

Symmetry Assembled Supramolecular Protein Cages: Investigating a Strategy for  
Constructing New Biomaterials.

by

Dustin P. Patterson

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Chemistry)  
in The University of Michigan  
2011

Doctoral Committee:

Professor E. Neil G. Marsh, Chair  
Professor Nils G. Walter  
Associate Professor Zhaohui Xu  
Assistant Professor Julie Biteen

## Dedication

I dedicate this dissertation to that which means the most to me—My Family. To my parents for their love, encouragement, and support throughout my life, to my wife for her love and patience, and to my daughter Trinity for the joy she brings to my life.

## Acknowledgements

I would like to thank all those who played a part in helping me along my journey in reaching this destination. First I would like to thank my advisor, Dr. Neil G. Marsh, for his patience and guidance and giving me the opportunity to work in his research group. I would also like to thank the members of the Marsh lab, both past and present, for their help, advice, insights, and friendships. There are also a number of people who have been instrumental in completing the work presented here. I would like to thank Dr. Manoj Cheriyon and Prof. Carol Fierke for the KDPG aldolase gene and technical assistance, Prof. Mark Banaszak Holl and members of his lab for access and help with AFM, HP-SEC, and SLS, Prof. Kumar Sinniah for AFM training, Mr. Ankur Desai performing the HP-SEC runs for me, MNIMBS for access to the HP-SEC, Prof. Zhaohui Xu and LSI for the use of the analytical ultracentrifuge, Dr. Titus Franzmann for his technical support and high resolution data analysis of my analytical ultracentrifugation data, and Prof. Yiorgo Skiniotis, Mr. Justin Schilling, and Dr. Min Su for their collaboration in obtaining cryo-EM images of the protein cages. I would also like to acknowledge with appreciation, the funding that I received during my studies, an NIH funded Chemistry Biology Interface Training Program fellowship and a dissertation fellowship from The University of Michigan Chemistry Department.

## Table of Contents

Dedication.....	.ii
Acknowledgements.....	.iii
List of Figures.....	.vi
List of Appendices.....	.ix
Abstract.....	.x
Chapter	
1. Introduction	
1.1 – Introduction.....	.1
1.2 – Natural Protein Cages.....	.4
1.3 – Protein Assembly.....	.8
1.4 – Peptide Assembly.....	.10
1.5 – DNA Assembly.....	.16
1.6 – Goals.....	.21
1.7 – References.....	.23
2. A Strategy for Assembling Protein Cages	
2.1 – Introduction.....	.28
2.2 – Cross-linking Proteins.....	.29
2.3 – Directing Assembly by Symmetry.....	.37
2.4 - Incorporating Flexibility Between the Protein Building Block and Peptide Linker.....	.40

2.5 – The Strategy.....	41
2.6 – Discussion.....	42
2.7 – References .....	43
3. Investigations of a Designed Coiled Coil to be Used for the Controlled Assembly of Protein Building Blocks	
3.1 – Introduction.....	46
3.2 – Design.....	48
3.3 – Experimental Procedures.....	50
3.4 – Results and Discussion.....	56
3.5 – Conclusions.....	72
3.6 – References.....	74
4. Proof of Concept: Investigations of Symmetry Directed Protein Cage Assembly of the Trimeric KDPG Aldolase	
4.1 – Introduction.....	82
4.2 – Experimental Procedures.....	84
4.3 – Results.....	95
4.4 – Discussion.....	124
4.5 – References.....	127
5. Conclusions and Future Directions.....	128
5.1– References.....	132
Appendices.....	134

## List of Figures

### Figure

1.1	Examples of closed and extended protein structures.....	3
1.2	Examples of natural protein cages being applied towards practical applications.....	7
1.3	Strategies for constructing higher order protein structures.....	10
1.4	Schematic representations and transmission electron microscopy (TEM) images of beta-sheet, beta-hairpin and alpha-helical peptide assemblies.....	13
1.5	Helical wheel depiction of a coiled coil.....	15
1.6	Assembly of DNA into cage structures.....	18
1.7	Assembly of DNA into extended structures.....	20
2.1	Covalent capture cross-linking methods.....	30
2.2	Schematic representation of chemical induced dimerization.....	32
2.3	Methods of linking proteins with DNA.....	34
2.4	Metal directed assembly of a protein.....	35
2.5	Protein subunit fusion assembly.....	36
2.6	Peptide directed assembly of proteins.....	37
2.7	Protein quaternary structure geometries and structures that could be constructed from their geometric shapes.....	39
2.8	Parallel vs. antiparallel coiled coil association for linking protein building blocks.....	41
2.9	Schematic representation for the assembly strategy of two protein geometries...	42
3.1	Helical wheel representation of a parallel dimeric coiled coil.....	47

3.2	Helical wheel representation and peptide sequences of the de novo designed peptides Helix(-)1 and Helix(+)1.....	50
3.3	HPLC chromatograms and ESI-MS spectra of Helix(-)1 and Helix(+)1 peptides.....	57
3.4	CD spectra and analysis of the designed peptides Helix(-)1, Helix(+)1, and their 1:1 mixture.....	59
3.5	Representative SE AUC data scan for Helix(-)1 with data fitted to a monomer-dimer equilibrium and residuals for the fit.....	61
3.6	Helical wheel representation and peptide sequences of the designed modified peptides Helix(-) and Helix(+).....	63
3.7	SDS-PAGE and ESI-MS data analysis for Helix(-) and Helix(+).....	64
3.8	CD spectra and analysis of the redesigned peptides Helix(-) and Helix(+).....	66
3.9	SE AUC data and fits with residuals for Helix(-).....	68
3.10	SE AUC data and fits with residuals for Helix(+).....	69
3.11	Denaturation of Helix(-) and determination of $\Delta G_{unfold}$ and m.....	71
4.1	KDPG aldolase crystal structure assembly model.....	84
4.2	SDS-PAGE analysis of KDPG aldolase and fusion protein constructs.....	97
4.3	Elution profiles and molecular weight analysis of aldolase and fusion proteins.	100
4.4	HP-SEC elution profiles and molecular weight analysis of 1:1 mixtures of A(-) and A(+). ....	101
4.5	Sedimentation analytical ultracentrifugation data and van Holde-Weischet analysis for KDPG aldolase.....	104
4.6	Sedimentation velocity analytical ultracentrifugation data and van Holde-Weischet analysis for A(-).....	105
4.7	Sedimentation velocity analytical ultracentrifugation data and van Holde-Weischet analysis for A(+). ....	106
4.8	Sedimentation velocity analytical ultracentrifugation data and van Holde-Weischet analysis for 1:1 A(-)/A(+) mix.....	107

4.9	SV AUC van Holde-Weischet distributions plot.....	109
4.10	Results of the 2DSA for A(-) samples.....	111
4.11	Results of the 2DSA for A(+) samples.....	112
4.12	Results of the 2DSA for A(-)/A(+) mixture samples.....	113
4.13	TEM images of A(-), A(+), and 1:1 A(-)/A(+) mix with schematic visual aides.....	115
4.14	Size exclusion chromatography separation of A(-)/A(+) mixture aggregates and TEM images of selected fractions.....	116
4.15	High resolution TEM images of SEC purified A(-)/A(+) mixture protein fractions.....	117
4.16	AFM images of protein samples. A) Two image perspectives from AFM imaging of A(-).....	120
4.17	AFM height distributions for images taken of A(-), A(+), and an A(-)/A(+) mixture (data for images in Figure 4.15).....	121
4.18	Atomic force microscopy imaging of KDPG aldolase and analysis.....	122
4.19	Results from cryo-electron microscopy of A(-)/A(+) mix samples separated by SEC.....	124

## List of Appendices

### Appendix

A. DNA and protein sequences of KDPG aldolase and fusion proteins.....	134
B. Cryo-EM image gallery.....	139
C. The Subunit Structure of Benzylsuccinate Synthase.....	141

## **ABSTRACT**

Symmetry Assembled Supramolecular Protein Cages: Investigating a Strategy for  
Constructing New Biomaterials.

by

Dustin P. Patterson

Chair: E. Neil G. Marsh

In nature, the assembly of individual protein subunits into larger quaternary structures allows new biological properties to emerge as consequence of the higher order structure. The biological function and physical properties exhibited by these higher order structures differ widely from those of the individual subunits of which they are composed. Inspired by nature, my research objective was to develop and test new strategies for creating novel biomaterials with useful properties by the directed assembly of natural proteins into larger artificial protein structures. In particular, my research has explored the assembly of proteins into “cages”, which have potential applications in drug delivery or gene therapy by encapsulating therapeutic agents within the “cage”. The strategy for making such higher order protein structures in a specific and directed manner is to use protein building blocks (PBBs) with well defined symmetrical quaternary

structures that can coordinate the assembly, and to utilize novel non-covalent cross-linking agents that bring the PBBs together and link them to form the “cage” structures. To demonstrate proof of concept, the protein KDPG aldolase from *T. maritima* was utilized as a PBB, which consists of three identical subunits that form a trimeric quaternary structure that closely resembles a triangle. Geometry dictates that triangles can form various highly symmetrical 3-D structures; for example tetrahedrons, octahedrons, and/or icosahedrons depending on the number of units in the assembly. As linkers, we exploited peptides designed *de novo* to form anti-parallel heterodimeric coiled coils. This method of cross-linking is useful because the heterodimeric nature allows only subunits with complementary peptides to be joined, providing control over assembly and maintaining adequate spatial arrangement of the PBBs. The results from our investigations show that fusion proteins of KDPG aldolase and peptide linkers assemble into cage structures, cage structures predicted to form based on the symmetry of the KDPG aldolase PBB, upon mixing of fusion proteins with complementary peptide linkers.

## Chapter 1

### Introduction

#### 1.1 - Introduction

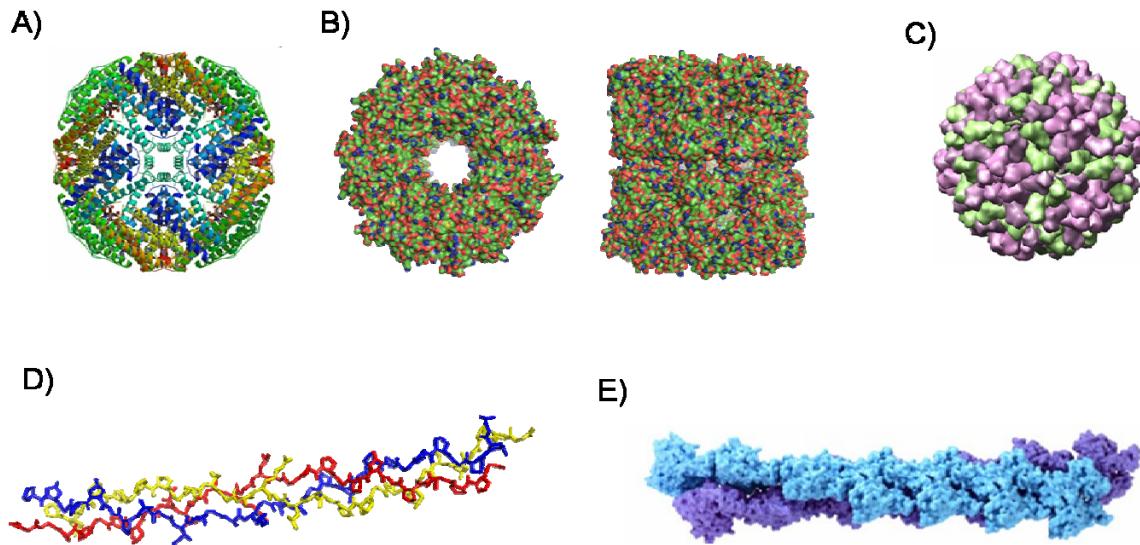
Nature utilizes what might be described as a modular “bottom-up” approach for constructing the essential materials of life. The primary components in this case are molecules such as amino acids, nucleic acids, or lipids that are assembled into higher order structures to form the functionally and structurally complex macromolecular modules from which cells are built. Biological function arises in many cases from the assembly of individual macromolecular modules into large complexes. Understanding the fundamental principles controlling assembly of biomacromolecules, in particular proteins, into defined higher order structures is an important question in structural biology. Furthermore, the ability to design and control higher order structures has important applications in the areas of synthetic biology and materials science. For example, the ability to assemble enzymes into complexes may lead to the design of highly efficient metabolic pathways that produce useful compounds. The field of nanotechnology is faced with major challenges in constructing smaller structures, especially for applications in electrical device fabrication. Natural biomaterials readily form nanometer-scale structures with well-defined architectures and biological molecules are promising candidates to incorporate into responsive “smart” materials.

