; Schiltz, M.; Pernot, L.; Colloc'h, N.; Longhi, S. *Proteins: Struct., Funct., Genet.* **1998**, *30*, 61-73. Exploring Hydrophobic Sites in Proteins With Xenon or Krypton.

(49) Evans, G.; Bricogne, G. *Acta Crystallogr., Sect. D* **2002**, *58*, 976-991. Triiodide Derivatization and Combinatorial Counter-Ion Replacement: Two Methods for Enhancing Phasing Signal Using Laboratory Cu Kα X-Ray Equipment.

(50) Dauter, Z.; Dauter, M. *Structure* **2001**, *9*, R21-R26. Entering a New Phase: Using Solvent Halide Ions in Protein Structure Determination.

(51) Dauter, Z.; Dauter, M.; Rajashankar, K. R. *Acta Crystallogr., Sect. D* **2000**, *56*, 232-237. Novel Approach to Phasing Proteins: Derivatization by Short Cryo-Soaking with Halides.

(52) Terwilliger, T. C.; Berendzen, J. *Acta Crystallogr., Sect. D* **1999**, *55*, 849-861. Automated Structure Solution for MIR and MAD.

(53) Sigler, P. B. *Biochemistry* **1970**, *9*, 3609-3617. Iodination of a Single Tyrosine in Crystals of α-Chymotrypsin.

(54) Brzozowski, A. M.; Derewenda, Z. S.; Dodson, E. J.; Dodson, G. G.; Turkenburg, J. P. *Acta Crystallogr., Sect. B* **1992**, *48*, 307-319. Structure and Molecular Refinement of Rhizomucor miehei Triacylglyceride Lipase: a Case Study of the Use of Simulated Annealing in Partial Model Refinement.

(55) Derewenda, U.; Swenson, L.; Green, R.; Wei, Y.; Morosoli, R.; Shareck, F.; Kluepfel, D.; Derewenda, Z. S. *J. Biol. Chem.* **1994**, *269*, 20811-4. Crystal Structure, at 2.6-A Resolution, of the Streptomyces Lividans Xylanase A, A Member of the F Family of beta-1,4-D-Glycanases.

(56) Spurlino, J. C.; Smallwood, A. M.; Carlton, D. D.; Banks, T. M.; Vavra, K. J.; Johnson, J. S.; Cook, E. R.; Falvo, J.; Wahl, R. C.; Pulvino, T. A.; et al. *Proteins* **1994**, *19*, 98-109. 1.56 A Structure of Mature Truncated Human Fibroblast Collagenase.

(57) *CRC Handbook of Chemistry and Physics*; 75 ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, 1995.

(58) Gloss, L. M.; Kirsch, J. F. *Biochemistry* **1995**, *34*, 3990-3998. Decreasing the Basicity of the Active Site Base, Lys-258, of *Escherichia coli* Aspartate Aminotransferase by Replacement with γ-Thialysine.

(59) Hopkins, C. E.; O'Connor, P. B.; Allen, K. N.; Costello, C.; Tolan, D. R. *Protein Sci.* **2002**, *11*, 1591-1599. Chemical-modification rescue assessed by mass spectrometry demonstrates that γ-thia-lysine yields the same activity as lysine in aldolase.

(60) Hendrickson, W. A.; Teeter, M. M. *Nature* **1981**, *290*, 107-113. Structure of the Hydrophobic Protein Crambin Determined Directly From the Anomalous Scattering of Sulphur.

(61) Lue, Z.-J.; Vystotski, E. S.; Chen, C.-J.; Rose, J. P.; Lee, J.; Wang, B.-C. *Protein Sci.* **2000**, *9*, 2085-2093. Structure of the Ca^{2+}-Regulated Photoprotein Obelin at 1.7 Å Resolution Determined Directly From Its Sulfur Substructure.

(62) Micossi, E.; Hunter, W. N.; Leonard, G. A. *Acta Crystallogr., Sect. D* **2002**, *58*, 21-28. *De Novo* Phasing of Two Crystal Forms of Tryparedoxin II Using the Anomalous Scattering From S Atoms: A Combination of Small Signal and Medium Resolution Reveals This to be a General Tool for Solving Protein Crystal Structures.

(63) Han, Y.; Albericio, F.; Barany, G. *J. Org. Chem.* **1997**, *62*, 4307-4312. Occurrence and Minimization of Cysteine Racemization During Stepwise Solid-Phase Peptide Synthesis.

(64) Smith, J. L.; Thompson, A. *Structure* **1998**, *6*, 815-819. Reactivity of Selenomethionine - Dents in the Magic Bullet?

(65) Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M. *EMBO J.* **1990**, *9*, 1665-1672. Selenomethionyl Proteins Produced for Analysis by Multiwavelength Anomalous Diffraction (MAD): A Vehicle For Direct Determination of Three-Dimensional Structure.

(66) Smith, G. D.; Nagar, B.; Rini, J. M.; Hauptman, H. A.; Blessing, R. H. *Acta Crystallogr., Sect. D* **1998**, *54*, 799-804. The Use of SnB to Determine an Anomalous Scattering Substructure.

(67) Paulmier, C. *Selenium Reagents and Intermediates in Organic Synthesis*; Pergamon Press: Oxford, 1986.

(68) Sharff, A. J.; Koronakis, E.; Luisi, B.; Koronakis, V. *Acta Crystallogr., Sect. D* **2000**, *56*, 785-788. Oxidation of Selenomethionine: Some MADness in the Method!

(69) Peluso, S.; Dumy, P.; Nkubana, C.; Yokokawa, Y.; Mutter, a. M. *J. Org. Chem.* **1999**, *64*, 7114-7120. Solid-Phase Strategies for the Assembly of Template-Based Protein Mimetics.

(70) Johnson, M. L.; Correia, J. C.; Yphantis, D. A.; Halvorson, H. R. *Biophys. J.* **1981**, *36*, 575-588. Analysis of Data from the Analytical Ultracentrifuge by Nonlinear Least-Squares Techniques.

(71) McRorie, D. K.; Voelker, P. J. *Self-Associating Systems in the Analytical Ultracentrifuge*; Beckman Instruments, Inc.: Fullerton, 1993.

(72) McRee, D. E. *J. Struct. Biol.* **1999**, *125*, 156-65. XtalView/Xfit--A Versatile Program for Manipulating Atomic Coordinates and Electron Density.

(73) Jones, T. A.; Zou, J.-Y.; Cowan, S. W.; Kjeldgaard, M. *Acta Crystallogr., Sect. A* **1991**, *47*, 110-119. Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models.

Chapter 3: A MAD solution to the phase problem

Introduction

The previous chapter related early efforts to find a solution to the phase problem for peptide **1**. Direct methods, molecular replacement, and introduction of heavy atoms by crystal soaks were investigated extensively and unsuccessfully, and appeared unlikely to yield a solution. Several attempts at the covalent incorporation of a heavy atom at the peptide level did not bear fruit, but encouraging preliminary results with a selenomethionine-substituted peptide, Gln18SeMet, directed our attention towards this avenue.

The Gln18SeMet mutant was found to have promising solution-phase behavior and excellent diffraction in the solid state. Although no phase information was obtained from a preliminary MAD experiment, we felt confident that phases could be obtained from a selenomethionine derivative of **1** given a suitably homologous selenomethionine derivative and a homogeneous selenium oxidation state.

This chapter will present a systematic study of six selenomethionine mutants of peptide **1**, two of which are identical to peptide **1** in solution-phase properties. These two peptides were crystallized and the structures independently solved via a MAD experiment. The refined structure of one of these selenomethionine derivatives was then used to obtain a molecular replacement solution of **1**. The nature of the oligomeric BBA structure will be examined in greater detail in the following chapter.

Design Strategy

Selenomethionine (Figure 3-1) was first proposed by Hendrickson et al. as a conservative substitution for methionine that introduces an anomalously scattering atom for phase determination by the multiple wavelength anomalous diffraction (MAD) method.[1] Selenomethionine-labeled proteins can be prepared by expression from bacterial strains

auxotrophic for methionine grown in selenomethionine-containing medium. Selenomethionine-substituted proteins generally have native-like properties, although they are occasionally less soluble.[2] A selenomethionine MAD strategy is particularly attractive as it guarantees that there will be an anomalously-scattering atom present in the crystal with complete occupancy, thereby eliminating the problems of non-isomorphism between derivatives and of incomplete occupancy of a soaked heavy atom.

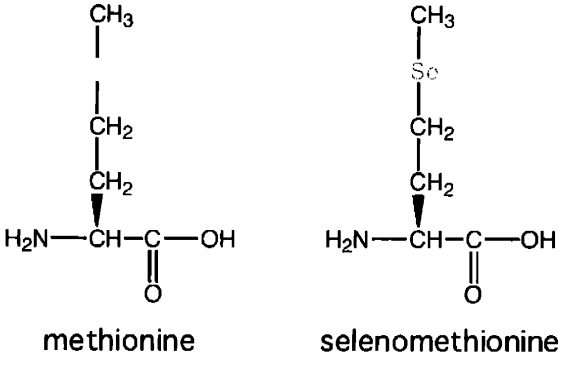


Figure 3-1. The chemical structures of methionine and selenomethionine.

It was necessary to substitute selenomethionine for another hydrophobic residue as the sequence of peptide **1** does not contain any methionines. The BBA hairpin is very tightly packed, each residue having specific interactions across the top and bottom faces, so it was deemed preferable to substitute selenomethionine for one of the hydrophobic residues along the BBA helix. No substitution was made for the unnatural amino acid DapBz, as NMR studies indicated that this residue forms important interactions with aromatic residues in the beta sheet region.[3] Six peptides were prepared (Table 3-1): Leu12SeMet, Ala13SeMet, Leu15SeMet, Leu16SeMet, Gln18SeMet, and Ala19SeMet. The helical wheel projection in Figure 3-2 indicates the relative placement of these residues with regard to the hydrophobic core.

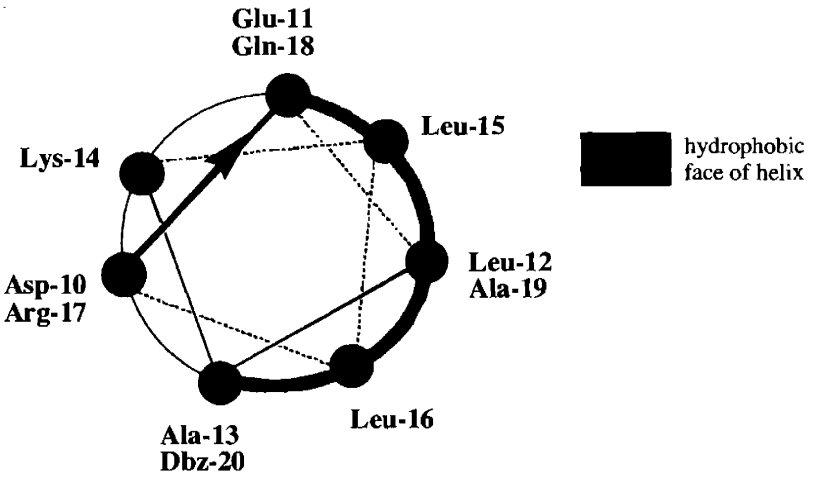


Figure 3-2. Helical wheel projection of the BBA helix, indicating the hydrophobic face of the helix and the relative placement of the six selenomethionine substitutions.

		helix												
	Peptide	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
1		a	D	E	L	A	K	L	L	R	Q	A	Z	G	NH_2
2	Leu12SeMet	a	D	E	SM	A	K	L	L	R	Q	A	Z	G	NH_2
3	Ala13SeMet	a	D	E	L	SM	K	L	L	R	Q	A	Z	G	NH_2
4	Leu15SeMet	a	D	E	L	A	K	SM	L	R	Q	A	Z	G	NH_2
5	Leu16SeMet	a	D	E	L	A	K	L	SM	R	Q	A	Z	G	NH_2
6	Gln18SeMet	a	D	E	L	A	K	L	L	R	SM	A	Z	G	NH_2
7	Ala19SeMet	a	D	E	L	A	K	L	L	R	Q	SM	Z	G	NH_2

Table 3-1. Sequences of the six selenomethionine derivatives of peptide **1**. All six peptides begin with the capped octapeptide sequence Ac-YRIpSYDF. Abbreviations: p=D-proline, a=D-alanine, Z=benzoylated L-α,β-diaminopropionic acid, SM=L-selenomethionine.

Oxidation

Selenoethers are prone to oxidation to the corresponding selenoxides or selenones.[4] In a protein, the susceptibility of selenomethonine selenoethers to oxidation is variable and dependent on the microenvironment of that residue. Partial oxidation introduces heterogeneity, which can prevent crystallization, although it often does not.[2] Incomplete oxidation also results in electronic heterogeneity at selenium, which can preclude obtaining phase information from a MAD experiment.[2,5]

We suspected that peptide **6**, the Gln18SeMet mutant of **1**, oxidizes to a significant, but incomplete, extent because of the failure of the initial MAD experiment, but did not know *a priori* how significant oxidation would be for the other five mutants. Moreover, it was not clear what, if any structural effect, oxidation could have. The oxidation of selenomethionine to methionine selenoxide results in an increase in side chain polarity, which could disrupt the oligomeric BBA structure, particularly if incorporated along the hydrophobic face of the BBA helix.

We thus planned to independently characterize both the oxidized and reduced forms of all six peptides, as we anticipated that there could be significant differences in their behavior. Whichever combination of peptide and oxidation state had the most native-like solution- and crystal-phase properties would be characterized further, crystallized, and used for MAD phasing.

Smith and Thompson recommend that selenomethionine-containing proteins be crystallized from buffer containing dithiothreitol or β-mercaptoethanol to keep the selenomethionine reduced if reduced selenomethionine derivatives are desired.[2] Other precautions include degassing crystallization solutions and buffers prior to use, or even crystallization under an inert atmosphere.[6]

Several groups have recently championed the use of oxidized selenomethionine for phasing purposes, especially for proteins that are particularly prone to oxidation. Methionine selenoxide (Figure 3-3) possesses somewhat greater phasing powers than does reduced selenomethionine due to its higher dispersive (f') and anomalous (f'') signals.[5,6] Working with methionine selenoxide-substituted peptides would be advantageous because adventitious oxidation in the time period between setup of the crystallization trays and flash-freezing of the crystal would not be a concern. Any increase in phasing power would be an additional bonus.

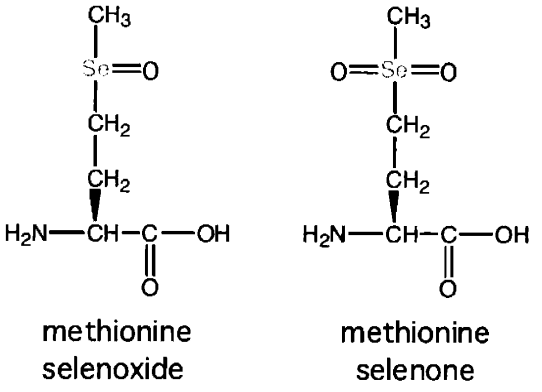


Figure 3-3. The chemical structures of methionine selenoxide and methionine selenone.

Characterization

The tendency of the selenomethionine peptides to oxidize was investigated by incubating a small sample of reduced peptide at room temperature and monitoring the appearance of any new species over time. Each of the selenomethionine peptides was found to oxidize to a significant extent to the selenoxide (MW+16 by electrospray mass spectrometry). Oxidation to methionine selenone was not observed. Incubation with dithiothreitol (DTT) for two days resulted in significant, but incomplete, reduction of the selenoxide in most cases.

Initial CD studies examined the (mostly) reduced peptides. The extent of oxidation was assessed by analytical HPLC of an aliquot of the sample before and after a CD wavelength scan. Each of these peptides, when reduced or predominantly reduced, exhibits some interhelical association, as evidenced by a ratio greater than one of ellipticities at 222 nm to 208 nm.[7,8]

It proved impractical to characterize the reduced state of these peptides. In the absence of a reducing agent, a fraction of the peptide would oxidize, and any experiment would be performed on a heterogeneous mixture. However, as most reducing agents absorb light at 280 nm, the introduction of a reducing agent made characterization of peptides by CD or AUC, or even determination of the peptide concentration, difficult.

If we were able to obtain and characterize a pure reduced selenomethionine peptide, and identify a native-like reduced selenomethionine derivative, the next step would be to crystallize that peptide. The addition of a reducing agent could affect crystallization, but in the absence of a reducing agent the peptide would be susceptible to adventitious oxidation between crystallization and data collection, unless crystallization trials and all other manipulations were performed under an inert atmosphere.

We deemed that it would be experimentally advantageous to work with a homogeneously-oxidized selenomethionine peptide, provided that a derivative could be found that was homologous to peptide **1**. We thus shifted our attention to characterization of the oxidized species. Homogeneously oxidized selenomethionine peptides were prepared by oxidation of purified SeMet-containing peptides with dilute hydrogen peroxide following a modified literature procedure.[5] As with air oxidation, the predominant product of hydrogen peroxide oxidation was found to be methionine selenoxide. Although minor peaks were occasionally detected by analytical HPLC, no species corresponding to the selenone was ever identified by mass spectrometry.

Those peptides having a selenomethionine substitution at an internal position (peptides **2**, **4**, **5**, **7**) are structured but primarily alpha helical when oxidized (Figure 3-4). Their CD spectra, like that of **BBA5**, and unlike that of **1**, indicate a lack of interhelical association. As they do not self-associate, they are not homologous to peptide **1**.

Oxidized peptides **3**, Ala13SeMet, and **6**, Gln18SeMet, have a selenomethionine located at the periphery of the hydrophobic core (Figure 3-2). Both peptides exhibit interhelical association when oxidized. Residues 13 and 18 define the interface between the hydrophobic and hydrophilic faces of the helix and hence are more accommodating of a charged residue with hydrophobic character. Oxidized peptides **3** and **6** were thus selected for further studies to determine whether they were indeed native-like as they appeared by CD.

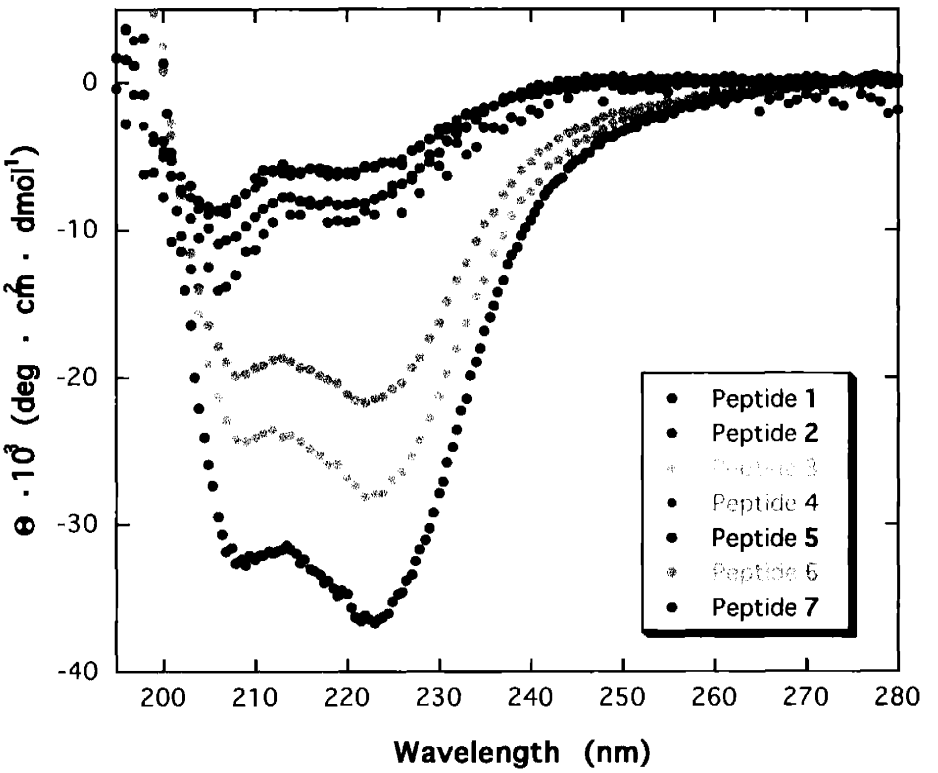


Figure 3-4. Circular dichroism spectra of peptide **1** and of oxidized peptides **2**-**7**.

Peptides **4** and **5** were studied by AUC in order to verify the CD-based observation of a monomeric state. It was found that oxidized peptides **4** and **5** are indeed monomeric (Table 3-2). As AUC analyses of the other four peptides were consistent with CD-based findings, the stoichiometries of peptides **2** and **7** were not determined by analytical ultracentrifugation. These four peptides were not characterized further, as they do not exhibit native-like behavior in solution.

The oligomeric states of oxidized peptides **3** and **6** were determined by equilibrium analytical ultracentrifugation for comparison to the native **1**. As predicted by the CD results, peptides **3** and **6** form stable oligomers in solution (Table 3-2). These oligomers were assigned

stoichiometries intermediate between three and four, consistent with values obtained for peptide **1** under the same conditions.

Peptide	Stoichiometry
1	3.3^{2}
3	3.4^{2}
4	1.0^{1}
5	$1.0^{1,2}$
6	3.7^{2}

Table 3-2. Results of sedimentation equilibrium AUC experiments with oxidized peptides **3**-**6** in 10 mM phosphate buffer at pH 7.2. Stoichiometries were determined by the analysis of data collected at three peptide concentrations and three angular velocities (40,000, 45,000, and 50,000 rpm). [1]Data were analyzed by a linear plot of ln(abs) vs $\left(r^2/2\right)$. [2]Data were analyzed using NONLIN and fit to a single ideal species model.

After establishing that oxidized peptides **3** and **6** have the same secondary structure content and stoichiometry as peptide **1**, and thus were likely to be useful for the structure determination of this family of oligomeric BBA peptides, the stabilities of the peptides were compared to the native by means of a CD melt experiment. The melting curves and temperatures of the three peptides are highly similar (Figure 3-5) and indicative of a very tight association. (The difference in ellipticities of peptides **3** and **6** between Figures 4 and 5 can be attributed to an inaccurate concentration determination in the former case due to residual contaminating oxidant.)

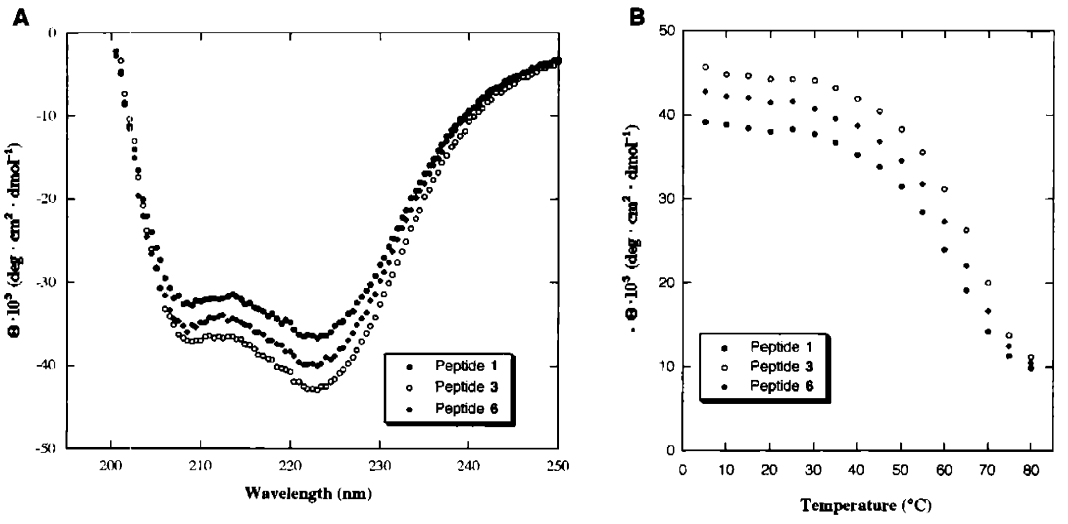


Figure 3-5. CD spectra of peptide **1** and of oxidized peptides **3** and **6** (50 μM peptide). A. Wavelength scans at 25 °C. B. Normalized molar ellipticity at 222 nm as a function of temperature.

We chose to proceed with the crystallization of both oxidized peptides **3** and **6** in the event that one peptide crystallized or diffracted significantly better than the other. The two crystallization conditions that yield diffraction-quality crystals of the native peptide (**a**: 25% *t*-butanol in 0.1M TRIS buffer, pH 8.5, **b**: 100 mM Na HEPES buffer pH 7.5, 10% v/v *i*-propanol, 20% w/v PEG 4000) were assayed prior to undertaking a sparse matrix screen. It was anticipated that one of these conditions would yield crystals because of earlier success in growing crystals of Gln18SeMet from condition **b** and because of the similarity of these peptides to **1**. Small needles of both peptides **3** and **6** grow from condition **b** upon overnight incubation at room temperature. Condition **a** did not yield useful crystals, nor did a short survey of commercial sparse-matrix crystallization formulations yield any better conditions.

Significant variation in morphology (bricks, rods, bipyramids and very thin plates) and in crystal quality is observed between batches, presumably due to fluctuations in ambient

temperature or humidity. Crystallizations were thus set up at regular intervals in order to maximize the yield of diffraction-quality crystals.

Crystals are less clumped, and more brick-like, when the peptide is repurified after the hydrogen peroxide oxidation (Figure 3-6). This second purification step removes all residual peroxide and any trace amounts of oxidation-damaged peptide. This step appears to be more critical for the growth of crystals of **3** than of crystals of **6**.

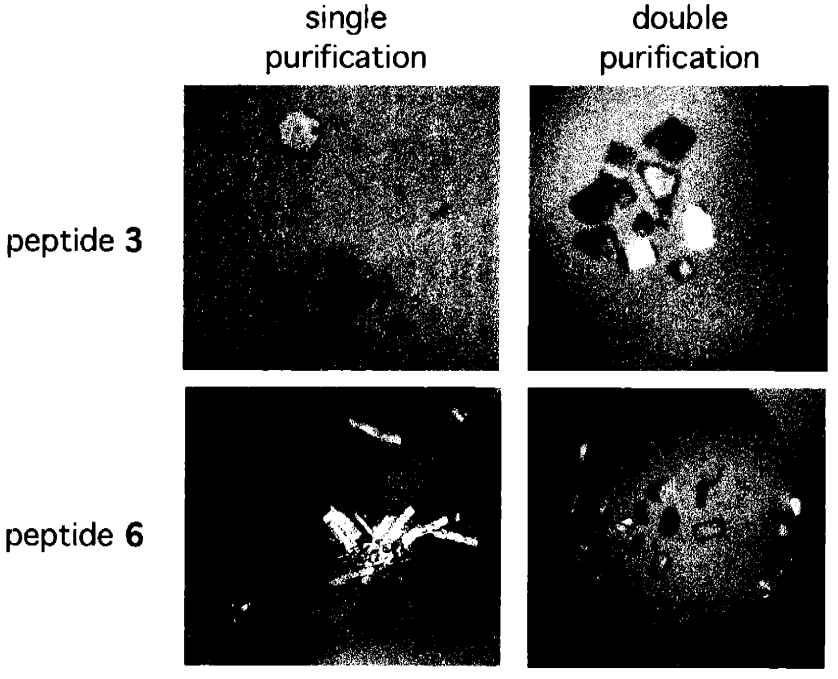


Figure 3-6. A second purification step following H_2O_2 oxidation results in better-quality crystals.

In previous experiments it was observed that small variations in pH (±1 pH unit) had little effect on crystallization, in keeping with the finding that there is little pH-dependence of the CD spectrum of the parent peptide **1** over the pH range 4-8. An attempt was made to optimize crystal growth by varying the precipitant (PEG 4000) concentration. A lower precipitant concentration should result in a slower equilibration time, less nucleation, and fewer and larger crystals. This prediction was upheld by a screen with condition **b** having between 10% and 30%

PEG 4000 (Figure 3-7). Crystallizations at 15% and 20% PEG 4000 were found to result in the best diffraction-quality rods.

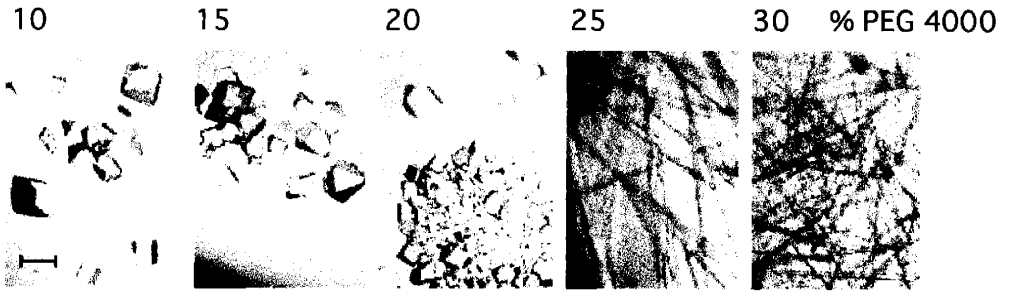


Figure 3-7. Effect of PEG 4000 concentration on growth of crystals of peptide **3**. Figures are to scale; bar indicates 0.1 mm.

Crystal density measurements were performed, as with peptide **1**, to measure the number of monomers per asymmetric unit.[9] We found that crystals of both peptides **3** and **6** are neutrally buoyant at ~45% w/w ficoll, leading to estimates of 3.1 monomers per asymmetric unit of **3**, and 1.5 monomers per asymmetric unit of **6**. There is some uncertainty in these values incidental to the use of an inappropriate value for the partial specific volume in their calculation (inappropriate as these peptides are in the crystalline state with a high percentage of organic cosolvent, rather than in dilute aqueous solution). These data suggest that peptide **6** has half as many monomers per asymmetric unit as do peptides **1** and **3**, thus suggesting that there are an even number of monomers per asymmetric unit in crystals of **1** and **3**.

Oxidized Peptide 3: Data Collection, Phasing, Refinement

Crystallization of peptide **3** is more sensitive to environmental conditions than is crystallization of peptides **1** and **6**. The selection of an appropriate cryoprotectant for peptide **3** also proved more challenging. Most standard choices resulted in visible damage to the crystals or extreme mosaicity to the extent of minimal diffraction or unusable streakiness. After extensive

screening, it was found that poly-3,3,3-trifluoropropylmethylsiloxane (FMS oil) is by far the best cryoprotectant, resulting in the least mosaicity and highest resolution in screening studies. Diffraction is correlated with crystal morphology; the best diffraction was obtained from rod-shaped crystals. There is considerable variation in diffraction between crystals of **3**, and it is thus necessary to screen for a crystal with the best diffraction.

Peptide **3** crystallizes in the monoclinic space group C2, with unit cell dimensions a = 50.0 Å, b = 46.9 Å, c = 31.4 Å, $\alpha = 90°$, $\beta = 98.6°$, and $\gamma = 90°$. The data exhibit two-fold noncrystallographic symmetry, as determined by means of a self-rotation function with the program GLRF (Figure 3-8).[10]

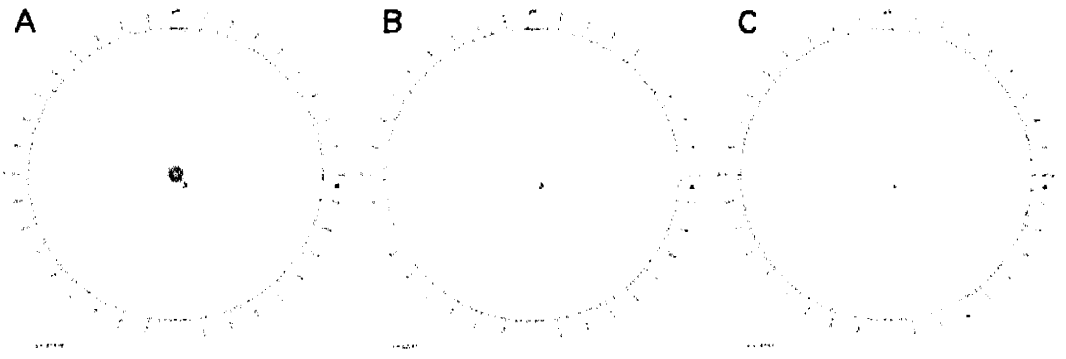


Figure 3-8. Results of the self-rotation function using GLRF from 10 to 4.0 Å. A. There is one prominent peak on a $\kappa = 180°$ plot indicative of two-fold symmetry. B. There are no peaks on a $\kappa = 120°$ plot suggestive of three-fold symmetry. C. There are no peaks on a $\kappa = 90°$ plot suggestive of four-fold symmetry. Contour levels start at 200.

A Matthews coefficient calculation[11] indicated that three monomers per asymmetric unit would result in a typical protein solvent content of 47%. Four monomers per asymmetric unit would result in a solvent content of 29%, which is relatively low for a protein.

A MAD experiment was performed by Dr. Ezra Peisach and Dr. Tracy L. Arakaki at the X12C beam line at the National Synchrotron Light Source at Brookhaven National Laboratory. Data were collected at three wavelengths (0.978240 Å, 0.977939 Å, 0.949359 Å) near the

selenium K absorption edge for each peptide: peak, edge, and remote (Figure 3-9, Table 3-3). Data were collected to ~1.5 Å resolution.

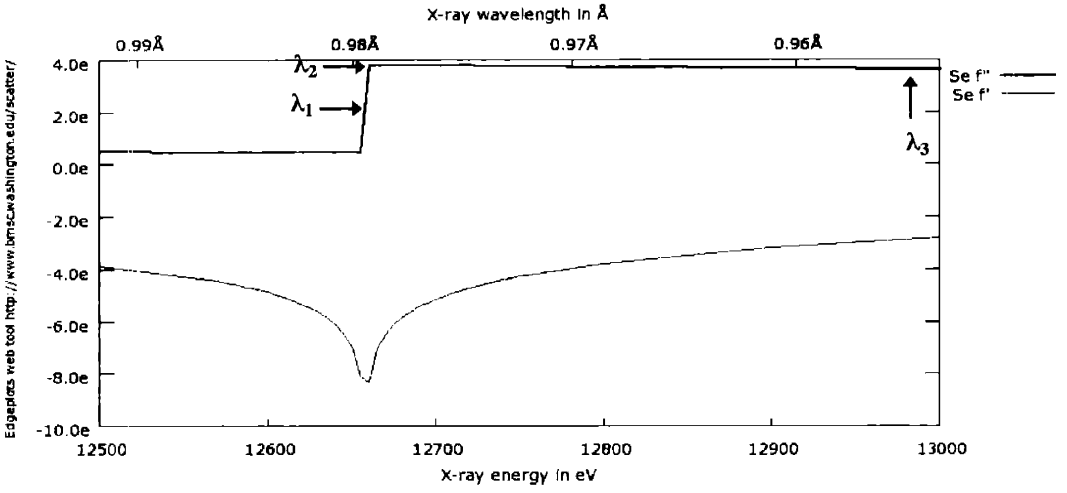


Figure 3-9. Calculated anomalous scattering coefficients, f′ and f″, near the K absorption edge of selenium. Plot generated from http://www.bmsc.washington.edu:80/scatter/AS_form.html.

Selenium sites were located and starting phases obtained by MAD phasing using the automated heavy atom solution program SOLVE.[12] The best starting phases, as judged by electron density map interpretability, were obtained from three selenomethionine sites found using resolution limits of 15 to 2 Å, and resulted in an initial figure of merit of 0.62. The initial electron density maps showed connected helical density. Selenium sites were covered by electron density.

Unit cell (Å, °)	a = 50.03 b = 46.87 c = 31.43 α = 90 β = 98.61 γ = 90		
Space group	C2		
X-ray source	BNL X12C		
Wavelength (Å)	0.978240	0.977939	0.949359
Resolution (Å)	∞ – 1.5	∞ – 1.5	∞ – 1.5
Total/	32,445	32,558	34,820
unique reflections	17,229	17,273	18,363
Completeness[1] (%)	76.0 (24.9)	76.0 (24.9)	80.4 (32.1)
I/σ(I)	33.8 (7.7)	32.3 (6.2)	30.6 (5.5)
R_{merge}[2] (%)	3.5 (6.6)	3.4 (8.6)	3.0 (11.0)
f'/f'' (e^-)	-10.0/3.4	-7.5/5.7	-2.5/3.0
Figure of merit[3]	0.62 (to 2.0 Å)		
Solvent content (%)	29		

Table 3-3. Data collection statistics for peptide **3**. [1]Values for the outermost shell (1.55-1.50 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]Reported by SOLVE.

Two helices were docked into the map generated from the initial MAD phases using the graphics program O[13] with the assistance of a skeletonized map generated by MAPMAN.[14] After a round of phase combination and rigid body refinement in CNS,[15] another helix and a hairpin were found.

Helices and hairpins were docked independently into appropriately-shaped density as we had no *a priori* knowledge of their relative orientation. A poly-alanine alpha helix was used for docking in order to avoid model bias. Side chains were introduced after aligning the sequence with correspondingly-shaped regions of electron density.

Figure 3-10 depicts electron density maps of the first two helices immediately after the first round of phase combination. Even at this early stage of the refinement, the electron density was quite clear. The helix and hairpin regions are connected by a D-alanine residue, the stereochemistry of which could already be discerned.

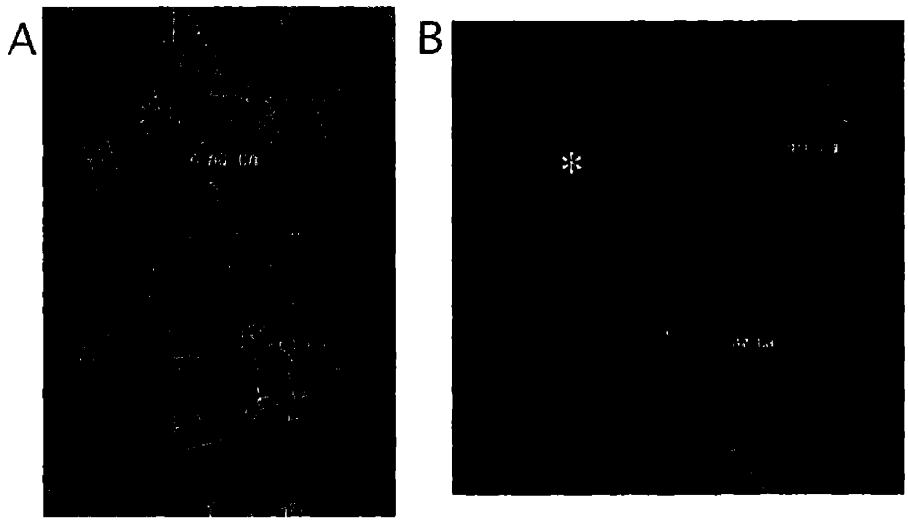


Figure 3-10. Electron density maps generated after one round of phase combination. A. Top view of the helix with side chains evident. B. Close-up view of the D-alanine hinge residue connecting helix and hairpin domains. An asterisk denotes the location of the D-alanine C_α.

Additional helical density was found after three helices had been located, suggesting the existence of a fourth monomer. An additional monomer was introduced to fit this density, leading to a total of four monomers per asymmetric unit.

Refinement consisted of alternate rounds of manual fitting to SigmaA weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps[16] in the graphics program O[13] followed by minimization and simulated annealing in CNS.[15] For statistical cross-validation purposes 10% of the data was reserved as a test set and excluded from refinement.[17,18] The current model was checked periodically against a $2F_o - F_c$ composite-omit map. At convergence of the R-factors, R was 16.1% and R_{free} was 19.9%.

Analysis of the Ramachandran plot defined by PROCHECK[19] showed a good final model for the structure, with 92% of residues in the most favorable regions and 8% of residues in

additionally allowed regions. All D-proline residues fell in the generously allowed regions, and were excluded from the above percentages. Final refinement statistics are given in Table 3-4.

Refinement Statistics	
Program	CNS[15]
Refinement	isotropic
Resolution (Å)	50 – 1.5
No. of reflections working/	1,5512
test set	1,604
R_{work}[1]/R_{free}[2] (%)	16.1 / 19.9
Number of Atoms per Asymmetric Unit	
Amino Acids	744
i-Propanol	4
Water	134
Average B-factors	
Wilson plot (Å^2)	13.7
Amino acids (Å^2)	15.2
i-Propanol (Å^2)	39.0
Water (Å^2)	19.9
RMS Deviations from Ideality	
Bond lengths (Å)	0.021
Angles (°)	2.3
Dihedral angles (°)	21.5
Ramachandran Plot[3]	
Most Favorable (%)	91.7
Additionally Allowed (%)	8.3

Table 3-4. Refinement statistics for peptide **3**. [1]$R_{work} = \sum_{hkl}|F_{obs} - F_{calc}| / \sum_{hkl}|F_{obs}|$. [2]$R_{free} = \sum_{hkl\ \sum T}|F_{obs} - F_{calc}| / \sum_{hkl}|F_{obs}|$, where the test set T includes 10% of the data. [3]Defined by PROCHECK. All D-proline residues fell in the generously allowed regions, and were excluded from the percentages calculated above.

There are two sets of parallel monomers in each asymmetric unit, offset by an angle of ~30°. Upon application of the appropriate symmetry operators, each pair of monomers forms, with the corresponding symmetry mates, a tetrameric structure (Figure 3-11).

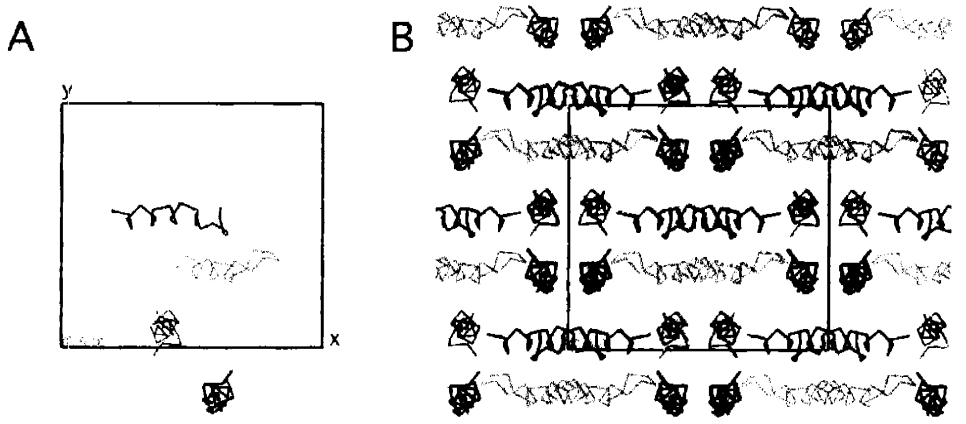


Figure 3-11. Crystal packing of peptide **3**. The unit cell is depicted as red box. This view is looking down the z-axis. A. The four crystallographically independent monomers are grouped in two sets of parallel monomers. B. Upon application of symmetry operators two distinct sets of tetramers are observed.

Two non-crystallographic two-fold axes can be seen in Figure 3-12. They are offset by a translation of half a unit cell along the x-axis. These axes are parallel to the y-axis and to each other. As they are parallel, they result in a single peak along the b-axis of a $\kappa = 180°$ GLRF plot.

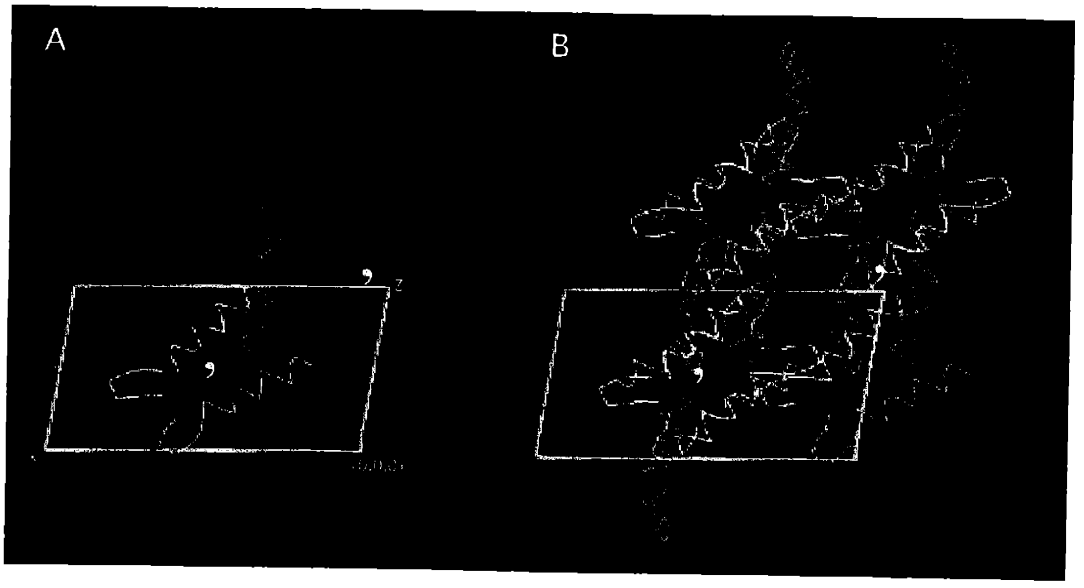


Figure 3-12. Non-crystallographic symmetry in crystals of peptide **3**. The two noncrystallographic two-fold axes present in crystals of peptide **3** are denoted by yellow inverted commas. The unit cell is denoted by a red parallelogram. This view is looking down the y-axis. A. The four crystallographically independent monomers in a unit cell and the two NCS axes. B. The four monomers and their symmetry mates. The first axis rotates a yellow monomer onto a yellow monomer and a red monomer onto a red monomer. The second axis rotates a green monomer onto a green monomer and blue monomer onto a blue monomer.

Oxidized Peptide 6: Data Collection, Phasing, Refinement

Crystals of peptide **6** exhibit less variation in their diffraction than do crystals of peptide **3**. As with peptide **3**, the best diffraction was obtained from rod-shaped crystals, and it was helpful to screen for a crystal with the best diffraction when collecting data. Both Paratone-N and FMS oil are effective at guarding against ice crystals.

Peptide **6** crystallizes in the monoclinic space group C2 with unit cell dimensions a = 57.9 Å, b = 21.3 Å, c = 31.9 Å, $\alpha = 90°$, $\beta = 121.3°$, and $\gamma = 90°$. The unit cell of peptide **6** is distinct from, and smaller than, that of peptide **3**, although both are monoclinic. The data for peptide **6** exhibit two-fold noncrystallographic symmetry (Figure 3-13), as determined by means of a self-rotation function with the program GLRF.[10]

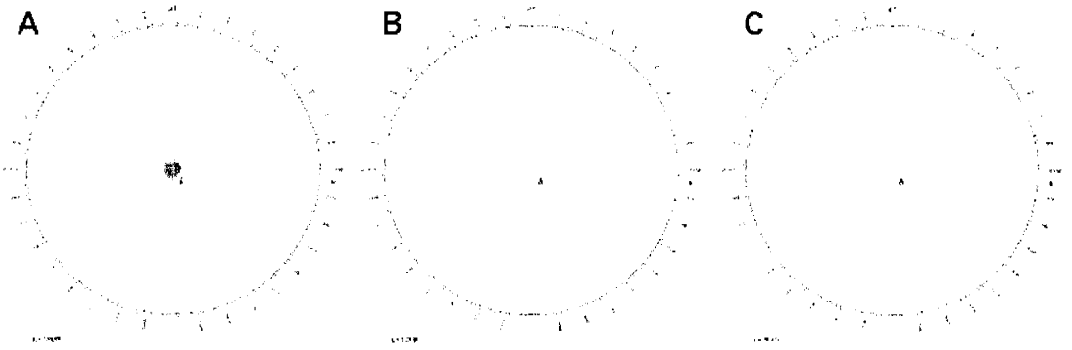


Figure 3-13. Results of the self-rotation function using GLRF from 10 to 4.0 Å. A. There is one prominent peak on a $\kappa = 180°$ plot indicative of two-fold symmetry. B. There are no peaks on a $\kappa = 120°$ plot suggestive of three-fold symmetry. C. There are no peaks on a $\kappa = 90°$ plot suggestive of four-fold symmetry. Contour levels start at 200.

A Matthews coefficient calculation[11] indicated that either one or two monomers per asymmetric unit were possible, giving rise to solvent contents of 62% and 23% respectively. The latter was deemed more likely as it is comparable to the solvent content of peptide **3**.

A MAD experiment was performed by Prof. Karen N. Allen at the X12C beam line at the National Synchrotron Light Source at Brookhaven National Laboratory. Data were collected at three wavelengths (0.978240 Å, 0.977939 Å, 0.949359 Å) near the selenium K absorption edge for each peptide: peak, edge, and remote (Table 3-5). Data were collected to ~1.5 Å resolution.

Unit cell (Å,°)	a = 50.03 b = 46.87 c = 31.43 α = 90 β = 98.61 γ = 90		
Space group		C2	
X-ray source		BNL X12C	
Wavelength (Å)	0.978240	0.977939	0.949359
Resolution (Å)	∞ – 1.5	∞ – 1.5	∞ – 1.5
Total/ unique reflections	32,445 17,229	32,558 17,273	34,820 18,363
Completeness[1] (%)	76.0 (24.9)	76.0 (24.9)	80.4 (32.1)
I/σ(I)	33.8 (7.7)	32.3 (6.2)	30.6 (5.5)
R_{merge}[2] (%)	3.5 (6.6)	3.4 (8.6)	3.0 (11.0)
f'/f'' (e^-)	-10.0/3.4	-7.5/5.7	-2.5/3.0
Figure of merit[3]		0.62 (to 2.0 Å)	
Solvent content (%)		23	

Table 3-5. Data collection statistics for peptide **6**. [1]Values for the outermost shell (1.55-1.50 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]Reported by SOLVE.

It is of interest to compare the data collected from the electronically homogeneous oxidized peptide **6** with the original MAD data set collected from crystals of mixed oxidation state. The electronically heterogeneous peptide crystallizes in either space group $P222_1$ or $P2_12_12$ (the position and number of screw axes was ambiguous) with unit cell dimensions of a = 48.9 Å, b = 31.9 Å, c = 21.85 Å, α = 90°, β = 90°, and γ = 90°. In this case the diffraction is significantly worse, with a high resolution limit of 2.3 Å. Homogeneous oxidation of the selenomethionine resulted in an alternative crystal packing and significantly improved diffraction.

Two large peaks are located on the v = 0.5 Harker section of an anomalous Patterson map of **6**, indicating that there are two selenium atoms per asymmetric unit. Selenium sites were located and starting phases obtained by MAD phasing using the automated heavy atom solution program SOLVE[12] with resolution limits of 15 to 2 Å. Two selenium sites were found, resulting in an initial figure of merit of 0.56. The initial electron density maps showed connected helical density. The two selenium sites were located in electron density.

It proved possible to dock one helix into the map generated from the initial MAD phases using the graphics program O[13] with the assistance of a skeletonized map generated by MAPMAN.[14] After a second round of phase combination two monomers were found per asymmetric unit.

Refinement consisted of alternate rounds of manual fitting to SigmaA weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps[16] in the graphics program O[13] followed by minimization and simulated annealing in CNS.[15] For statistical cross-validation purposes 10% of the data was reserved as a test set and excluded from refinement.[17,18] The current model was checked periodically against a $2F_o - F_c$ composite-omit map. At convergence of the R-factors, R was 16.9% and R_{free} was 18.3%.

Analysis of the Ramachandran plot defined by PROCHECK[19] showed a good final model for the structure, with 92% of residues in the most favorable regions and 8% of residues in additionally allowed regions. All D-proline residues fell in the generously allowed regions, and were excluded from the above percentages. Final refinement statistics are given in Table 3-6.

Refinement Statistics	
Program	CNS
Refinement	isotropic
Resolution (Å)	50 – 1.5
No. of reflections working/	1,5512
test set	1,604
R_{work}[1]/R_{free}[2] (%)	16.1 / 19.9
Number of Atoms per Asymmetric Unit	
Amino Acids	364
i-Propanol	4
Water	69
Average B-factors	
Wilson plot (Å^2)	13.7
Amino acids (Å^2)	15.2
i-Propanol (Å^2)	39.0
Water (Å^2)	19.9
RMS Deviations from Ideality	
Bond lengths (Å)	0.021
Angles (°)	2.3
Dihedral angles (°)	21.5
Ramachandran Plot[3]	
Most Favorable (%)	91.7
Additionally Allowed (%)	8.3

Table 3-6. Refinement statistics for peptide **6**. [1]$R_{work} = \sum_{hkl}|F_{obs} - F_{calc}| / \sum_{hkl}|F_{obs}|$. [2]$R_{free} = \sum_{hkl \in T}|F_{obs} - F_{calc}| / \sum_{hkl}|F_{obs}|$, where the test set T includes 10% of the data. [3]Defined by PROCHECK. All D-proline residues fell in the generously allowed regions, and were excluded from the percentages calculated above.

Application of symmetry operators to the two crystallographically independent monomers gives rise to an array of monomers aligned along the same axis (Figure 3-14). Columns of peptide monomers, all oriented in the same direction, pack alongside columns of peptide monomers all oriented in the opposite direction. There are no discontinuities in the crystal packing that would allow for a ready identification of the solution-phase oligomeric unit.

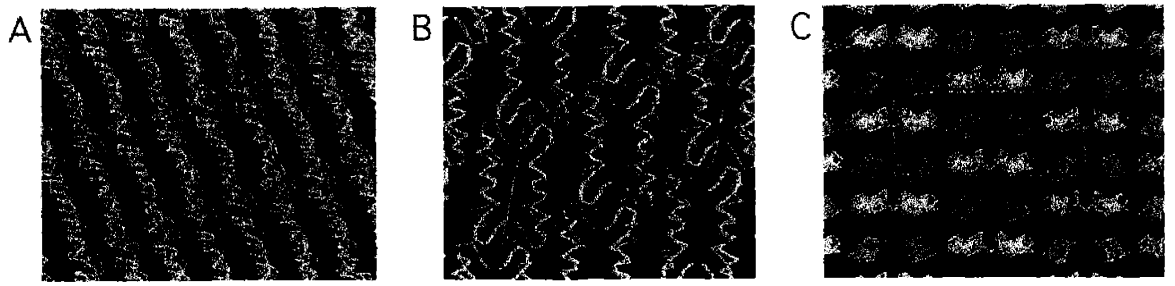


Figure 3-14. Packing of monomers in crystals of peptide **6**. A. View looking down x-axis. B. View looking down y-axis. C. View looking down z-axis.

The solution-phase tetramer can be deduced from the orientation of the hydrophobic residues comprising the core. The solution-phase tetramer can also be determined by an examination of the waters of crystallization (Figure 3-15).

Figure 3-15. A. The two crystallographically independent monomers of peptide **6**. B. The two monomers after application of symmetry operators. C. The location of bound water molecules.

Peptide 1: Phasing, Refinement

The primary obstacle to a molecular replacement solution to the structure of **1** had been the lack of a highly homologous search model. With the X-ray structures of peptides **3** and **6** in hand, three highly homologous search models (two tetramers from **3**, one tetramer from **6**) were available. These were expected to lead to a facile molecular replacement solution for **1**.

As an examination of the systematic absences did not allow an unambiguous determination of the space group, molecular replacement was attempted, using MOLREP,[20] in the eight possible permutations of P222 having 0, 1, 2, and 3 screw axes.

The results from a number of search models, including a peptide monomer, dimer, and tetramer were compared. The best results were obtained using the refined tetramer of **6**, including side chains, as a search model. Of the eight possible space groups, $P2_12_12$, with axes oriented as a = 38.94 Å, b = 56.17 Å, c = 32.32 Å, stood out as the correct solution. This solution had a correlation coefficient of 0.53 and an R-factor of 0.43.

Refinement consisted of alternate rounds of manual fitting to SigmaA weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps[16] in the graphics program O[13] followed by minimization and simulated annealing in CNS.[15] For statistical cross-validation purposes 10% of the data was reserved as a test set and excluded from refinement.[17,18] The current model was checked periodically against a $2F_o - F_c$ composite-omit map. The refinement of peptide **1** was continued using SHELXL,[21] extending the resolution from 1.5 Å to 1.2 Å. Anisotropic B-factor refinement resulted in a drop in both R and R_{free}.

Refinement of this structure proceeded more slowly than did refinement of peptides **3** and **6**, presumably because of the lack of independent, experimental phases. Moreover, the value of R_{free} was considerably (~8-10%) higher than the value of R throughout the refinement of **1**, suggesting that there could be a problem with the model or the data. Several attempts were made to discover the source of the discrepancy between the values of R and R_{free}.

The discrepancy does not appear to stem from an incorrect space group. Further rounds of molecular replacement in the eight possible orthorhombic space groups confirmed the assignment of the correct space group. Several cycles of refinement were performed in the lower-symmetry space group $P2_1$, in case the data were of lower symmetry. Refinement in $P2_1$ did not lead to any improvement in R_{free}. Furthermore, the data in $P2_1$ were checked to exclude the possibility of twinning by merohedry.

The discrepancy does not appear to be due to poor data at high resolution. Data quality does not deteriorate at higher resolutions. Moreover, refinement at lower resolution does not improve the difference between R_{free} and R_{work}.

The discrepancy does not appear to be due to poor data quality. The images were re-processed and re-scaled so as to eliminate any possible overloads in the low-resolution bins. Several rounds of refinement were performed with the newly-processed data without any improvement in R_{free}. A new data set was collected from crystals of **1**. These data were of high quality but lower resolution. Molecular replacement followed by refinement and rebuilding of these data also did not converge to a lower R_{free}.

The discrepancy does not appear to be due to a few bad reflections in the test set. An examination of the test set did not reveal any overloaded or otherwise suspicious reflections. Reducing the test set to 5% did not have any effect, nor did selecting a new 10% test set.

The discrepancy does not appear to be due to an incorrect register of side chains. Maps generated from a poly-alanine model, debiased by simulated annealing, and thus free of bias from the side chains, showed no slipping of register of the side chains.

The model does not appear to be caught in a local minimum. The high resolution limit was cut back to 2.0 Å on two occasions so as to increase the radius of convergence and give the model more freedom to shift during simulated annealing steps. Resolution was gradually increased in subsequent refinement steps. This procedure did not reduce R_{free}.

The discrepancy does not appear to be due to improperly-placed waters, which can have a significant effect in a structure of such a small size. Waters of crystallization were located from scratch in a map generated without any waters on multiple occasions. No significant difference in R_{free} was found.

The discrepancy may be due to a number of small differences between the partially-refined working model and the true structure. However, we were unable to determine what those changes might be, as both the SigmaA-weighted $2F_o - F_c$ and composite omit maps, beyond a

certain stage of refinement, reflected the partially-refined model used to calculate phases. Attempts to improve the electron density maps using the Prime-and-Switch[22] protocol in SOLVE[12] and SHELXE[23,24] were unsuccessful due to the relatively low solvent content of crystals of **1**.

At convergence, R was 18.2% and R_{free} was 27.8%. We attribute the relatively high value of R_{free} to an inability to overcome the model bias due to a lack of independent phase information inherent in a molecular replacement solution. It is likely that the high value of R_{free} is indicative of a number of small deviations of the model from the actual structure, rather than any single large flaw. The value of R_{free} does indicate that the model of peptide **1** is correct in its essential features. Together with the two selenomethionine structures, the structure of peptide **1** demonstrates that the selenomethionine derivatives do indeed share the native structure.

Analysis of the Ramachandran plot defined by PROCHECK[19] showed a good final model for the structure, with 92% of residues in the most favorable regions and 8% of residues in additionally allowed regions. All D-proline residues fell in the generously allowed regions, and were excluded from the percentages calculated above. Final refinement statistics are given in Table 3-7.

Refinement Statistics	
Program	SHELXL
Refinement	anisotropic
Resolution (Å)	8-1.2
No. of reflections working/	17,629
test set	1,977
R_{work}[1]/R_{free}[2] (%)	18.8[3] / 28.5[3] 18.2[4] / 27.8[4]
Number of Atoms per Asymmetric Unit	
Amino Acids	722
Water	118
Average B-factors	
Wilson plot (Å^2)	15.8
Amino acids (Å^2)	22.3
Water (Å^2)	43.5
RMS Deviations from Ideality	
Bond lengths (Å)	0.032
Angles (°)	2.3
Dihedral angles (°)	23.7
Ramachandran Plot[5]	
Most Favorable (%)	92
Additionally Allowed (%)	8

Table 3-7. Refinement statistics for peptide **1**. [1]$R_{work} = \sum_{hkl}|F_{obs} - F_{calc}| / \sum_{hkl}|F_{obs}|$. [2]$R_{free} = \sum_{hkl\ \in T}|F_{obs} - F_{calc}| / \sum_{hkl}|F_{obs}|$, where the test set T includes 10% of the data. [3]Values calculated using all data. [4]Values calculated using data for which $F > 4\sigma(F)$. [5]Defined by PROCHECK. All D-proline residues fell in the generously allowed regions, and were excluded from the percentages calculated above.

The refined structure of peptide **1** was used to solve a data set collected from crystals of **1** having hexagonal symmetry (described in the previous chapter). The structure was solved in the lower-symmetry space group $P3_1$ using MOLREP.[20] Peptide **1** packs as a hexagonal array of tetramers (Figure 3-16). A few cycles of refinement were performed in CNS[15] in order to determine whether the oligomeric structure of **1** differed in a crystal with hexagonal crystal packing. The overall features of the tetrameric structure were identical to the refined structures of **1**, **3**, and **6**. This data set was not refined further as it did not appear that it would yield any additional information.

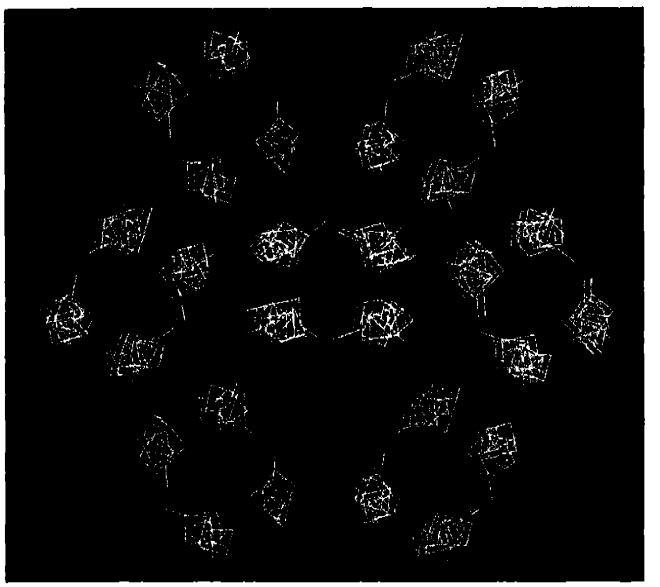


Figure 3-16. Packing arrangement of peptide **1** in crystals with hexagonal symmetry.

Conclusions

Six mutants of peptide **1** were prepared having selenomethionine substitutions at hydrophobic positions along the BBA helix. It proved impractical to characterize the fully reduced selenomethionine peptides. Two selenomethionine peptides, **3** and **6**, when homogeneously oxidized, have native-like properties. The structures of oxidized peptides **3** and **6** were solved by MAD phasing and determined to be tetrameric. The refined structure of **6** was used to phase the data of peptide **1** by molecular replacement.

Acknowledgements

I am very grateful to Prof. Karen N. Allen and Dr. Ezra Peisach for assistance throughout this project, and to Professor Karen N. Allen, Dr. Ezra Peisach, and Dr. Tracy L. Arakaki for collecting the synchrotron data sets for all three peptides. I thank Dr. Nicholas R. Silvaggi for his

assistance with SHELXL. I thank Professor George M. Sheldrick for helpful suggestions regarding the refinement of peptide **1**.

Materials and Methods

Peptide Synthesis

Peptides were prepared by standard Fmoc-based solid-phase peptide synthesis. Unless otherwise noted, all peptide synthesis reagents were purchased from Applied Biosystems or Novabiochem. The *C*-terminal dipeptide was coupled by hand using HBtU/HOBt/DIPEA chemistry. DapBz was incorporated as Fmoc-Dap(alloc)-OH (Bachem), deprotected on the solid phase resin,[25] and side chain capped with benzoic acid (Aldrich). The synthesis was then completed using a 9050 PepSynthesizer™ or an Advanced ChemTech 396W peptide synthesizer. Fmoc-SeMet-OH (AnaSpec) was coupled by hand with PyBOP/HOBt/DIPEA.

Cleavage and Purification

Peptides were cleaved from the resin for 2-6 hours with 2.5% 1,2-ethanedithiol, 2.5% water, 1% triisopropylsilane, and 94.5% trifluoroacetic acid, after which the cleavage cocktail was filtered to remove the resin and the filtrate concentrated under a gentle stream of nitrogen. When the volume was between ~1-2 ml the peptide was precipitated from ice-cold 2:1 ether:hexane, and centrifuged. The supernatant was decanted and the pellet resuspended in ice-cold 2:1 ether:hexane and washed 1-2 more times. The supernatant was then decanted, and the pellet dissolved in ~10% acetonitrile and immediately purified or lyophilized. High performance liquid chromatography was performed using a Waters Prep LC 4000 System equipped with a Waters 2487 absorbance detector (set at 228 and 280 nm), a C_{18} reverse-phase analytical column (Beckman Ultrasphere, 5 μm, 4.6 x 150 mm) and a C_{18} reverse-phase preparative column (YMC-pack, 250 x 20 mm). A standard gradient of 80:20 to 20:80 (water:acetonitrile, 0.1% TFA) over

25 minutes (with a flow rate of 1 ml/min on the analytical column, and 15 ml/min on the preparative column) was employed in all cases. Purity was determined by electrospray mass spectroscopy (PerSeptive Biosystems Mariner™ BioSpectrometry Workstation, using the Turbo Ion Source) and by analytical HPLC. Retention times for oxidized peptides **2**, **3**, **4**, **5**, **6** and **7**, on the analytical column are 20.5 min, 26.0 min, 23.5 min, 22.5 min, 26.5 min, and 24.5 min, respectively.

Oxidation

Method A. Peptides were oxidized using a modified literature procedure.[5] The peptide was dissolved in 3% H_2O_2 in 10 mM phosphate buffer at pH 7.2. After incubation at room temperature for 20 minutes the solution was loaded onto a 1.0 kD MWCO DispoDialyzer™ (Harvard Biosciences) and dialyzed at room temperature against ~250 ml buffer under ambient conditions to remove the peroxide. After several buffer exchanges the oxidized peptide was collected and an aliquot injected onto a C18 analytical column to confirm the peptide's purity and extent of oxidation. Peptide stock concentrations were determined either by absorbance at 280 nm or by quantitative amino acid analysis.

Method B. Peptides were oxidized as above for 20 minutes, diluted with distilled water, filtered through a 0.45 μm syringe filter, and immediately HPLC purified on a YMC C18 preparative column with a Waters Prep LC 4000 System using a 20-80% acetonitrile gradient.

Extinction Coefficient Determination

An attempt was made to directly measure the extinction coefficient of peptide **6** by correlating the absorbance of a concentrated stock solution at 280 nm (measured with a Shimadzu UV160U spectrophotomer) with its concentration, determined by duplicate quantitative amino acid analysis measurements, as was done for peptide **1**. This approach failed because of precipitation of peptide **6** in the few hours between measurement of the absorbance of the stock solution and its

hydrolysis for amino acid analysis. Consequently, we chose to prepare a model peptide with better solubility with which to determine the contribution of the selenoxide to the extinction coefficient. The pentapeptide Ac-EA(SeMet)AK-NH_2 was synthesized, purified, oxidized, repurified, and used to determine the extinction coefficient of the selenoxide group (38.5 cm^{-1} M^{-1}). The extinction coefficient of the selenoxide peptides was thus determined to be 3129 cm^{-1} M^{-1}.

Circular Dichroism

Peptides were dissolved in degassed 10 mM phosphate buffer at pH 7.2. Stock solution concentrations were determined based on absorbance at 280 nm, and appropriate dilutions used to prepare 50 μM samples for analysis. Data were collected with an Aviv Circular Dichroism Spectrometer Model 202 using strain-free quartz cells having a path length of 0.1 cm. Wavelength scan experiments were performed in duplicate or triplicate, from 300 to 195 nm, at 25 °C, with an averaging time of 0.5 s and a step size of 1.0 or 0.5 nm. Melting experiments were performed as wavelength scans (260 nm to 205 nm, in duplicate, 0.5 s averaging time, 1 nm step size) in temperature increments of 5 °C from 5 °C to 80 °C.

Analytical Ultracentrifugation

Concentrated stock solutions of peptides were prepared in 10 mM phosphate buffer at pH 7.2. Each solution was then loaded onto a 1.0 kDa MWCO DispoDialyzer™ (Harvard Biosciences) and dialyzed at room temperature. Stock solution concentrations were determined based on absorbance at 280 nm, and dilutions made to cover a range of concentrations. Peptide solutions were loaded into standard six-sector epon centerpieces and spun, at 5 °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. Data were analyzed fitting selected scans to ln(abs) vs $\left(r^2/2\right)$ plots, abs vs r plots,

and with NONLIN.[26] Molecular weights were obtained using a partial specific volume, $\bar{\nu}$, calculated from the peptide sequence by SEDNTERP.[27] DapBz was approximated as phenylalanine and methionine selenoxide as methionine.

Crystallization

Peptides were dissolved at ≥8 mg/ml in 10 mM phosphate at pH 7.2 for crystallization. Crystals were grown using vapor diffusion with hanging-drop geometry by mixing 1.5 μl of protein with an equal volume of reservoir solution. The reservoir contained between 200 and 300 μl of precipitant. Peptides **3** and **6** crystallize overnight as small rods from 100 mM Na HEPES buffer at pH 7.5, 10% v/v *i*-propanol, 20% w/v PEG 4000.

Crystal Density Measurements

Individual solutions were prepared as described in the literature.[9] Solution densities were measured or calculated following the literature reference. One to three crystals were gently placed in clear centrifuge tubes containing ficoll solutions of varying densities. The crystals were allowed to equilibrate overnight on the benchtop or were gently spun at low speed in a benchtop centrifuge in order to reach equilibrium. The tubes were then carefully examined under a microscope to determine the location of the crystal in the tube.

Data Collection

Crystals were frozen in a stream of N_2 gas cooled to -180 °C using FMS oil (Hampton Research) as a cryoprotectant. Data sets were collected at the Boston University Core Facility for Macromolecular Crystallography and at the National Synchrotron Light Source, Brookhaven National Laboratory. The DENZO and SCALEPACK package[28] was used for data indexing, reduction, and scaling.