Proteins represent the most structurally and functionally diverse group of all biomacromolecules utilized by nature. The complex architecture of proteins makes the

prediction of structure based on amino acid sequence extremely challenging. The problem of designing a sequence that will adopt a specified structure is correspondingly difficult. This problem requires an understanding of how multiple weak non-covalent interactions are coordinated based on amino acid sequence. Computational methods have been developed that can predict protein structure with some success and the *de novo* design and redesign of proteins has met some success.<sup>1-3</sup> However, the protein folding problem is still far from solved. Adding to the complexity is the fact that protein subunits often associate to form higher order quaternary structures that constitute their functional states.

Relatively few proteins exist as monomeric species, but instead associate with other protein subunits to form multimeric complexes, known as the quaternary structure. Often, quaternary structures of proteins are composed from multiple copies of only a single subunit or a few different subunits. In these cases, assembly results in structures that exhibit a high degree of symmetry. The function of proteins often only emerges after they assemble into higher order quaternary structures. Multimeric protein structures comprised of regular subunit repeats can be divided into two general categories, either closed (cage) or extended (filamentous). Closed structures, referred to as cages henceforth, function as containers that encapsulate biological cargo for protection, storage or transport within the cellular environment. Examples of multimeric proteins with cage structures include icosahedral viral capsids, ferritins, the chaperone GroEL/ES complex, the pyruvate dehydrogenase core, clathrin, vault proteins, and some heat shock proteins. Extended protein structures generally make up structural or mechanical components of cellular systems. Examples of multimeric proteins exhibiting extended

structures include the protein actin, collagen, fibrin, and tubulin. Depictions of different closed and extended protein structures are provided in Figure 1.1.



**Figure 1.1.** Examples of closed and extended protein structures. A) Ferritin (PDB ID: 3GVY), B) GroEL (top and side views, PDB ID: 3E76), C) Cowpea chlorotic mottle virus (PDB ID: 1ZA7), D) Collagen (PDB ID: 1BKV), and E) Actin (taken from <http://ghr.nlm.nih.gov/handbook/illustrations/actin>).

The functional and structural diversity of proteins present them as potentially useful building blocks for constructing new biomaterials. The structures of a large number of proteins are now known in atomic detail and proteins with many different quaternary structures representing different symmetries are all commonly observed. The ever-expanding large repository of protein structures provides a “Lego” or “Erector” set from which a vast number of protein architectures can potentially be constructed. In addition, proteins are known to be stimuli responsive, that is, they respond to changes in the physical and chemical environment, which could be exploited for making responsive “smart” materials. Changes in pH, temperature, the presence of metals or other ligands,

are among a number of factors that can affect the oligomeric state, enzymatic activity, and conformational properties of proteins. Furthermore, proteins from organisms that live in extreme environments, such as in areas of high temperature (hot springs, deep sea vents, etc.), could be used to construct robust materials resistant to harsh conditions.

Strategies for assembling proteins into higher order multimeric assemblies will provide an important tool for constructing protein-based biomaterials. Research aimed at designing protein assemblies could provide a greater understanding of factors controlling oligomeric state and overall structure formation, topics of great importance in structural biology and biochemistry. A discussion of progress made in assembling proteins into ordered structures is provided below. In addition, discussions of assembling systems made from peptides and DNA are also included as the strategies used to assemble those systems and the goals of this research are in many ways comparable. The assembly of lipids into closed structures such as micelles and liposomes has been an area of intense study for many decades.<sup>4</sup> However, their assembly and the structures they are able to form are quite different from those of proteins, peptides, and DNA; therefore they will not be discussed further. Although examples of both extended and cage structure assembly are highlighted, the major focus of this dissertation is on the assembly of protein cages. As such, a discussion on natural protein cages, their unique properties, and research in using them toward practical applications will be presented first. In closing, I will lay out the goals of my research in directing protein cage assembly.

## 1.2 - Natural Protein Cages

Compartmentalization is an important feature of biological systems as biological components are often not mutually compatible when confined to the same area or space.

For example, proteases degrade proteins and are essential for food digestion and important for signaling pathways (i.e. blood clotting), however they have the ability to destroy other proteins essential to normal biological function. In addition, protection against changing environmental conditions pose a potential risk to biological components, such as proteins and DNA, which function under limited conditions and/or are damaged if conditions are too harsh. To circumvent such problems, Nature has evolved structures that separate important biological materials and protect them from elements that could compromise their essential function. Protein cages are an important class of “biological containers” which have evolved to separate, protect, and store important biological cargoes.

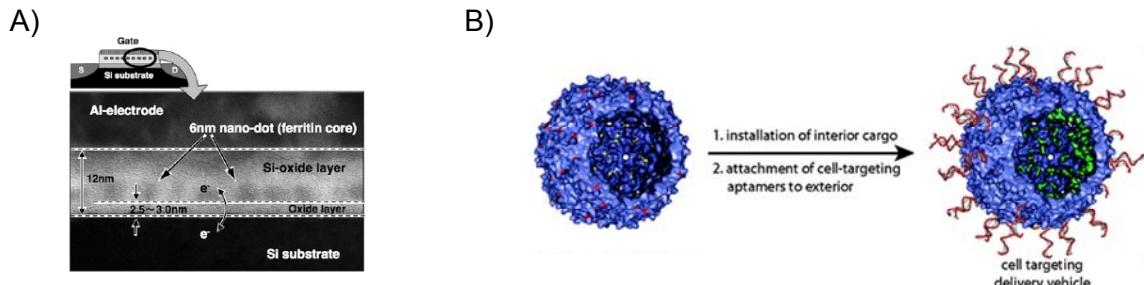
Protein cages have numerous functions in nature including the storage, protection and delivery of viral genomes, metal storage, and protein refolding. Protein cages are multimeric protein assemblies that form a protein shell about a hollow interior cavity in which important biological molecules can be sequestered. Due to their significance in pathology and disease, for example viral infection, protein cages have been extensively studied. More recently, protein cage assembly has especially attracted the attention of researchers in nanotechnology who are trying to synthetically construct materials of similar dimensions and architectures and find inspiration in natural structures.<sup>5</sup>

An innovative application of protein cages has been their use as size constrained reaction chambers for synthesizing inorganic and polymeric nanoparticles of homogenous size and composition.<sup>5-7</sup> An array of “spherical” protein cages, derived mainly from viral capsids, ferritins, and chaperonins/heat shock proteins, has been characterized with sizes ranging from 9-31 nm in diameter that provide a library of

templates from which various sized particles can be synthesized.<sup>5</sup> Reaction reagents are entrapped inside the protein cage interior cavity either through disassembly of the protein cage and then reassembly in the presence of the reagents, or through a gating method where pores in the cage can be opened and closed allowing entry. These applications exploit the natural ability of proteins such as ferritins to biomimic mineralize inorganic salts; the interiors of the cage contain coordination sites through which nucleation can occur. Protein cages have also been genetically modified to provide nucleation sites by either mutagenesis to yield new nucleation sites for coordination or by adding genes encoding metal-nucleating peptides that enable nucleation of targeted metals.<sup>5,8</sup> In addition to making particles of specific size, the protein-nanoparticles often have unique properties not observed for nanoparticles synthesized by other means. Varpness et al. synthesized Pt nanoparticles on the interior of the small heat-shock protein from *M. jannaschii* that showed significantly higher activity for the reduction of H<sup>+</sup> to H<sub>2</sub> than alternatively synthesized particles.<sup>9</sup> Other examples include protein-inorganic particles that exhibit unique magnetic properties, which have potential in MRI as contrast agents, and use of ferritin to construct nanoelectrical devices (Figure 1.2A).<sup>10-16</sup>

Protein cages have been used to encapsulate a host of biological and non-biological species including proteins, RNA, DNA, drug molecules, preformed inorganic nanoparticles (e.g. gold nanoparticles and quantum dots), synthetic polymers, MRI contrast agents, and nanoemulsion droplets.<sup>5,17-23</sup> Encapsulation of proteins in the confines of protein cage interiors provides a way to study protein activity under crowded conditions with the aim of mimicking the conditions proteins experience in cellular environments.<sup>17</sup> Results from these studies may provide insights into effects that the

cellular environment have on protein activity and function. Protein cages have been more widely studied for use as potential delivery systems for therapeutic reagents.<sup>24</sup> The natural function of viral capsids as agents for packaging, protecting, and delivering DNA and RNA to cells, has made them attractive protein cages for study as delivery agents of therapeutic cargoes. Researchers have incorporated drug molecules into viral capsid interiors and observed the time dependant release of the drug.<sup>25</sup> In addition, exterior surfaces of viral capsids have been modified to display a number of ligands including antibodies, proteins, peptides, DNA, MRI contrast agents, PEG, fluorophores, redox-active moieties, gold nanoparticles, and quantum dots.<sup>5,26-29</sup> Researchers have shown selective targeting of protein cages affixed with cell targeting moieties *in vitro* (Figure 1.2B).<sup>25,30,31</sup>



**Figure 1.2.** Examples of natural protein cages being applied towards practical applications. A) TEM image and schematics showing a floating gate transistor derived from ferritin nanoparticles. The floating gate transistor image was taken from reference 5. B) MS2 capsid functionalized for targeted drug delivery. DNA aptamers were conjugated to the outside surface for targeting and drug molecules were attached to internal residues. The schematic of the functionalized capsid was taken from reference 30.

These examples of how natural protein cages are being exploited illustrate their importance as potential solutions to problems faced in nanotechnology. They provide the impetus to the design of synthetic protein cages from protein building blocks. These

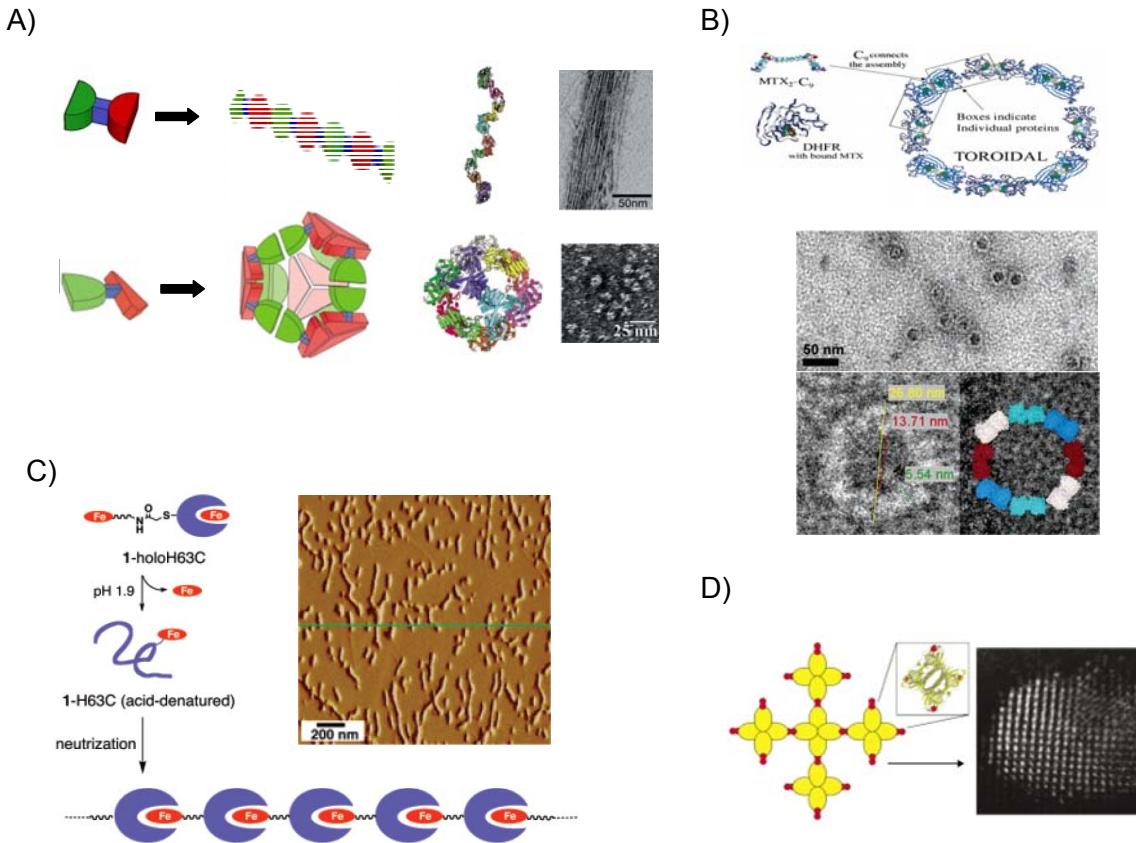
could provide further advances in nanotechnology, both in understanding principles of assembly and for making cages with new functionalities.