Phasing

Heavy-atom sites were searched for using the automated heavy atom solution program SOLVE.[12] Patterson maps were generated using XtalView.[29] Molecular replacement was performed using MOLREP.[20] Electron density maps were viewed in the graphics program O.[13]

Refinement.

Manual fitting was performed using SigmaA weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps[16] in the graphics program O.[13] Resolution limits were gradually extended as the model improved. Refinement, using an MLHL (peptides **3** and **6**) or MLF (peptide **1**) 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. Peptides **3** and **6** were refined using an MLF target in the later stages of the refinement. $2F_o - F_c$ composite-omit maps were used periodically to check the current model. Waters were added by hand and using the automated water-picking program in CNS using a threshold value of 2.5 σ on an $F_o - F_c$ map. All waters were checked by hand at each cycle of the refinement. For statistical cross-validation purposes 10% of the data was reserved as a test set and excluded from refinement.[17,18] Topology and parameter files were created for the nonstandard groups acetyl,[30,31] amide,[30-33] DapBz,[34-40] and methionine selenoxide[41-45] using bond lengths and angles from the literature. Values for D-Ala and D-Pro were identical to CNS and SHELXL defaults for their L-enantiomers with an inversion of chirality. S chirality was arbitrarily assigned at the oxidized selenium. Wilson plot values were calculated using the CCP4 program TRUNCATE[46] with resolution limits 2.5 – 1.5 Å. Coordinates of the refined structures have been deposited with the Protein Data Bank (accession codes 1SN9, 1SNA and 1SNE for **1**, **3**, and **6**).

References

(1) Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M. *EMBO J.* **1990**, *9*, 1665-1672. Selenomethionyl Proteins Produced for Analysis by Multiwavelength Anomalous Diffraction (MAD): A Vehicle For Direct Determination of Three-Dimensional Structure.

(2) Smith, J. L.; Thompson, A. *Structure* **1998**, *6*, 815-819. Reactivity of Selenomethionine - Dents in the Magic Bullet?

(3) Mezo, A. R.; Ottesen, J. J.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 1002-1003. Discovery and Characterization of a Discretely Folded Homotrimeric Peptide.

(4) Paulmier, C. *Selenium Reagents and Intermediates in Organic Synthesis*; Pergamon Press: Oxford, 1986.

(5) Sharff, A. J.; Koronakis, E.; Luisi, B.; Koronakis, V. *Acta Crystallogr., Sect. D* **2000**, *56*, 785-788. Oxidation of Selenomethionine: Some MADness in the Method!

(6) Thomazeau, K.; Curien, G.; Thompson, A.; Dumas, R.; Biou, V. *Acta Crystallogr., Sect. D* **2001**, *57*, 1337-1340. MAD on Threonine Synthase: The Phasing Power of Oxidized Selenomethionine.

(7) Zhou, N. E.; McKay, C.; Hodges, R. S. *Biochemistry* **1992**, *31*, 5739-5746. 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.

(8) Zhou, N. E.; Zhu, B.-Y.; Kay, C. M.; Hodges, R. S. *Biopolymers* **1992**, *32*, 419-426. The Two-Stranded α-Helical Coiled-Coil is an Ideal Model for Studying Protein Stability and Subunit Interactions.

(9) Westbrook, E. M. *Methods Enzymol.* **1985**, *114*, 187-196. Crystal Density Measurements Using Aqueous Ficoll Solutions.

(10) Tong, L.; Rossman, M. G. *Acta Crystallogr., Sect. A* **1990**, *46*, 783-792. The Locked Rotation Function.

(11) Matthews, B. W. *J. Mol. Biol* **1968**, *33*, 491-497. Solvent Content of Protein Crystals.

(12) Terwilliger, T. C.; Berendzen, J. *Acta Crystallogr., Sect. D* **1999**, *55*, 849-861. Automated Structure Solution for MIR and MAD.

(13) Jones, T. A.; Zou, J.-Y.; Cowan, S. W.; Kjeldgaard, M. *Acta Crystallogr., Sect. A* **1991**, *47*, 110-119. Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models.

(14) Kleywegt, G. J.; Jones, T. A. *Acta Crystallogr., Sect. D* **1996**, *52*, 826-828. xdlMAPMAN and xdlDATAMAN - Programs For Reformatting, Analysis and Manipulation of Biomacromolecular Electron-Density Maps and Reflection Data Sets.

(15) Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; L.Warren, G. *Acta Crystallogr., Sect. D* **1998**, *54*, 905-921. *Crystallography & NMR System*: A New Software Suite for Macromolecular Structure Determination.

(16) Read, R. J. *Methods Enzymol.* **1997**, *278*, 110-128. Model Phases: Probabilities and Bias.

(17) Brünger, A. T. *Nature* **1992**, *355*, 472-475. Free R Value: A Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures.

(18) Brünger, A. T. *Methods Enzymol.* **1997**, *277*, 366-396. Free R Value: Cross-Validation in Crystallography.

(19) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Cryst.* **1993**, *26*, 283-291. Procheck - A Program to Check the Stereochemical Quality of Protein Structures.

(20) Vagin, A.; Teplyakov, A. *J. Appl. Cryst.* **1997**, *30*, 1022-1025. MOLREP: An Automated Program For Molecular Replacement.
(21) Sheldrick, G. M. *Acta Crystallogr., Sect. A Suppl. S* **1984**, *40*, C440-C440. SHELX-84 - A Program System for Crystal-Structure Solution and Refinement.
(22) Terwilliger, T. C. *Acta Crystallogr., Sect. D* **2001**, *57*, 1763-1775. Map-Likelihood Phasing.
(23) Usón, I.; Sheldrick, G. M. *Curr. Opin. Struct. Biol.* **1999**, *9*, 643-648. Advances in Direct Methods for Protein Crystallography.
(24) Schneider, T. R.; Sheldrick, G. M. *Acta Crystallogr., Sect. D* **2002**, *58*, 1772-1779. Substructure Solution with SHELXD.
(25) Peluso, S.; Dumy, P.; Nkubana, C.; Yokokawa, Y.; Mutter, a. M. *J. Org. Chem.* **1999**, *64*, 7114-7120. Solid-Phase Strategies for the Assembly of Template-Based Protein Mimetics.
(26) Johnson, M. L.; Correia, J. C.; Yphantis, D. A.; Halvorson, H. R. *Biophys. J.* **1981**, *36*, 575-588. Analysis of Data from the Analytical Ultracentrifuge by Nonlinear Least-Squares Techniques.
(27) Laue, T. M.; Shah, B.; Ridgeway, T. M.; Pelletier, S. L. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Horton, J. C., Ed.; Royal Society of Chemistry: Cambridge, 1992.
(28) Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307-326. Processing of X-ray Diffraction Data Collected in Oscillation Mode.
(29) McRee, D. E. *J. Struct. Biol.* **1999**, *125*, 156-65. XtalView/Xfit--A Versatile Program for Manipulating Atomic Coordinates and Electron Density.
(30) Puliti, R.; Mattia, C. A.; Sena, C. D.; Barone, G. *J. Mol. Struct.* **1998**, *442*, 1-9. *N*-Acetyl–L-prolyl–L-leucyl–glycinamide: X-Ray Structure, Energy Minimization and Calorimetric Determinations.
(31) Puliti, R.; Mattia, C. A. *Acta Crystallogr., Sect. C* **1995**, *51*, 336-339. *N*-Acetylglycyl-L-alaninamide and *N*-Acetyl-L-alanyl-L-alaninamide.
(32) Kojima, T.; Tanaka, I.; Ashida, T. *Acta Crystallogr., Sect. B* **1982**, *38*, 221-225. The Structure of *tert*-Butoxycarbonyl-L-prolyl-L-alanylglycinamide.
(33) Bardi, R.; Piazzesi, A. M.; Toniolo, C.; Jensen, O. E.; Andersen, T. P.; Senning, A. *Tetrahedron* **1988**, *44*, 761-769. Molecular and Crystal Structures of Two β-Bend Forming Monothiated Analogues of Melanostatin.
(34) Hibbert, F.; Mills, J. F.; Nyburg, S. C.; Parkins, A. W. *J. Chem. Soc., Perkin Trans. 2* **1998**, 629-634. Hydrogen Bonding and Structure of 2-hydroxy-N-acylanilines in the Solid State and in Solution.
(35) Harrisonn, W.; Rettig, S.; Trotte, J. *J. Chem. Soc., Perkin Trans. 2* **1972**, 1036-1040. Crystal and Molecular Structure of Hippuric Acid.
(36) Palmer, A.; Brisse, F. *Acta Crystallogr., Sect. B* **1980**, *36*, 1452-1455. La Structure du *N,N'*-Ethylène-dibenzamide.
(37) Huber, C.; Carey, P. R.; Hsi, S.-C.; Lee, H.; Storer, A. C. *J. Am. Chem. Soc.* **1984**, *106*, 8263-8268. Conformational Study of *N*-Acyl Amino Acid Esters and Thiol Esters by FT-IR and X-Ray Crystallography: Evidence for a N••• S Interaction in Thiol Esters.
(38) Ringertz, H. *Acta Crystallogr., Sect. B* **1971**, *27*, 185-291. The Molecular and Crystal Structure of Hippuric Acid.
(39) Vrielink, A.; Obel-Jorgensen, A.; Codding, P. W. *Acta Crystallogr., Sect. C* **1996**, *52*, 1300-1302. Hippuryl-L-histidyl-L-leucine, a Substrate for Angiotensin Converting Enzyme.

(40) Brisson, J.; Brisse, F. *Can. J. Chem.* **1985**, *63*, 3390-3397. Étude Comparative Dans la Série des Dibenzamido-alcanes, Composés Modèles des Poly(oligométhylènes téréphtalamides).

(41) Dobado, J. A.; Henar-Martínez-García; Molina, J. M.; Sundberg, M. R. *J. Am. Chem. Soc.* **1999**, *121*, 3156-3164. Chemical Bonding in Hypervalent Molecules Revised. 2.‡ Application of the Atoms in Molecules Theory to Y_2XZ and Y_2XZ_2 (Y = H, F, CH_3; X = O, S) Compounds.

(42) Dikarev, E. V.; Petrukhina, M. A.; Li, X.; Block, E. *Inorg. Chem.* **2003**, *42*, 1966-1972. Small Organoselenium Molecules. 1. Dimethyl Selenoxide: Structure, Complexation, and Gas-Phase Transformation.

(43) Iwaoka, M.; Takahashi, T.; Tomoda, S. *Heteroat. Chem.* **2001**, *12*, 293-299. Syntheses and Structural Characterization of Water-Soluble Selenium Reagents for the Redox Control of Protein Disulfide Bonds.

(44) Takahashi, T.; Kurose, N.; Kawanami, S.; Arai, Y.; Koizumi, T.; Shiro, M. *J. Org. Chem.* **1994**, *59*, 3262-3264. Optically Pure Haloselenuranes. First Synthesis and Nucleophilic Substitutions.

(45) Fujiwara, T.; Tanaka, N.; Ooshita, R.; Hino, R.; Mori, K.; Toda, F. *Bull. Chem. Soc. Jpn.* **1990**, *63*, 249-251. Crystal and Molecular Structure of the Crystalline Host-Guest Complex Between (*R*)-(+)-2,2'-1,1'-Binaphthyl and (*S*)-(-)-(Ethyl *m*-Tolyl Selenoxide).

(46) French, G. S.; Wilson, K. S. *Acta Crystallogr., Sect. A* **1978**, *34*, 517-525. On The Treatment of Negative Intensity Observations.

Chapter 4: Structural Analysis of a Tetrameric ββα Mini-Protein

Introduction

The X-ray crystal structures of the homooligomeric ββα mini-protein **1**, and of two selenomethionine analogs, **3** and **6**, were determined as described in Chapter 3. These constitute the first reported structures of a mixed α/β oligomeric mini-protein. The crystal structures of peptides **1**, **3**, and **6** are highly homologous. This chapter will focus on a structural analysis of the oligomeric BBA motif exhibited by this family of peptides.

The X-ray crystal structures reveal that the oligomeric BBA motif has a domain-swapped architecture that supports a protein-like and water-exclusive core. The structures elucidate the unique role of unnatural amino acids in conferring native secondary structure in a short peptide sequence (21 residues per monomer). Furthermore, the crystal structures reveal that the stoichiometry of the oligomer is tetrameric, rather than trimeric, as originally proposed.[1] Finally, the X-ray crystal structures offer several indications that the oligomeric state in solution is also tetrameric.

A review of all prior sedimentation equilibrium analytical ultracentrifugation (AUC) data was undertaken. Improved conditions for the collection and interpretation of AUC data of the oligomeric BBA family of peptides were developed. Higher quality data were achieved when peptides were sedimented in two-sector velocity cells and in the presence of salt to reduce the effects of nonideality. Data analysis was improved by testing various association models and incorporating nonideality in the fitting model as appropriate. The partial specific volume of peptide **1** was experimentally determined by densitometry. It was found that the calculated value for the partial specific volume of **1** was in excellent agreement with the measured value. A tetrameric solution-phase stoichiometry for the peptides **BBAT1**, **1**, **3**, and **6** was confirmed by these more rigorous analytical ultracentrifugation experiments.

Overall Features of the Tetrameric BBA Structure

The three crystal structures yielded four crystallographically-independent ββα tetramers: one from peptide **1**, which exists as a tetramer in the asymmetric unit, two from peptide **3**, which packs as two half-tetramers per asymmetric unit, and one from peptide **6**, which has one half-tetramer per asymmetric unit. Although each peptide crystallizes in a unique unit cell, the structures of the four tetramers are highly similar (Figure 4-1), having rms deviations less than 0.52 Å. This indicates that the tetrameric structure revealed by these studies is unlikely to be an artifact of crystallization, but rather is likely to reflect the true conformation of this family of peptides in both aqueous and crystalline states. The high degree of structural homology also confirms that the substitution of an oxidized SeMet at an interface position is structurally benign.

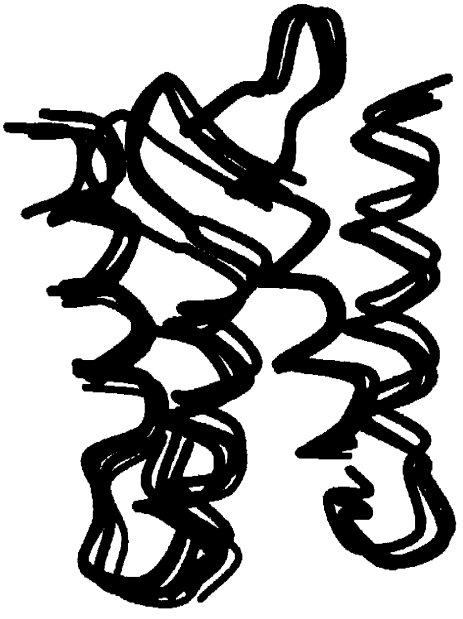


Figure 4-1. Overlaid crystal structures of the four crystallographically-independent tetramers of peptides **1**, **3**, and **6**.

The tetramer consists of four linear monomers arranged in an antiparallel topology (Figure 4-2). Each monomer within the tetramer is flanked by two monomers oriented in the

opposite direction. Each subunit makes helix-helix, helix-hairpin, and hairpin-helix interactions with these adjacent monomers, as predicted by the domain-swapping strategy. A central hydrophobic core is formed by contributions from each of the four monomers. The tetramer possesses approximate, but inexact, four-fold symmetry. Each monomer is slightly closer to one of its neighbors than to the other.

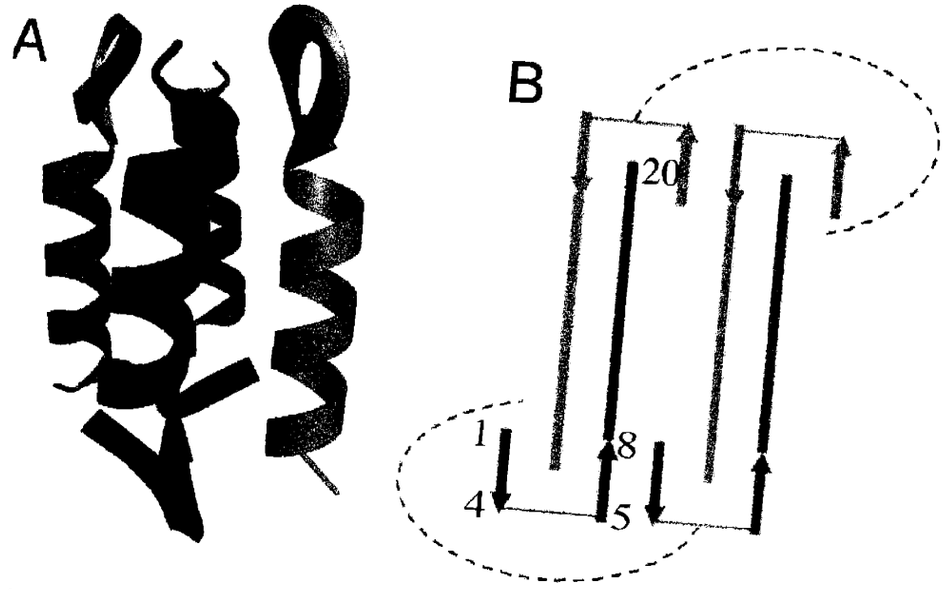


Figure 4-2. A. Ribbon diagram of tetrameric BBA (side view). B. Topology diagram of tetrameric BBA indicating regions of alpha and beta structure and intermonomer association. Dashed line indicates inter-monomer interaction.

The term domain-swapping was originally coined by Eisenberg in 1994 to describe the behavior of proteins having alternate oligomeric states with same inter-domain interactions.[2] The term is often used more broadly to encompass engineered proteins with minor mutations resulting in a domain-exchanged topology. In particular, designed oligomeric proteins having the same exchanged, inter-domain interactions as the monomeric protein from which they derived are commonly referred to as domain-swapped.[3-8]

Peptides **BBA5** and **1** have the same number and type of inter-domain contacts between helix and hairpin secondary structural elements. Furthermore, these hydrophobic contacts are made across the same face of the BBA helix, as indicated in Figure 4-3, although the specific residues in contact are different, as deletion of two residues from the **BBA5** loop changes the helical register of the linker between helix and hairpin domains. Peptide **1** is thus a domain-swapped oligomer in the broad sense in which the term is commonly employed.

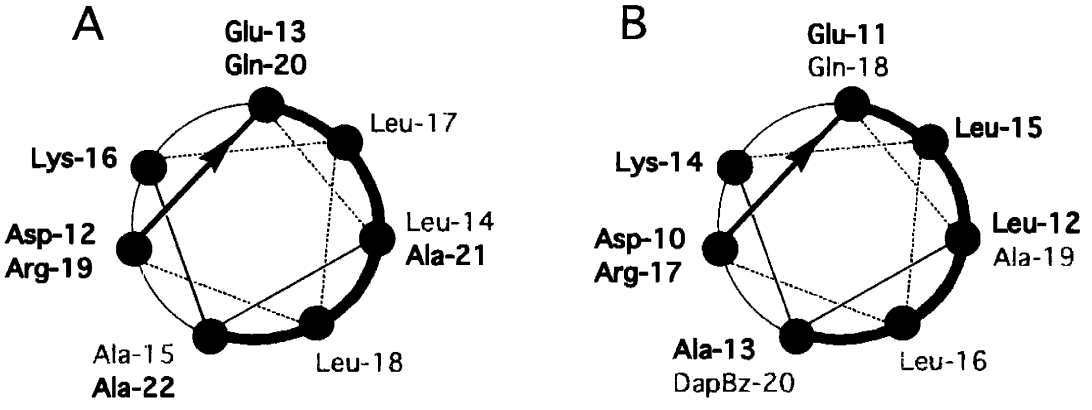


Figure 4-3. A. Helical wheel projection of **BBA5**. The hydrophobic face of the **BBA5** helix and residues involved in inter-domain contacts are indicated in green. B. Helical wheel projection of peptide **1**. The hydrophobic face of **1** and residues involved in inter-domain contacts are indicated in green.

The domain-swapping interaction between alpha and beta elements is allowed by the fact that each monomer is approximately linear (Figure 4-4). A consequence of the linearity of the structural elements in the monomer, and of the overall symmetry of the oligomer, is that the oligomeric structure is constrained to multimers of even number. It would be impossible to construct a symmetrical domain-swapped trimer from these monomer building blocks.

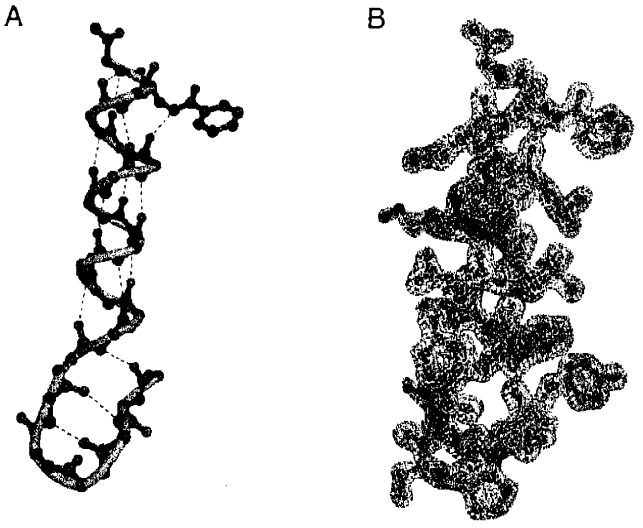


Figure 4-4. A. Monomer model indicating hydrogen bonding. B. Model of chain A of peptide **2** with corresponding electron density at 1σ contour level.

The crystal structures of peptides **1**, **3**, and **6** fully account for the biophysical data collected for this family of peptides. The individual domains are highly similar to those in **BBA5**,[9] as was predicted on the basis of their high sequence identity and of NMR studies of **BBAT1**.[1] Each monomer exists in a symmetrical environment, as suggested by the degeneracy of the NMR signals in spectra of peptide **BBAT1**.[1]

The residues predicted to be in close proximity by NOESY spectroscopy of peptide **BBAT1** are indeed in close contact in the crystal structures. NOE interactions observed between helix and hairpin residues (Ile-3/Dbz-20, Tyr-6/Dbz-20, Phe-8/Dbz-20) in **BBAT1** were taken as proof of the success of the domain-swapping design strategy. These residues are adjacent in the crystal structure, and do indeed interact via a domain-swapping interaction.

The central region of helix-helix interactions results in the characteristic CD signature of interacting helices[10,11] observed for these peptides. The helices associate with a slight right-handed superhelical twist (Figure 4-5), the reverse hand of what is commonly seen with coiled coils.[12] (There are a few coiled coils having a right-handed superhelical twist, including a

designed tetrameric peptide[13] and a naturally occurring tetrameric coiled coil.[14]) The angle of interhelical association, and thus the superhelical twist, is probably determined by the linkage between alpha and beta domains.

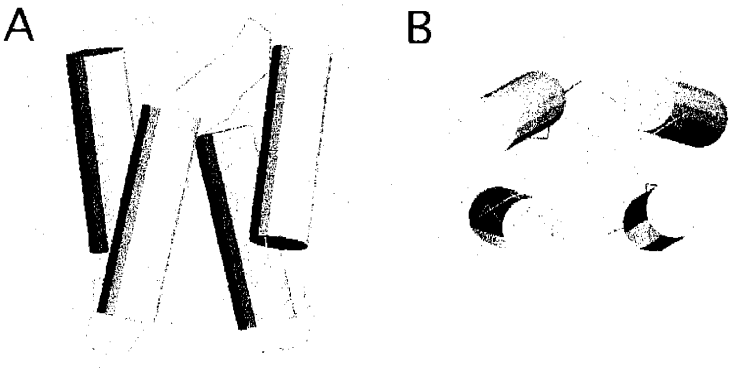


Figure 4-5. The helical domains associate in a right-handed fashion. The helical regions of the structure are represented by cylinders and the beta sheet regions by a C_α trace. A. Side view. B. Top view.

The crystallographic packing suggests that the peptide associates as a tetramer prior to crystallization. Each tetramer is enveloped by a sheath of waters of crystallization (Figure 4-6). There is a clearly demarcated zone of water exclusion in the central core region of the tetramer, indicating that no waters can pass into this hydrophobic core. In addition, Connolly surface diagrams[15,16] of the tetrameric structure show that solvent is excluded from the hydrophobic core.

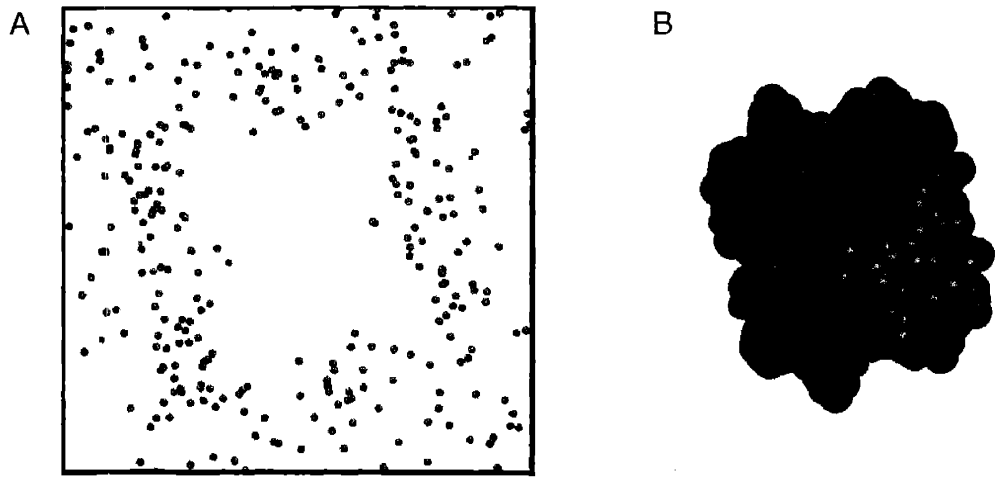


Figure 4-6. A. Crystallographic waters, with void in the shape of the tetramer. B. CPK model of peptide **1** indicating space occupied by tetramer.

The tetrameric structure is highly protein-like in its degree of complexity and well-organized packing of side chains. Moreover, the relationships between surface area, volume, and molecular weight fit well to a series of empirical relationships developed by Chothia for monomeric and oligomeric proteins of varying sizes (Table 4-1),[17,18] further suggesting that the BBA tetramer has a protein-like organization.

	Measured	**Calculated**
Tetramer Surface Area (Å^2)	5969 ± 427	6027
Monomer Surface Area (Å^2)	2382 ± 71	2102
Total Buried Surface (Å^2)	9559 ± 695	9501
Surface Buried in Interface (Å^2)	890 ± 78	595

Table 4-1. Tetramer and monomer surface areas measured from the fully-refined models of peptides **1**, **3**, and **6** by the CCP4 program SURFACE are in excellent agreement with the calculated values determined from empirical relationships for oligomeric proteins developed by Chothia and coworkers. Reported values are the average of four crystallographically independent tetramers.

Hydrophobic Core

The well-defined central hydrophobic core of the tetramer is formed by the residues from one monomer packed against core residues from the three other monomers. Those residues on the periphery of the hydrophobic interface interact with residues of the antiparallel monomer nearest them. The hydrophilic residues are solvent-exposed.

The packing of the hydrophobic core of the BBA tetramer is highly specific, in contrast to the molten globule state common to many designed peptide oligomers, in which hydrophobic residues form a core but lack specific interactions amongst residues in the core.[12] The tight packing of the hydrophobic core accounts for the cooperative unfolding and high melting temperatures of these peptides. This packing is remarkable in light of the very short linear sequence (21 amino acids in total, with 8 residues in the beta hairpin and 12 residues in the helix).

In many designed coiled-coil peptides specificity has been achieved by the inclusion of a buried polar residue, designed to mimic the role of the buried asparagine in GCN4[19,20] in setting the register of the association and in preventing non-specific interactions.[21,22] The buried polar residue increases specificity by destabilizing interactions that are slid out of register or have an unfavorable relative orientation of the helices.

The hydrophobic core of the BBA tetramer does not contain any buried polar residues that may confer packing specificity. Instead, the hairpin region of the oligomeric peptides may serve to set the register of the monomers with respect to one another, and to define the anti-parallel packing. The relative prevalence of bulky and conformationally restricted aromatic residues in the core (*vide infra*) may be another determinant of specificity. A third contribution may arise from the fact that a loss of registration in such a small peptide would result in a large energetic cost as a greater relative percentage of the hydrophobic surface area would be exposed to solvent.

The hydrophobic interface of peptide **1** is defined by five palindromic layers including residues from both the α-helix and the β-sheet (Figure 4-7, Figure 4-8): Ile-3/DapBz-20 (layer A), Phe-8/Leu-16 (layer B), Leu-12/Leu-12 (layer C), Leu-16/Phe-8 (layer B'), and DapBz-20/Ile-3 (layer A'). Two of the five residues involved in these contacts (Phe and DapBz) are aromatic amino acids. This is consistent with the results of a survey of which found a relatively high proportion of aromatic amino acids in oligomeric protein interfaces.[23] The three inner layers roughly resemble an antiparallel four-stranded coiled coil as they are composed of residues from four, antiparallel, associating alpha helices.

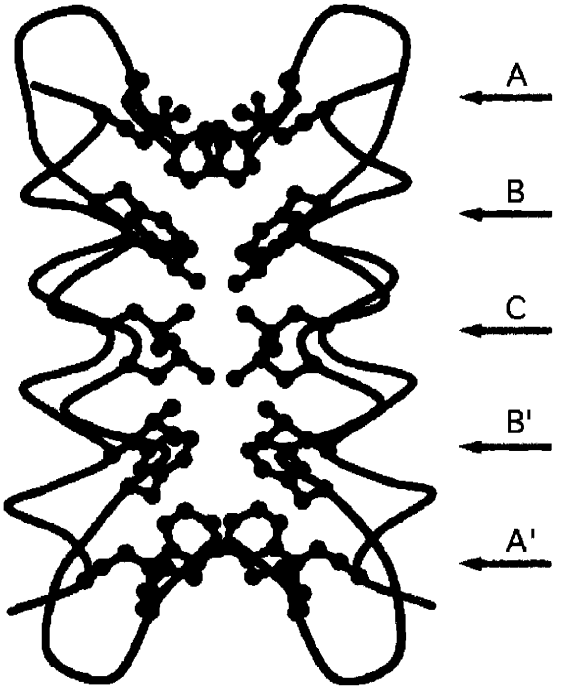


Figure 4-7. Five palindromic layers comprise the hydrophobic core of the tetramer.

Whereas in an antiparallel tetrameric coiled coil the angles at which side chains reach into the core, as measured by the vector connecting C_α and C_β atoms, are regular, in the BBA tetramer each of three unique layers differs with regard to the C_α-C_β vectors. There is no clear

hydrophobic patterning analogous to a coiled-coil heptad repeat, although every fourth residue of the helix is a hydrophobic residue forming the core. Ile-3, located on the β-sheet just before the type II' turn, does not conform to this repeat pattern.

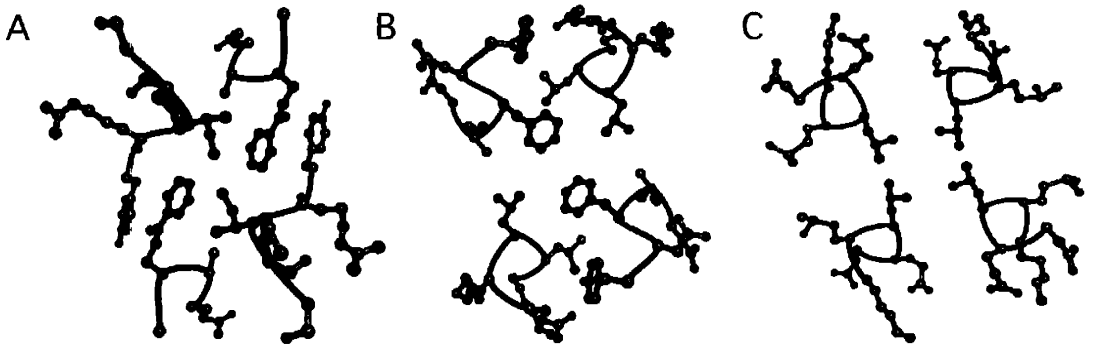


Figure 4-8. Cross-sectional slices through layers A, B, and C. Residues comprising the hydrophobic core are highlighted.

The outermost layers, A and A' (Figure 4-9), are composed of two large DapBz residues from the two monomers with one orientation and two smaller isoleucines from the two monomers with the opposite orientation. DapBz is located at the C-terminal end of the BBA helix and isoleucine near the N-terminus of the beta sheet. This close proximity in the oligomer is the source of the NOEs observed between DapBz and Ile in the NMR studies of **BBAT1**.[1]

DapBz was not introduced in **BBAT1** as a design feature, but as a remnant of the fluorescence-quenching screen used for the identification of oligomerization.[1,24] Fortuitously, the structural and conformational properties of DapBz appear uniquely suited for its location. The large size of the DapBz side chain enables it to approach an isoleucine from a neighboring chain, and the aromatic head group imparts structural rigidity. With the aliphatic base, it is both larger and more flexible than the standard aromatic amino acids. The polar elements of the DapBz side chain also have interesting structural roles. These contributions will be addressed in a later section.

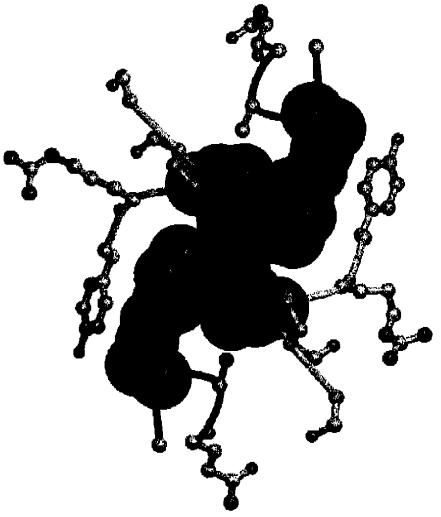


Figure 4-9. Cross-sectional slice through layer A. Space-filling models are shown of the residues comprising the hydrophobic core.

Layers A and A' are packed fairly efficiently, although a CPK model indicates that there is some unfilled space between the residues. To investigate whether this packing could be improved by mutating either site, Christina M. Taylor and Professor Amy E. Keating of the MIT Biology Department applied the Keating lab protein design algorithm,[25,26] which is described further in the following chapter, to investigate whether a pair of natural amino acids could be used to repack the A and A' layers.

The highest-ranking solutions found replaced DapBz with lysine (or less frequently, arginine) and isoleucine with a similarly-sized nonpolar residue (Ile, Leu, Val) and in two instances, Lys. The hydrophobic body of the lysine (or arginine) packed into the hydrophobic core, whereas the charged group reached into the solvent. Neither arginine nor lysine packs as efficiently in position 20 as DapBz. The results of the calculations were taken to indicate that a large and hydrophobic amino acid is necessary in position 20, and a smaller hydrophobic residue in position 3. The large size and mixed aliphatic and aromatic character of DapBz appear to render it uniquely suited for position 20.

The crystal structures reveal that *m*- or *p*-substituted benzoic acid caps, such as those prepared in Chapter 2, would be likely to result in a steric clash with an isoleucine residue, and thus preclude a specific tetrameric interaction. Nonspecific burial of hydrophobic residues in these mutants may have been responsible for the observed aggregation and insolubility.

A number of mutations of peptide **1** at DapBz-20 were made and characterized by Dr. Kevin A. McDonnell.[27] Substituting DapBz with a natural aromatic amino acid (Trp, Phe, Tyr) resulted in peptides that were poorly soluble. The Trp mutant was too insoluble to study by AUC. It may lack the flexibility to replace DapBz in a tightly-packed core. The Phe and Tyr mutants are soluble at low concentrations, at which they displayed intermediate, non-discrete stoichiometry. Some association may be possible, but it would be weaker because neither phenylalanine nor tyrosine is large enough to pack tightly against isoleucine.

Substituting DapBz with the more flexible aliphatic amino acids leucine, isoleucine, and norleucine resulted in peptides with increased solubility.[27] Some association was noted by AUC, but none of these peptides sedimented as a single species. This may be because these peptides lacked the steric bulk for close-packing with adjacent isoleucine residues.

Layers B and B' are composed of two phenylalanines (Phe-8, the last residue of the BBA hairpin) and two leucines (Leu-16, along the BBA helix). A space-filling view of these layers reveals that they are very tightly packed (Figure 4-10). In the NMR studies of **BBAT1** an NOE interaction was observed between Phe-8 and one of the three leucines (12, 15, or 16), which were not distinguished.[1] It is likely that that NOE was between Phe-8 and Leu-16.

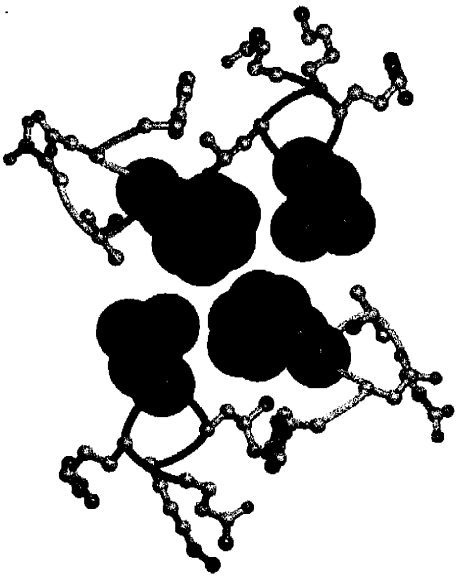


Figure 4-10. A cross-sectional slice through layer B. Space-filling models are shown of the residues comprising the hydrophobic core.

A mutant of **1** having a tyrosine at position 8, prepared by Dr. McDonnell,[27] was found to be quite insoluble. At the low concentrations at which it is soluble, it is completely monomeric, suggesting that the environment around Phe-8 is either too closely packed to accommodate an extra hydroxyl group, too hydrophobic to accommodate a polar hydroxyl group, or both. The crystal structure reveals that the environment of Phe-8 is indeed both very hydrophobic and very closely packed.

The central layer, layer C (Figure 4-11), is composed of leucine-12 residues from each of the monomers. This layer has the least optimal packing of the five layers comprising the core. The four leucines do not extend far enough to fill the center of the layer, resulting in a small cavity. This cavity affords the leucines significant conformational mobility. Two of the four crystallographic tetramers contain one Leu-12 with an alternate conformation populated highly enough to model.

Cavities between subunit interfaces are common in oligomeric proteins and occur in approximately three quarters of oligomeric structures.[28] Tetrameric oligomers are particularly prone to channel formation as each monomer has close contacts with the two monomers adjacent to it, but is often too far to have significant contact with the remaining distal monomer.[29] This suggests that the cavity in the central layer of the tetrameric BBA is not necessarily a design flaw particular to mini-proteins, but may reflect a general feature of oligomeric proteins.

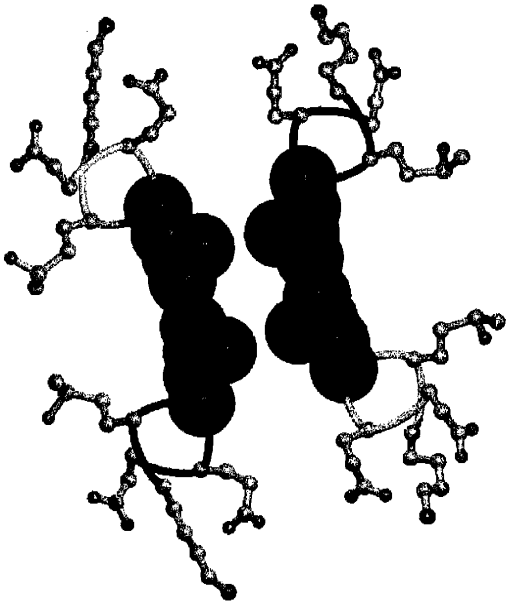


Figure 4-11. A cross-sectional slice through layer C. Space-filling models are shown of the residues comprising the hydrophobic core.

Christina M. Taylor and Professor Amy E. Keating also applied their protein design algorithm[25,26] toward layer C of the tetrameric BBA. Using the crystal structures of **1**, **3**, and **6**, with a fixed backbone, they found that, in a homooligomeric tetramer, leucine, isoleucine, and valine pack together approximately equally well, and pack better than any other naturally-occurring residues (Figure 4-12). Their calculations predicted that substituting four larger

residues for Leu-12 would create a steric clash in the tetrameric core, whereas substituting four smaller residues would result in insufficient bulk for significant inter-monomer interactions.

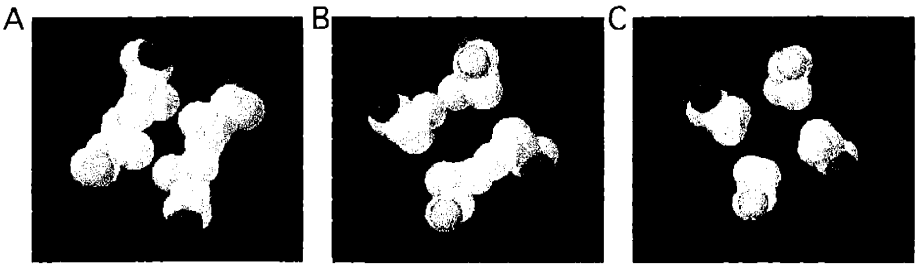


Figure 4-12. Minimized structures of the best solutions for the repacking of layer C, by Christina M. Taylor and Prof. Amy E. Keating. A. Leucine residues at position 12. B. Isoleucine residues at position 12. C. Valine residues at position 12. Figure courtesy of Christina M. Taylor.

Unnatural and Otherwise Noteworthy Amino Acids

The first non-standard amino acid in the sequence is D-proline. The DPro-Ser sequence nucleates a type II' turn in the short BBA hairpin sequence.[30,31] All D-proline residues in the three structures fell in generously allowed regions of a Ramachandran plot defined by PROCHECK.[32]

The last two residues of the hairpin, Asp-7 and Phe-8, simultaneously participate in the β-hairpin and nucleate the α-helix via *(i,i+4)* hydrogen bonds between the carbonyl groups and the amides of residues one turn above (Figures 4-4 and 4-13). The dihedral angles of Asp-7 (ϕ = −84.3° ± 4.6°, ψ = 80.7° ± 7.9°) are typical of beta secondary structure, but those of Phe-8 (ϕ = −60.0° ± 3.7°, ψ = −45.8° ± 4.1°) are α-helical. Phe-8 could thus be considered the first residue of the BBA helix as well as the last residue of the hairpin.

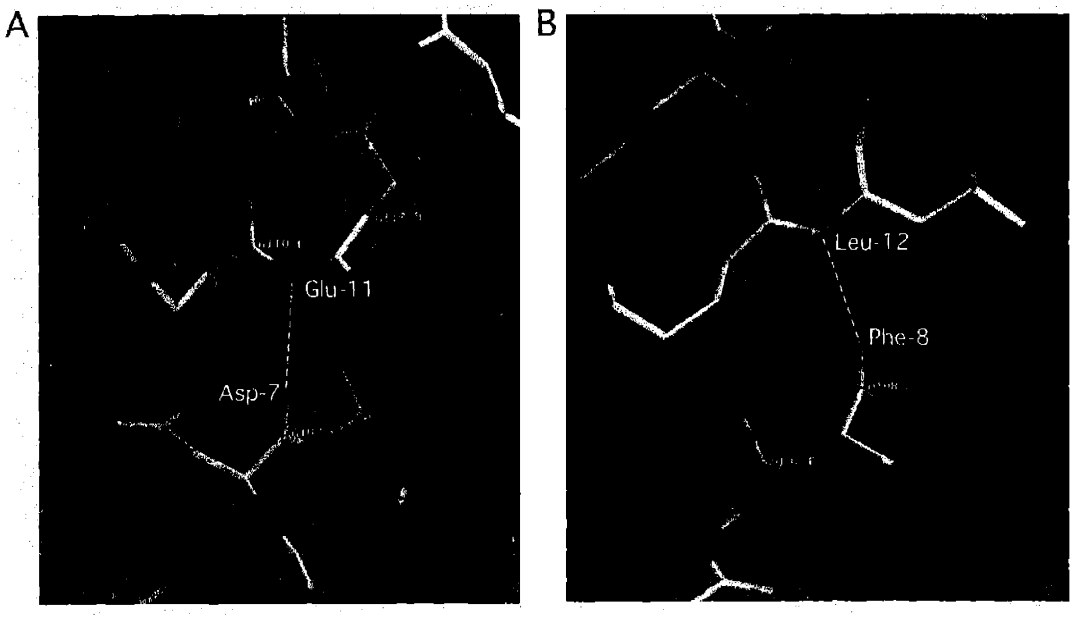


Figure 4-13. Asp-7 and Phe-8, the last residues of the beta hairpin, nucleate the alpha helix via (*i,i+4*) hydrogen bonds to Asp-11 and Leu-12 (A and B, respectively). Electron density is contoured at 1.2σ.

The D-alanine hinge residue participates in the first turn of the helix and compactly reverses the direction of the hairpin to produce a linear monomer. The dihedral angles ($\phi = -51.8° \pm 2.5°$, $\psi = -52.1° \pm 2.0°$) fall in the right-handed α-helical region for an L-amino acid on a Ramachandran plot (Figure 4-14).

These dihedrals fall in the "inverted α-left" region (−60°, −45°), defined by Mitchell and Smith,[33] allowed for D-amino acids in an L-amino acid environment. Intriguingly, this region, although allowed, is sparsely populated in known structures of amino acids in LDL environments. (The corresponding α_L region is also sparsely populated in L-amino acid environments.)

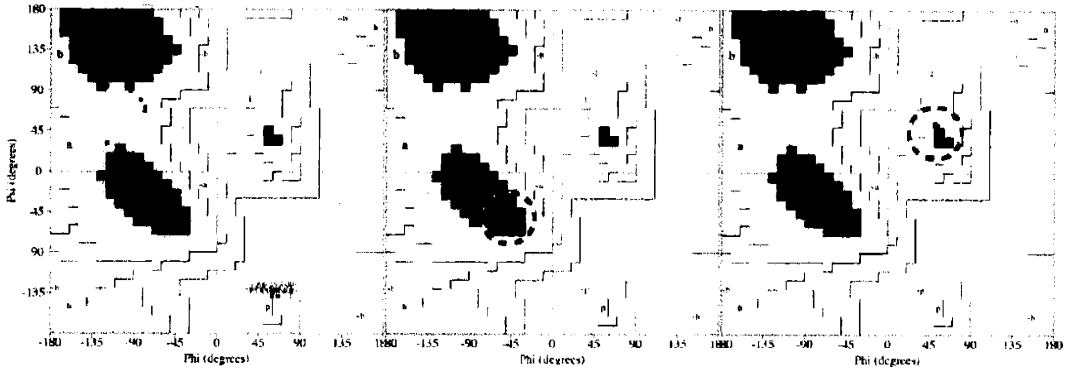


Figure 4-14. Left: Ramachandran plot of the refined structure of peptide **6**. Middle: The D-alanine residues fall in the allowed right-handed alpha helical region of a standard Ramachandran plot. Right: The D-alanine residues, with inverted chirality, would fall in the allowed left-handed alpha helical region of a standard Ramachandran plot. The locations of the D-alanine residues are circled in the middle and right panels.

The winning peptide identified from the fluorescence quenching assay had a capped L-cysteine, Cys(SSMe), as an inter-domain linker.[1,24] Although this peptide exhibited fluorescence quenching indicative of self-association, it was relatively insoluble, and was replaced by the more soluble **BBAT1**, having a glycine hinge residue, for further characterization.[1]

McDonnell and Imperiali found that replacing the glycine of **BBAT1** with D-alanine resulted in a more stable complex with the same oligomeric state.[27,34] By contrast the substitution of an L-alanine resulted in a species with mixed oligomeric state. It was postulated that the flexible glycine hinge residue could access dihedral angles favored by D-amino acids, and that these D-dihedral angles were optimal for the hinge position. The greater stability of **1** relative to **BBAT1** could be attributed to the increased conformational rigidity of D-alanine relative to glycine.

The crystal structures reveal that the D-alanine residue does not access dihedral angles that an L-amino acid cannot. However, the conformational space available to it in the inverted α-left region is quite limited (Figure 4-14), which may impart significant structural specificity. This may account for the observed stabilization of **1** relative to **BBAT1**. In contrast, glycine or an L-

amino acid at position 9 could exhibit dihedral angles anywhere in the right-handed alpha helical space.

A number of studies with model peptides have shown that the introduction of a D-alanine residue in the center of a helix results in significant destabilization of the helical structure.[35-37] However, none of these studies address the role of the sequence position of the D-alanine. This structure shows that a D-alanine situated at the start of a helix can participate in, and help to stabilize, a right-handed alpha helix.

DapBz is the last completely ordered residue of the helix and plays a key structural role as a helix cap. Notably, the amide nitrogen of the benzoyl group forms an (*i*,*i*+*4*) hydrogen bond to the carbonyl of Leu-16 that is the penultimate structural feature of the helix, as shown in Figure 4-15.

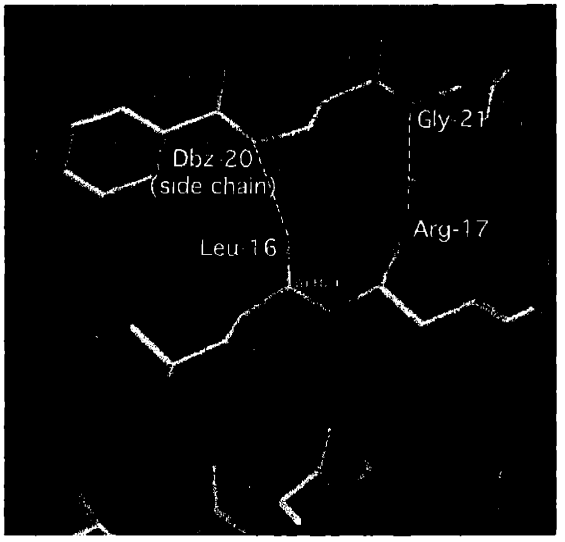


Figure 4-15. The side chain of DapBz serves as a helix cap by accepting a hydrogen bond from the carbonyl of Leu-16. The amine of the terminal glycine also accepts a hydrogen bond from the carbonyl of Arg-17.

Dr. McDonnell replaced the benzoyl group of DapBz with acetyl and benzyl caps in order to probe the contributions of the amine, carbonyl, and aromatic portions.[27] He found that

the acetyl-capped peptide was monomeric at low concentrations, but exhibits some association as the concentration is increased. The benzyl-capped peptide self-associates more strongly than does the acetyl-capped peptide, but the association is not as stable as the native benzoyl-capped peptide. These data suggested that both the side chain amide and the aromatic ring have important structural roles. These findings are borne out by the X-ray crystal structures.

The last residue of the sequence is a glycine designed to terminate the helix. The *C*-terminal glycine is flexible, exhibiting alternate conformations in several of the monomers. However, the nitrogen of the glycine is rigid, and forms an (*i,i+4*) hydrogen bond to the carbonyl groups of residues Arg-17 and Gln-18.

Sedimentation Equilibrium Analytical Ultracentrifugation

Sedimentation equilibrium analytical ultracentrifugation (AUC) is a standard and thermodynamically rigorous method for the determination of molecular weight.[38] In a sedimentation equilibrium experiment a protein sample is spun in an analytical ultracentrifuge, where it is subject to centrifugal forces F_c, counteracted by a buoyant force, F_b, and by frictional drag, F_d (Figure 4-16).[38-40]

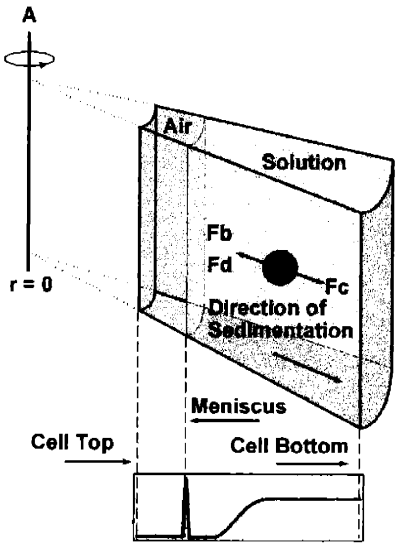


Figure 4-16. The forces at work in a sedimentation equilibrium experiment. Figure courtesy of Professor Borries Demeler, The University of Texas Health Science Center at San Antonio.

After a period of time the sample reaches equilibrium, and there is no net force acting upon it. The centrifugal force is balanced by the buoyant force and the frictional drag: $F_c = F_b + F_d$. As there is no net force, the concentration distribution of the protein does not change after reaching equilibrium.[38-40]

A concentration distribution of a sample at equilibrium is entirely independent of molecular shape and depends only on molecular weight. This provides an absolute and thermodynamically rigorous means of determining the molecular weight and/or association state of a protein of interest.[38-40]

The concentration distribution of a single ideal species at equilibrium can be described by equation (1), where δ is a baseline offset, c_{r_o} is the concentration of the species at the reference radius r_o, $\bar{\nu}$ is the partial specific volume of the protein, and ρ is the density of the solvent.[40]

$$(1)\ C(r) = \delta + C_{r_o} \exp\left[\frac{\omega^2}{2RT} M(1-\bar{\nu}\rho)(r^2 - r_o^2)\right]$$

The concentration distribution of a self-associating species at sedimentation equilibrium is the sum of the concentration distributions of each of the oligomeric species present, which are related by the equilibrium constants of the associations. It can be described by equation (2).[40]

$$(2)\ C(r) = \delta + \sum_{i=1}^{n} C_i(r)$$

Equation (2) can be expanded to equation (3), where n_2 is the stoichiometry of the second species, c_{ro} is the concentration of the monomer at reference radius r_o, and $K_{a_{n_2}}$ is the association constant for the monomer-n_2-mer equilibrium.[40]

$$(3)\ C(r) = \delta + c_{r_o} \exp\left[\frac{\omega^2}{2RT} M(1-\bar{\nu}\rho)(r^2 - r_o^2)\right] + (c_{r_o})^{n_2} K_{a_{n_2}} \exp\left[\frac{\omega^2}{2RT} M(1-\bar{\nu}\rho)(r^2 - r_o^2)\right] + \ldots$$

Proteins can exhibit nonideal behavior that is not adequately described by the above equation because they are neither uncharged nor dimensionless, but are subject to charge-charge repulsions and to steric crowding. This nonideal behavior can be modeled by the introduction of an additional nonideality term. Nonideal behavior can be minimized or eliminated by studying protein samples in salt-containing buffers (~100 mM NaCl) that reduce the effects of charge interactions and at low concentrations (< 1 mg/ml) to minimize crowding.[39]

Sedimentation equilibrium data can be analyzed by fitting data to equation (4), the linearized form of equation (1). The data from an experiment are plotted in the form ln(C) vs $\left(r^2/2\right)$.[40,41] The slope of the line tangent to the curve of ln(C) vs $\left(r^2/2\right)$ at any radial point is proportional the mass-average molecular weight at that radial position. If the system is not polydisperse, the plot will be linear and the slope of the tangent at any position will be equal to

the slope of the line. The linearity of the data is used as an indication of how well the data fits to a single ideal species model.[40,41]

$$(4)\quad \frac{d\ln[C(r)]}{d(r^2/2)} = \frac{M(1-\bar{\nu}\rho)\omega^2}{RT} = \sigma$$

Although this method can indicate more complex behavior, it is limited by the ability to fit only one model (a single ideal species). Moreover, as only a single scan, at a single concentration and wavelength can be fit at a time, limited information can be obtained. It is difficult to detect a small fraction (≤10%) of the sample that aggregates by this method.[40] Moreover, if both polydispersity and nonideality are present, the effects can be disguised by a linear plot.[40]

A more powerful approach is to globally fit data from a sample collected at multiple speeds and concentrations.[42] A variety of models (single or multiple species, ideal or nonideal) can be fit using software such as NONLIN[42] or SEDPHAT.[43,44]

Solution-Phase Stoichiometry of BBA Peptides

The oligomeric BBA peptides **1** and **BBAT1** were originally assigned as trimeric in aqueous solution on the basis of sedimentation equilibrium AUC experiments.[1,34] In contrast, the oligomerization state for peptides **1**, **3**, and **6** in the crystalline state is clearly tetrameric.

Several pieces of evidence indicated that the tetramer observed in the crystalline form was also the true oligomeric state in solution. (1) The crystallographic data are consistent with all of the biophysical solution-phase data, including the AUC finding of a single discrete oligomeric state, with the exception of the solution-phase stoichiometry of that oligomer. (2) The joint constraints of a symmetrical, domain-swapped, oligomer and a linear monomer preclude a

trimeric association with inter-domain interactions. (3) The crystal packing suggests that the peptides associate as tetramers prior to crystallization. (4) The same tetrameric structure was obtained from three different crystal forms, suggesting that it is not an artifact of crystallization.

A depressed apparent molecular weight is often the result of charge-based nonideality that is not accounted for in the fitting model. There are several examples in the literature of unexpectedly low molecular weights being obtained for designed mini-proteins, particularly due to charge-based nonideality under salt-free conditions.[45-47] This suggested that unaccounted nonideality could be a significant factor accounting for the initially-assigned trimeric oligomeric state.

In the earliest AUC studies from this group[1,27,34] data were collected from peptides in salt-free buffers and were analyzed with a linear plot following equation (4). Moreover, molecular weights were determined using a calculated partial specific volume that may not have adequately accounted for the non-standard amino acids in the oligomeric BBA sequence.

We suspected that the discrepancy in oligomeric state between the early AUC studies and the crystallographic findings was due to any one, or combination, of the following factors: (1) suboptimal data collection conditions, (2) suboptimal data analysis protocols, or (3) the use of an inappropriate partial specific volume. Optimization of data collection and analysis protocols and the experimental determination of the partial specific volume were undertaken in order to obtain a better measure of solution-state stoichiometry.

Data collection conditions were investigated first. As this family of peptides has six charges per monomer, BBA oligomers are susceptible to charge-charge repulsions and consequent nonideal behavior, despite the net neutrality. All subsequent data were collected under conditions of increased ionic strength: 50 mM phosphate buffer, 100 mM salt, pH 7.2. Even with the addition of salt, the introduction of a nonideality term was found to result in improved fits to the data in many cases.

Sedimentation equilibrium data are typically collected at three rotor speeds in order to accurately model the behavior of the sample. These three speeds are chosen such that the equilibrium concentration distribution at each speed of the sample under study is significantly different, thus providing unique information for data fitting.[40,48] Recommended differences in rotor speed can be calculated from equations (5) and (6), where ω_1, ω_2, and ω_3 are the first, second, and third angular velocities.[40]

$$(5)\ \frac{\omega_2^2}{\omega_1^2} \geq 1.4$$

$$(6)\ \frac{\omega_3^2}{\omega_1^2} \geq 3$$

It is not possible to achieve the recommended difference between rotor speeds at the high rotor speeds necessary for the sedimentation of these low molecular weight peptides. The simultaneous constraints of measurable sedimentation at the low end and of the ratings of the cells and rotor at the high end preclude a large distribution in angular velocities. This limitation results in somewhat less information from an experiment than is recommended.

Additional resolution was obtained by using longer solution columns, as suggested by Jeffrey W. Lary of the University of Connecticut's National Analytical Ultracentrifugation Facility. This was accomplished by using two-sector velocity cells rather than standard six-sector equilibrium cells. The use of two-sector cells was found to result in a significant improvement in data quality.

Data analysis conditions were optimized next. Analysis of data for this family of peptides using an ideal species model in NONLIN, rather than a linear plot of ln(C) vs $\left(r^2/2\right)$, resulted in stoichiometries intermediate between three and four, as shown in Table 3-2. These findings

suggested that at least part of the discrepancy between the crystallographic and early AUC results was due to the use of a less powerful fitting model.

An additional source of error may have stemmed from poor fitting of the data in NONLIN due to correlation of the baseline and sigma parameters for low molecular weight species.[49] In NONLIN the baseline of each scan is fit individually, although the baseline offsets of scans of the same cell at different speeds are not independent. This can be a significant source of error for low-molecular weight species such as peptides.[49]

It is possible to constrain the baseline offsets for a given cell to have the same value in SEDPHAT, thus excluding one potential source of error. All data were fit with both NONLIN and SEDPHAT; results from the two programs were consistent in all cases. As the current release of SEDPHAT cannot model nonideality for sedimentation equilibrium data, the final model (and stoichiometry) was taken from NONLIN.

The partial specific volume is generally calculated as the weighted average of tabulated values per amino acid residue.[40,50,51] For proteins, this typically results in very low errors in the molecular weights obtained.[51-53] However, the validity of these values for smaller peptides containing non-standard amino acids has not been established by direct measurement.

In order to obtain confirmation that the tabulated values regularly used for macromolecules were appropriate for this mini-protein family, the partial specific volume of peptide **1** was measured directly by densitometry. This resulted in a value of 0.752 cm^3/g at 20 °C in 10 mM sodium phosphate buffer, pH 7.2, as compared to our calculated value of 0.738 cm^3/g. This value is in excellent agreement with the calculated value.

Optimization of the data collection and analysis procedures, as described above, resulted in more ideal behavior of the peptide, higher-quality data, and better fitting of the data. Analysis of these data, using corrected values for the partial specific volume and including nonideality in the model as appropriate, indicated that peptides **BBAT1**, **1**, **3**, and **6** are best described by a single species model. In each case, the single species has a tetrameric stoichiometry, as shown in

Table 4-2. The tetrameric structure revealed by the X-ray crystallographic studies thus accounts for all of the biophysical data collected from this family of homooligomeric peptides, including the stoichiometry of association.

Peptide	**BBAT1**[2]	**1**[3]	**3**[3]	**6**[2]
Monomer MW	2544	2558	2681	2624
Sigma[1]	1.684	1.778	1.982	1.863
Calculated $\bar{\nu}$[4]				
$\bar{\nu}$	0.738	0.740	0.740	0.744
MW	9,285	9,928	11,088	10,569
Stoichiometry	3.7	3.9	4.1	4.0
Measured $\bar{\nu}$[4,5]				
$\bar{\nu}$	0.754	0.754	0.754	0.754
MW	9,921	10,517	11,721	11,016
Stoichiometry	3.9	4.1	4.4	4.2

Table 4-2. Results of sedimentation equilibrium experiments.
1. Sigma is related to the molecular weight by the equation $M = \frac{\sigma RT}{(1-\bar{\nu}\rho)\omega^2}$, and determined using the program NONLIN.
2. Single species model, B=0.
3. Single species model, B≠0.
4. Partial specific volume calculated using SEDNTERP (25 °C).
5. Vbar of all four peptides are assumed to be approximately equal.
6. Extrapolated from data at 20 °C to 25 °C using the formula $\bar{\nu}(T) = \bar{\nu}(20°C) - [4.25 \bullet 10^{-4}(T - 298.15)]$

Conclusions

The X-ray crystal structures of peptides **1**, **3**, and **6** are structurally homologous. They reveal a tetrameric and highly protein-like oligomeric BBA motif. Each monomer within the tetramer is approximately linear, and is flanked by two monomers oriented in the opposite direction. Each monomer makes both interhelical and domain-swapped interactions with its neighbors. Residues from the four monomers form a central hydrophobic core, which can be divided into five palindromic layers of interaction. The unnatural amino acid DapBz plays a key structural role in forming the hydrophobic core of the tetramer and as a helix cap.

The tetrameric BBA structure is consistent with all the biophysical data previously collected from this family of oligomeric peptides with the exception of the initial assignment of a trimeric oligomeric state. However, evidence from the crystal structures suggested that the peptides associate as tetramers in the aqueous phase as well as in the crystalline state.

Improved sedimentation equilibrium analytical ultracentrifugation data collection and data analysis protocols were developed for this family of peptides. The partial specific volume of peptide **1** was experimentally determined by densitometry and the experimental value found to agree with the approximate partial specific volume calculated from tabulated values. The results of the sedimentation equilibrium experiments confirmed that peptides of this family are tetrameric in solution.

Acknowledgements

I am grateful to Christina M. Taylor and Professor Amy E. Keating of the MIT Biology Department for performing design calculations on the core of the tetramer to optimize stability. I thank a number of people for assistance with analytical ultracentrifugation experiments and data analysis: Jeffrey W. Lary and Professor James L. Cole of the University of Connecticut National Analytical Ultracentrifugation Facility, Dr. Peter Schuck of the National Institutes of Health, Dr. Walter F. Stafford of the Boston Biomedical Research Institute, and Dr. James D. Lear of the University of Pennsylvania. I am also very grateful to Dr. K. Jebrell Glover, without whom the densitometry experiments would not have been possible.

Materials and Methods

Surface Area Calculations

Measured. Surface areas were calculated by the CCP4[54] program SURFACE[55,56] with a 1.40 Å radius water probe. Van der Waals radii were taken from Chothia.[55] Reported values are the average of the four crystallographically independent tetramers.

Calculated. Surface areas were calculated using empirical relationships developed by Chothia and coworkers.[17,18] The surface area as a function of molecular weight is given by the equation $A_S = 5.3M^{0.76}$ or $A_s = 5.3m^{0.76}$, where A_s is the surface area in $Å^2$, M is the molecular weight of the oligomer in Daltons, and m is the molecular weight of the subunit in Daltons. The total buried surface in the oligomer, A_b, is given by the equation $A_b = 1.48M - 5.3M^{0.76}$. The surface buried in an oligomeric interface, A_{int}, is the difference between the surface area of a subunit in isolation and in an oligomer, and is given by the equation $A_{int} = 3.8m^{0.76} - 5.3m^{0.76}$.

Analytical Ultracentrifugation

Stock solutions of peptides were prepared in 50 mM phosphate buffer with 100 mM NaCl at pH 7.2. Each solution was then loaded onto a 1.0 kDa MWCO DispoDialyzer™ (Harvard Biosciences) and dialyzed at room temperature. Dilutions were made from the dialyzed stock solution. Peptide solutions were loaded into standard two-sector epon centerpieces and 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 contents of each cell were confirmed to be at equilibrium using WINMATCH prior to increasing the speed. Data were analyzed with the programs NONLIN[42] and SEDPHAT.[43,44] Molecular weights were determined using a partial specific volume, $\bar{\nu}$, calculated using SEDNTERP[57] or determined by densitometry.

Partial Specific Volume Determination

Peptide **1** was dissolved in 10 mM phosphate buffer (pH 7.2), filtered through a 0.45 μM syringe filter, and extensively dialyzed in MWCO 500 dialysis tubing (Spectrum® Laboratories). The concentration of the dialyzed solution was determined based on absorbance at 280 nm. The dialyzed solution of **1** and a sample of dialysate were immediately brought to the National

Analytical Ultracentrifugation Facility at the University of Connecticut. The densities of the stock solution and four dilutions thereof (80, 60, 40, and 20% of the stock solution) were measured at 20 °C using a DMA 60 Density Meter (Anton Paar).

References

(1) Mezo, A. R.; Ottesen, J. J.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 1002-1003. Discovery and Characterization of a Discretely Folded Homotrimeric Peptide.

(2) Bennett, M. J.; Choe, S.; Eisenberg, D. *Protein Sci.* **1994**, *3*, 1444-1463. Refined Structure of Dimeric Diphtheria Toxin at 2.0 Å Resolution.

(3) Byeon, I.-J. L.; Louis, J. M.; Gronenborn, A. *J. Mol. Biol.* **2003**, *333*, 141-152. A Protein Contortionist: Core Mutations of GB1 That Induce Dimerization and Domain Swapping.

(4) Chen, D.-T.; Lin, A. *Protein Eng.* **2002**, *15*, 997-1003. Domain-Swapping in Ribonuclease T1 Allows the Acquisition of Double-Stranded Activity.

(5) O'Neill, J. W.; Kim, D. E.; Johnsen, K.; Baker, D.; Zhang, K. Y. J. *Structure* **2001**, *9*, 1017-1027. Single-Site Mutations Induce 3D Domain Swapping in the B1 Domain of Protein L From *Peptostreptococcus magnus*.

(6) Green, S. M.; Gittis, A. G.; Meeker, A. K.; Lattman, E. E. *Nat. Struct. Biol.* **1995**, *2*, 746-751. One-Step Evolution of a Dimer From a Monomeric Protein.

(7) Kuhlman, B.; O'Neill, J. W.; Kim, D. E.; Zhang, K. Y. J.; Baker, D. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10687-10691. Conversion of Monomeric Protein L to an Obligate Dimer By Computational Protein Design.

(8) Ogihara, N. L.; Ghirlanda, G.; Bryson, J. W.; Gingery, M.; DeGrado, W. F.; Eisenberg, D. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 1404-1409. Design of Three-Dimensional Domain-Swapped Dimers and Fibrous Oligomers.

(9) Struthers, M. D.; Ottesen, J. J.; Imperiali, B. *Fold. Des.* **1998**, *3*, 95-103. Design and NMR Analyses of Compact, Independently Folded BBA Motifs.

(10) Zhou, N. E.; Zhu, B.-Y.; Kay, C. M.; Hodges, R. S. *Biopolymers* **1992**, *32*, 419-426. The Two-Stranded α-Helical Coiled-Coil is an Ideal Model for Studying Protein Stability and Subunit Interactions.

(11) Zhou, N. E.; McKay, C.; Hodges, R. S. *Biochemistry* **1992**, *31*, 5739-5746. 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.

(12) Kohn, W. D.; Hodges, R. S. *Trends Biotechnol.* **1998**, *16*, 379-389. *De Novo* Design of α-Helical Coiled Coils and Bundles: Models for the Development of Protein-Design Principles.

(13) Harbury, P. B.; Plecs, J. J.; Tidor, B.; Alber, T.; Kim, P. S. *Science* **1998**, *282*, 1462-1467. High-resolution protein design with backbone freedom.

(14) Stetefeld, J.; Jenny, M.; Schulthess, T.; Landwehr, R.; Engel, J.; Kammerer, R. A. *Nat. Struct. Biol.* **2000**, *7*, 772-776. Crystal Structure of a Naturally Occurring Parallel Right-Handed Coiled Coil Tetramer.

(15) Connolly, M. L. *J. Appl. Cryst.* **1983**, *16*, 548-558. Analytical Molecular Surface Calculation.

(16) Connolly, M. L. *Science* **1983**, *221*, 709-713. Solvent-Accessible Surfaces of Proteins and Nucleic Acids.