### 1.3 - Protein Assembly

Research in directing the assembly of proteins into higher order structures has been relatively limited. A study by Padilla et al. provides one of the few examples for constructing cage structures from proteins.<sup>32</sup> Their study described a basic strategy for constructing protein cages, 2-D layers, crystals, and filaments by exploiting the symmetry of the quaternary structures of proteins whose subunits were utilized in constructing the modular protein building blocks. In this strategy structural architecture would be determined by the angle generated between two protein subunits (fusion domains) linked by a rigid, helical peptide segment. Extended networks would result from a 180° angle between fusion domains, whereas cage structures would result by constraining the angles to less than 180° to generate curvature for cage formation whose symmetry would depend upon the quaternary structures formed by the fusion domains. As proof of concept, they constructed two fusion proteins: a fusion of a subunit from a dimeric protein and trimeric protein that they predicted would assemble into a tetrahedral cage; and a fusion protein of subunits from two dimeric proteins that they predicted would form filaments (Figure 1.3A). Transmission electron micrographs showed that indeed, extended filaments formed from the fusion protein containing the subunits from the dimeric proteins, whereas micrographs showed morphologies and sizes consistent with a cage structure for fusion proteins constructed from the subunits of the trimeric and dimeric proteins. Light scattering and analytical ultracentrifugation data also supported the formation of cages.<sup>32</sup>

Few studies have explored anything other than assembly of extended networks of proteins. However, Carlson et al. provided insight into the construction of ring structures from proteins. In their study a set of dihydrofolate reductase fusion proteins with varying inter-domain linker lengths were constructed and various bivalent methotrexate analogs were used as chemical inducers of protein dimerization to promote assembly of the fusion proteins. Depending on the length of the inter-domain linker and the length of the bivalent methotrexate analog used, rings of proteins could be constructed (Figure 1.3B).<sup>33</sup> Kitagashi et al. linked modified cytochrome b<sub>562</sub> (Cyt b<sub>562</sub>) proteins together to form chain structures of varying lengths. A surface exposed cysteine placed opposite the heme binding pocket of Cyt b<sub>562</sub> was used to conjugate a modified heme group to the surface of the protein. The surface exposed heme allowed linking of proteins into chains upon binding of the heme in the heme binding site of surface modified apo- Cyt b<sub>562</sub> (Figure 1.3C).<sup>34</sup> More elaborately, Doten et al. constructed tetrahedral 3D protein arrays using the tetrameric lectin concanavalin A and a bivalent carbohydrate ligand (Figure 1.3D).<sup>35</sup> More recently, Radford et al. have described a system for assembling porous crystalline frameworks from engineered metal binding proteins.<sup>36</sup> A few other examples of assembling proteins into networks or small clusters exist, but the overall body of work exploring assembly of proteins into higher order structures is dwarfed by assembly systems using other biologic materials.<sup>37-43</sup>

Notably none of the design strategies that have appeared in the literature have been applied to other protein systems (i.e. alternative protein building blocks), suggesting that they are not generalizable and/or lack general utility.



**Figure 1.3.** Strategies for constructing higher order protein structures. A) Schematics showing the strategy proposed by Padilla et al. for assembling filaments of proteins (top) or protein cages (bottom). Next to each assembly schematic are crystal structure models and TEM images of the experimental results. The images were taken from reference 32. B) Assembly of ring structures using a bivalent substrate analog ( $\text{MTX}_2\text{-C}_9$ ) to link the dimeric dihydrofolate reductase. TEM images show ring structures and docking of the protein crystal structure into the ring image structure. Images were taken from reference 33. C) Linking of hemoproteins through a surface displayed heme to create protein chains. An Experimental AFM image of the resulting hemoprotein chains is shown at the right. The images were taken from reference 34. D) Construction of 3D arrays from a tetrameric carbohydrate binding protein using a bivalent ligand. Tetrameric protein ribbon structure provided in inset along with experimental TEM image of resulting protein structure. The images were taken from reference 35.

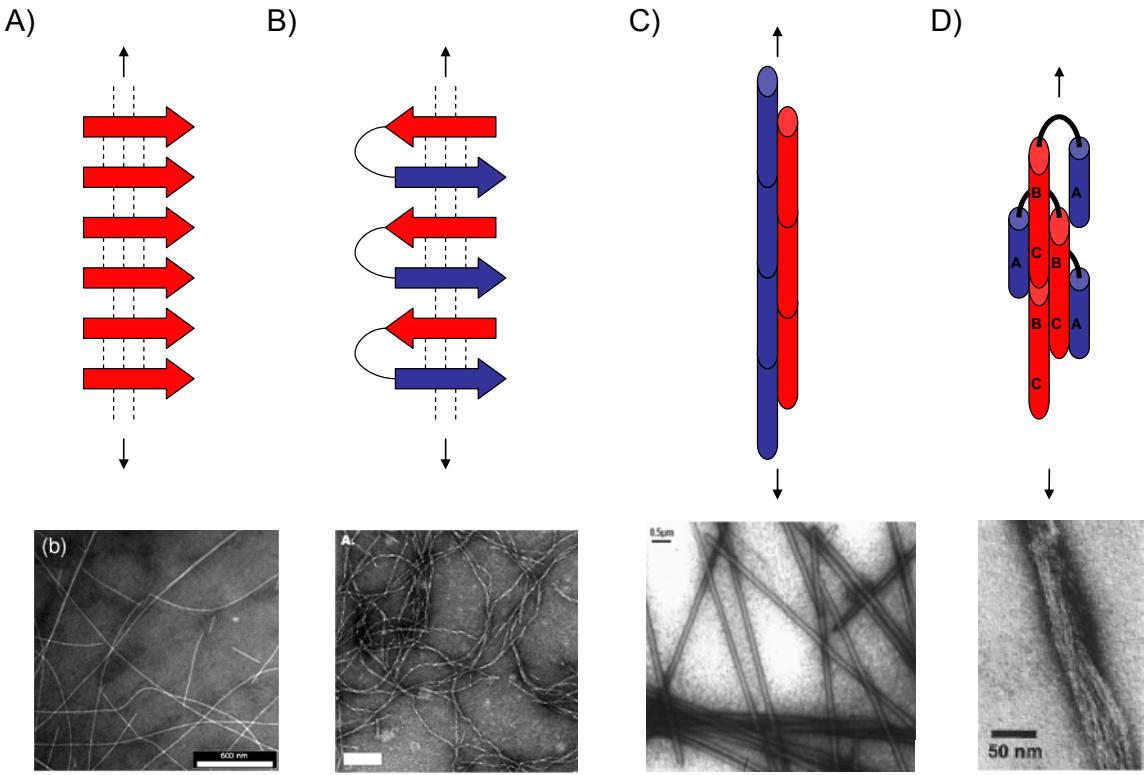
#### 1.4 - Peptide Assembly

Research in designing peptides that assemble into higher order structures is well established. Strategies for assembly exploit peptide sequences that form beta-sheet or alpha-helical secondary structures upon contact with complementary peptides. This

drives the assembly of multiple peptides to “weave” extended arrays of peptide fibers. Peptide systems described to date exclusively assemble into extended structures, a product of the rigid linear structure of beta-sheets and alpha-helices. In addition, peptides have been shown to assemble into peptide nanotubes.<sup>44</sup> Furthermore, there are classes of modified peptides incorporating non-natural amino acid moieties, such as peptide amphiphiles, cyclic peptides, etc., that can assemble into unique rod structures.<sup>45</sup> Here the discussion will focus on assembly of peptides composed of natural amino acids.

Beta-sheet forming peptides have been the most thoroughly studied assembly system (Figure 1.4A). This is due in part to the importance of beta-sheet fibril formation in diseases such as Parkinson’s and Alzheimer’s. Beta-sheet forming peptides are characterized by a repeating motif of alternating hydrophilic and hydrophobic amino acids.<sup>46</sup> The alternating pattern gives rise to a hydrophobic and hydrophilic face of the peptide, causing assembly of sheets through burial of the hydrophobic faces between sheets in aqueous environments. Typically ionizable amino acids are incorporated in alternating pattern, e.g. (IKIE)<sub>x</sub>, to provide complementary charges between peptides that help in directing assembly.<sup>47</sup> Beta-sheet peptides form a number of different hierarchical structures including tapes, ribbons, fibrils and fibres. Distinctions between classes are based on the number of sheets that pack together to form the final structure.<sup>48</sup> The formation of higher order structures is particularly sensitive to the concentrations of peptide and salt. Higher peptide concentrations yield assembly of larger structures, such as the fibrils or fibres. Higher salt concentrations mask charge-charge interactions allowing random associations by peptides. Other key factors effecting beta-sheet peptide assembly include hydrophobicity and peptide length which effect the critical

concentration for assembly.<sup>47</sup> Peptides as short as two amino acids can assemble into beta-sheet structures. For example, Reches and Gazit showed that phenylalanine dipeptides, Phe-Phe is an important sequence motif in beta-amyloid polypeptide, assembles into nanotubes.<sup>44</sup> An inherent problem in the design of beta-sheet peptides is the ability of the peptides to align out of register, hindering control over structure formation. Peptides that form beta-hairpins, which consist of two short beta-sheets connected through a turn (Figure 1.4B), can be designed to prevent out of register alignment.<sup>49-52</sup> However in this case, hydrogen bonding associated with beta-sheet structure make control of assembly difficult.