(17) Miller, S.; Lesk, A. M.; Janin, J.; Chothia, C. *Nature* **1987**, *328*, 834-6. The Accessible Surface Area and Stability of Oligomeric Proteins.
(18) Janin, J.; Miller, S.; Chothia, C. *J. Mol. Biol.* **1988**, *204*, 155-64. Surface, Subunit Interfaces and Interior of Oligomeric Proteins.
(19) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401-1407. A Switch Between Two-, Three-, and Four-Stranded Coiled Coils in GCN4 Leucine Zipper Mutants.
(20) Gonzales, L., Jr.; Brown, R. A.; Richardson, D.; Alber, T. *Nat. Struct. Biol.* **1996**, *3*, 1002-1009. Crystal Structures of a Single Coiled-Coil Peptide in Two Oligomeric States Reveal the Basis for Structural Polymorphism.
(21) Lumb, K. J.; Kim, P. S. *Biochemistry* **1995**, *34*, 8642-8648. A Buried Polar Interaction Imparts Structural Uniqueness in a Designed Heterodimeric Coiled Coil.
(22) Schneider, J. P.; Lear, J. D.; DeGrado, W. F. *J. Am. Chem. Soc.* **1997**, *119*, 5742-5743. A Designed Buried Salt Bridge in a Heterodimeric Coiled Coil.
(23) Argos, P. *Protein Eng.* **1988**, *2*, 101-113. An Investigation of Protein Subunit and Domain Interfaces.
(24) Mezo, A. R.; Cheng, R. P.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 3885-3891. Oligomerization of Uniquely Folded Mini-Protein Motifs: Development of a Homotrimeric Beta Beta Alpha Peptide.
(25) Park, D. J. M. Eng. Thesis, Massachusetts Institute of Technology, 2002. Computational Tools For Including Specificity In Protein Design.
(26) Grigoryan, G.; Keating, A. E. Unpublished results.
(27) McDonnell, K. A. Ph.D. Thesis, Massachusetts Institute of Technology, 2001. Towards Incorporation of Catalytic Function Into Small Folded Peptide Scaffolds.
(28) Hubbard, S. J.; Argos, P. *Protein Sci.* **1994**, *3*, 2194-2206. Cavities and Packing at Protein Interfaces.
(29) Miller, S. *Protein Eng.* **1989**, *3*, 77-83. The Structure of Interfaces Between Subunits of Dimeric and Tetrameric Proteins.
(30) Hutchinson, E. G.; Thornton, J. M. *Protein Sci.* **1994**, *3*, 2207-2216. A Revised Set of Potentials for β-Turn Folding in Proteins.
(31) Haque, T. S.; Little, J. C.; Gellman, S. H. *J. Am. Chem. Soc.* **1994**, *116*, 4105-4106. "Mirror Image" Reverse Turns Promote β-Hairpin Formation.
(32) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Cryst.* **1993**, *26*, 283-291. Procheck - A Program to Check the Stereochemical Quality of Protein Structures.
(33) Mitchell, J. B. O.; Smith, J. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 563-571. D-Amino Acid Residues in Peptides and Proteins.
(34) McDonnell, K. A.; Imperiali, B. *J. Am. Chem. Soc.* **2002**, *124*, 428-433. Oligomeric Beta Beta Alpha Miniprotein Motifs: Pivotal Role of Single Hinge Residue in Determining the Oligomeric State.
(35) Fairman, R.; Anthony-Cahill, S. J.; DeGrado, W. F. *J. Am. Chem. Soc.* **1992**, *114*, 5458-5459. The Helix-Forming Propensity of D-Alanine in a Right-Handed α-Helix.
(36) Krause, E.; Bienert, M.; Schmieder, P.; Wenschuh, H. *J. Am. Chem. Soc.* **2000**, *122*, 4865-4870. The Helix-Destabilizing Propensity Scale of D-Amino Acids: The Influence of Side Chain Steric Effects.
(37) Chen, Y.; Mant, C. T.; Hodges, R. S. *J. Peptide Res.* **2002**, *59*, 18-33. Determination of Stereochemistry Stability Coefficients of Amino Acid Side-Chains in an Amphipathic α-Helix.
(38) Cole, J. L.; Hansen, J. C. *J. Biomol. Tech.* **1999**, *10*, 163-176. Analytical Ultracentrifugation as a Contemporary Biomolecular Research Tool.

(39) Lebowitz, J.; Lewis, M. S.; Schuck, P. *Protein Sci.* **2002**, *11*, 2067-2079. Modern Analytical Ultracentrifugation in Protein Science: A Tutorial Review.
(40) McRorie, D. K.; Voelker, P. J. *Self-Associating Systems in the Analytical Ultracentrifuge*; Beckman Instruments, Inc.: Fullerton, 1993.
(41) Stafford, W. F.; Schuster, T. M. In *Introduction to Biophysical Methods for Protein and Nucleic Acid Research*; Glasel, J. A., Deutscher, M. P., Eds.; Academic Press, Inc.: San Diego, 1995.
(42) Johnson, M. L.; Correia, J. C.; Yphantis, D. A.; Halvorson, H. R. *Biophys. J.* **1981**, *36*, 575-588. Analysis of Data from the Analytical Ultracentrifuge by Nonlinear Least-Squares Techniques.
(43) Schuck, P. *Anal. Biochem.* **2003**, *320*, 104-124. On the Analysis of Protein Self-Association by Sedimentation Velocity Analytical Ultracentrifugation.
(44) Vistica, J.; Dam, J.; Balbo, A.; Yikilmaz, E.; Mariuzza, R. A.; Roualt, T. A.; Schuck, P. *Anal. Biochem.* **2004**, *326*, 234-256. Sedimentation Equilibrium Analysis of Protein Interactions with Global Implicit Mass Conservation Constraints and Systematic Noise Decomposition.
(45) de Alba, E.; Santoro, J.; Rico, M.; Jiménez, M. A. *Protein Sci.* **1999**, *8*, 854-865. *De Novo* Design of a Monomeric Three-Stranded Antiparallel β-Sheet.
(46) Carulla, N.; Woodward, C.; Barany, G. *Protein Sci.* **2002**, *11*, 1539-1551. BetaCore, a Designed Water Soluble Four-Stranded Antiparallel β-Sheet Protein.
(47) Schenk, H. L.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4869-4870. Use of a Designed Triple-Stranded Antiparallel β-Sheet to Probe β-Sheet Specificity in Aqueous Solution.
(48) Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Harding, S. E., Rowe, A. J., Horton, J. C., Eds.; Royal Society of Chemistry: Cambridge, 1992.
(49) Lear, J. D., Personal Communication.
(50) Cohn, E. J.; Edsall, J. T. In *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*; Cohn, E. J., Edsall, J. T., Eds.; Hafner Publishing Company: New York, 1965.
(51) Durchschlag, H. In *Thermodynamic Data for Biochemistry and Biotechnology*; Hinz, H.-J., Ed.; Springer-Verlag: Berlin, 1986.
(52) Charlwood, P. A. *J. Am. Chem. Soc.* **1956**, *79*, 776-781. Partial Specific Volumes of Proteins in Relation to Composition and Environment.
(53) McMeekin, T. L.; Marshall, K. *Science* **1952**, *116*, 142-143. Specific Volumes of Proteins and the Relationship to their Amino Acid Contents.
(54) 4, C. C. P. N. *Acta Crystallogr., Sect. D* **1994**, *50*, 760-763. The *CCP4* Suite: Programs for Protein Crystallography.
(55) Chothia, C. *Nature* **1975**, *254*, 304-308. Structural Invariants in Protein Folding.
(56) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379-400. The Interpretation of Protein Structures: Estimation of Static Accessibility.
(57) Laue, T. M.; Shah, B.; Ridgeway, T. M.; Pelletier, S. L. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Horton, J. C., Ed.; Royal Society of Chemistry: Cambridge, 1992.

Chapter 5: A Heterooligomeric BBA

Introduction

The design of complex and native-like mini-proteins is a longstanding goal of this group. The two-fold symmetry of the tetrameric BBA motif, revealed by the X-ray crystal structures, suggested an opportunity to increase the structural complexity of the mini-protein through the design of a heterotetrameric analog made of two dissimilar components. A heterooligomeric alpha/beta mini-protein would be a valuable minimal model for the study of complex natural proteins and of protein-protein associations.

This chapter will describe the design and characterization of heterooligomeric [A_2B_2] BBA peptides in collaboration with Christina M. Taylor and Professor Amy E. Keating of the MIT Biology Department. Two general design principles were employed. (1) Acidic and basic residues were substituted along the inter-monomer interface in order to disfavor homoassociation and favor heteroassociation. (2) Specific steric interactions in the core were designed that would disfavor homoassociation and favor heteroassociation.

A number of heterotetrameric BBA peptides have been designed, synthesized, and characterized. The introduction of like charges along the inter-monomer interface results in a preference for heteroassociation, but does not appear to be sufficient to preclude homooligomerization. However, specific steric interactions in the middle layer of the hydrophobic core, in combination with electrostatic factors, significantly destabilize homoassociation and greatly favor heteroassociation. Heterotetramers that are comparable to the homotetramer formed by peptide **1** in terms of structure and stoichiometry, and that approach the native homotetramer in terms of stability, have been characterized by a variety of biophysical techniques.

Heterooligomeric Mini-Proteins

A number of heterodimeric,[1-4] heterotrimeric,[5-8] and heterotetrameric[9-11] coiled-coil peptides have been described in the literature. No designed heterooligomeric beta or alpha/beta mini-proteins have been reported, which likely reflects the relative paucity of designed discretely homooligomeric beta and alpha/beta mini-proteins.

Certain features of the tetrameric BBA structure are reminiscent of a coiled coil. The central region of the hydrophobic core is composed of four interacting helices. The hydrophobic core can be divided into discrete layers composed of one side chain from each monomer. Residues along the inter-subunit interface can be polar or nonpolar. It was thus anticipated that strategies employed in the design of heterooligomeric coiled coils would be applicable to the BBA mini-protein.

Coiled coils consist of two to five associating, interwound, helices. The interhelical association is specified by a heptad repeat, the positions of which are typically denoted **abcdefg**.[1,12] Positions **a** and **d** comprise the hydrophobic core. Positions **e** and **g** are solvent-exposed and along the interhelical interface, whereas positions **b**, **c**, and **f** define the hydrophilic face of the helix.[1,12]

Heterospecificity can be achieved in coiled coils by the selective placement of charged residues at the peripheral **e** and **g** positions. In the case of a heterodimer, one peptide is designed with positively-charged residues at **e** and **g** positions, and the partner is designed with negatively-charged residues at **e** and **g** positions.[13-15] The charges are situated so as to disfavor the homooligomeric state, in which the like charges are in close proximity. By contrast, association with a partner peptide bearing oppositely-charged residues relieves the electrostatic repulsions and may result in inter-monomer salt bridges. The extent to which surface salt bridges contribute to the stability of hetrooligomers has been debated.[16,17] However, a number of experimental[9,14,18] findings indicate that inter-monomer salt bridges can in some cases result in moderately increased heterooligomer stability.

Steric complementarity is a second general strategy for the positive and negative design of heterooligomeric coiled coils. A peptide with a large residue at a specific core position is juxtaposed with a partner having a small residue at that position. Homoassociation of the large partner is disfavored by steric clashes, and homoassociation of the small partner is disfavored by poor packing in the core. The combination of large and small peptides results in a tightly-packed core. Kennan's laboratory has designed heterotrimeric coiled-coil peptides differing only in a single core residue that associate as a 2:1 mixture of alanine- and cyclohexylalanine-substituted peptides.[8] In other examples steric complementarity has been used in addition to charge complementarity.[7]

Design Overview

The two-fold symmetry of the BBA tetramer lends itself to the design of an $[A_2B_2]$ heterotetramer by desymmetrization of the pseudo four-fold axis (Figure 5-1). Residues at the hydrophobic/hydrophilic interface of peptide **1** can accommodate charged residues, as demonstrated in Chapter 3. Furthermore, as recounted in the previous chapter, attempts to repack the tetrameric BBA hydrophobic core revealed that it was unfavorable to have four large or four small residues in the central layer. These factors suggested that both electrostatic and steric design elements could be exploited for the design of a heterotetrameric BBA.

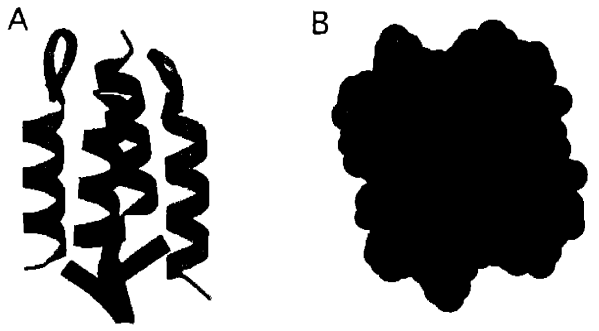


Figure 5-1. The goal of this project is the design of a heterotetrameric BBA with 1:1 stoichiometry of the components. One peptide was designed to have acidic residues along the inter-monomer interface and the other to have basic residues along the inter-monomer interface.

A bias towards heterooligomerization was created by the introduction of negative charges at the inter-subunit interfaces of one peptide and of positive charges at the inter-subunit interfaces of the other peptide. The charged residues were situated so as to have the potential to form stabilizing salt bridges, as well as to ward against self-association.

An additional level of differentiation was achieved by varying the residues in the central layer of the hydrophobic core, composed of four leucine-12 residues. Although computational efforts by Ms. Taylor to repack this layer in a more efficient manner revealed that it could not accommodate four large residues, nor would it be well packed with four small residues, these calculations indicated that this layer could be packed more efficiently with two large and two small residues.

The desired outcome was a pair of peptides that did not self-associate and that formed a stable tetramer in combination. Each peptide was characterized with regard to structure and oligomerization state, both singly and in combination with its partner.

Design Algorithm

All computational aspects of this project were performed by Ms. Taylor, in the laboratory of Prof. Keating. Design software, written in the Keating laboratory by Daniel J. Park,[19] optimizes

for a particular state over another by maximizing a user-defined equation reflecting the desired stability or specificity. The program uses a Monte Carlo search algorithm to sample through user-defined sequence space using an empirical energy function described below. Once a sequence is defined, side chains from the defined sequence are placed into the lowest-energy conformation by application of the dead-end elimination[20,21] and A* algorithms.[21] The solution is then evaluated based upon the user-defined equation and the resultant energy compared to the energy of the previous solution. If the energy is greater, the sequence is recorded; if it is less the sequence is discarded with some probability. Beyond a certain probability level, lower energies are accepted to avoid getting caught in local minima, and to allow greater sampling of sequence space.

All calculations were performed with Keating Lab Energy Scheme 1 using Version 1.0 or 2.0.[22] The Dunbrack backbone-dependent rotamer library was used in all cases.[23] Energies were calculated following equation (1). E_{vdW}, the van der Waals energy, E_{tor}, the torsional energy, and E_{solv}, the solvation energy, were calculated using CHARMM.[24,25] E_{vdW}, the van der Waals energy, was calculated using van der Waals radii scaled to 90% to compensate for the use of discrete rotamers. E_{elec}, the electrostatic energy was calculated with a distance-dependent dielectric method. The dielectric constant was scaled by 4 (polar/polar interactions), 8 (polar/charged interactions), or 16 (charged/charged interactions). Solvation energies were calculated using the Effective Energy Function[24-26] with some modified reference free energies. All calculations were performed starting with two or more X-ray crystal structures of peptides **1**, **3**, and **6**.

$$(1)\quad E = E_{vdW} + E_{tor} + E_{elec} + E_{solv}$$

Two general calculations were used in the evaluation of heterotetramers. A specificity score was calculated for every pair of peptides, comparing the energy of the heterotetrameric state to the two corresponding homotetramers. A stability calculation was also performed to compare

the folded heterotetramer or homotetramer to the corresponding unfolded peptides. The stability calculation is important for the elimination of peptide pairs that score highly with respect to specificity, but are unlikely to form a stable heterotetramer.

Specificity calculations were based upon the equilibrium described in equation (2). Peptides were assumed to exist in an equilibrium between a homotetrameric and heterotetrameric state.

$$(2)\ A_4 + B_4 \rightleftarrows 2(A_2B_2)$$

The combinations of mutations at user-selected positions that maximized the energy difference between two homooligomers and the corresponding heterooligomer (equation 3) received the highest specificity score. In the calculations, structural relaxation was allowed in surrounding sites in order to accommodate the mutations being evaluated.

$$(3)\ \max f, f = E_{A_4} + E_{B_4} - 2_{A_2B_2}$$

The stabilities of all peptide pairs identified by the specificity function were determined by means of a stability function maximizing the difference between the folded and unfolded state of the heterotetramer (equation 4). The unfolded state was modeled by calculating and summing the energy of each residue flanked by diglycine units; these pentapeptides were modeled using the local backbone structure observed in tetrameric BBA.

$$(4)\ \max f, f = -E_{A_2B_2} + E_{2A+2B,unfolded}$$

The predicted structures of the folded states were constructed, then minimized using CHARMM19 with van der Waals radii scaled to 100%. The minimized structures of the peptide combinations were then inspected visually in order to eliminate any unfavorable combinations. Structures deemed unsatisfactory, e.g. by virtue of unpaired hydrogen bond donors or acceptors, were not pursued further.

First Generation

The first generation of peptides differed from peptide **1** only in the interfacial residues. Residues 11, 13, and 18 of peptide **1** were selected for mutation to charged residues (Figure 5-2). Position 11 of peptide **1** is an aspartate, and positions 13 and 18 had been mutated to methionine selenoxide in peptides **3** and **6** (discussed further in Chapter 3). All three positions were thus expected to accommodate polar residues without structural disruption and so were optimized in the protein design calculations.

Residues at position 13 are situated to form an intramolecular salt bridge on one face of the helix, and residues at positions 11 and 18 are situated to form two salt bridges on the opposite face. There is thus the potential to form six stabilizing salt bridges per tetramer.

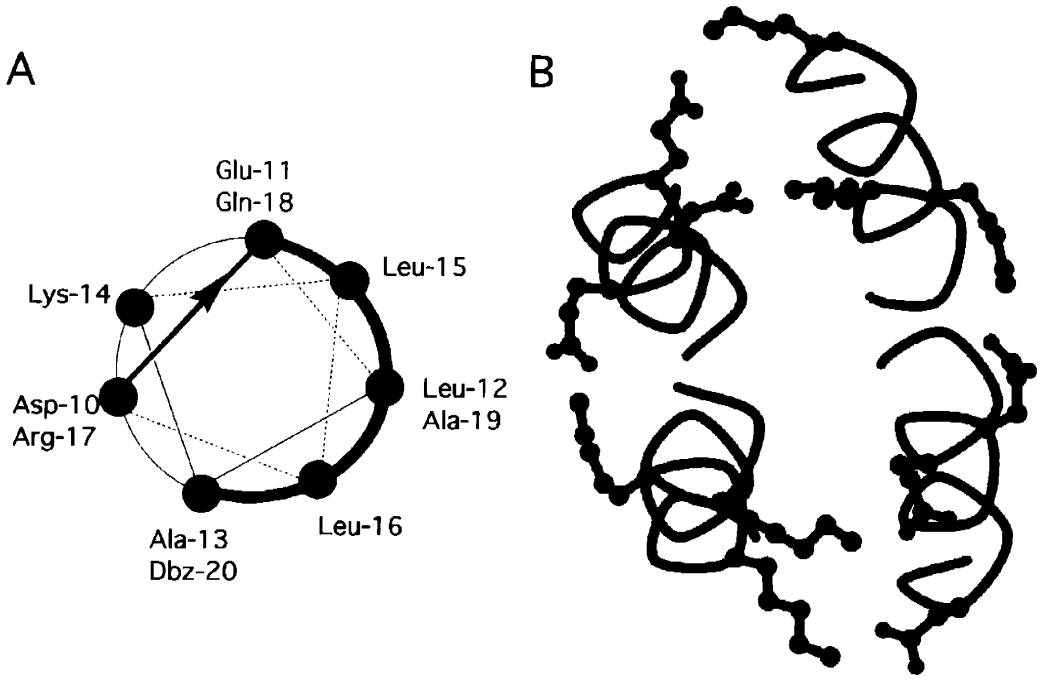


Figure 5-2. A. Helical wheel projection of peptide **1**. The hydrophobic face of the helix is indicated in green. B. Top view of a model of a heterotetrameric BBA with ball-and-stick models of the residues (11, 13, 18) designed to participate in salt bridges. C_α traces of acidic monomers are colored red and C_α traces of basic monomers blue.

A general acid/base strategy emerged from the computational selections as a means of selecting for heterotetramers over homotetramers. Lysine and glutamic acid were selected at position 13 on the basis of calculations and visual inspection of the minimized structures. Lysine was also selected for positions 11 and 18 of the basic peptide. Position 18 of the acidic peptide was mutated to aspartate because modeling suggested that an aspartate at 18 would be better able than a glutamate to form a salt bridge with lysine at 11 of the basic peptide. Position 11 of the acidic peptide was mutated to aspartate. These mutations resulted in peptides **A1** and **B1**, respectively (Table 5-1). Basic peptide **B1** has a calculated charge of +4 at neutral pH and acidic peptide **A1** has a calculated charge of –2 at neutral pH.

		hairpin								*helix*												
Peptide		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
A1	Ac	Y	R	I	p	S	Y	D	F	a	D	E	L	E	K	L	L	R	D	A	Z	G	NH_2
B1	Ac	Y	R	I	p	S	Y	D	F	a	D	K	L	K	K	L	L	R	K	A	Z	G	NH_2

Table 5-1. Sequences of peptides **1**, **A1**, and **B1**. Abbreviations: p=D-proline, a=D-alanine, Z=benzoylated L-α,β-diaminopropionic acid. Acidic residues at interface positions 11, 13, and 18 are indicated in red, and basic residues at interface positions 11, 13, and 18 are indicated in blue.

The conformational properties of these peptides were assessed by circular dichroism spectroscopy. Both peptides were found to self-associate in aqueous solution at 50 μM, suggesting that the charge-charge repulsion negative design strategy was insufficient to specify a monomeric state. Peptide **A1** was found to have near-native structure and stability, whereas the self-association of **B1** is less pronounced, as predicted by its greater net charge. The high stability of peptide **A1** is somewhat surprising as the peripheral negative charges were expected to significantly destabilize the homooligomeric state. It can be rationalized by assuming a variable protonation state and/or sufficient side-chain flexibility to avoid repulsive interactions.

The results of CD spectroscopy studies suggesting that peptide **A1** is quite similar to peptide **1** were confirmed by sedimentation equilibrium AUC experiments. It was found that peptide **A1** associates as a tetramer in solution (Table 5-2).

Peptide	**A1**
$\bar{\nu}$ (cm^3/g)	0.7306
Sigma (NONLIN)	1.6766
Stoichiometry (NONLIN)	3.5
Stoichiometry (SEDPHAT)	3.6

Table 5-2. Sedimentation equilibrium AUC results of peptide **A1**. Data collected at 50, 150, 309 μM. The best association model is a single ideal species.

Peptides **A1** and **B1**, when combined, show interhelical association greater than the average of their individual spectra, indicating that they interact in solution (Figure 5-3). The complex they form is as structured as the native peptide **1**. This indicates that mutations at the

periphery of the hydrophobic core are well tolerated, as was predicted based on the selenomethionine mutations described in Chapter 3.

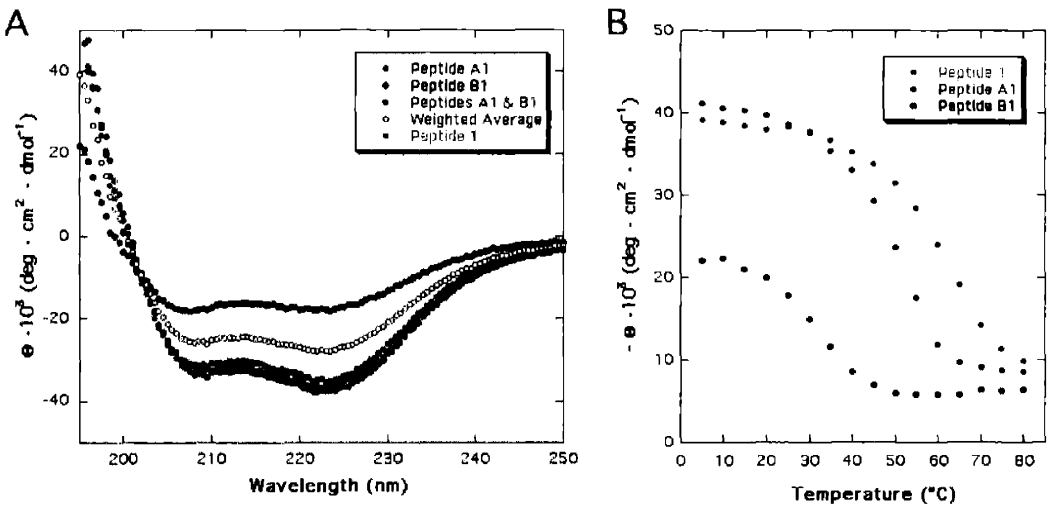


Figure 5-3. CD spectra of peptides **1**, **A1**, and **B1** at 50 μM peptide. A. Wavelength scans of peptides **1**, **A1**, and **B1** at 25 °C. B. Normalized molar ellipticity at 222 nm as a function of temperature.

A fluorescence quenching assay was performed to probe the heteroassociation was formed by peptides **A1** and **B1**. Peptide **A1** was synthesized with a 3-nitrotyrosine label at position 6 and peptide **B1** with an anthranilic acid cap at position 20 (Figure 5-4). These substitutions were expected to be non-perturbing based on their location in the crystal structure. Furthermore, as these were the substitutions that led to the identification of **BBAT1** from the original fluorescence quenching screen,[27] minimal structural disruption was expected.

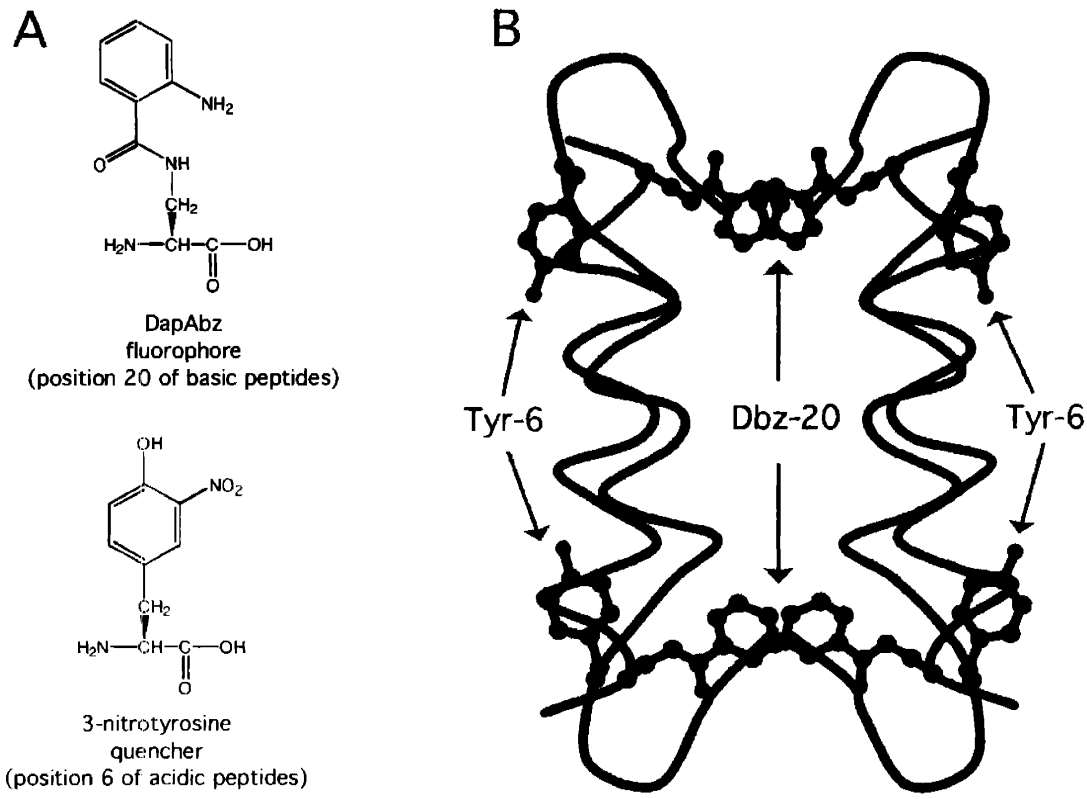


Figure 5-4. A. Chemical structures of DapAbz fluorophore (introduced at position 20 of basic peptides) and 3-nitrotyrosine quencher (introduced at position 6 of acidic peptides). B. Locations of residues 6 and 20 in the tetrameric BBA motif. Residues 6 and 20 are in close proximity, permitting quenching of the anthranilic acid fluorescence. The structure suggests that a *meta* nitro group at position 6 and an *ortho* amino group at position 20 can be incorporated without significant distortion.

Significantly less fluorescence was found in an equimolar solution of **A1** and **B1** than in a solution of **B1** alone (Figure 5-5). The observation of intramolecular fluorescence quenching confirms that peptides **A1** and **B1** associate in solution, and that the electrostatic-based design strategy favoring a heterooligomer over a homooligomer was successful.

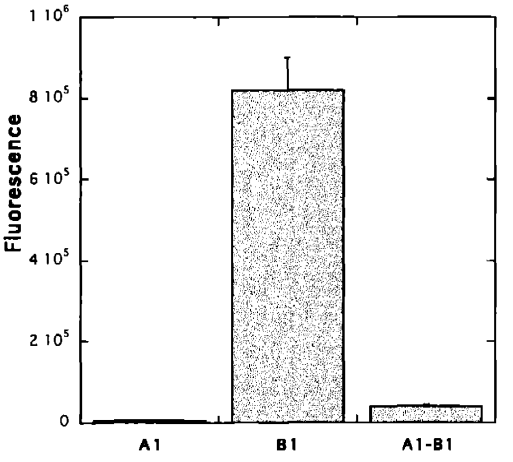


Figure 5-5. Results of fluorescence quenching experiment. Peptide **A1** is labeled with a 3-nitrotyrosine quencher and peptide **B1** with an anthranilic acid fluorophore. Significant quenching of fluorescence is observed in an equimolar mixture of **A1** and **B1**.

The results from the first generation of peptides indicate that the relief of charge-charge repulsions at the inter-subunit interface can significantly favor the heterooligomeric state over the homooligomeric state. Furthermore, the introduction of charged residues at the periphery of the BBA core is not disruptive, but can result in a structured heterooligomeric complex.

Second Generation

Steric complementarity was incorporated into the next design in order to further favor the heterooligomeric state and disfavor the homooligomeric state. A computational search was made through various possibilities at layer C. The calculations revealed a "large-small" strategy. The highest-ranking outcomes had a large residue at position 12 in one peptide and a small residue at position 12 in the other peptide. Some of the high-scoring results had peptide pairs with unsatisfied hydrogen bonding as the design program does not simultaneous optimize for heterospecificity and heterooligomer stability. These were not pursued further.

The best solution found consists of phenylalanine at position 12 in one peptide and alanine in the other. On visual inspection of the predicted minimized structure of the designed heterooligomer, these side chains packed relatively efficiently (Figure 5-6). By contrast, the homooligomeric state having four phenylalanines in the core is predicted to suffer from significant steric repulsion, whereas the homooligomeric state having four alanines in the core is predicted to be destabilized by a large cavity in the center of the tetramer.

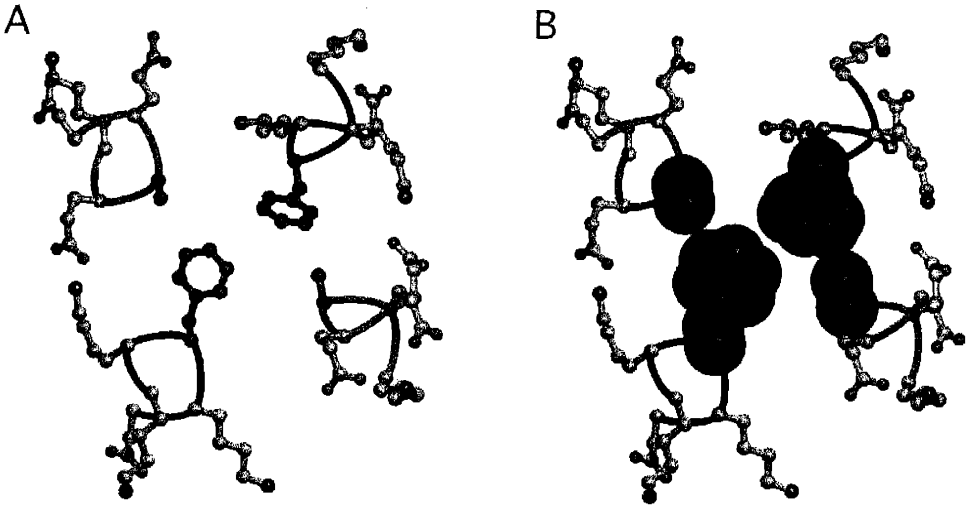


Figure 5-6. Minimized model structures of layer C of the **A2-B2** heterotetramer. A. Ball-and-stick models of the core residues indicated in bold. B. Space-filling models of core residues.

Calculations indicated that the arrangements having phenylalanine on the basic peptide and alanine on the acidic peptide and vice versa were isoenergetic. The former arrangement was chosen arbitrarily, resulting in the sequences of peptides **A2** and **B2** (Table 5-3).

		hairpin								*helix*												
Peptide		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
A2	Ac	Y	R	I	p	S	Y	D	F	a	D	E	A	E	K	L	L	R	D	A	Z	G	NH_2
B2	Ac	Y	R	I	p	S	Y	D	F	a	D	K	I	K	K	L	L	R	K	A	Z	G	NH_2

Table 5-3. Sequences of peptides **1**, **A1**, **B1**, **A2**, and **B2**. Abbreviations: p=D-proline, a=D-alanine, Z=benzoylated L-α,β-diaminopropionic acid. Acidic residues at interface positions 11, 13, and 18 are indicated in red, and basic residues at interface positions 11, 13, and 18 are indicated in blue. Hydrophobic residues at position 12 are indicated in green.

Mutation of the hydrophobic core to disfavor homooligomerization, in conjunction with the introduction of charged residues along the helical periphery, results in a peptide with a reduced tendency for self-association. Peptides **A2** and **B2** are monomeric in solution at 50 μM by CD spectroscopy (Figure 5-7). The CD spectra of 50 μM solutions of peptides **A2** and **B2** indicate weak, alpha helical, non-interacting structure. Peptide **A2** is more structured than peptide **B2**, presumably because of the lower net charge, as was seen with peptides **A1** and **B1**.

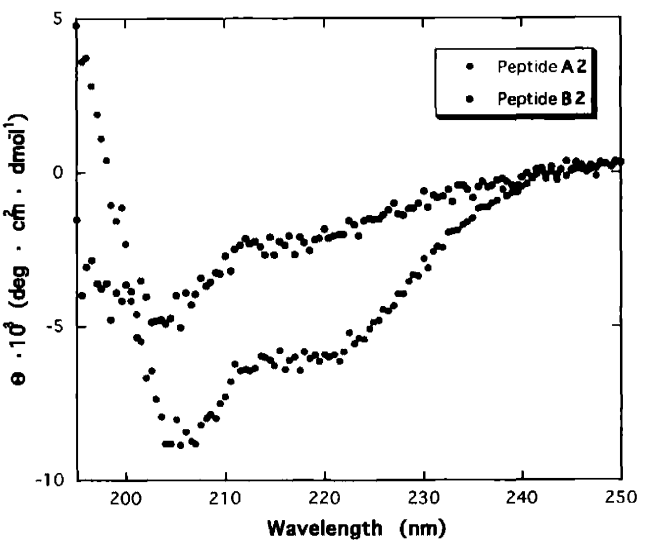


Figure 5-7. CD spectra of peptides **A2** and **B2** (50 μM total peptide) indicate non-associating alpha structure. The weakness of the signal indicates a relative lack of structure.

The CD spectrum of an equimolar mixture of peptides **A2** and **B2** is considerably more structured than that of the individual components (Figure 5-8A). Moreover, whereas the peptides individually appear monomeric and poorly structured, the spectrum of the mixture is qualitatively different and indicative of interhelical association, as evidenced by a ratio greater than one of ellipticities at 222 nm to 208 nm.[28,29] This finding strongly supports the formation of a heterooligomeric interaction. The **A2-B2** complex is both less structured and less stable than the native peptide. Its melting temperature is about 20 °C lower than that of peptide **1** (Figure 5-8B).

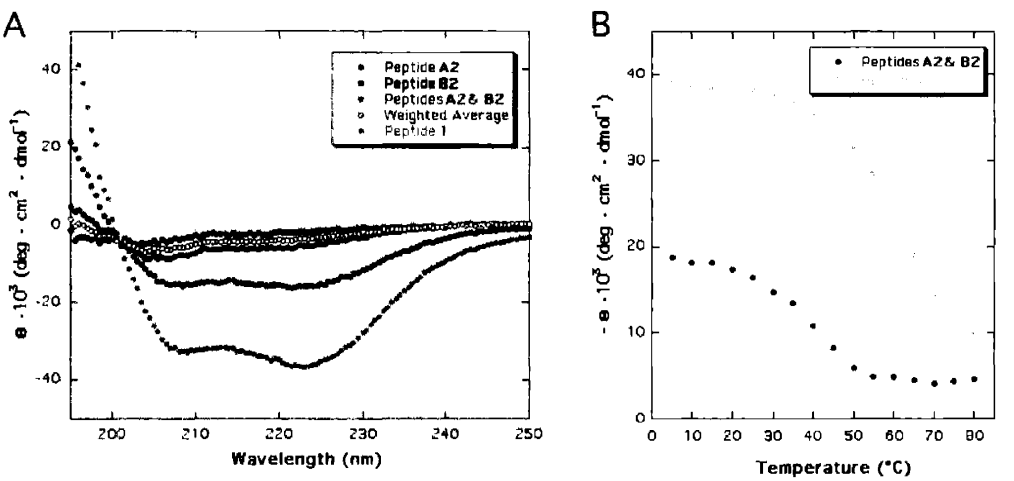


Figure 5-8. CD spectra of peptides **1**, **A2**, and **B2** at 50 μM peptide. A. Wavelength scans of peptides **1**, **A2**, and **B2** at 25 °C. B. Normalized molar ellipticity at 222 nm as a function of temperature.

Peptides **A1** and **B2**, when combined, show association greater than the average of their individual spectra, indicating that they associate (Figure 5-9A). The spectrum of the **A1-B2** mixture is less structured than peptide **1**, but is also indicative of interhelical interactions (Figure 5-9B). The less structured association may indicate some distortion due to the juxtaposition of two leucine and two phenylalanine residues in the core. The CD spectrum of the mixture of **A2** and **B1** also reveals some interhelical association. The **A2-B1** combination appears to be roughly

as structured as the weighted average of the individual spectra, suggesting that heteroassociation is disfavored for this pair. This may be due to inefficient packing in a core composed of two alanine residues and two leucine residues.

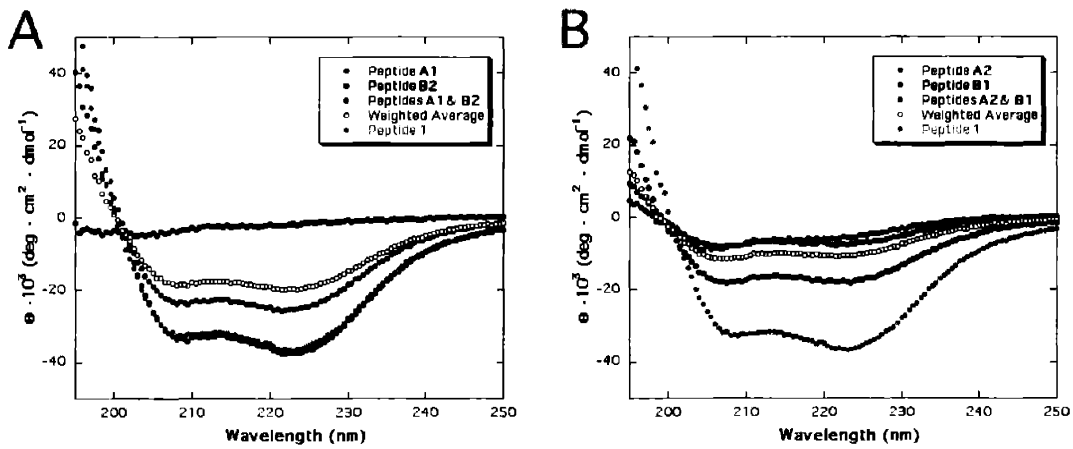


Figure 5-9. A. CD spectra of peptides **A1** and **B2**. B. CD spectra of peptides **A2** and **B1**. Total peptide concentrations are 50 μM and all mixtures are equimolar.

Fluorescence quenching experiments were performed with all pair-wise combinations of peptides, revealing that all four combinations of peptides form complexes that bring the fluorophore and quencher into proximity (Figure 5-10). Peptides **A2** and **B2** appear to have a weaker association than do **A1** and **B1**, which is consistent with the CD results. Peptides **A1** and **B2** form a slightly more structured complex than **A2** and **B2** by CD and by fluorescence quenching. Peptides **A2** and **B1**, which associate weakly by CD, appear comparable to **A2** and **B2** by fluorescence.

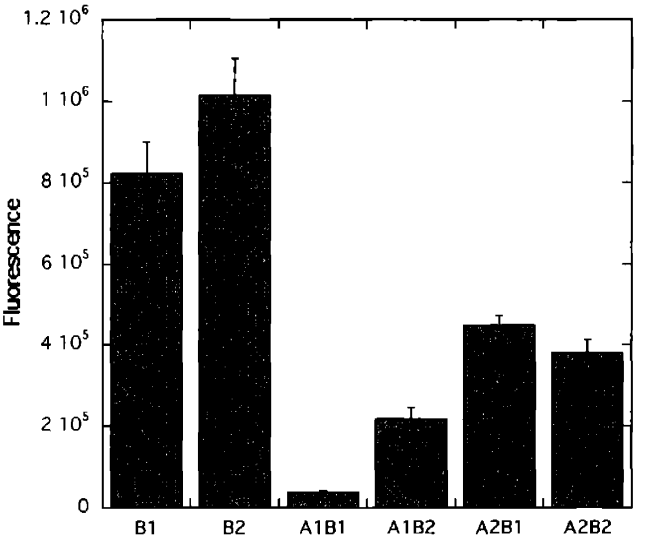


Figure 5-10. Results of fluorescence quenching experiments. Peptides **A1** and **A2** are labeled with a 3-nitrotyrosine quencher and peptides **B1** and **B2** with an anthranilic acid fluorophore.

The stoichiometries of peptides **A2** and **B2**, as well as 3-nitrotyrosine-labeled **A2** and anthranilic acid-labeled **B2**, were determined individually by analytical ultracentrifugation (Table 4). As predicted by the CD results, peptide **A2** does not self-associate at concentrations up to 266 μM. **A2** with a 3-nitrotyrosine label is also monomeric in solution, indicating that 3-nitrotyrosine is non-perturbing.

We were surprised to find that peptide **B2** self-associates to form a tetrameric species by AUC, as does the anthranilic acid-labeled **B2**. By contrast, circular dichroism results revealed peptide **B2** to be both monomeric and less structured in solution than peptide **A2**, a finding consistent with its higher net charge. It may be that peptide **A2** self-associates appreciably only at concentrations higher than those used in the CD experiments (CD: 50 μM, AUC: 150, 287 μM). Another contributing factor may be attenuation of charge-charge repulsions in the higher ionic strength buffer used for the AUC studies (CD: 10 mM phosphate buffer, pH 7.2; AUC: 50 mM phosphate buffer, 100 mM salt, pH 7.2). The AUC findings imply that the bulkier hydrophobic face of the **B2** peptide makes it more prone to self-association than the smaller, but less charged

peptide **A2**. They also suggest that a hole in the core of a homotetramer may be more destabilizing than steric congestion in the core, which could be accommodated by a flexible backbone. It is also possible that perturbation of the pK_a values of one or more of the peripheral lysines may reduce the net charge and thus help to stabilize the homooligomer.

Peptide[1]	**A2**	**A2 (NO_2-Tyr)**	**B2**[2]	**B2**	**B2 DapAbz**
$\bar{\nu}^2$ (cm^3/g)	0.7234	0.7234	0.7510	0.7510	0.7510
NONLIN					
Best Model[2]	a	b	c	a	a
Sigma[3]	0.5012	0.4964	0.49188	1.6040	1.6096
Stoichiometry	1.0	1.0	1.1 + 4mer	3.5	3.5
SEDPHAT					
Stoichiometry	1.1	1.1	3.2	3.5	3.7

Table 5-4. Sedimentation equilibrium analytical ultracentrifugation results.

1. Peptide concentrations are **A2**: 40, 100, 266 μM, **A2** (NO_2-Tyr): 186 μM, **B2**: 150 μM, **B2**: 287 μM, **B2** (DapAbz): 657 μM.
2. Models: a: Single species, B = 0. b: Single species, B $\neq$ 0. c: Monomer-tetramer equilibrium, B = 0.
3. Sigma is related to the molecular weight by the equation $M = \frac{\sigma RT}{(1-\bar{\nu}\rho)\omega^2}$.

The stoichiometry of the complex formed by **A2** and **B2** was investigated by sedimentation equilibrium experiments on equimolar mixtures of the peptides over a range of concentrations. The data are best described by a single species model with a tetrameric stoichiometry (Table 5-5). As peptide **A2** is monomeric by AUC under these conditions, it should only be possible to adequately describe the data as a single tetrameric species if peptide **A2** associates with peptide **B2** to form a heterotetramer.

Peptide[1]	A2 & B2
$\bar{\nu}$ (cm^3/g)	0.7372[1]
Sigma[2] (NONLIN)	1.6758
Stoichiometry (NONLIN)	3.6
Stoichiometry (SEDPHAT)	3.6

Table 5-5. Sedimentation equilibrium AUC results of peptides **A2** and **B2**. Data collected at 50, 150, 320 μM total concentration. The best association model is a single ideal species.

1. Average of the calculated partial specific volumes for **A2** and **B2**.
2. Sigma is related to the molecular weight by the equation $M = \frac{\sigma RT}{(1-\bar{\nu}\rho)\omega^2}$.

A common strategy in studying heterogeneous interactions by analytical ultracentrifugation is to dissect the contributions of each component by monitoring sedimentation at different wavelengths.[30-32] If the components do not have inherently different absorption properties one component can be labeled to distinguish it spectroscopically.[32,33] The association of a constant amount of peptide **A2** labeled with 3-nitrotyrosine with varying concentrations of peptide **B2** was monitored in the analytical ultracentrifuge at 360 nm. It was found that the stoichiometry of **A2** increases from a monomer to a tetramer as the number of equivalents of **B2** is increased, further confirming that the peptides associate to form a heterotetramer (Table 5-6).

[A2 (NO_2-Tyr)] (μM)	[B2] / [A2 (NO_2-Tyr)]	Sigma[1]	MW	Stoichiometry
186	0	0.7234	2,975	1.1
120	0.5	1.5881	8,769	3.4
120	1.0	1.7843	9,853	3.8
120	1.5	1.8355	10,135	3.9
120	3.0	1.8254	10.080	3.9

Table 5-6. Results of tracer sedimentation equilibrium AUC study of **A2** and **B2** interaction.

1. Sigma is related to the molecular weight by the equation $M = \frac{\sigma RT}{(1-\bar{\nu}\rho)\omega^2}$, and determined from data collected at 360 nm using the program NONLIN. Data were fit to a single ideal species model.
2. Molecular weights were calculated using $\bar{\nu} = 0.7372$ cm^3/g.

It was anticipated that a crystal structure of the **A2-B2** heterotetramer would provide valuable information for the design of future generations of heterooligomeric BBA peptides. Moreover, a crystal structure would confirm the predicted 1:1 interaction stoichiometry.

Crystals were grown from an equimolar mixture of peptides **A2** and **B2** (Figure 5-11). These crystals diffract to ~1.95 Å with a unit cell of a = b = 41.69 Å, c = 51.33 Å, $\alpha = \beta = 90°$, $\gamma = 120°$. A data set was collected (Table 5-7). Although the structure has not yet been solved due to time limitations, it is anticipated that a molecular replacement solution should be facile.

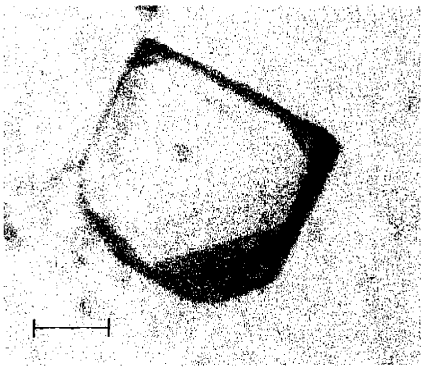


Figure 5-11. Crystal grown from an equimolar mixture of peptides **A2** and **B2**. Bar indicates 0.1 mm.

Unit cell (Å,°)	a = b = 41.69 c = 51.33 α = β = 90° γ = 120°
Space group	P3
X-ray source	Rigaku RU300
Wavelength (Å)	1.54
Resolution (Å)	1.95
Total/ unique reflections	32,702 7,010
Completeness[1] (%)	95.6 (100.0)
I/σ(I)	20.3 (5.4)
R_{merge}[2] (%)	5.7 (36.5)
Solvent content (%)	24.9

Table 5-7. Data collection statistics for crystal grown from **A2+B2** solution. [1]Values for the outermost shell (2.01-1.95 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.

A Matthews coefficient (V_M) calculation[34] (Table 5-8) suggests that there are probably six monomers per asymmetric unit. Four or five monomers per asymmetric unit would also be possible, but would result in a higher solvent content than was observed for crystals of peptides **1**, **3**, and **6**.

Monomers per ASU	V_M	Solvent Content (%)
1	9.9	87.5
2	5.0	75.0
3	3.3	62.5
4	2.5	49.9
5	2.0	37.4
6	1.7	24.9
7	1.4	12.4

Table 5-8. Matthews coefficient determination for crystal grown from an equimolar stock of peptides **A2** & **B2**.

Third Generation

The goal of the next generation was to increase the stability of the heterooligomer by improving the packing of the core. It was postulated that the reduction in heterooligomer stability

in the **A2-B2** pair relative to **A1-B1** was due to poor packing in the C layer, and that improved packing would result in a more stable heterooligomer. Peptide **A3** (Table 5-9) features an α-aminobutyric acid (Abu) residue, having an ethyl group side chain, juxtaposed with a phenylalanine partner. It was predicted that this larger side chain would result in more effective packing (Figure 5-12); this prediction was supported by design calculations.

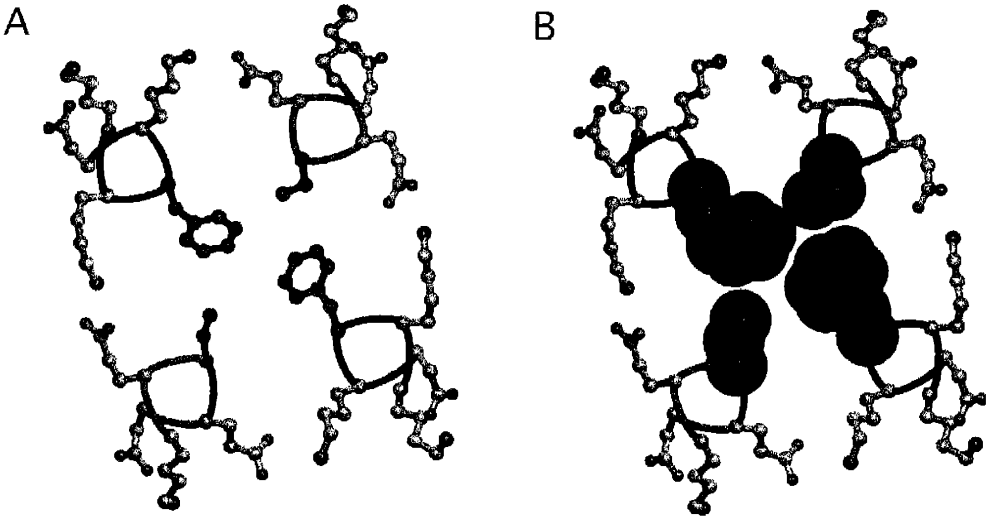


Figure 5-12. Minimized model structures of layer C of the **A3-B2** heterotetramer. A. Ball-and-stick models of the core residues indicated in bold. B. Space-filling models of core residues.

		hairpin								*helix*													
Peptide		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	
A3	Ac	Y	R	I	p	S	Y	D	F	a	D				K	L	L	R		A	Z	G	NH_2

Table 5-9. Sequence of peptide **A3**. Abbreviations: p=D-proline, a=D-alanine, Z=benzoylated L-α,β-diaminopropionic acid, B = L-α-aminobutyric acid. Acidic residues at interface positions 11, 13, and 18 are indicated in red. The hydrophobic residue at position 12 is indicated in green.

Peptide **A3** exhibits non-associating alpha structure by circular dichroism, as does **A2**, indicating that the additional methylene group does not result in self-association under the conditions of the CD experiment (Figure 5-13A). In association with **B2**, it forms a highly-structured oligomer with considerably greater stability than the **A2-B2** complex, as indicated by

its ellipticity and melting temperature, thus suggesting that improved packing in the hydrophobic core can significantly increase heterooligomer stability (Figure 5-13B). The **A3-B2** complex is almost as stable as the native peptide **1**.

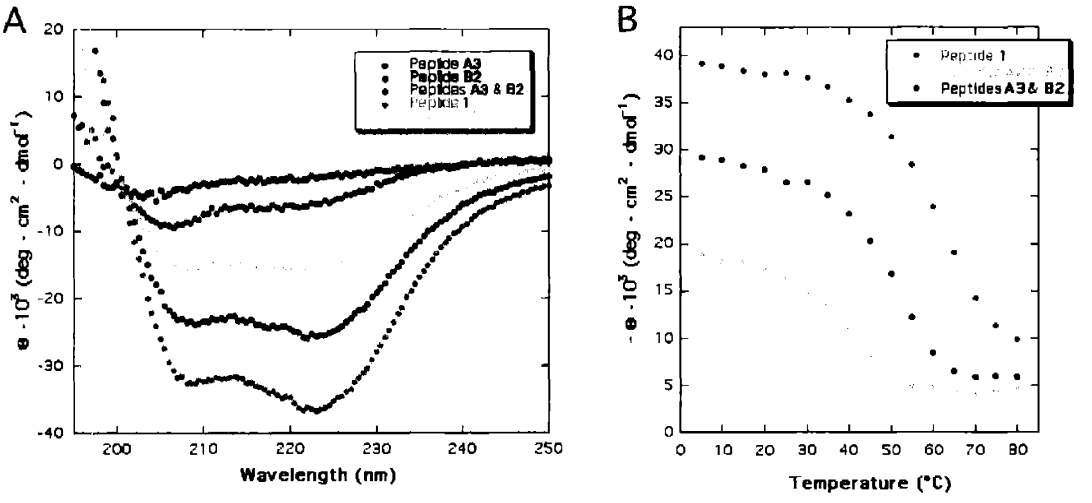


Figure 5-13. CD spectra of peptides **A3** and **B2** at 50 μM peptide with spectra of peptides **1** and **A2+B2** for comparison. A. Wavelength scans of peptides **1**, **A3**, and **B2** at 25 °C. B. Normalized molar ellipticity at 222 nm as a function of temperature.

Peptide **A3** was confirmed to form a heterospecific interaction with peptides **B1** and **B2** by fluorescence quenching experiments (Figure 4-14). Peptide **A3** appears to be as effective as peptide **A1** at quenching fluorescence in a basic partner, and more effective than peptide **A2**.

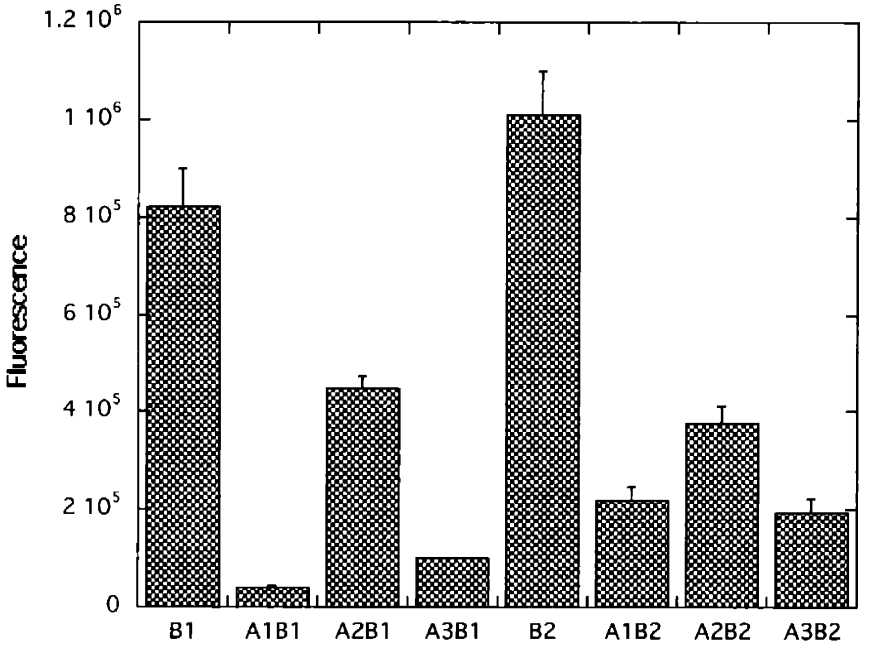


Figure 5-14. Results of fluorescence quenching experiments. Peptides **A1**, **A2** and **A3** are labeled with a 3-nitrotyrosine quencher and peptides **B1** and **B2** with an anthranilic acid fluorophore.

Peptide **A3** was confirmed to be monomeric in solution by AUC (Table 5-10). Peptide **A3**, labeled with 3-nitrotyrosine, and studied at higher concentrations, shows a slight tendency to self-associate when analyzed with SEDPHAT.[35,36] (However, the same data fit well to a single, non-ideal, monomeric species with NONLIN.[37]) This tendency to self-associate, if legitimate, is likely due to the higher concentrations at which it was studied relative to **A3**. It is also possible that the additional methylene group in the core of **A3** relative to **A2** makes it more prone to self-association, although further studies would be necessary to support this conclusion.

Peptide[1]	A3	A3 (NO_2-Tyr)
$\bar{\nu}^2$ (cm^3/g)	0.7248	0.7248
NONLIN		
Best Model[2]	b	B
Sigma[3]	0.4452	0.4595
Stoichiometry	0.9	0.9
SEDPHAT		
Stoichiometry	1.0	1.8

Table 5-10. Sedimentation equilibrium analytical ultracentrifugation results.

1. Peptide concentrations are **A3**: 50, 100, 200 μM, **A3** (NO_2-Tyr): 369, 738, 1476 μM.
2. Models: a: Single species, B = 0. b: Single species, B ≠ 0. c: Monomer-tetramer equilibrium, B = 0.
3. Sigma is related to the molecular weight by the equation $M = \frac{\sigma RT}{(1-\bar{\nu}\rho)\omega^2}$.

The stoichiometry of the complex formed by **A3** and **B2** was investigated by sedimentation equilibrium experiments on equimolar mixtures of the peptides over a range of concentrations (Table 5-11). The data were best described by a single, tetrameric, species model with nonideality. As peptide **A3** is monomeric at these concentrations, the fit to a single species with the molecular weight of a tetramer suggests that **A3** and **B2** associate as a heterotetramer.

Peptide[1]	A3 & B2
$\bar{\nu}$ (cm^3/g)	0.7379[1]
Sigma[2] (NONLIN)	1.8010
Stoichiometry (NONLIN)	3.8
Stoichiometry (SEDPHAT)	3.6

Table 5-11. Sedimentation equilibrium AUC results of peptides **A3** and **B2**. Data collected at 50, 150, 320 μM total concentration. The best association model is a single nonideal species.

1. Average of the calculated partial specific volumes for **A2** and **B2**.
2. Sigma is related to the molecular weight by the equation $M = \frac{\sigma RT}{(1-\bar{\nu}\rho)\omega^2}$.

Peptide **A3**, with an Abu residue at position 12, is predominantly monomeric. It forms heterospecific complexes with **B1** and **B2**. Its complex with **B2** is a heterospecific tetramer with

increased, and near-native, stability. This stability is almost certainly due to improved packing of the C layer of the heterooligomer.

Fourth Generation

The design of a non-associating basic peptide having a large and hydrophobic residue at position 12 was the goal of the fourth generation of peptides. Three negative design strategies disfavoring homoassociation were investigated. In two of these strategies mutations were made to the charged residues at the inter-monomer interface, and in the third strategy a mutation was made to the hydrophobic core.

The first strategy addressed the possibility that pK_a perturbation is a significant factor reducing the net charge on the peptides **B1** and **B2**, and thus facilitating self-association. If so, substitution of the surface lysines with more basic arginine residues would further disfavor the homomeric state. The initial computational selection for an acid/base pair had indicated that lysine and arginine were roughly equivalent with regard to specificity and heterooligomer stability. Lysine was selected over arginine as the geometry appeared better suited for salt bridge interactions. Peptide **B3** was derived from **B2** by replacement of lysine residues at positions 11 and 18 with arginine residues (Table 5-12).

The second strategy investigated the role of flexibility of the basic side chain. The side chain of lysine is long and flexible, which may allow the terminal amino groups on adjacent monomers in a homotetramer to avoid close contact and hence repulsive interactions. Substitution of a less flexible basic residue for lysine would result in decreased freedom of motion and increased proximity of the positive charges on adjacent monomers, further disfavoring the homooligomeric state.[13] Peptide **B4** has ornithine at positions 11 and 18 (Table 5-12). Peptide **B5** has ornithine at position 11 and L-α,β-diaminobutyric acid (Dab) at position 18 (Table 5-12) because modeling studies indicated that Dab would have the correct geometry to form a good salt

bridge with Asp-11 of the partner peptide. Computations indicated that, in combination with **A3**, peptide **B2** has greater specificity than **B4** or **B5**, and that the stabilities of **B4** and **B5** are comparable to or greater than **B2**.

The third strategy investigated the role of the steric crowding in the hydrophobic core. Mutation of phenylalanine to a methyl-substituted phenylalanine was expected to increase destabilizing steric crowding in the homooligomer and to improve packing in the heterooligomer. Calculations indicated that *m*-methylphenylalanine has a significantly improved specificity score relative to **B2**, and would have a somewhat lower stability in the absence of backbone relaxation. Calculations on *o*-methylphenylalanine resulted in moderate improvements in both specificity and stability. By contrast, *p*-methylphenylalanine scored modestly with regard to both specificity and stability.

The *m*-methylphenylalanine isomer was chosen for experimental studies because of its high specificity score, as the primary goal of this design generation was the destabilization of the homooligomeric state. Peptides **B6**, **B7**, and **B8** (Table 5-12) include *m*-methylphenylalanine at position 12 and at positions 11 and 18 either lysine (**B6**), ornithine (**B7**) or ornithine/Dab (**B8**). Peptides **B7** and **B8** were designed in anticipation of an additive increase in homotetramer destabilization upon application of both negative design strategies.

		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	O	F	K	K	L	L	R	O	A	Z	G	NH_2
B5	Ac	Y	R	I	p	S	Y	D	F	a	D	O	F	K	K	L	L	R	X	A	Z	G	NH_2
B6	Ac	Y	R	I	p	S	Y	D	F	a	D	K	mT	K	K	L	L	R	K	A	Z	G	NH_2
B7	Ac	Y	R	I	p	S	Y	D	F	a	D	O	mT	K	K	L	L	R	O	A	Z	G	NH_2
B8	Ac	Y	R	I	p	S	Y	D	F	a	D	O	mT	K	K	L	L	R	X	A	Z	G	NH_2

Table 5-12. Sequences of peptides **B3**, **B4**, **B5**, **B6**, **B7**, and **B8**. Abbreviations: p=D-proline, a = D-alanine, Z= benzoylated L-α,β-diaminopropionic acid, mT= *m*-methylphenylalanine, X= L-α,β-diaminobutyric acid. Basic residues at interface positions 11, 13, and 18 are indicated in blue. Hydrophobic residues at position 12 are indicated in green.

By CD spectroscopy (Figure 5-15), peptides **B3** through **B8** appear to be monomeric with weak alpha structure comparable to peptide **B2**. When mixed with peptide **A2** or **A3**, all exhibit interhelical association indicating the formation of a heterooligomer. The heterooligomer formed with **A3** is slightly more structured than that formed with **A2** in all cases, but the difference is slight, whereas for **B2** the difference in structure is significant. The complex each of the peptides **B3** through **B8** forms with **A3** is more structured than the **A3-B2** complex in all cases, and the complex formed with **A2** is comparable to, or slightly more structured than, the **A3-B2** complex. The most structured association appears to be that between **A3** and **B6**, the pair computationally predicted to form the most stable complex, but the results from all six peptides are roughly comparable.

The **A3-B3** complex is somewhat more structured than **A3-B2**, as are **A3-B4** and **A3-B5**, suggesting that the surface mutations have resulted in a moderate increase in stability. By contrast, a comparison of **A3-B7** to **A3-B4**, and **A3-B8** to **A3-B5**, suggests that the substitution of *m*-methylphenylalanine results in a slight decrease in structure, as was predicted by the calculations. (However, this is not true of the **A3-B6** to **A3-B2** comparison, where *m*-methylphenylalanine appears to cause a modest increase in ellipticity.)

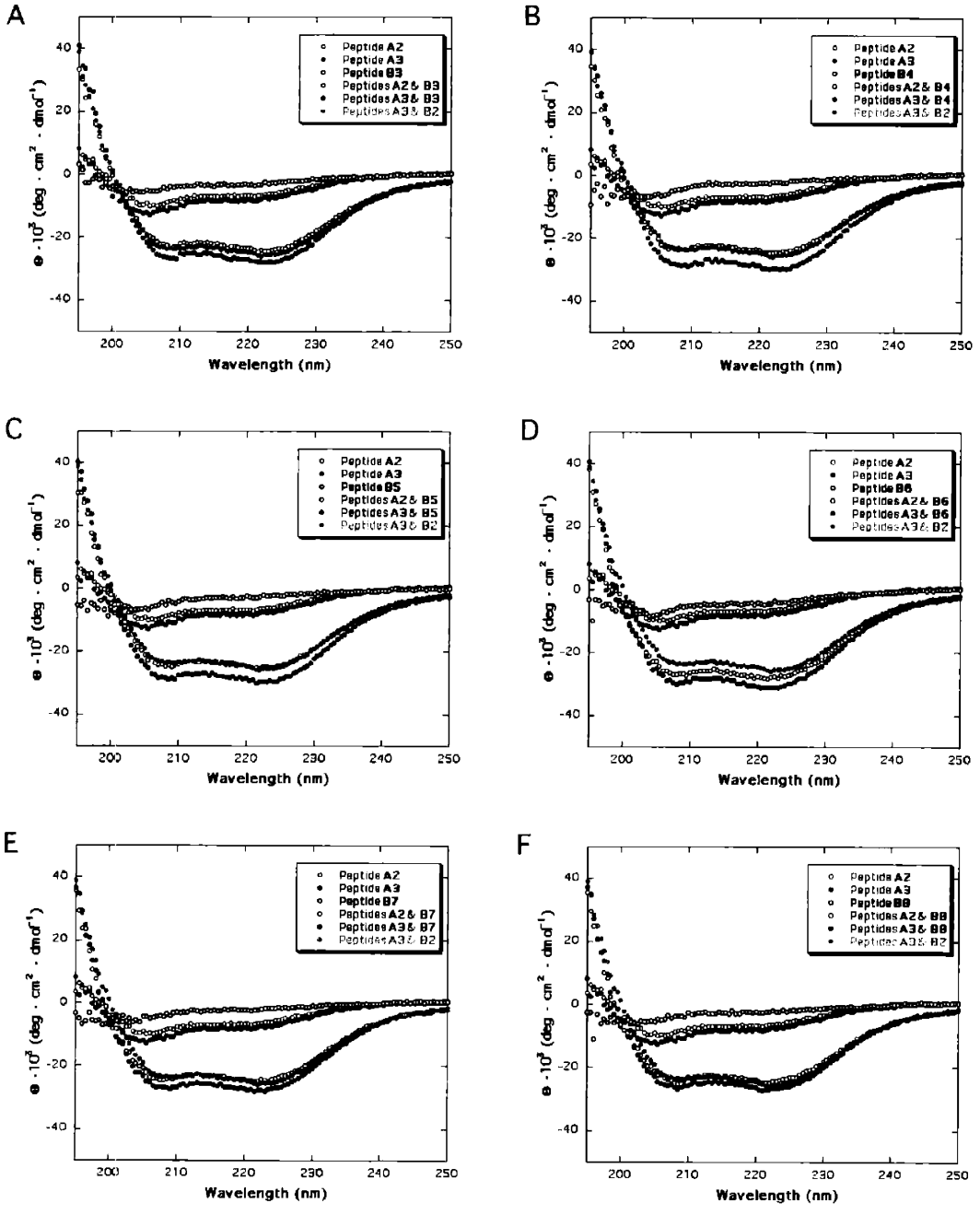


Figure 5-15. CD spectra of fourth-generation peptides singly and in combination with **A2** and **A3**. The spectrum of the **A3-B2** complex is shown for comparison. A: Peptide **B3**. B: Peptide **B4**. C: Peptide **B5**. D: Peptide **B6**. E: Peptide **B7**. F: Peptide **B8**.

Fluorescence quenching was observed in all pair-wise combinations, indicating that all peptide combinations were capable of interacting in a specific manner placing the fluorophore an quencher in proximity, presumably via a BBAT-like heterotetramer (Figure 16). The fluorescence quenching assay did not reveal a clear "winner." In most cases the greatest quenching was observed in combination with peptide **A3**, and the least in combination with **A2**. This is consistent with the hypothesis that greater steric bulk in the core results in a more stable interaction. The combination **A3-B6**, the most structured by CD, was not one of the strongest interactions by fluorescence. The greatest fluorescence quenching was observed for the **A3-B8** pair.