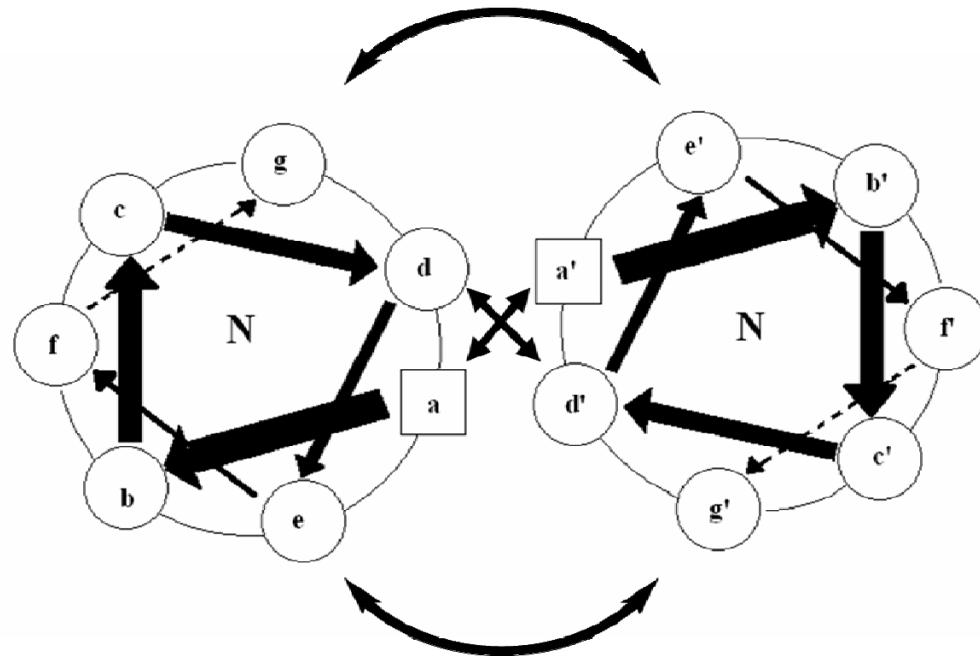


**Figure 1.4.** Schematic representations and transmission electron microscopy (TEM) images of beta-sheet, beta-hairpin and alpha-helical peptide assemblies. A) Beta-sheet peptide assembly (TEM image from reference 61). The scale bar (white) represents 500 nm. B) Beta-hairpin peptide assembly (TEM image from reference 49). The scale bar (white) represents 100 nm. C) Alpha-helical SAF system (TEM image from reference 55). The scale bar (top, left corner) represents 500 nm. D) Domain swap assembly (TEM image from reference 59). The scale bar (black) represents 50 nm.

Alpha-helical peptide systems provide a much higher degree of control over assembly than beta-sheet systems. Backbone hydrogen bonding requirements are largely contained within alpha-helix structures and are unavailable for higher order interactions. Assembly of alpha-helices is directed by side chain interactions that are best understood for coiled-coils. The coiled-coil is a structural motif characterized by a heptad repeat, denoted **abcdefg** by convention, of amino acids whereby two or more helices wrap around one another in a left-handed manner to generate a superhelix (Figure 1.5). The **a** and **d** positions consist of hydrophobic amino acids (L, I, and V) that associate between

peptide chains to generate a hydrophobic core that generates the helix-helix structure. Amino acids in the **e** and **g** positions are generally ionic (E, K, D, and R) and participate in Coulombic interactions at the peptide interfaces. Residues in the **b**, **c**, and **f** positions are solvent exposed and therefore are hydrophilic residues or small amino acids (A) that promote helical structure.<sup>53</sup> In general, extended structures are formed from peptides associating in staggered alignments extending the structure linearly in two directions. Kojima et al. described the first alpha-helical based peptide assembly of fibers from a peptide of sequence (LETLAKA)<sub>3</sub>, which formed fibrils 5-10 nm thick and several microns long.<sup>54</sup> Other single-component, alpha-helical systems have been described with alternative designs.<sup>55</sup> A major breakthrough in this area has been by the work of Woolfson and coworkers who developed the first rationally designed heterodimeric, sticky ended, coiled coil building block for the assembly of fibrous biomaterials, which they term self-assembling fiber (SAF) systems (Figure 1.4C).<sup>56</sup> The heterodimeric design allows assembly only when complementary peptides are present, providing control over assembly not seen for single-component systems. The SAF system has been extensively studied and modified to control thickness and stability of peptide fibers.<sup>57</sup> In addition, Woolfson has created several three component SAF systems that enable branch points, kinks, and other features to be incorporated for altering fiber morphology.<sup>58</sup> Oghara et al. described a slightly different approach through domain swapping.<sup>59</sup> In their strategy a single polypeptide that formed an intramolecular three helix bundle through a helix-loop-helix-loop-helix motif was redesigned so that it no longer contained the final loop connecting helix B and helix C. Deletion of the loop caused projection of helix C away from and parallel to the bundle core. Assembly occurred through the intermolecular

association of the extended helix C domain of one peptide with helix A and B of another peptide to reform a three helix bundle, which would propagate to form fibers (Figure 1.4D).<sup>59</sup>



**Figure 1.5.** Helical wheel depiction of a coiled coil. Amino acids in the **a** and **d** positions are hydrophobic residues, in general, that pack next to each other in a “knobs in holes” fashion to form a hydrophobic core. Residues in the **e** and **g** positions typically contain charged side chains (eg. K, E, R) that charge pair to form inter-chain salt bridges. Prime symbols are used to indicate different peptide strands.

Peptides appear to be well suited for the construction of fibrous materials. They are readily synthesized and as noted above, a number of strategies for assembling them into regular structures have been developed. In addition, a variety of moieties have been added to fiber surfaces of both beta-sheet and alpha-helical systems including: fluorophores, metal clusters, functional moieties, and proteins.<sup>45,60-62</sup> A beta-ribbon structure was used to encapsulate guest molecules and shown to provide intracellular delivery of the molecules.<sup>63</sup> Nanotubes assembled from the F-F dipeptides were used to synthesize nanowires of defined sizes.<sup>44</sup> In addition, a beta-sheet has been used as a light

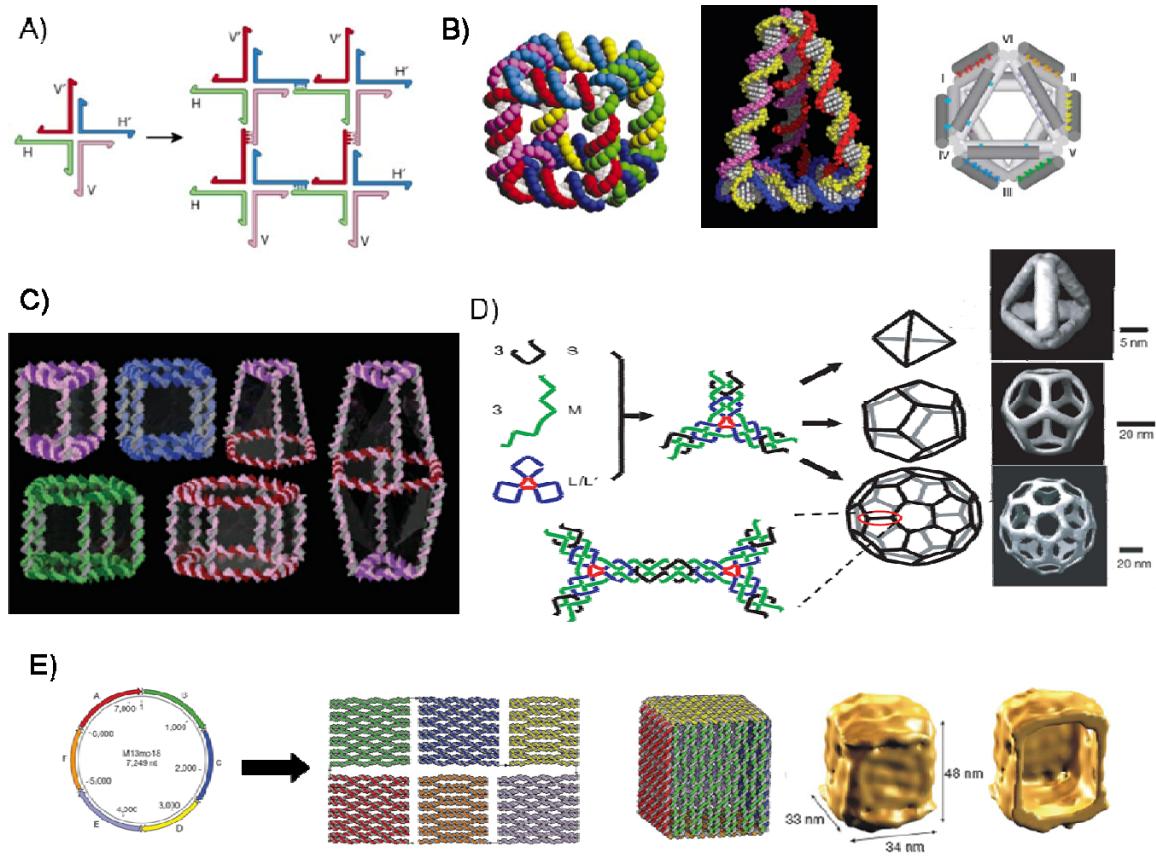
harvesting scaffold.<sup>62</sup> A number of other applications-based studies for peptides have been described. However, peptides are limited to mainly extended structures and are not suited for constructing cages.

### 1.5 - DNA Assembly

Although DNA does not possess the functional abilities of proteins, it has proved to be a surprisingly versatile medium for assembling higher order structures. The success of DNA for constructing well-defined higher order architectures can be attributed to the simple rules for invoking specific associations through Watson-Crick base pairing and by the limited structural conformations that DNA forms. In addition, discovery of branched “junction” DNA structures *in vivo*, primarily Holliday junctions, and subsequent studies in constructing such structures *in vitro* opened the door for constructing 2D and 3D structures from DNA.<sup>64-66</sup> Synthetic DNA complexes having fixed junctions containing between 3 and at least 8 arms have been described, giving researchers access to constructing a wide range of 2D and 3D structures. The basic strategy for constructing higher order architectures is to synthesize “junction point” DNA complexes containing “sticky ends” at the termini of the branched arms which can base pair to assemble the structure (Figure 1.6A). The “sticky ends” are used to control the mode of assembly by encoding unique sequences at each branch terminus. The number of branches and the flexibility of the DNA dictate the structure formed. Stiffer motifs that restrict assembly to linear assembly are required for extended structures, whereas more flexible motifs are required for constructing cage structures.

In 1991, Seeman described the assembly of the first three-dimensional structure constructed from DNA, a cube (Figure 1.6B left).<sup>67</sup> Seeman later described the assembly