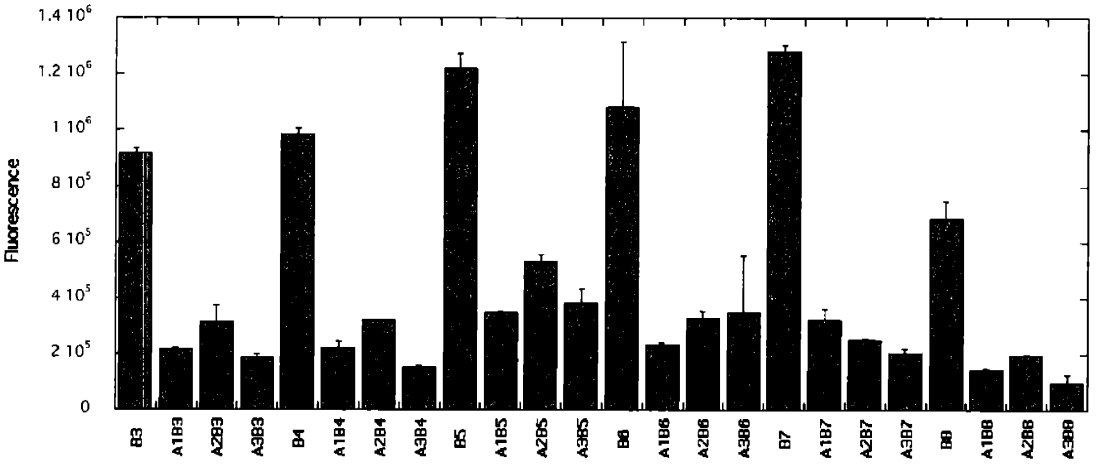


Figure 5-16. Results of fluorescence quenching experiments. Peptides **A1**, **A2** and **A3** are labeled with a 3-nitrotyrosine quencher and peptides **B3**, **B4**, **B5**, **B6**, **B7** and **B8** with an anthranilic acid fluorophore.

The spectroscopic findings were taken as indication that the mutations did not substantially alter the heterooligomeric interaction. The goal of this generation of peptides was destabilization of the homooligomeric state by selective mutation of core and interface residues. It was anticipated some of the changes made could also result in modest stabilization of

heterooligomer, as was observed by CD. The success of the negative design strategy was assessed by sedimentation equilibrium AUC.

All six peptides were found to self-associate appreciably (Table 5-13). AUC results indicated that more than one species was present in most cases. As the data, collected in six-sector cells, were of insufficient quality for confidence in fitting complex models, weight-average molecular weights were determined for the peptides in a global fit. The weight-average molecular weight was also determined for the lowest of the three concentrations examined. A weight-average molecular weight for the lowest concentration that is lower than the overall weight-average molecular weight can indicate the presence of monomer at low concentrations, and thereby reveal directions to pursue in following generations. A weight-average molecular weight that is higher can reveal that the peptide forms aggregates that precipitate at high concentrations, thus reducing the apparent molecular weight in a global fit of all the data.

Peptide[1]	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 5-13. Results of AUC experiments with fourth-generation peptides. 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.

The AUC results indicated that the changes made in the fourth generation were not sufficient to preclude homotetramerization. Peptide **B3**, having arginine in place of two lysines, has the most favorable properties with regard to self-association, with an average stoichiometry of roughly 2 at 100 μM. An analog with an additional arginine at position 13 may exhibit a correspondingly reduced tendency to self-associate.

The substitution of Orn and/or Dab does not appear to have significantly affected the tendency to self-associate in **B4** and **B5** relative to **B2**, suggesting that the second design strategy was ineffective in this system. Indeed, peptides **B7** and **B8** show considerably more self-association than does peptide **B6** (a comparison of the global fit to the lowest concentration fit indicates that both peptides aggregate and precipitate from solution at high concentrations), leading to the conclusion that Orn and/or Dab are sometimes less effective than Lys (or Arg) at disfavoring homooligomerization. By contrast, these substitutions have been reported by Kim and coworkers to reduce homooligomerization in the case of peptide "Velcro."[13]

A comparison of **B6** to **B2** suggests that the introduction of *m*-methylphenylalanine has little effect on association state. It is of interest that the introduction of *m*-methylphenylalanine in the core, in conjunction with the substitution of Orn and/or Dab at the interface, appears to have increased the tendency of the peptide to associate, as suggested by a comparison of **B7** to **B4** and **B8** to **B5**. Both **B7** and **B8** form aggregates at increased concentrations.

The additional methyl group, rather than disfavoring the homooligomeric state, may, by virtue of its greater hydrophobicity, increase the tendency of the peptide to self-associate. This finding suggests that adding bulk to the "large" partner is not an effective negative design strategy in this peptide system, and that efforts to reduce homooligomerization in the "large" partner should focus on the inter-subunit interface residues or expand the "large-small" strategy to additional layers.

Conclusions and Future Directions

The crystal structure of peptide **1**, described in detail in Chapter 4, will greatly facilitate further design work on oligomeric BBA peptides. Moreover, knowledge of the structure will be an invaluable aid in the design of a tetrameric BBA scaffold with improved catalytic function. Other design goals are increased homooligomer stability and the substitution of natural amino acids for nonstandard amino acids. It may also be possible to design a dimeric BBA from peptide

1 by reducing the size of the hydrophobic face. Furthermore, the structure of the homotetrameric α/β mini-protein **1** provides a valuable model, arguably the most protein-like in the literature, for the computational study of protein folding and stability.

This chapter relates efforts to increase the structural complexity of peptide **1** by the design of an $[A_2B_2]$ heterotetrameric BBA peptide. Two distinct components were designed by desymmetrization of the pseudo four-fold axis. These components were differentiated by the placement of opposite charges at the inter-monomer interface, and by the adoption of a "large-small" steric complementarity strategy in the C layer of the hydrophobic core.

The substitution of oppositely charged residues along the inter-monomer interface was found to specify a preference for the heterooligomeric state, but did not preclude self-association of the component peptides. Replacement of a core leucine in layer C with a small residue (Ala, Abu) or a large residue (Phe, *m*-methylphenylalanine) was found to destabilize the corresponding homooligomers. These substitutions result in a monomeric state in the "small" partner, whereas the "large" partner retains a tendency to self-associate at high concentrations and in the presence of salt.

Heterooligomer stability, as assessed by circular dichroism melt studies, appears to be determined primarily by the packing of the C layer. Design calculations suggest that peptides **A3** and **B2**, having Abu and Phe, respectively, at position 12, form a well-packed hydrophobic core. The heterotetramer formed by **A3** and **B2** approaches peptide **1** in stability.

Three strategies to reduce oligomerization of the basic partner with a large residue at position 12 were investigated. The most effective of these strategies, the substitution of arginine for lysine at two interface positions, resulted in a significant, but incomplete, reduction in homoassociation. This finding suggests that pK_a perturbation may enable homooligomerization in peptide **B2**.

Significant progress has been made on this design project, but several challenges remain. One challenge is the design of a basic peptide with a large residue at position 12 that is not prone to self-association. Another is the design of a heterotetramer, the components of which do not self-associate, with stability comparable to that of peptide **1**.

Certain features of the heterooligomeric interaction remain to be established. It would be of interest to obtain quantitative values for the association constants of the acidic and basic peptides. Furthermore, the stoichiometry of the interaction has not been rigorously established. Finally, a crystal structure of a heterooligomeric BBA would be an invaluable aid to the design of further heterooligomeric BBA peptides.

In conclusion, the studies described in this thesis have provided detailed structural information about a family of homooligomeric α/β mini-proteins. This information will greatly facilitate the design of other oligomeric BBA mini-proteins. Heterooligomeric BBA mini-proteins have been designed and characterized, yielding insights into the determinants of mini-protein specificity and stability. These heterooligomeric BBA mini-proteins constitute valuable minimal models for the study of heterooligomeric α/β proteins.

Acknowledgements

I thank Christina M. Taylor and Professor Amy E. Keating for undertaking the design and computational aspects of this project and for many helpful discussions. I am also very grateful to Ms. Taylor for experimental assistance.

Materials and Methods

Peptide Synthesis

Peptides were prepared by standard Fmoc-based solid-phase peptide synthesis. Unless otherwise noted, all peptide synthesis reagents were purchased from Applied Biosystems or Novabiochem.

The *C*-terminal dipeptide was manually coupled using HBtU/HOBt/DIPEA chemistry. DapBz (DapAbz) was incorporated as Fmoc-Dap(alloc)-OH (Bachem), deprotected on the solid phase resin, and side chain capped with benzoic acid (Boc-anthranilic acid) purchased from Aldrich (Advaned ChemTech). The synthesis was then completed using an Advanced ChemTech 396Ω peptide synthesizer. Fmoc-3-nitrotyrosine was purchased from Advanced ChemTech, Fmoc-*m*-methylphenylalanine was purchases from Fluka, and Fmoc-Dab(Boc)-OH was purchased from Bachem.

Cleavage and Purification

Peptides were cleaved from the resin for 2-6 hours with 2.5% triisopropylsilane, 2.5% water, and 95% trifluoroacetic acid, after which the cleavage cocktail was filtered to remove the resin and the filtrate concentrated under a gentle stream of nitrogen. When the volume was between ~1-2 ml the peptide was precipitated from ice-cold 2:1 ether:hexane, and centrifuged. The supernatant was decanted and the pellet resuspended in ice-cold 2:1 ether:hexane and washed 1-2 more times. The supernatant was then decanted, and the pellet dissolved in ~10% acetonitrile and immediately purified or lyophilized. High performance liquid chromatography was performed using a Waters Prep LC 4000 System equipped with a Waters 2487 absorbance detector (set at 228 and 280 nm), a C_{18} reverse-phase analytical column (Beckman Ultrasphere, 5 μm, 4.6 x 150 mm) and a C_{18} reverse-phase preparative column (YMC-pack, 250 x 20 mm). A standard gradient of 90:10 to 30:70 (water:acetonitrile, 0.1% TFA) over 25 minutes (with a flow rate of 1 ml/min on the analytical column, and 15 ml/min on the preparative column) was employed in all cases. Purity was determined by electrospray mass spectroscopy (PerSeptive Biosystems Mariner™ BioSpectrometry Workstation, using the Turbo Ion Source) and by analytical HPLC. Retention times for peptides **A1**, **A2**, **A3**, **B1**, and **B2** are 31 min, 29 min, 29 min, 28 min, and 26.5 min, respectively. For peptides **B3**, **B4**, **B5**, **B6**, **B7**, and **B8** purity was confirmed by injection onto a

Beckman System Gold HPLC system equipped with a Beckman System Gold 166 Absorbance Detector (monitoring at 229 nm), Beckman System Gold 125 Solvent Module, Beckman System Gold 508 Autosampler, a C18 reverse-phase analytical column (Vydac, 5 μm, 4.6 x 250 mm) and a C18 reverse-phase pre-column (Beckman Coulter Ultrasphere 4.6 x 45 mm) with 20 minutes equilibration at 90:10 (water:10% acetonitrile, 0.1% TFA) followed by a standard gradient of 90:10 to 20:80 (water:10% acetonitrile, 0.1% TFA) over 60 minutes at a flow rate of 1 ml/min. Retention times for peptides **B3**, **B4**, **B5**, **B6**, **B7**, and **B8** are 48.5 min, 48 min, 47.5 min, 49.5 min, 49 min, and 48.5 min, respectively.

	Calculated			Observed		
Peptide	**M**	**$(MH_2)^{2+}$**	**$(MH_3)^{3+}$**	**M**	**$(MH_2)^{2+}$**	**$(MH_3)^{3+}$**
A1	2601.3	1301.7	868.1	2602.0	1302.0	868.3
A1 (NO_2-Y)	2647.3	1324.7	883.4	2647.0	1324.5	883.3
B1	2612. 5	1307.2	871.8	2613.2	1307.6	872.1
B1 (DapAbz)	2626.5	1314.2	876.5	2628.2	1315.1	877.1
A2	2559.2	1280.6	854.1	2560.0	1281.0	854.3
A2 (NO_2-Y)	2605.2	1303.6	869.4	2605.2	1303.7	869.4
B2	2646.4	1324.2	883.2	2647.2	1324.6	883.4
B2 (DapAbz)	2660.4	1331.2	887.8	2662.2	1332.1	888.4
A3	2573.3	1287.6	858.8	2574.0	1288.0	859.0
A3 (NO_2-Y)	2619.2	1310.6	874.1	2619.0	1310.4	874.0
B3	2702.5	1352.2	901.8	2703.0	—	902.0
B3 (DapAbz)	2717.5	1359.7	906.8	2718.0	—	907.0
B4	2618.4	1310.2	873.8	2619.0	1310.5	874.0
B4 (DapAbz)	2632.4	1317.2	878.5	2634.0	1318.0	879.0
B5	2604.4	1303.2	869.1	2604.9	1303.5	869.3
B5 (DapAbz)	2619.4	1310.7	874.1	2619.9	1311.0	874.3
B6	2661.5	1331.7	888.2	2661.0	1331.5	888.0
B6 (DapAbz)	2676.5	1339.2	893.2	2676.0	1339.0	893.0
B7	2633.4	1317.7	878.8	2633.1	1317.5	878.7
B7 (DapAbz)	2648.4	1325.2	883.8	2648.1	1325.1	883.7
B8	2619.4	1310.7	874.2	2619.0	1310.5	874.0
B8 (DapAbz)	2634.4	1318.2	879.2	2634.0	1318.0	879.0

Table 5-14. Calculated and observed mass spectroscopic data for peptides characterized in this chapter.

Circular Dichroism

Peptides were dissolved in degassed 10 mM phosphate buffer, pH 7.2. Stock solution concentrations were determined based on absorbance at 280 nm, and appropriate dilutions used to prepare 50 μM samples for analysis. All combination samples were equimolar with a total peptide concentration of 50 μM. Data were collected with an Aviv Circular Dichroism Spectrometer Model 202 using strain-free quartz cells having a path length of 0.1 mm. Wavelength scan experiments were performed in duplicate or triplicate, from 300 to 195 nm, at 25 °C, with an averaging time of 0.5 s and a step size of 1.0 or 0.5 nm. Melting experiments were performed as wavelength scans (260 nm to 205 nm, in duplicate, 0.5 s averaging time, 1 nm step size) in temperature increments of 5 °C from 5 °C to 80 °C.

3-Nitrotyrosine Extinction Coefficient Determination.

A 2.7 mM stock solution of 3-nitrotyrosine (Sigma) was prepared by dissolving 30.2 mg in a total volume of 50 ml in a volumetric flask. The extinction coefficient was determined by correlating the absorbance of a concentrated stock solution at 360 nm (measured with a Shimadzu UV160U spectrophotometer) with its concentration. The extinction coefficient of 3-nitrotyrosine at 360 nm was found to be 1874 cm^{-1} M^{-1}.

Fluorescence

Fluorophore- or quencher-labeled peptides solutions were prepared in 10 mM phosphate buffer, pH 7.2. Fluorophore-labeled peptide stock solution concentrations were determined based on absorbance at 315 nm (ε = 2000 cm^{-1} M^{-1}).[38] Quencher-labeled peptide stock solutions were determined based on absorbance at 360 nm (ε = 1874 cm^{-1} M^{-1}). 12.5 μM samples were prepared of fluorophore- or quencher-labeled peptides. Combination samples comprised 12.5 μM fluorophore-labeled peptide + 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

Stock solutions of peptides were prepared in 50 mM phosphate buffer with 100 mM NaCl at pH 7.2. Each solution was then loaded onto a 1.0 kDa MWCO DispoDialyzer™ (Harvard Biosciences) and dialyzed at room temperature. Dilutions were made from the dialyzed stock solution. Peptide solutions were loaded into standard two- or six-sector epon centerpieces and 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 contents of each cell were confirmed to be at equilibrium using WINMATCH prior to increasing the speed. Data were analyzed using the programs NONLIN[37] and SEDPHAT.[35,36] Molecular weights were determined using a partial specific volume, $\bar{\nu}$, calculated by SEDNTERP.[39]

Crystallization

An equimolar mixture of peptides **A2** and **B2** was dissolved at 11 mg/ml in 10 mM phosphate at pH 7.2 for crystallization. Crystals were grown using vapor diffusion with hanging-drop geometry 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).

Data Collection

Crystals were frozen in a stream of N_2 gas cooled to -180 °C using FMS oil (Hampton Research) as a cryoprotectant. A 1.2 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. Data were collected in two sweeps of varied exposure time to ensure high completeness of both high resolution and low resolution data. The DENZO and SCALEPACK package[40] was used for data indexing, reduction, and scaling.

References

(1) Hodges, R. S. *Biochem. Cell Biol.* **1996**, *74*, 133-154. *De Novo* Design of α-Helical Proteins: Basic Reseach to Medical Applications.

(2) Litowski, J. R.; Hodges, R. S. *J. Pept. Res.* **2001**, *58*, 477-492. Designing Heterodimeric Two-Stranded Alpha-Helical Coiled-Coils: The Effect of Chain Length on Protein Folding, Stability and Specificity.

(3) Litowski, J. R.; Hodges, R. S. *J. Biol. Chem.* **2002**, *277*, 37272-37279. Designing Heterodimeric Two-Stranded Alpha-Helical Coiled-Coils. Effects of Hydrophobicity and Alpha-Helical Propensity on Protein Folding, Stability, and Specificity.

(4) McClain, D. L.; Binfet, J. P.; Oakley, M. G. *J. Mol. Biol* **2001**, *313*, 371-383. Evaluation of the Energetic Contribution of Interhelical Coulombic Interactions for Coiled Coil Helix Orientation Specificity.

(5) Lombardi, A.; Bryson, J. W.; DeGrado, W. F. *Biopolymers* **1996**, *40*, 495-504. *De Novo* Design of Heterotrimeric Coiled Coils.

(6) Nautiyal, S.; Woolfson, D. N.; King, D. S.; Alber, T. *Biochemistry* **1995**, *34*, 11645-11651. A Designed Heterotrimeric Coiled Coil.

(7) Schnarr, N. A.; Kennan, A. J. *J. Am. Chem. Soc.* **2003**, *125*. Specific Control of Peptide Assembly with Combined Hydrophilic and Hydrophobic Interfaces.

(8) Schnarr, N. A.; Kennan, A. J. *J. Am. Chem. Soc.* **2001**, *123*, 11081-11082. Coiled-Coil Formation Governed by Unnatural Hydrophobic Core Side Chains.

(9) Fairman, R.; Chao, H.-G.; Lavoie, T. B.; Villafranca, J. J.; Matsueda, G. R.; Novotny, J. *Biochemistry* **1996**, *35*, 2824-2829. Design of Heterotetrameric Coiled Coils: Evidence for Increased Stabilization by Glu^{-}-Lys^{+} Ion Pair Interactions.

(10) Sia, S. K.; Kim, P. S. *Biochemistry* **2001**, *40*, 8981-8989. A Designed Protein with Packing between Left-Handed and Right-Handed Helices.

(11) Lumb, K. J.; Kim, P. S. *Biochemistry* **1995**, *37*, 13042. A Buried Polar Interaction Imparts Structural Uniqueness in a Designed Heterdimeric Coiled Coil.

(12) Kohn, W. D.; Hodges, R. S. *Trends Biotechnol.* **1998**, *16*, 379-389. *De Novo* Design of α-Helical Coiled Coils and Bundles: Models for the Development of Protein-Design Principles.

(13) O'Shea, E. K.; Lumb, K. J.; Kim, P. S. *Curr. Biol.* **1993**, *3*, 658-667. Peptide 'Velcro': Design of a Heterodimeric Coiled Coil.

(14) Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Mol. Biol* **1994**, *237*, 500-512. The Role of Interhelical Ionic Interactions in Controlling Protein Folding and Stability. *De Novo* Designed Synthetic Two-Stranded α-Helical Coiled Coils.

(15) Burkhard, P.; Ivaninski, S.; Lustig, A. *J. Mol. Biol.* **2002**, *2002*, 901-910. Improving Coiled-Coil Stability by Optimizing Ionic Interactions.

(16) Lavigne, P.; Sönnichsen, F. D.; Kay, C. S.; Hodges, R. S. *Science* **1996**, *271*, 1136-1137. Interhelical Salt Bridges, Coiled-Coil Stability, and Specificity of Dimerization.

(17) Lumb, K. J.; Kim, P. S. *Science* **1996**, *271*, 1137-1138. Interhelical Salt Bridges, Coiled-Coil Stability, and Specificity of Dimerization.

(18) Spek, E. J.; Bui, A. H.; Lu, M.; Kallenbach, N. R. *Protein Sci.* **1998**, *7*, 2431-2437. Surface Salt Bridges Stabilize the GCN4 Leucine Zipper.
(19) Park, D. J. M. Eng. Thesis, Massachusetts Institute of Technology, 2002. Computational Tools For Including Specificity In Protein Design.
(20) Desmet, J.; Maeyer, M. D.; Hazes, B.; Lasters, I. *Nature* **1992**, *256*, 539-542. The Dead-End Elimination Theorem and Its Use in Protein Side-Chain Positioning.
(21) Leach, A. R.; Lemon, A. P. *Proteins: Struct., Funct., Genet.* **1998**, *1998*, 227-239. Exploring the Conformational Space of Protein Side Chains Using Dead-End Elimination and the A* Algorithm.
(22) Grigoryan, G.; Keating, A. E. Unpublished results.
(23) Dunbrack, R. L., Jr.; Karplus, M. *Nat. Struct. Biol.* **1994**, *1*, 334-340. Conformational Analysis of the Backbone-Dependent Rotamer Preferences of Protein Sidechains.
(24) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217. CHARMM.
(25) MacKerell, A. D., Jr.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *The Encyclolpedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; John Wiley: Chichester, 1998; Vol. 1.
(26) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133-52. Effective Energy Function For Proteins In Solution.
(27) Mezo, A. R.; Ottesen, J. J.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 1002-1003. Discovery and Characterization of a Discretely Folded Homotrimeric Peptide.
(28) Zhou, N. E.; McKay, C.; Hodges, R. S. *Biochemistry* **1992**, *31*, 5739-5746. 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.
(29) Zhou, N. E.; Zhu, B.-Y.; Kay, C. M.; Hodges, R. S. *Biopolymers* **1992**, *32*, 419-426. The Two-Stranded α-Helical Coiled-Coil is an Ideal Model for Studying Protein Stability and Subunit Interactions.
(30) Rivas, G.; Minton, A. P. *Biochem. Soc. Trans.* **2003**, *31*, 1015-1019. Tracer Sedimentation Equilibrium: A Powerful Tool For The Quantitative Characterization of Macromolecular Self- and Hetero-Associations in Solution.
(31) Rivas, G.; Stafford, W.; Minton, A. P. *Methods* **1999**, *19*, 194-212. Characterization of Heterologous Protein-Protein Interactions Using Analytical Ultracentrifugation.
(32) Lebowitz, J.; Lewis, M. S.; Schuck, P. *Protein Sci.* **2002**, *11*, 2067-2079. Modern Analytical Ultracentrifugation in Protein Science: A Tutorial Review.
(33) Laue, T. M.; Senear, D. F.; Eaton, S.; Ross, J. B. A. *Biochemistry* **1993**, *32*, 2469-2472. 5-Hydroxytryptophan as a New Intrinsic Probe for Investigating Protein-DNA Interactions by Analytical Ultracentrifugation. Study of the Effect of DNA on Self-Assembly of the Bacteriophage λ *c*I Repressor.
(34) Matthews, B. W. *J. Mol. Biol* **1968**, *33*, 491-497. Solvent Content of Protein Crystals.
(35) Schuck, P. *Anal. Biochem.* **2003**, *320*, 104-124. On the Analysis of Protein Self-Association by Sedimentation Velocity Analytical Ultracentrifugation.
(36) Vistica, J.; Dam, J.; Balbo, A.; Yikilmaz, E.; Mariuzza, R. A.; Roualt, T. A.; Schuck, P. *Anal. Biochem.* **2004**, *326*, 234-256. Sedimentation Equilibrium Analysis of Protein Interactions with Global Implicit Mass Conservation Constraints and Systematic Noise Decomposition.
(37) Johnson, M. L.; Correia, J. C.; Yphantis, D. A.; Halvorson, H. R. *Biophys. J.* **1981**, *36*, 575-588. Analysis of Data from the Analytical Ultracentrifuge by Nonlinear Least-Squares Techniques.

(38) Mezo, A. R.; Cheng, R. P.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 3885-3891. Oligomerization of Uniquely Folded Mini-Protein Motifs: Development of a Homotrimeric Beta Beta Alpha Peptide.

(39) Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Harding, S. E., Rowe, A. J., Horton, J. C., Eds.; Royal Society of Chemistry: Cambridge, 1992.

(40) Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307-326. Processing of X-ray Diffraction Data Collected in Oscillation Mode.

Appendix: An Improved Method for the Palladium-Catalyzed Amination of Aryl Iodides

Abstract

Aryl iodides are coupled with amines to give the corresponding arylamines in high yield in the presence of palladium, a suitable ligand, and NaO*t*-Bu. Functionalized aryl iodides give good yields of the corresponding arylamines when $Cs_2CO_3$is substituted as the base.

Introduction

Arylamines are attractive targets for chemical synthesis because of their prevalence and wide utility. One of their earliest applications was in the production of brightly colored synthetic dyes, introduced in the late nineteenth century.[1] Arylamines have a large number of other applications and are thus attractive targets for chemical synthesis. They are found in biologically active compounds such as pharmaceuticals[2-4] and agrochemicals.[5] Several commonly occurring DNA lesions are arylamines, and they have been the target of recent synthetic efforts.[6-8] Arylamines have also been employed as ligands for transition metals,[9] and in the design of conductive polymers[10] and other electronically interesting materials.[11]

Traditional routes for the synthesis of these compounds such as electrophilic nitration and subsequent reduction, nucleophilic aromatic substitution, and Ullmann-type couplings often suffer from relatively harsh conditions and limited generality.[12,13] By contrast, the transition metal-mediated coupling of amines with aryl halides is regioselective, does not require activating groups, and occurs under relatively mild conditions. Since its discovery by Migita and co-workers in 1983,[14] and the initial breakthrough of a tin-free reaction,[15,16] this reaction has evolved into a general synthesis of substituted and unsubstituted anilines. There are a number of reviews

highlighting recent progress in this area.[12,17-19] Much of this work has focused on the reactions of industrially relevant aryl chlorides and bromides.

It is also important to develop appropriate reaction conditions for aryl iodides, as they are widely used as substrates for palladium-catalyzed cross-coupling reactions. They are readily accessible intermediates for use in synthetic organic chemistry, e.g., by electrophilic iodination,[20] directed ortho metalation,[21] or by conversion of organometallic intermediates.[22] Moreover, a large selection is available from commercial sources. Although their cost is a deterrent on a production scale, they are invaluable both in academic and industrial laboratories.

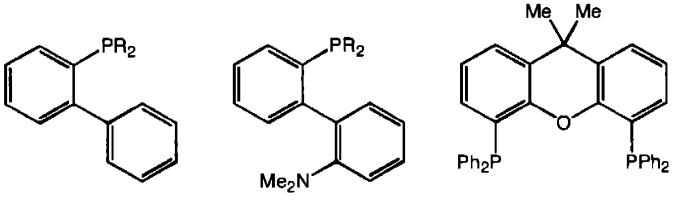
Exceptional among palladium-catalyzed cross-coupling reactions, aryl iodides have historically been less effective substrates for the amination reaction than their bromide counterparts.[14,15,23] In the initial procedure developed in our group, it was found that aryl iodides were ineffective in the intermolecular amination reaction, whereas they outperformed aryl bromides in intramolecular aminations.[15]

The earliest work from these laboratories to focus on the intermolecular amination of aryl iodides employed P(*o*-tol)$_3$ and NaO*t*-Bu (the first generation improvement upon the Migita procedure).[24] The use of dioxane as solvent was critical to the success of this undertaking. Yields, on average, were lower than those of the corresponding aryl bromide substrates. After the discovery that chelating bisphosphines enhanced both substrate scope and reaction rates relative to the monodentate phosphine P(*o*-tol)$_3$,[25,26] it was found that a Pd/BINAP mixture, with NaO*t*-Bu as base, could catalyze the amination of aryl iodides at room temperature in good yield in the presence of stoichiometric 18-crown-6.[27]

In 1996, Hartwig described five examples in which aryl iodides were coupled with amines in high yield using 5 mol % of (DPPF)PdCl$_2$ at 100 °C.[26] He later reported that aryl iodides underwent the amination reaction at room temperature when the ligand D*t*BPF was employed; however, the yields obtained were low in two of the three examples cited.[28] More

recently, Nolan, in an application of Arduengo carbene-type ligands, has shown one instance of an aryl iodide undergoing the amination reaction in high yield at room temperature.[29]

One aim of this study was to obtain general conditions, not requiring the use of toxic additives, for the cross-coupling of aryl iodides with all classes of amines. Another goal was to optimize mild conditions for the reaction of aryl iodides and amines containing base-sensitive functional groups. A class of biphenyl-based ligands (e.g., **1a**, **1b**, **2**) developed by our group[30,31] provided a starting point for these explorations. Xantphos (3), a chelating bisphosphine developed by van Leeuwen,[32] has been found to be a very effective ligand for amination[33-35] and amidation[36,37] reactions. Xantphos was screened alongside the biphenyl-based ligands in this study, and was the optimal ligand for certain substrate combinations.



1a: R=*t*-Bu
1b: R=Cy

2

3
xantphos

Figure A-1. Chemical structures of ligands **1a**, **1b**, **2**, and **3**.

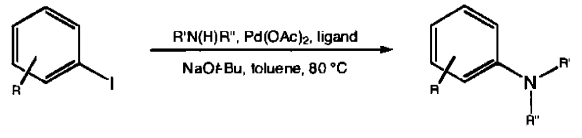
Results and Discussion

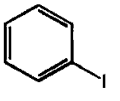
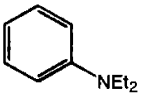
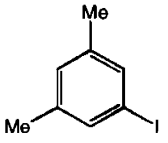
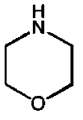
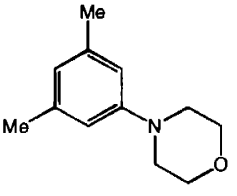
Most of the methodological advances in the palladium-catalyzed amination reaction have resulted from the development of increasingly effective ligands.[12,17-19,38] The monophosphinobiphenyl ligands developed in our group (e.g., **1a**, **1b**, **2**) are constituents of highly effective catalyst systems that have been applied in a number of palladium-catalyzed cross-coupling reactions, including Suzuki reactions,[31,39] amination of aryl halides,[31,40] diaryl ether formation,[41] and arylation of enolates.[42] These ligands are notable for being both bulky and electron-rich: their bulk speeds reductive elimination and favors coordinative unsaturation, and

their high electron density facilitates oxidative addition.[39] Other ligands currently used for palladium-catalyzed amination (e.g., PCy_3,[43] $P(t\text{-}Bu)_3$,[38,44] PPF-OMe, and related ferrocenyl ligands,[28,45] palladacycles,[46] and carbene-type ligands[29,47]) share many of these attributes. One distinguishing feature of ligands **1a**, **1b** and **2** is their exceptional air stability, even in solution.[40]

Initial studies on the amination of aryl iodides utilized ligand **1a** in light of its general utility for the amination of aryl bromides and aryl chlorides.[40] The results proved disappointing: low conversions were observed, and free ligand was often detectable by GC. A switch to the smaller cyclohexyl analogue, **1b**, provided more satisfactory results. Although in certain very facile reactions the *tert*-butyl ligand was as effective, or almost as effective, as the cyclohexyl, in most cases **1b** performed significantly better. Two such examples are given in Table 6-1.

Table A-1. Cyclohexyl and *tert*-Butyl Biphenyl Ligands for the Amination of Aryl Iodides



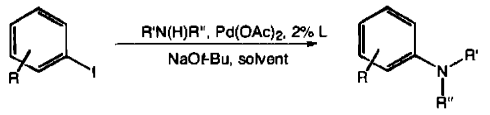
Entry	Aryl	Amine	Product	Ligand	Yield (%)
1		Et_2NH		**1a**	85%
2				**1b**	trace product
3				**1a**	86%
4				**1b**	trace product

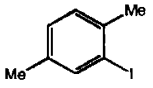
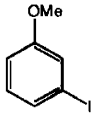
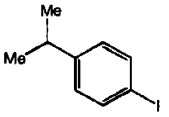
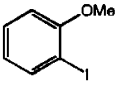
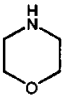
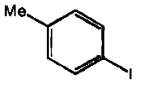
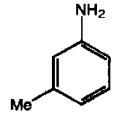
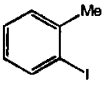
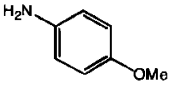
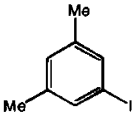
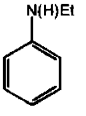
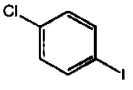
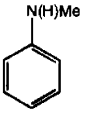
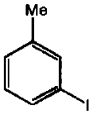
As further experiments revealed large disparities in reaction times, temperatures, and optimal ligand between substrates, a systematic optimization of the reaction conditions (including ligand, temperature, and catalyst loading) was undertaken for each type of amine being coupled.

A preliminary screen indicated the best ligand for a given class of amines. Initial experiments indicated that **1b**, **2**, and **3** were the most generally effective ligands; they were screened in each case. Other ligands, e.g., binaphthyl-based, ferrocenyl-based, and *tert*-butyl biphenyl phosphines, were periodically reexamined and found to be less effective than this basis set. Reactions were run in dioxane and in a 2:1 dioxane/*tert*-butyl alcohol mixture for the first round of experiments, as preliminary results indicated that dioxane was the most generally effective solvent (as in previous studies with aryl iodides[24]). The addition of *tert*-butyl alcohol led to enhanced reaction rates and conversion in most cases.[48] The use of nonpolar solvents resulted in slower conversion, whereas employing highly polar solvents such as NMP or DMA led to significant reduction of aryl iodide starting material to arene (presumably via β-hydride elimination and reductive elimination from the amido complex).

Once an optimal ligand was found, further screening was used to optimize reaction temperatures and solvents. The results are shown in Table 2. Yields are very good to excellent for most substrate combinations and higher than those obtained with the BINAP/18-crown-6 protocol. In most instances, the ligand of choice is **2**, but **1b** and **3** are useful in certain cases.

Table A-2. Amination of Aryl Iodides with NaO*t*-Bu as Base[a]



Entry	Aryl	Amine	Ligand	Solvent[b]	Temp (°C)	Yield[c] (%)
1		*n*-HexNH$_2$	**1b**	a	100	70
2		BnNH$_2$	**1b**	a	100	79
3		MeN(H)Bn	**2**	a	60	89
4			**3**	b	60	94
5			**2**	b	45	99
6			**2**	b	45	99
7			**2**	c	RT	>99
8			**2**	c	RT	97
9		Ph_2NH	**2**	b	100	98

[a] Reaction times varied from 2 to 12 h.[b] Solvents: (a) 2:1 dioxane/*t*-BuOH, (b) dioxane, (c) tetrahydrofuran.[c] Yields refer to the average isolated yield of two runs. All new compounds were characterized by ^{1}H NMR, ^{13}C NMR, and elemental analysis.

It is necessary to heat reactions involving primary aliphatic amines (entries 1 and 2) to 100 °C to achieve complete conversion. This may be due to the formation of catalytically inactive amido-bridged complexes and/or bis(amine) complexes.[49] The somewhat lower yields obtained with primary amines relative to other classes of amines stems in part from the formation of diarylated side products. When optimizing ligands, we sought to minimize the extent of diarylation: **2** exhibited comparable reactivity to **1b** but led to higher incidences of diarylation, and hence **1b** was deemed the ligand of choice. The use of 1.4 equiv of amine (rather than 1.2 equiv) also helps to increase the ratio of desired monoarylated to undesired diarylated products.

The reactions of secondary acyclic amines are usually inefficient when BINAP is employed as ligand.[50] Nonetheless, with this improved protocol, they give the corresponding anilines in good yields upon heating to 60 °C (entry 3). Secondary cyclic amines (entry 4) were also coupled in high yield.

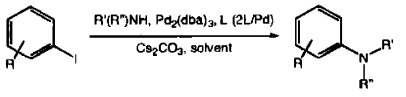
Particularly reactive substrates such as *N*-alkylanilines are converted readily at room temperature to afford the corresponding products in nearly quantitative yield (entries 7 and 8). Tetrahydrofuran was found to be the best solvent for this transformation. Although primary anilines can also be coupled at room temperature, it was found that gentle heating accelerates their reaction (entries 5 and 6).

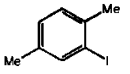
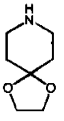
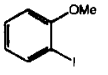
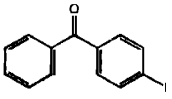
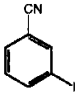
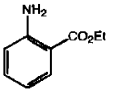
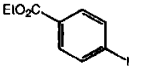
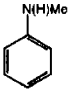
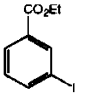
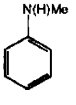
The reaction of diphenylamine (entry 9) requires heating to 100 °C to achieve complete conversion. This may be a consequence of its relative bulk, which would inhibit its substitution of an iodide-bridged palladium dimer complex[51] or of a bulky palladium alkoxide complex.[52] An alternative explanation is that their reduced basicity inhibits the bonding to the Pd(II) center, which is necessary to facilitate their deprotonation.

The palladium-catalyzed cross-coupling of aryl iodides with amines in the presence of NaO*t*-Bu is an efficient and high-yielding reaction appropriate for the reactions of many substrates. However, NaO*t*-Bu is incompatible with many commonly encountered functional groups such as esters, enolizable ketones,[42] and other base-sensitive groups. An effort was thus

made to extend the substrate scope of the process by the use of a milder base, as has been reported for the amination of aryl bromides and chlorides.[30,53] Among the bases screened were K_3PO_4, Cs_2CO_3, $CsHCO_3$, CsOAc, NaOAc, and proton sponge, the most effective of which was Cs_2CO_3. By contrast, when using our biphenyl-based ligands for the coupling of aryl bromides and chlorides, the best results were generally obtained with K_3PO_4. Accordingly, a ligand screen analogous to that described above for NaO*t*-Bu was undertaken with Cs_2CO_3.

Table A-3. Amination of Aryl Iodides with Cs_2CO_3 as Base[a]



Entry	Aryl	Amine	mol % Pd	Ligand	Solvent[b]	Temp (°C)	Yield[c] (%)
1		*n*-HexNH$_2$	3	2	a	120	74
2			1	2	b	100	81
3			2	2	b	120	62
4		(*n*-Pr$_2$)NH	5	2	c	120	62
5			1	3	a	100	91
6			1	3	c	120	86
7			1	3	c	120	84

[a] All reactions were run overnight; reaction times were not minimized.[b] Solvents: (a) 2:1 dioxane/*t*-BuOH, (b) 1:1 dioxane/*t*-BuOH, (c) 2:1 dioxane/Et_3N. [c] Yields refer to the average isolated yield of two runs. All new compounds were characterized by ^1H NMR, ^{13}C NMR, and elemental analysis.

Appropriate conditions were thus found for the functional group-tolerant reaction of various amines with aryl iodides in good to excellent yield, as shown in Table 3. Yields, although satisfactory, are somewhat lower than those obtained with NaO*t*-Bu. This can be attributed to the greater incidence of reduction observed under the mild base conditions. Elevated reaction temperatures (100 or 120 °C) are required in order to achieve high conversions with Cs_2CO_3. Reaction times are somewhat longer than those utilizing NaO*t*-Bu, although the protocol can be conveniently performed overnight. These differences may be a consequence of the lower solubility of Cs_2CO_3 in organic solvents relative to NaO*t*-Bu. The use of triethylamine as a cosolvent provided significant rate enhancements when running reactions with cesium carbonate (but not with NaO*t*-Bu). As with NaO*t*-Bu, ligand **2** is the most generally effective, although 3 is superior with primary and substituted aniline substrates.

Primary amines (entry 1) are readily coupled in good yields, although a higher catalyst loading (3% Pd) is required. As with NaO*t*-Bu, the requisite temperature is higher than for more activated substrates (vide supra). Secondary cyclic amines are coupled efficiently at 100 or 120 °C (entries 2 and 3).

Anilines and *N*-alkylanilines give high yields of the corresponding diarylamines when xantphos (**3**) is employed as ligand. Entry 3 of Table 3 (62%) can be compared to entry 4 of Table 2 (94%); the lower yield is due to the increased reduction to arene observed with Cs_2CO_3. The last two examples show that *N*-methylaniline reacts with activated and mildly activated aryl iodides in very good yield (comparable to those obtained with NaO*t*-Bu).

Two of the more difficult substrate classes, secondary acyclic amines and diarylamines, continue to give low conversions, decomposition products, and high incidences of reduction of aryl iodide to arene when Cs_2CO_3 is employed, despite much effort at optimization. The good yield obtained from the reaction of di(*n*-propyl)amine with 4-iodobenzophenone (entry 4) appears to be somewhat exceptional. Other substrates were less efficiently coupled. Efforts to find suitable conditions for these two classes of amine substrates are still in progress.

Conclusions

Conditions have been developed that allow for the coupling of aryl iodides with all classes of amines in good to excellent yields with NaO*t*-Bu as base. Milder conditions, utilizing Cs_2CO_3 as a base, have been optimized that permit the amination of aryl iodides in good yields in the presence of base-sensitive functional groups.

Experimental Section

General Considerations. All reactions were performed in oven- or flame-dried resealable Schlenk tubes under an atmosphere of argon. Aryl iodides were purchased from Aldrich Chemical Co., Lancaster Synthesis Limited or Trans World Chemicals Inc., and used without further purification. Amines were also purchased from commercial sources - those available in anhydrous form were used without further purification; all others were filtered through basic alumina prior to use. Tris-(dibenzylideneacetone)dipalladium(0) was purchased from Strem Chemical Company, as were ligands **1a** and **1b**. Ligands **2** and **3** were prepared following literature procedures.[30,32,54] Sodium *tert*-butoxide (Aldrich Chemical Co.) and cesium carbonate (Chemetall) were stored under nitrogen in a Vacuum Atmospheres glovebox. Small portions were removed every few weeks and stored under air in a benchtop desiccator. Dioxane and 2-methyl-2-propanol were purchased anhydrous from Aldrich Chemical Co. and were used without further purification. Triethylamine was distilled under argon from calcium hydride and stored over potassium carbonate. THF was distilled under argon from sodium benzophenone ketyl. Yields in Tables 2 and 3 refer to the average of two runs with isolated compounds of greater than 95% purity as determined by ^{1}H NMR and capillary GC. New compounds were characterized by combustion analysis; all elemental analyses were obtained from Atlantic Microlabs Incorporated, Norcross, GA. IR spectra of neat samples were acquired with the DiComp probe of an ASI REACTIR *in situ* IR instrument.

General Procedure. $Pd_2(dba)_3$ (4.6 mg, 0.005 mmol, 1 mol %), ligand **2** (7.9 mg, 0.02 mmol, 2 mol %), and base were weighed in air and transferred to a flame-dried resealable Schlenk tube. For Table 2, the base used was NaO*t*-Bu (135 mg, 1.4 mmol) and for Table 3 the base used was Cs_2CO_3 (456 mg, 1.4 mmol). Amine (1.2 mmol) and aryl iodide (1.0 mmol), if solid, were added at this point. The tube was then evacuated and backfilled with argon. The flask was capped with a rubber septum under an argon purge, and the liquid reagents (amine and/or aryl iodide) added. The sides of the flask were rinsed with solvent (3.0 mL), and the septum replaced by a Teflon screwcap. The Schlenk tube was then sealed and stirred at the appropriate temperature. When GC analysis indicated disappearance of aryl iodide, the reaction mixture was cooled to room temperature, diluted with ether, and filtered through a plug of Celite. The crude product was concentrated and purified directly by flash chromatography on silica gel.

***N*-(2,5-Xylyl)hexylamine (Table 2, Entry 1).**[24] The general procedure, with 0.02 mmol (7.0 mg) of ligand **1b**, 1.0 mmol (145 mL) of 2-iodo-*p*-xylene, and 1.4 mmol (185 mL) of *n*-hexylamine in 2:1 dioxane/*t*-BuOH at 100 °C, yielded 159 mg (77%) of a yellow oil.

***N*-(3-Methyoxyphenyl)benzylamine (Table 2, Entry 2).** The general procedure, with 0.02 mmol (7.0 mg) of ligand **1b**, 1.0 mmol (120 mL) of 3-iodoanisole, and 1.4 mmol (153 mL) of benzylamine in 2:1 dioxane/*t*-BuOH, yielded 166 mg (78%) of a yellow oil: ^{1}H NMR (300 MHz, $CDCl_3$) 7.38-7.25 (m, 5H), 7.10-7.04 (app t, 1 H), 6.29-6.23 (m, 2H), 6.19-6.18 (m, 1H), 4.31 (s, 2H), 4.04 (s, 1H), 3.74 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) 160.9, 149.6, 139.4, 130.1, 128.8, 127.6, 127.4, 106.1, 102.8, 99.0, 55.3, 48.6; IR (neat, cm^{-1}) 3415, 3062, 3029, 3000, 2954, 2834, 1613, 1598, 1495, 1453. Anal. Calcd for $C_{14}H_{15}NO$: C, 78.84; H, 7.09. Found: C, 78.96; H, 7.17.

***N*-Benzyl-*N*-methyl-4-isopropylaniline (Table 2, Entry 3).** The general procedure, with 0.02 mmol (7.9 mg) of ligand **2**, 1.0 mmol (166 mL) of 4-iodoisopropylbenzene, and 1.2 mmol (155

mL) of *N*-methylbenzylamine in 2:1 dioxane/*t*-BuOH at 60 °C, yielded 210 mg (88%) of an orange oil: ^{1}H NMR (300 MHz, $CDCl_3$) 7.33-7.20 (m, 5H), 7.10-7.06 (m, 2H), 6.73-6.68 (m, 2H), 4.48 (s, 2H), 2.97 (s, 3H), 2.88-2.75 (sept, 1H, J = 6.9 Hz), 1.22-1.20 (d, 6H, J = 7.1 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) 148.2, 139.5, 137.1, 128.6, 127.1, 127.0, 126.9, 112.6, 57.3, 38.8, 33.3, 24.5; IR (neat, cm^{-1}) 3029, 2958, 2927, 2867, 2809, 1615, 1519, 1495, 1453. Anal. Calcd for $C_{17}H_{21}N$: C, 85.31; H, 8.84. Found: C, 85.08; H, 8.89.

***N*-(2-methoxyphenyl)morpholine (Table 2, Entry 4).** The general procedure, with 0.02 mmol (11.6 mg) of ligand **3**, 1.0 mmol (130 mL) of 2-iodoanisole, and 1.2 mmol (105 mL) of morpholine in dioxane at 60 °C, yielded 184 mg (95%) of a brown oil: ^{1}H NMR (500 MHz, $CDCl_3$) 7.04-7.00 (m, 1H), 6.94-6.93 (m, 2H), 6.89-6.87 (m, 1H), 3.91-3.86 (m, 4H), 3.87 (s, 3H), 3.08-3.07 (m, 4H); ^{13}C NMR (75 MHz, $CDCl_3$) 152.3, 141.1, 123.3, 121.1, 118.1, 111.3, 67.4, 55.6, 51.4; IR (neat, cm^{-1}) 3062, 2956, 2854, 2833, 2817, 1594, 1499, 1447, 1378, 1299, 1237. Anal. Calcd for $C_{11}H_{15}NO_2$: C, 68.37; H, 7,82. Found: C, 68,13; H, 7.79.

3-Methyl-4'-methyldiphenylamine (Table 2, Entry 5). The general procedure, with 0.02 mmol (7.9 mg) of ligand **2**, 1.0 mmol (218 mg) of 4-iodotoluene, and 1.2 mmol (129 mL) of *m*-toluidine in dioxane at 45 °C, yielded 198 mg (>99%) of a yellow oil: ^{1}H NMR (300 MHz, $CDCl_3$) 7.15 (t, 1H, J = 7.5 Hz), 7.12-7.10 (d, 2H, J = 8.0 Hz), 7.03-7.01 (m, 2H), 6.86-6.84 (m, 2H), 6.74-6.72 (d, J = 7.0 Hz), 5.58 (br s, 1H), 2.33 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) 144.0, 140.4, 139.3, 130.9, 129.9, 129.3, 121.3, 119.0, 114.1, 21.8, 21.0; IR (neat, cm^{-1}) 3394, 3025, 2919, 2861, 1605, 1590, 1515, 1490, 1318, 1304. Anal. Calcd for $C_{14}H_{15}N$: C, 85.24; H, 7.66. Found: C, 85.01; H, 7.72.

2-Methyl-4'-methoxydiphenylamine (Table 2, Entry 6). The general procedure, with 0.02 mmol (7.9 mg) of ligand **2**, 1.0 mmol (127 mL) of 2-iodotoluene, and 1.2 mmol (148 mg) of *p*-

anisidine in dioxane at 45 °C, yielded 211 mg (99%) of a yellow-white solid: mp 78-81 °C; ^{1}H NMR (500 MHz, $CDCl_3$) 7.17-7.16 (m, 1H), 7.11-7.06 (m, 1H), 7.05-7.01 (m, 3H), 6.89-6.86 (m, 2H), 6.84-6.81 (td, 1H, J = 7.3 Hz, 1.2 Hz), 5.23 (br s, 1 H), 3.82 (s, 3H), 2.27 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) 155.2, 143.4, 136.3, 130.8, 126.9, 125.3, 122.3, 120.1, 115.2, 114.8, 55.8, 18.1; IR (neat, cm^{-1}) 3394, 2968, 2927, 2834, 1584, 1505, 1478, 1468, 1453, 1312, 1293, 1237. Anal. Calcd for C14H15NO: C, 78.84; H, 7.09. Found: C, 79.04; H, 7.13.

***N*-(3,5-Xylyl)-*N*-ethylaniline (Table 2, Entry 7).**[45] The general procedure, with 0.02 mmol (7.9 mg) of ligand **2**, 1.0 mmol (144 mL) of 5-iodo-*m*-xylene, and 1.05 mmol (132 mL) of *N*-ethylaniline in THF at rt, yielded 226 mg (>99%) of an orange oil.

***N*-Methyl-*N*-phenyl-4-chloroaniline (Table 2, Entry 8).**[24] The general procedure, with 0.02 mmol (7.9 mg) of ligand **2**, 1.0 mmol (238 mg) of 1-chloro-4-iodobenzene, and 1.0 mmol (108 mL) of *N*-methylaniline in THF at rt yielded 211 mg (97%) of a yellow oil.

***N*-(3-Methylphenyl)diphenylamine (Table 2, Entry 9).**[55] The general procedure, with 0.02 mmol (7.9 mg) of ligand **2**, 1.0 mmol (128 mL) of 3-iodotoluene, and 1.05 mmol (178 mg) of diphenylamine in 0.5 M dioxane at 80 °C, yielded 254 mg (98%) of a white solid: mp 62-65 °C (lit.55 mp 66-67 °C); ^{1}H NMR (500 MHz, $CDCl_3$) 7.25-7.22 (m, 4H), 7.15-7.12 (m, 1H), 7.09-7.06 (m, 4H), 7.01-6.98 (m, 2H), 6.90 (m, 1H), 6.89-6.87 (m, 1H), 6.84-6.83 (m, 1H), 2.26 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) 148.0, 147.8, 139.2, 129.3, 129.1, 125.1, 124.2, 123.8, 122.6, 121.7, 21.7.

***N*-(2,5-Xylyl)hexylamine (Table 3, Entry 1).**[24] The general procedure, with 0.0075 mmol (6.9 mg) of $Pd_2(dba)_3$, 0.03 mmol (11.8 mg) of ligand **2**, 0.5 mmol of 2-iodo-*p*-xylene, and 0.7 mmol

(92 mL) of *n*-hexylamine in 1.5 mL of 2:1 dioxane/*t*-BuOH at 120 °C, yielded 81 mg (79%) of a yellow oil.

***N*-(3-Carbethoxyphenyl)-1,4-dioxa-8-azaspiro[4.5]decane (Table 3, Entry 2). The** general procedure, with 0.02 mmol (7.9 mg) of ligand **2**, 1.0 mmol (165 mL) of ethyl 3-iodobenzoate, and 1.2 mmol (154 mL) of 1,4-dioxa-8-azasprio[4.5]decane in 2.0 mL of 1:1 dioxane/*t*-BuOH at 100 °C, yielded 237 mg (81%) of a yellow oil: ^{1}H NMR (300 MHz, $CDCl_3$) 7.61 (m, 1H), 7.51-7.48 (dd, 1H, J = 7.6 Hz, 1.0 Hz), 7.32-7.26 (m, 1H), 7.13-7.09 (dt, 1H, J = 82 Hz, 1.3 Hz), 4.40-4.32 (qd, 2H, J = 7.1 Hz, 1.0 Hz), 4.00 (s, 3H), 3.37 (t, 4H, J = 5.5 Hz), 1.84 (t, 4H, J = 5.5 Hz), 1.41 (td, 3H, J = 7.1 Hz, 1.1 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) 167.1, 150.9, 131.4, 129.1, 120.9, 120.4, 117.4, 107.2, 64.6, 61.3, 47.8, 34.7, 14.7; IR (neat, cm^{-1}) 2960, 2933, 2885, 2834, 1713, 1600, 1580, 1492, 1443. Anal. Calcd for $C_{16}H_{21}NO_4$: C, 65.96; H, 7.27. Found: C, 66.12; H, 7.15.

***N*-(2-Methoxyphenyl)morpholine (Table 3, Entry 3).** The general procedure, using 0.01 mmol (9.2 mg) of $Pd_2(dba)_3$, 0.04 mmol (15.7 mg), 1.0 mmol (130 mL) of 2-iodoanisole, and 1.2 mmol (105 mL) of morpholine in 2.0 mL of 1:1 dioxane/*t*-BuOH at 120 °C, yielded 123 mg (64%) of a yellow oil.

***N*-(4-Benzoylphenyl)dipropylamine (Table 3, Entry 4).** The general procedure, with 0.025 mmol (11.4 mg) of $Pd_2(dba)_3$, 0.10 mmol (19.7 mg) of ligand **3**, 0.5 mmol (154 mg) of 4-iodobenzophenone, and 0.6 mmol (82 mL) of di(*n*-propyl)amine in 1.5 mL of 2:1 dioxane/Et_3N at 100 °C yielded 101 mg (72%) of a yellow solid: mp 96-99 °C; ^{1}H NMR (300 MHz, $CDCl_3$) 7.78-7.73 (m, 2H), 7.73-7.70 (m, 2H), 7.54-7.41 (m, 3H), 6.63-6.60 (m, 2H), 3.35-3.29 (t, 4H, J = 7.7 Hz), 1.70-1.60 (m, 4H), 0.98-0.93 (t, 6H, J = 7.4 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) 194.8, 151.6, 139.5, 133.1, 131.0, 129.5, 128.1, 124.0, 110.3, 53.0, 20.7, 11.7; IR (neat, cm^{-1}) 2960, 2931,

2873, 1636, 1584, 1574, 1542, 1524. Anal. Calcd for $C_{19}H_{23}NO$: C, 81.0; H, 8.24. Found: C, 80.81; H, 8.24.

3-Cyano-2'-carboethyoxydiphenylamine (Table 3, Entry 5). The general procedure, with 0.02 mmol (11.6 mg) of ligand **3**, 1.0 mmol (229 mg) of 3-iodobenzonitrile, and 1.05 mmol (155 mL) of ethyl 2-aminobenzoate in 2:1 dioxane/*t*-BuOH at 100 °C, yielded 242 mg (91%) of a white solid: mp 78-79 °C; ^{1}H NMR (300 MHz, $CDCl_3$) 9.64 (s, 1H), 8.04-8.01 (ddd, 1H, J = 8.0 Hz, 1.7 Hz, 0.5 Hz), 7.53-7.52 (m, 1H), 7.45-7.37 (m, 3H), 7.32-7.26 (m, 2H), 6.89-6.83 (td, 1H, J = 7.5 Hz, 1.2 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) 168.4, 146.0, 142.2, 134.2, 131.9, 130.4, 126.1, 125.4, 123.6, 119.0, 114.7, 113.9, 113.4, 61.2, 14.6; IR (neat, cm^{-1}) 3294, 3245, 3188, 2977, 2231, 1679, 1603, 1571, 1453. Anal. Calcd for $C_{16}H_{14}N_2O_2$: C, 72.17; H, 5.30. Found: C, 71.89; H, 5.43.

***N*-(4-Carbethoxyphenyl)-*N*-methylaniline (Table 3, Entry 6).** The general procedure, with 0.02 mmol (11.6 mg) of ligand **3**, 1.0 mmol (171 mL) of ethyl 4-iodobenzoate, and 1.0 mmol (108 mL) of *N*-methylaniline in 2:1 dioxane/Et_3N at 120 °C, yielded 234 mg (92%) of a yellow oil: ^{1}H NMR (300 MHz, $CDCl_3$) 7.89-7.84 (m, 2H), 7.41-7.35 (m, 2H), 7.22-7.16 (m, 3H), 6.79-6.74 (m, 2H), 4.35-4.28 (q, 2H, J = 7.2 Hz), 3.60 (s, 3H), 1.38-1.34 (t, 3H, J = 6.9 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) 166.8, 152.5, 147.6, 131.1, 129.9, 125.9, 125.4, 119.7, 114.0, 60.5, 40.5, 14.8; IR (neat, cm^{-1}) 3060, 2981, 2935, 2904, 2819, 1702, 1609, 1590, 1515, 1493, 1351. Anal. Calcd for $C_{16}H_{17}NO_2$: C, 75.27; H, 6.71. Found: C, 75.44; H, 6.89.

***N*-(3-Carbethoxyphenyl)-*N*-methylaniline (Table 3, Entry 7).** The general procedure, with 0.02 mmol of ligand **3** (11.6 mg), 1.0 mmol (165 mL) of ethyl 3-iodobenzoate, and 1.0 mmol (108 mL) of *N*-methylaniline in 2:1 dioxane/Et_3N at 120 °C, yielded 217 mg (85%) of a yellow oil: ^{1}H NMR (300 MHz, $CDCl_3$) 7.66 (app. t, 1H), 7.60-7.56 (m, 1H), 7.33-7.25 (m, 3H), 7.15-7.11 (ddd, 1H, J = 8.2 Hz, 2.5 Hz, 1.1 Hz), 7.08-7.00 (m, 3H), 4.36 (q, 2H, J = 7.1 Hz), 3.35 (s,

3H), 1.38 (t, 3H, J = 7.1 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) 166.9, 149.2, 148.7, 131.7, 129.6, 129.1, 123.7, 122.7, 122.1, 121.7, 119.7, 61.3, 40.7, 14.7; IR (neat, cm^{-1}) 3064, 3037, 2981, 2904, 2815, 1717, 1592, 1582, 1495, 1443. Anal. Calcd for $C_{16}H_{17}NO_2$: C, 75.27; H, 6.71. Found: C, 75.49; H, 6.68.

Acknowledgment

We are grateful to the National Institutes of Health (GM58160) for generous funding of this work and to Pfizer and Merck for additional unrestricted support. We thank Dr. Jingjun Yin for a gift of xantphos (ligand **3**) and Dr. Hiroshi Tomori for a sample of ligand **2**.

References

1. Waring, D. R.; Hallas, G. *The Chemistry and Application of Dyes*; Plenum: New York, 1990.
2. *The Merck Index*, 12th ed.; Budavari, S., Ed.; Merck Research Laboratories: Whitehouse Station, 1996.
3. Kleemann, A.; Engel, J.; Kutscher, B.; Reichert, D. *Pharmaceutical Substances*, 3rd ed.; Thieme: Stuttgart, 1999.
4. Negwer, M. *Organic-chemical drugs and their synonyms: (an international survey)*, 7th ed.; Akademie Verlag: Berlin, 1994.
5. *CRC Handbook of Pesticides*; Milne, G. W. A., Ed.; CRC Press: Boca Raton, 1994.
6. Harwood, E. A.; Sigurdsson, S. T.; Edfeldt, N. B. F.; Reid, B. R.; Hopkins, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 5081-5082.
7. Lakshman, M. K.; Keeler, J. C.; Hilmer, J. H.; Martin, J. Q. *J. Am. Chem. Soc.* **1999**, *121*, 6090-6091.
8. De Riccardis, F.; Johnson, F. *Org. Lett.* **2000**, *2*, 293-295.
9. Greco, G. E.; Popa, A. I.; Schrock, R. R. *Organomet.* **1998**, *17*, 5591-5593.
10. D'Aprano, G.; Leclerc, M.; Zotti, G.; Schiavon, G. *Chem. Mater.* **1995**, *7*, 33-42.
11. Ito, A.; Taniguchi, A.; Yamabe, T.; Tanaka, K. *Org. Lett.* **1999**, *1*, 741-743.
12. Belfield, A. J.; Brown, G. R.; Foubister, A. J. *Tetrahedron* **1999**, *55*, 11399-11428.
13. Lindley, J. *Tetrahedron* 1984, 40, 1433-1456.
14. Kosugi, M.; Kameyama, M.; Migita, T. *Chem. Lett.* **1983**, 927-928.
15. Guram, A. S.; Rennels, R. A.; Buchwald, S. L. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1348-1350.
16. Louie, J.; Hartwig, J. F. *Tetrahedron Lett.* **1995**, *36*, 3609-3612.
17. Yang, B. H.; Buchwald, S. L. *J. Organomet. Chem.* **1999**, *576*, 125-146.
18. Wolfe, J. P.; Wagaw, S.; Marcoux, J.-F.; Buchwald, S. L. *Acc. Chem. Res.* **1998**, *31*, 805-818.
19. Hartwig, J. F. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 2046-2067.
20. Larock, R. C. *Comprehensive Organic Transformations*, 2nd ed.; Wiley-VCH: New York, 1999.
21. Snieckus, V. *Chem. Rev.* **1990**, *90*, 879-933.
22. Frid, M.; Perez, D.; Peat, A. J.; Buchwald, S. L. *J. Am. Chem. Soc.* **1999**, *121*, 9469-9470.

23. Bolm, C.; Hildebrand, J. P. *J. Org. Chem.* 2000, 65, 169-175.
24. Wolfe, J. P.; Buchwald, S. L. *J. Org. Chem.* **1996**, *61*, 1133-1135.
25. Wolfe, J. P.; Wagaw, S.; Buchwald, S. L. *J. Am. Chem. Soc.* **1996**, *118*, 7215-7216.
26. Driver, M. S.; Hartwig, J. F. *J. Am. Chem. Soc.* **1996**, *118*, 7217-7218.
27. Wolfe, J. P.; Buchwald, S. L. *J. Org. Chem.* **1997**, *62*, 6066-6068.
28. Hamann, B. C.; Hartwig, J. F. *J. Am. Chem. Soc.* **1998**, *120*, 7369-7370.
29. Huang, J.; Grasa, G.; Nolan, S. P. *Org. Lett.* **1999**, *1*, 1307-1309.
30. Old, D. W.; Wolfe, J. P.; Buchwald, S. L. *J. Am. Chem. Soc.* **1998**, *120*, 9722-9723.
31. Wolfe, J. P.; Buchwald, S. L. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 2413-2416.
32. Kranenburg, M.; van der Burgt, Y. E. M.; Kamer, P. C. J.; van Leeuwen, P. W. N. M.; Goubitz, K.; Fraanje, J. *Organomet.* **1995**, *14*, 3081-3089.
33. Guari, Y.; van Es, D. S.; Reek, J. N. H.; Kamer, P. C. J.; van Leeuwen, P. W. N. M. *Tetrahedron Lett.* **1999**, *40*, 3789-3790.
34. Wagaw, S.; Yang, B. H.; Buchwald, S. L. *J. Am. Chem. Soc.* **1999**, *121*, 10251-10263.
35. Harris, M. C.; Geis, O.; Buchwald, S. L. *J. Org. Chem.* **1999**, *64*, 6019-6022.
36. Yang, B. H.; Buchwald, S. L. *Org. Lett.* **1999**, *1*, 35-37.
37. Yin, J.; Buchwald, S. L. *Org. Lett.* **2000**, *2*, 1101-1104.
38. Nishiyama, M.; Yamamoto, T.; Koie, Y. *Tetrahedron Lett.* **1998**, *39*, 617-620.
39. Wolfe, J. P.; Singer, R. A.; Yang, B. H.; Buchwald, S. L. *J. Am. Chem. Soc.* **1999**, *121*, 9550-9561.
40. Wolfe, J. P.; Tomori, H.; Sadighi, J. P.; Yin, J.; Buchwald, S. L. *J. Org. Chem.* **2000**, *65*, 1158-1174.
41. Aranyos, A.; Old, D. W.; Kiyomori, A.; Wolfe, J. P.; Sadighi, J. P.; Buchwald, S. L. *J. Am. Chem. Soc.* **1999**, *121*, 4369-4378.
42. Palucki, M.; Buchwald, S. L. *J. Am. Chem. Soc.* **1997**, *119*, 11108-11109.
43. Reddy, N. P.; Tanaka, M. *Tetrahedron Lett.* **1997**, *38*, 4807-4810.
44. Yamamoto, T.; Nishiyama, M.; Koie, Y. *Tetrahedron Lett.* **1998**, *39*, 2367-2370.
45. Marcoux, J.-F.; Wagaw, S.; Buchwald, S. L. *J. Org. Chem.* **1997**, *62*, 1568-1569.
46. Beller, M.; Riermeier, T. H.; Reisinger, C.-P.; Herrmann, W. A. *Tetrahedron Lett.* **1997**, *38*, 2073-2074.
47. Stauffer, S. R.; Lee, S.; Stambuli, J. P.; Hauck, S. I.; Hartwig, J. F. *Org. Lett.* **2000**, *2*, 1423-1426.
48. Harris, M. C.; Buchwald, S. L. Unpublished results.
49. Widenhoefer, R. A.; Buchwald, S. L. *Organometallics* **1996**, *15*, 3534-3542.
50. Wolfe, J. P.; Buchwald, S. L. *J. Org. Chem.* **2000**, *65*, 1144-1157.
51. Widenhoefer, R. A.; Buchwald, S. L. *Organometallics* **1996**, *15*, 2755-2763.
52. Mann, G.; Hartwig, J. F. *J. Am. Chem. Soc.* **1996**, *118*, 13109-13110.
53. Wolfe, J. P.; Buchwald, S. L. *Tetrahedron Lett.* **1997**, *38*, 6359-6362.
54. Tomori, H.; Fox, J. M.; Buchwald, S. L. *J. Org. Chem.* **2000**, *65*, 5334.
55. Gauthier, S.; Fréchet, J. M. J. *Synthesis* **1987**, 383-385.

MAYSSAM HANI ALI

EDUCATION

Massachusetts Institute of Technology
Ph.D. in Organic Chemistry, 2004.

Harvard College
A.B. in Chemistry, *magna cum laude*, 1998.

PROFESSIONAL EXPERIENCE

RESEARCH

Graduate Research Assistant Advisor: Professor Barbara Imperiali — *MIT*
Design and Structural Characterization of Oligomeric Beta Beta Alpha Peptides. — *2000-20004*

Graduate Research Assistant Advisor: Professor Stephen L. Buchwald — *MIT*
Development of Palladium-Catalyzed Amination Reaction of Aryl Iodides — *1998-2000*

Undergraduate Research Assistant Advisor: Professor Charles N. Serhan — *Harvard*
Study of Pharmacokinetic Properties of Lipoxins. — *1997*

Undergraduate Research Assistant Advisor: Professor Yoshito Kishi — *Harvard*
Preparation of Siderophores and Saccharides for Conformational Studies. — *1996-1997*

TEACHING

Incoming Teaching Assistant Trainer — *MIT*
Trained and mentored new chemistry department teaching assistants. — *2000-2002*

Teaching Assistant — *MIT*
Chemistry 5.451 (Graduate Bioorganic Chemistry) — *Fall 2000*

Head Teaching Assistant — *MIT*
Chemistry 5.13 (Organic Chemistry 2) — *Spring 1999*

Teaching Assistant — *MIT*
Chemistry 5.12 (Organic Chemistry 1) — *Fall 1998*

Laboratory Teaching Fellow — *Harvard*
Chemistry 27 (Organic Chemistry 2 for non-chemistry majors) — *Spring 1998*

Teaching Fellow — *Harvard*
Chemistry 30 (Organic Chemistry 2 for chemistry majors) — *Fall 1997*

Teaching Fellow — *Harvard*
Chemistry S-20 (Intensive Summer Organic Chemistry 1 & 2) — *Summer 1997*

AWARDS

Sigma Xi Honor Society — *Spring 2004*

Department of Chemistry Award for Outstanding Teaching by Graduate Students — *Fall 1999*

Phi Beta Kappa Honor Society — *Spring 1998*

Certificate of Distinction in Teaching — *Spring 1998*

Certificate of Distinction in Teaching — *Fall 1997*

MAYSSAM HANI ALI

ACTIVITIES

Member, American Chemical Society	*1998-present*
Member, MIT Chemistry Department Safety Committee	*2000-2002*
Member, Organic Chemistry Graduate Student Seminar Committee	*2001-2002*
President, MIT Women in Chemistry	*1999-2001*
Member, Women in Science at Harvard/Radcliffe	*1994-1998*

PUBLICATIONS AND PRESENTATIONS

X-ray Structure Analysis of an Oligomeric Mini-Protein Reveals a Compact Domain-Swapped Architecture. Ali, M.; Peisach, E.; Allen, K. N.; Imperiali, B. submitted for publication.

Design and Characterization of Oligomeric Mini-Protein Motifs. Ali, M. H.; McDonnell, K. A.; Mezo, A. R.; Imperiali, B. Abstract #38-BIOL, *224th National Meeting of the American Chemical Society,* Boston, MA, August, 2002.

Enantioselective Synthesis and Application of the Highly Fluorescent and Environment-Sensitive Amino Acid 6-(2-dimethylaminonaphthoyl) alanine (DANA). Nitz, M., Mezo, A. R., Ali, M. H.; Imperiali, B. *Chem. Commun.*, **2002**, *17*, 1912-1913.

An Improved Method for the Palladium-Catalyzed Amination of Aryl Iodides. Ali, M. H.; Buchwald, S. L. *J. Org. Chem.*, **2001**, *66*, 2560-2565.