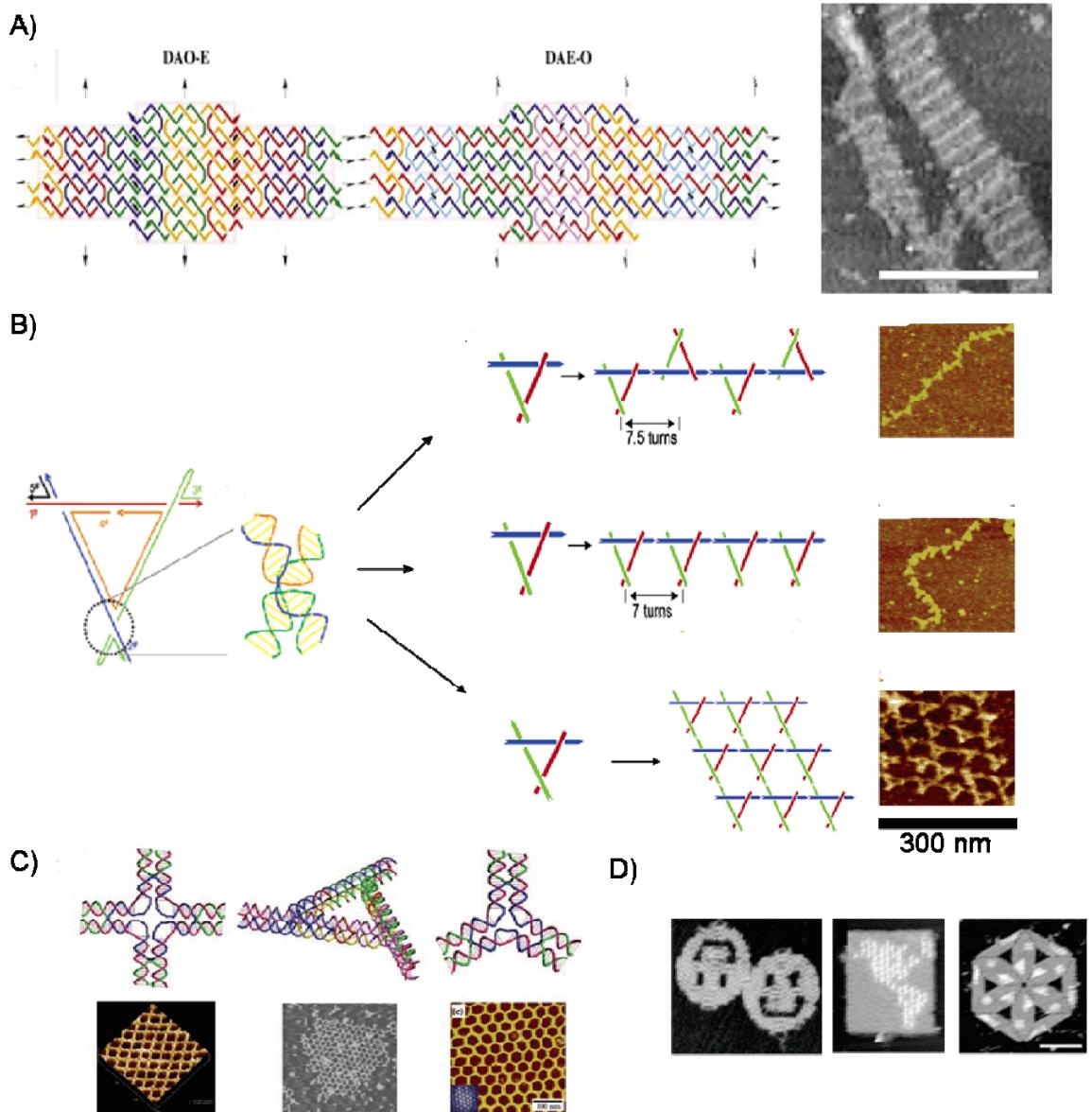
of a truncated octahedron, modifying the strategy used in cube assembly to increase product yield.<sup>68</sup> Almost a decade later Shih et al. described the construction of an octahedron from a DNA molecule that was readily replicated by polymerase (Figure 1.6B right).<sup>69</sup> This study provided the first visual images of a DNA cage structure using cryo-transmission electron microscopy (TEM). Assembly of tetrahedron was first shown by Goodman et al., who later demonstrated that cytochrome c could be encapsulated within the interior of the tetrahedron (Figure 1.6B center).<sup>70,71</sup> Work by Mao and coworkers demonstrated the assembly of tetrahedron, dodecahedron, and icosadodecahedron from a minimal set of DNA building blocks (Figure 1.6C).<sup>72</sup> By controlling annealing temperatures and concentrations of the components they could bias the assembly towards one structure or another, a great simplification over previous studies. By breaking down 3D DNA cage structures into 2D shape components, Sleiman showed that a large number of DNA cage structures could be constructed by joining synthesized DNA molecules of the specific 2D shapes (Figure 1.6D). Triangular prisms, a cube, a pentameric and hexameric prism, a heteroprism, and a biprism were all constructed in Sleiman's study, which also showed that the dimensions of the structures could be modified by strand switching.<sup>73</sup> A new and novel approach for constructing DNA cages utilizes DNA origami. DNA origami is the folding of a single strand of viral DNA into unique 2D and 3D shapes using short strands of DNA that base pair with the viral DNA at specified positions to fold and "staple" the viral DNA into the desired structure. Recently, Anderson et al. constructed a DNA box from the genome of M13 bacteriophage, which is openable by externally supplied DNA keys (Figure 1.6E).<sup>74</sup> A number of other examples of assembling DNA cages have also been reported.<sup>75-77</sup>



**Figure 1.6.** Assembly of DNA into cage structures. A) Schematic showing a Holliday junction DNA molecule and the general scheme for assembling higher order structures through associations of the junctions sticky ends. The schematic was taken from reference 80. B) Models depicting the cube, tetrahedron, and octahedron structures that have been constructed from DNA. The schematic was taken from reference 76. C) Structures constructed from 2D shape DNA building blocks. The schematic was taken from reference 76. D) Approach for constructing tetrahedron, octahedron, and bucky ball from the same small set of DNA building components. Cryo-EM reconstruction models obtained from the experiment are shown next to the schematic representations of each structure. The schematic was taken from reference 72. E) DNA origami construction of a DNA box from the M13 bacteriophage genome. Cryo-EM was used to verify the construction and reconstruction maps of the box and a cutaway map showing the hollow container are shown (right). The schematic was taken from reference 74.

Extended DNA structures were only fabricated several years after assembly of the DNA cube. The flexibility of the individual branched junctions hindered the formation of extended structures. Seeman solved this problem by constructing DNA double crossover molecules, which are less conformationally flexible.<sup>78</sup> DNA double crossover molecules

are DNA structures containing two Holliday junctions connected by two double helical arms. Extended 2-D DNA structures could be constructed using two DNA “tiles” whereby sticky ends on one DNA tile were encoded to associate with sticky ends of the other tile causing propagation of an extended structure. Images from atomic force microscopy (AFM) showed linear structures formed containing periodic elements with sizes and shapes constituent with the predicted model structure (Figure 1.7A).<sup>79</sup> Since Seeman’s ground breaking experiment, tiles of various types have been developed and used to construct numerous types of 1D and 2D extended DNA structures. 2D arrays with square, triangular, and hexagonal cavities have been constructed (Figure 1.7B & C). A thorough examination of the various tiles and the various structures constructed is available elsewhere.<sup>80-86</sup> DNA nanotubes have been constructed from helix bundles. In addition, scaffolded DNA origami has been used to construct depictions of smiley faces, North and South America, and other elaborate shapes (Figure 1.7D).<sup>86</sup>



**Figure 1.7.** Assembly of DNA into extended structures. A) Lattice topologies of the first DNA extended networks described by Seeman and coworkers. Double crossover units are highlighted by the grey rectangles and colors indicate different strand types. Arrowheads represent 3' end of strand. AFM of the extended network formed from the construct is shown at the right. The scale bar (white) represents 300 nm. Images were taken from reference 79. B) Strategy for constructing multiple triangular structures from limited set of DNA strands. Images were taken from reference 82. C) Samples of the broad range of extended 2D morphologies constructed from DNA. The scale bar in the right-most image (black, bottom right corner) represents 100 nm in that image only. D) Examples of DNA origami to construct images of smiley faces, North and South America, and a hexagonal star shape structure. The scale bar in the rightmost origami figure represents 100 nm. Images in C and D were taken from reference 86.

Methods for constructing nanoarchitectures, both closed and extended, using DNA are well advanced. Among the challenges faced in further advancing DNA nanotechnology are correcting errors that arise in DNA assembly and replication and scale-up of DNA nanostructures. Furthermore, to be useful, DNA nanostructures must be transformed into functional materials. Extended 2D structures are being investigated as scaffolds for arranging and displaying various ligands, which could be used to make functional surfaces.<sup>87</sup> Cages have been suggested as agents for drug delivery and studying enzymes in constrained areas. In addition, researchers are looking to construct DNA machines such as clocked walkers and molecular motors.<sup>88</sup> Advances in DNA nanotechnology continue to accumulate and the reality of DNA based nanomaterials for practical applications grows close.

### 1.6 - Goals

Methods for assembling higher order structures from peptides and DNA are well advanced and self-assembled biological materials are yielding very promising results for advancing nanotechnology. However, strategies and methodology for assembling proteins into higher order structures have been severely limited. The lack of methodology is somewhat stifling as proteins arguably hold greater potential for constructing new valuable materials due to their structural and functional prowess compared with other biomacromolecules. New strategies for assembling proteins into higher order structures must be developed to advance research in constructing protein-based materials.

Of the two categories of symmetrical multimeric structures, protein cage structures represent one of the most intriguing with respect to practical applications.

Natural protein cages are being utilized as size-constrained reaction chambers for synthesizing nanoparticles, as constrained areas mimicking cellular environments in enzymatic studies, and as delivery vessels for therapeutic agents. Rationally designed assemblies of proteins into cage structures could be used to construct cages with added functionality and be used to create new cage geometries not observed in natural protein cages. Protein cages endowed with catalytic function could be used as drug delivery vessels that activate their cargo upon deliver. In addition, new catalytic properties may arise through assembly of proteins into cage structures for example by providing protected micro-environments for molecules to be transformed by multiple activities. Furthermore, cages constructed with alternative geometries could be used for synthesizing nanoparticles with sizes and geometries different from those attainable through natural protein cages, providing nanoparticles with new properties. Directing the assembly of proteins into cage structures presents exciting possibilities for constructing new biomaterials with novel functions and further advancing nanotechnology.

The research presented here attempts to provide a foundation from which assembling of proteins into higher order cage structures can be advanced. The goals of the research discussed herein are as follows:

- 1) Develop a general strategy for assembling protein building blocks into higher order protein cage structures.
- 2) Design and characterize heterodimeric coiled coil peptides to be used as protein linking agents.
- 3) Show that the designed coiled coil peptides can link proteins together.

- 4) Show proof of concept for the assembly of protein cages using our strategy.

### 1.7– References

- 1 Havranek, J. J. Specificity in computational protein design. *J Biol Chem* **285**, 31095-31099 (2010).
- 2 Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H. & Meiler, J. Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry* **49**, 2987-2998 (2010).
- 3 Floudas, C. A., Fung, H. K., McAllister, S. R., Monnigmann, M. & Rajgaria, R. Advances in protein structure prediction and de novo protein design: A review. *Chem Eng Sci* **61**, 966-988 (2006).
- 4 Jesorka, A. & Orwar, O. Liposomes: technologies and analytical applications. *Annu Rev Anal Chem (Palo Alto Calif)* **1**, 801-832 (2008).
- 5 Uchida, M. *et al.* Biological Containers: Protein Cages as Multifunctional Nanoplatforms. *Advanced Materials* **19**, 1025-1042 (2007).
- 6 Abedin, M. J., Liepold, L., Suci, P., Young, M. & Douglas, T. Synthesis of a cross-linked branched polymer network in the interior of a protein cage. *J Am Chem Soc* **131**, 4346-4354 (2009).
- 7 Douglas, T. & Young, M. Host-guest encapsulation of materials by assembled virus protein cages. *Nature* **393**, 152-155 (1998).
- 8 Kramer, R. M., Li, C., Carter, D. C., Stone, M. O. & Naik, R. R. Engineered protein cages for nanomaterial synthesis. *J Am Chem Soc* **126**, 13282-13286 (2004).
- 9 Varpness, Z., Peters, J. W., Young, M. & Douglas, T. Biomimetic synthesis of a H<sub>2</sub> catalyst using a protein cage architecture. *Nano Lett* **5**, 2306-2309 (2005).
- 10 Klem, M. T. *et al.* Synthetic control over magnetic moment and exchange bias in all-oxide materials encapsulated within a spherical protein cage. *J Am Chem Soc* **129**, 197-201 (2007).
- 11 Shapiro, M. G., Szablowski, J. O., Langer, R. & Jasanoff, A. Protein nanoparticles engineered to sense kinase activity in MRI. *J Am Chem Soc* **131**, 2484-2486 (2009).
- 12 Srivastava, S. *et al.* Integrated Magnetic Bionanocomposites through Nanoparticle-Mediated Assembly of Ferritin. *J Am Chem Soc* **129**, 11776-11780 (2007).
- 13 Uchida, M. *et al.* A human ferritin iron oxide nano-composite magnetic resonance contrast agent. *Magn Reson Med* **60**, 1073-1081 (2008).
- 14 Miura, A. *et al.* Non-volatile flash memory with discrete bionanodot floating gate assembled by protein template. *Nanotechnology* **19**, 255201 (2008).
- 15 Miura, A. *et al.* Floating Nanodot Gate Memory Devices Based on Biomineralized Inorganic Nanodot Array as a Storage Node. *Jpn. J. Appl. Phys.* **45**, L1-L3 (2006).

- 16 Miura, A., Uraoka, Y., Fuyuki, T., Yoshii, S. & Yamashita, I. Floating nanodot  
gate memory fabrication with biomineralized nanodot as charge storage node. *J.  
Appl. Phys.* **103**, 74503-74512 (2008).
- 17 Comellas-Aragones, M. *et al.* A virus-based single-enzyme nanoreactor. *Nat  
Nanotechnol* **2**, 635-639 (2007).
- 18 Minten, I. J., Hendriks, L. J., Nolte, R. J. & Cornelissen, J. J. Controlled  
encapsulation of multiple proteins in virus capsids. *J Am Chem Soc* **131**, 17771-  
17773 (2009).
- 19 Seebeck, F. P., Woycechowsky, K. J., Zhuang, W., Rabe, J. P. & Hilvert, D. A  
simple tagging system for protein encapsulation. *J Am Chem Soc* **128**, 4516-4517  
(2006).
- 20 Loo, L., Guenther, R. H., Lommel, S. A. & Franzen, S. Encapsidation of  
nanoparticles by red clover necrotic mosaic virus. *J Am Chem Soc* **129**, 11111-  
11117 (2007).
- 21 Loo, L., Guenther, R. H., Basnayake, V. R., Lommel, S. A. & Franzen, S.  
Controlled encapsidation of gold nanoparticles by a viral protein shell. *J Am  
Chem Soc* **128**, 4502-4503 (2006).
- 22 Chen, C. *et al.* Nanoparticle-templated assembly of viral protein cages. *Nano Lett*  
**6**, 611-615 (2006).
- 23 Chang, C. B., Knobler, C. M., Gelbart, W. M. & Mason, T. G. Curvature  
dependence of viral protein structures on encapsidated nanoemulsion droplets.  
*ACS Nano* **2**, 281-286 (2008).
- 24 Lee, L. A. & Wang, Q. Adaptations of nanoscale viruses and other protein cages  
for medical applications. *Nanomedicine: Nanotechnology, Biology, and Medicine*  
**2**, 137-149 (2006).
- 25 Flenniken, M. L. *et al.* Selective attachment and release of a chemotherapeutic  
agent from the interior of a protein cage architecture. *Chem Commun (Camb)*,  
447-449 (2005).
- 26 Gillitzer, E., Suci, P., Young, M. & Douglas, T. Controlled ligand display on a  
symmetrical protein-cage architecture through mixed assembly. *Small* **2**, 962-966  
(2006).
- 27 Barnhill, H. N., Claudel-Gillet, S., Ziessel, R., Charbonniere, L. J. & Wang, Q.  
Prototype protein assembly as scaffold for time-resolved fluoroimmuno assays. *J.  
Am Chem Soc* **129**, 7799-7806 (2007).
- 28 Barnhill, H. N., Reuther, R., Ferguson, P. L., Dreher, T. & Wang, Q. Turnip  
yellow mosaic virus as a chemoaddressable bionanoparticle. *Bioconjug Chem* **18**,  
852-859 (2007).
- 29 Suci, P. A., Kang, S., Young, M. & Douglas, T. A streptavidin-protein cage Janus  
particle for polarized targeting and modular functionalization. *J Am Chem Soc*  
**131**, 9164-9165 (2009).
- 30 Gary J. Tong, S. C. H., Zachary M. Carrico, and Matthew B. Francis\*. Viral  
Capsid DNA Aptamer Conjugates as Multivalent Cell-Targeting Vehicles. *J Am  
Chem Soc* **131**, 11174-11178 (2009).
- 31 Uchida, M. *et al.* Targeting of cancer cells with ferrimagnetic ferritin cage  
nanoparticles. *J Am Chem Soc* **128**, 16626-16633 (2006).

- 32 Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc Natl Acad Sci U S A* **98**, 5 (2001).
- 33 Carlson, J. C. T. *et al.* Chemically Controlled Self-Assembly of Protein Nanorings. *J Am Chem Soc* **128**, 9 (2006).
- 34 Kitagishi, H. *et al.* Supramolecular Hemoprotein Linear Assembly by Successive Interprotein Heme-Heme Pocket Interactions. *J Am Chem Soc* **129**, 2 (2007).
- 35 Dotan, N., Arad, D., Frolov, F. & Freeman, A. Self-Assembly of a Tetrahedral Lectin into Predesigned Diamondlike Protein Crystals. *Angew Chem Int Ed* **38**, 4 (1999).
- 36 Radford, R. J., Lawrenz, M., Nguyen, P. C., McCammon, J. A. & Tezcan, F. A. Porous protein frameworks with unsaturated metal centers in sterically encumbered coordination sites. *Chemical Communications*, 3 (2011).
- 37 Dabrowski, M. J. *et al.* Supramolecular Self-Assembly of Glutamine Synthetase: Mutagenesis of a Novel Intermolecular Metal Binding Site Required for Dodecamer Stacking. *Biochemistry* **33**, 8 (1994).
- 38 Dietz, H., Bornschlogl, T., Heym, R., Konig, F. & Rief, M. Programming protein self assembly with coiled coils. *New Journal of Physics* **9**, 8 (2007).
- 39 Fegan, A., White, B., Carlson, J. C. T. & Wagner, C. R. Chemically Controlled Protein Assembly: Techniques and Applications. *Chemical Reviews* **110**, 11 (2010).
- 40 Liu, Q. *et al.* Hierarchical and Helical Self-Assembly of ADP-Ribosyl Cyclase into Large-Scale Protein Microtubes. *The Journal of Physical Chemistry B* **112**, 5 (2008).
- 41 Radford, R. J. & Tezcan, F. A. A Superprotein Triangle Driven by Nickel(II) Coordination: Exploiting Non-Natural Metal Ligands in Protein Self-Assembly. *J Am Chem Soc* **131**, 2 (2009).
- 42 Salgado, E. N., Lewis, R. A., Faraone-Mennella, J. & Tezcan, F. A. Metal-Mediated Self-Assembly of Protein Superstructures: Influence of Secondary Interactions on Protein Oligomerization and Aggregation. *J Am Chem Soc* **130**, 3 (2008).
- 43 Salgado, E. N., Radford, R. J. & Tezcan, F. A. Metal-Directed Protein Self-Assembly. *Accounts of Chemical Research* **43**, 12 (2010).
- 44 Reches, M. & Gazit, E. Casting Metal Nanowires Within Discrete Self-Assembled Peptide Nanotubes. *Science* **300**, 625-627 (2003).
- 45 Lim, Y.-b. & Lee, M. Nanostructures of Beta-sheet peptides: steps towards bioactive functional materials. *Journal of Materials Chemistry* **18**, 723-727 (2008).
- 46 Zhang, S., Holmes, T., Lochshin, C. & Rich, A. Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proc Natl Acad Sci U S A* **90**, 3334-3338 (1993).
- 47 Ulijn, R. V. & Smith, A. M. Designing peptide based nanomaterials. *Chem. Soc. Rev.* **37**, 664-675 (2008).
- 48 Aggeli, A. *et al.* Hierarchical self-assembly of chiral rod-like molecules as a model for peptide b-sheet tapes, ribbons, fibrils, and fibers. *Proc Natl Acad Sci U S A* **98**, 11857-11862 (2001).

- 49 Nagarkar, R. P., Hule, R. A., Pochan, D. J. & Schneider, J. P. De Novo Design of Strand-Swapped  $\beta$ -Hairpin Hydrogels. *J Am Chem Soc* **130**, 4466-4474 (2008).
- 50 Aggeli, A. *et al.* Responsive gels formed by the spontaneous self-assembly of peptides into polymeric beta-sheet tapes. *Nature* **386**, 259-262 (1997).
- 51 Schneider, J. P. *et al.* Responsive Hydrogels from the Intramolecular Folding and Self-Assembly of a Designed Peptide. *J Am Chem Soc* **124**, 15030-15037 (2002).
- 52 Brown, C. L., Aksay, I. A., Saville, D. A. & Hecht, M. H. Template-Directed Assembly of a de Novo Designed Protein. *J Am Chem Soc* **124**, 6846-6848 (2002).
- 53 Mason, J. M. & Arndt, K. M. Coil Domains: Stability, Specificity, and Biological Implications. *Chembiochem* **5**, 170-176 (2004).
- 54 Kojima, S., Kuriki, Y., Yoshida, T., Yazaki, K. & Miura, K.-i. Fibril Formation by an Amphipathic  $\alpha$ -Helix-Forming Polypeptide Produced by Gene Engineering. *Proceedings of the Japan Academy, Ser. B, Physical and Biological Sciences* **73**, 7-11 (1997).
- 55 Woolfson, D. N. Building fibrous biomaterials from alpha-helical and collagen-like coiled-coil peptides. *Peptide Science* **94**, 118-127 (2010).
- 56 Pandya, M. J. *et al.* Sticky-end assembly of a designed peptide fiber provides insight into protein fibrillogenesis. *Biochemistry* **39**, 8728-8734 (2000).
- 57 Papapostolou, D., Bromley, E. H., Bano, C. & Woolfson, D. N. Electrostatic control of thickness and stiffness in a designed protein fiber. *J Am Chem Soc* **130**, 5124-5130 (2008).
- 58 Ryadnov, M. G. & Woolfson, D. N. MaP peptides: programming the self-assembly of peptide-based mesoscopic matrices. *J Am Chem Soc* **127**, 12407-12415 (2005).
- 59 Ogihara, N. L. *et al.* Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc Natl Acad Sci U S A* **98**, 1404-1409 (2001).
- 60 Ryadnov, M. G. & Woolfson, D. N. Fiber recruiting peptides: noncovalent decoration of an engineered protein scaffold. *J Am Chem Soc* **126**, 7454-7455 (2004).
- 61 Channon, K. J. *et al.* Modification of Fluorophore Photophysics through Peptide-Driven Self-Assembly. *J Am Chem Soc* **130**, 5487-5491 (2008).
- 62 Channon, K. J., Devlin, G. L. & MacPhee, C. E. Efficient Energy Transfer within Self-Assembling Peptide Fibers: A Route to Light-Harvesting Nanomaterials. *J Am Chem Soc* **131**, 12520-12521 (2009).
- 63 Lim, Y.-b., Lee, E. & Lee, M. Cell-Penetrating-Peptide-Coated Nanoribbons for Intracellular Nanocarriers<sup>†</sup>. *Angew Chem Int Ed Engl* **46**, 3475-3478 (2007).
- 64 Seeman, N. C. Nucleic Acids Junctions and Lattices. *J Theor Biol* **99**, 11 (1982).
- 65 Kallenbach, N. R., Ma, R.-I. & Seeman, N. C. An imobile nucleic acid junction constructed from oligonucleotides. *Nature* **305**, 3 (1983).
- 66 Chen, J.-H., Kallenbach, N. R. & Seeman, N. C. A Specific Quadrilateral Synthesized from DNA Branched Junctions. *J Am Chem Soc* **111**, 6 (1989).
- 67 Chen, J. & Seeman, N. C. Synthesis from DNA if a molecule with the connectivity of a cube. *Nature* **350**, 3 (1991).
- 68 Zhang, Y. & Seeman, N. C. Construction of a DNA-Truncated Octahedron. *J Am Chem Soc* **116**, 9 (1994).

- 69 Shih, W. M., Quispe, J. D. & Joyce, G. F. A 1.7-kilobase single-stranded DNA  
that folds into a nanoscale octahedron. *Nature* **427**, 4 (2004).
- 70 Goodman, R. P. *et al.* Rapid Chiral Assembly of Rigid DNA Building Blocks for  
Molecular Nanofabrication. *Science* **310**, 6 (2005).
- 71 Erben, C. M., Goodman, R. P. & Turberfield, A. J. Single Molecule  
Encapsulation in a Rigid DNA Cage. *Angew Chem Int Edit* **45**, 4 (2006).
- 72 He, Y. *et al.* Hierarchical self-assembly of DNA into symmetric supramolecular  
polyhedra. *Nature* **452**, 5 (2008).
- 73 Aldaye, F. A. & Sleiman, H. F. Modular Access to Structurally Switchable 3D  
Discrete DNA Assemblies. *J Am Chem Soc* **129**, 2 (2007).
- 74 Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable  
lid. *Nature* **459**, 5 (2009).
- 75 Matsuura, K., Yamashita, T., Igami, Y. & Kimizuka, N. ‘Nucleo-nanocages’:  
designed ternary oligodeoxyribonucleotides spontaneously form nanosized DNA  
cages. *Chem Commun*, 2 (2003).
- 76 Aldaye, F. A., Palmer, A. L. & Sleiman, H. F. Assembling Materials with DNA as  
the Guide. *Science* **321**, 6 (2008).
- 77 Goodman, R. P. *et al.* Reconfigurable, braced, three dimensional DNA  
nanostructures. *Nat Nanotechnol* **3**, 4 (2008).
- 78 Li, X., Yang, X., Qi, J. & Seeman, N. C. Antiparallel DNA Double Crossover  
Molecules As Components for Nanoconstruction. *J Am Chem Soc* **118**, 10 (1996).
- 79 Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of  
two-dimensional DNA crystals. *Nature* **394**, 6 (1998).
- 80 Seeman, N. C. DNA in a material world. *Nature* **421**, 5 (2003).
- 81 Liu, F., Sha, R. & Seeman, N. C. Modifying the Surface Features of Two-  
Dimensional DNA Crystals. *J Am Chem Soc* **121**, 6 (1999).
- 82 Mao, C., Sun, W. & Seeman, N. C. Designed Two-Dimensional DNA Holliday  
Junction Arrays Visualized by Atomic Force Microscopy. *J Am Chem Soc* **121**, 7  
(1999).
- 83 LaBean, T. H. *et al.* Construction, Analysis, Ligation, and Self-Assembly of DNA  
Triple Crossover Complexes. *J Am Chem Soc* **122**, 13 (2000).
- 84 Liu, D., Wang, M., Deng, Z., Walulu, R. & Mao, C. Tensegrity: Construction of  
Rigid DNA Triangles with Flexible Four-Arm DNA Junctions. *J Am Chem Soc*  
**126**, 2 (2004).
- 85 He, Y., Tian, Y., Ribbe, A. E. & Mao, C. Highly Connected Two-Dimensional  
Crystals of DNA Six-Point-Stars. *J Am Chem Soc* **128**, 2 (2006).
- 86 Lin, C., Liu, Y., Rinker, S. & Yan, H. DNA Tile Based Self-Assembly:  
BuildingComplex Nanoarchitectures. *ChemPhysChem* **7**, 7 (2006).
- 87 Li, H., Park, S. H., Reif, J. H., LaBean, T. H. & Yan, H. DNA-Templated Self-  
Assembly of Protein and Nanoparticle Linear Arrays. *J Am Chem Soc* **126**, 2  
(2004).
- 88 Bath, J. & Turberfield, A. J. DNA nanomachines. *Nat Nanotechnol* **2**, 10 (2007).

## Chapter 2

### A Strategy for Assembling Protein Cages

#### 2.1 Introduction

To date, few strategies have been described for assembling proteins into higher order structures of defined architecture. In particular, strategies for constructing protein cages, which have a number of potential applications in nanotechnology, have been very limited with only a few examples found in the literature.<sup>1,2</sup> These strategies have not been generalizable; they have been limited to specific proteins and so lack general utility. Advancing protein cage nanotechnology requires new strategies to be formulated that are broadly applicable to a larger number of proteins and provide predictable outcomes for cage assembly.

Herein, I describe the design of a general strategy for assembling proteins into higher order cage structures. The formulation of an assembly strategy requires a mode of linking proteins together to drive assembly be devised and that the protein building components specify formation of the desired cage structure. When formulating our strategy we identified five criteria to be incorporated into the design, those being:

- 1) Assembly should be genetically encoded to eliminate the need for additional cross-linking steps and allow *in vivo* assembly if desired.
- 2) Assembly should only occur under specified conditions and be reversible.
- 3) The strategy should allow a broad range of proteins to be incorporated as building components.

4) Assembly should lead to the formation of a limited number of predictable cage architectures.

5) Assembled proteins should retain their structure and biological activity.

Below I discuss methods for linking proteins together, the drawbacks associated with each, and our design strategy that involves incorporating *de novo* designed coiled coil peptides as linking units. Further discussion will be given on a means for directing the assembly to give predictable cage architectures. In closing, I will lay out a general strategy for assembling protein building blocks (PBBs) into cage structures.

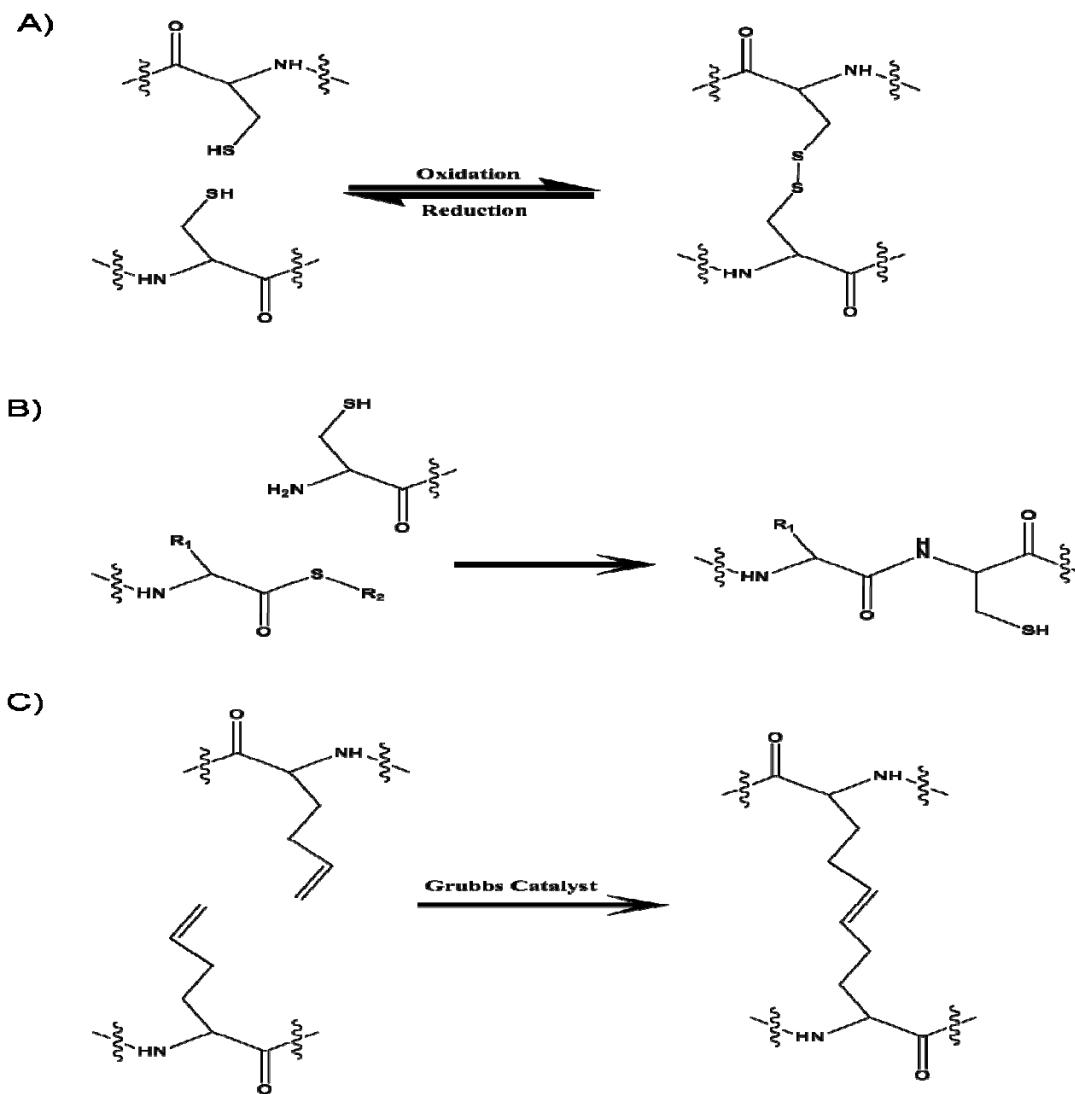
## 2.2 – Methods for Linking Proteins

A number of methods have been developed for linking proteins together. Methods for cross-linking are particularly important for understanding protein-protein interactions and effects of those interactions. For our purposes, a method that can be readily applied to any protein, is reversible, does not perturb protein structure/activity, provides a high degree of specificity of linking, potentially allows *in vivo* assembly, and contains few steps for cross-linking is desired. The following common methods of protein cross-linking are discussed with respect to their suitability for the design criteria outlined above.

### Covalent Capture

Covalent capture exploits protein side chains that react to form a covalent bond linking the proteins together. The most common method of covalent capture is disulfide bridge formation between two cysteine residues (Figure 2.1A). Another common method of covalent capture is native chemical ligation, which is typically used for creating semi-synthetic proteins that contain non-natural amino acids. In native chemical ligation a C-

terminal thioester selectively reacts with an N-terminal cysteine to form a standard peptide bond between two peptides/proteins (Figure 2.1B). Other methods of covalent capture incorporate unnatural amino acids to link proteins together. Incorporation of olefin containing amino acid residues allows one to perform olefin metathesis to cross link proteins in the presence of a Grubbs catalyst (Figure 2.1C). In addition, amino acids that perform radical and lysine cross-linking covalent capture reactions have been described.<sup>3</sup>



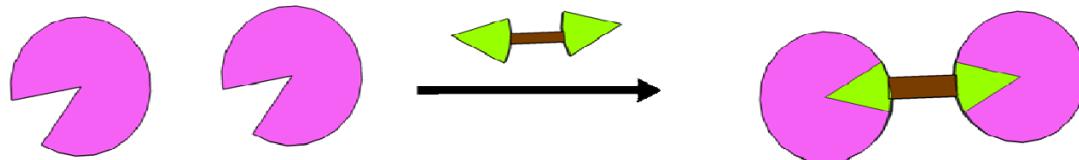
**Figure 2.1.** Covalent capture cross-linking methods. A) Disulfide bridge cross-linking. B) Native chemical ligation. C) Olefin metathesis.

The incorporation of unnatural amino acids is a major drawback for using covalent capture as a general cross-linking method. In addition, cross-linking only occurs when the reactive side chains are brought in close contact with one another, necessitating an additional moiety to promote localization of the reactive amino acids near one another. Covalent capture also suffers from a lack of specificity—cysteine can form disulfides with any cysteine—and many of the reactive unnatural amino acids are quite promiscuous, for example the radical forming photoreactive L-p-benzoylphenylalanine reacts with nearly any adjacent amino acid upon photoactivation.<sup>3</sup> The lack of specificity renders covalent capture unsuitable for assembly *in vivo*. Furthermore, formation of a permanent covalent cross-link between proteins does not meet the criteria that assembly be reversible, which provides an avenue for cargoes to be loaded and released from the cage.

#### Chemical Induced Dimerization

Chemically induced dimerization (CID) is a method that exploits the binding of protein substrates to link proteins. Typically a bivalent substrate molecule is synthesized allowing simultaneous binding of the substrate analog by two proteins (Figure 2.2). One of first examples came in 1993 in a landmark paper by Schreiber, Crabtree, and co-workers for binding of a bivalent drug derivative, FK1012, which reversibly dimerized its protein target FK506 binding protein and could be used to drive biological function.<sup>4</sup> Probably the most utilized CID system is the streptavidin biotin pair, which has been extensively used in molecular biology and nanotechnology. The affinity of biotin for streptavidin is one the strongest non-covalent interactions known. In general biotin is covalently attached to another substrate molecule (either small molecule or protein) and

assembly is observed upon mixing with streptavidin.<sup>5</sup> Several other examples of the use of CID to construct extended structures from proteins, including 3-D arrays, rings, and linear chains, were discussed previously in Chapter 1.<sup>6-8</sup>



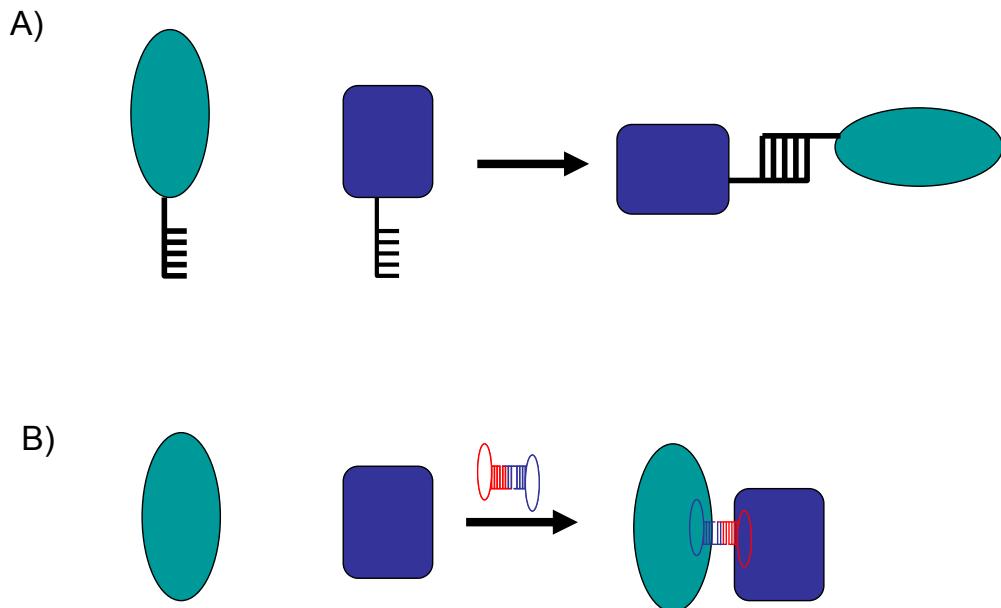
**Figure 2.2.** Schematic representation of chemical induced dimerization. Protein (purple) is linked by a bivalent substrate analog (green wedges connected by brown rectangle).

Use of the CID method relies on tight binding of a protein to a substrate molecule, however many proteins do not bind tightly to their substrates. In addition, many proteins do not have small chemical substrates that can be used for CID. The CID method requires synthesis of bivalent substrates or covalent attachment of substrates to proteins, which does not fit the criteria that genetically encoded to eliminate additional cross-linking steps. Furthermore, assembly *in vivo* may be inhibited if endogenous substrates are present. These features greatly limit the general applicability of CID for assembling proteins and fall well short of the criteria outlined earlier.

#### DNA Cross-linking

DNA is another potential agent that can be used to link proteins together. There are two strategies for linking proteins together with DNA (Figure 2.3). First complementary base paired strands can be appended to different protein building blocks. Recognition and base paring by the DNA strands would occur only upon mixing, giving highly specific control over assembly. In addition, the noncovalent association of the DNA would allow reversible association, meeting one of the design criteria identified

earlier. One method for linking DNA to proteins is to use thiolated DNA which can form disulfide bridges with cysteines recombinantly engineered onto the protein.<sup>9-11</sup> However, a less chemically labile linkage than disulfide linking would be desirable to maintain structural integrity. Alternative methods for covalently linking DNA to proteins rely on the use of hetero-bispecific cross-linkers and require multi-step chemical coupling and purification steps.<sup>12,13</sup> The second approach is to make DNA aptamers that selectively bind proteins together.<sup>14-16</sup> However, this method requires that many DNA sequences be screened to find an aptamer that binds exclusively to the protein of interest. This would require a new screen every time a new protein was to be used in construction and it is unclear whether aptamers could be found for all proteins. The use of DNA, either by base pairing of single stranded DNA or aptamer binding, for linking proteins and driving assembly has potential for allowing assembly *in vivo*. However, for *in vivo* assembly the cross-linked protein-DNA or DNA aptamers must be incorporated into the cell through some method of transfection, which could be somewhat difficult for large proteins and aptamers. A method that already places all the required building components within the cells for assembly would be more desirable.

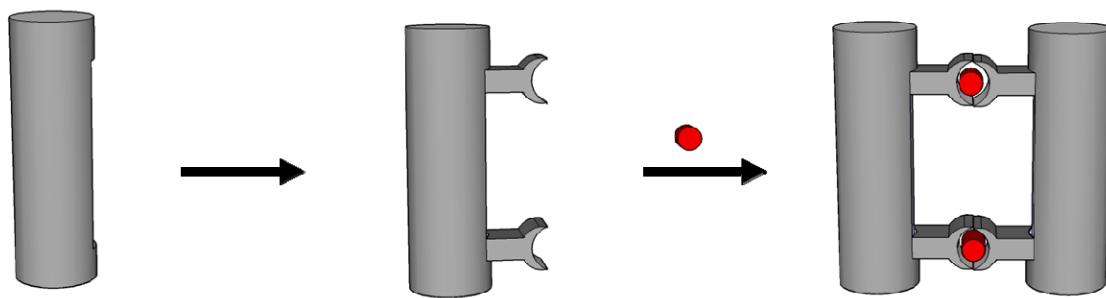


**Figure 2.3.** Methods of linking proteins with DNA. A) Proteins covalently attached to DNA with complementary sequences oligomerize upon base pairing between DNA strands. B) Proteins are linked by a DNA (or RNA) aptamer that binds to the proteins.

#### Metal Induced Assembly

A method that is enjoying some success in *de novo* designed assembly of proteins is metal directed assembly. Early studies of the highly symmetric dodecameric protein glutamine synthetase (GS) found that the protein assembled into highly regular stacks in the presence of  $Zn^{2+}$  and other divalent ions.<sup>17,18</sup> Dabrowski et al. determined that a cluster of histidines and methionine displayed on N-terminal helices of GS were responsible for coordinating assembly by binding of metal ions between adjacent GS proteins.<sup>18</sup> Work by Tezcan and co-workers has focused on the use of incorporating metal coordination sites onto proteins to control assembly into higher order structures (Figure 2.4).<sup>19</sup> They have introduced metal coordination sites into the small four helix bundle protein cytochrome cb<sub>562</sub> and shown that they can coordinate assembly of the protein into dimeric and tetrameric species, protein triangles, porous protein frameworks, and a tetrahedral protein cage by properly positioning the coordination sites on the

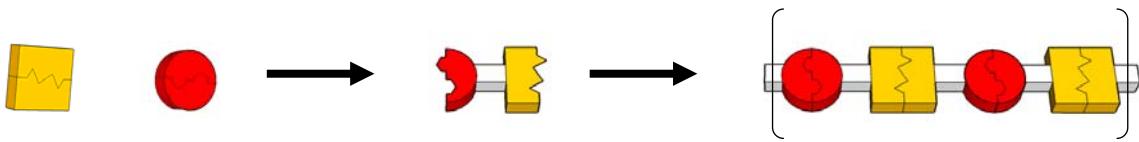
surface of the helices of the protein.<sup>2,20-22</sup> Metal induced assembly would allow assembly *in vivo*, depending on the metal chosen for the binding sites and the cellular availability or permeability/toxicity of that metal used to coordinate assembly. However, the strategies for metal directed assembly have not been translated to other protein systems. Often the divalent metal ions used for assembly can cause precipitation of proteins. A more broadly applicable method of assembly was desired for our strategy.



**Figure 2.4.** Metal directed assembly of a protein. Metal binding sites are incorporated onto the protein (gray cylinder) through mutagenesis. Upon addition of metal (red circle) protein subunits associate through metal binding.

#### Protein-Domain Assembly

Another genetically encodable assembly method is the use of protein domains that induce dimerization through recombinant expression of a fusion protein (Figure 2.5). Protein subunits are fused together, by genetic manipulation, and formation of the quaternary structure of the subunits drives the assembly. This method was successfully employed by Padilla et al. to construct a protein cage and protein fibers.<sup>1</sup> Protein-domain assembly would allow *in vivo* assembly, with protein building blocks ready for *in vivo* assembly upon expression. However, the use of protein domains for dimerization generally limits control over assembly; oligomerization of domains occurs upon expression. In addition, expression of linked protein domains can potentially lead to misfolding of the proteins.

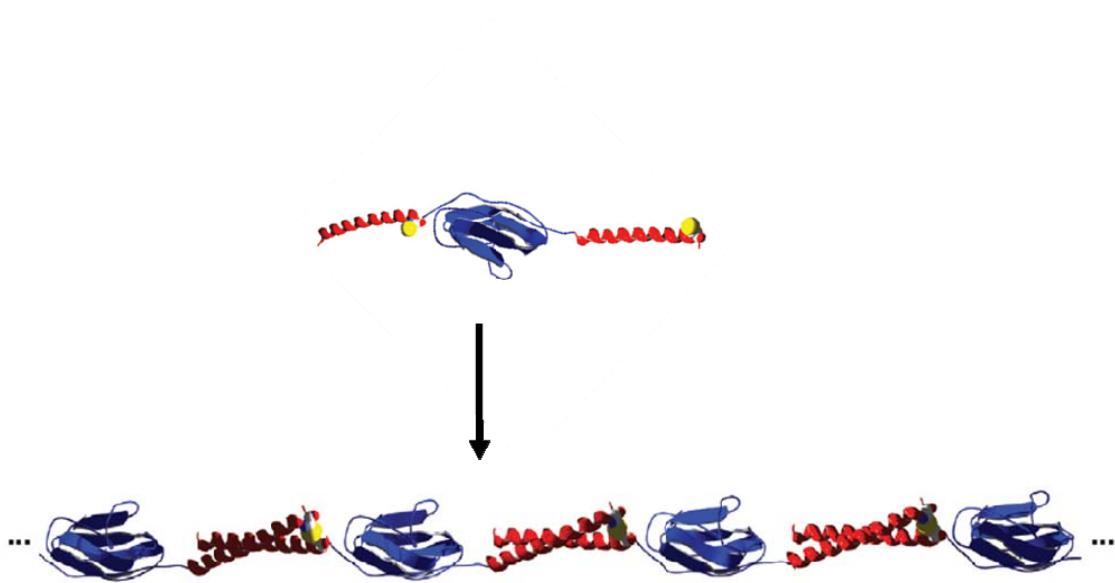


**Figure 2.5.** Protein subunit fusion assembly. Protein subunits or domains (red and yellow) that form higher order quaternary structures are covalently linked together. Formation of the quaternary structures by the subunits drives assembly.

### Peptide Directed Assembly

The use of dimerizing peptides is an alternative assembly method that is genetically encodable, with peptides being added to proteins through recombinant expression of a fusion protein, and allows for assembly *in vivo*. The smaller size of peptides relative to protein domains make them less prone to cause misfolding when appended to the termini of protein subunits. Peptides have been described that can assemble upon binding of metals.<sup>23-25</sup> However, such peptides pose the same problems as described for metal directed protein assembly. A better system is peptides that form dimeric coiled coils. The coiled coil is a structural motif formed by, in the simplest case, two  $\alpha$ -helices that coil around each other in a left handed manner to form a super helix.<sup>26</sup> Coiled coil peptides have been used as linking/assembly agents to form hydrogels, aggregate cells, construct sensors, link proteins together and for a number of other applications.<sup>27-37</sup> Design principles are well established, and a thorough discussion of their design is given in Chapter 3. Properties such as specificity (homo- or hetero-associating), stability, association orientation (antiparallel or parallel), and a number of other properties can be defined in the design, making the coiled coil a very versatile linking agent that has potential for making “smart” materials. Of the methods used for linking proteins, the coiled coil motif fits well with the criteria we wanted to incorporate

into our protein assembly design strategy. A model depicting cross-linking by a coiled coil, described by Dietz et al., is shown in Figure 2.6.<sup>31</sup>



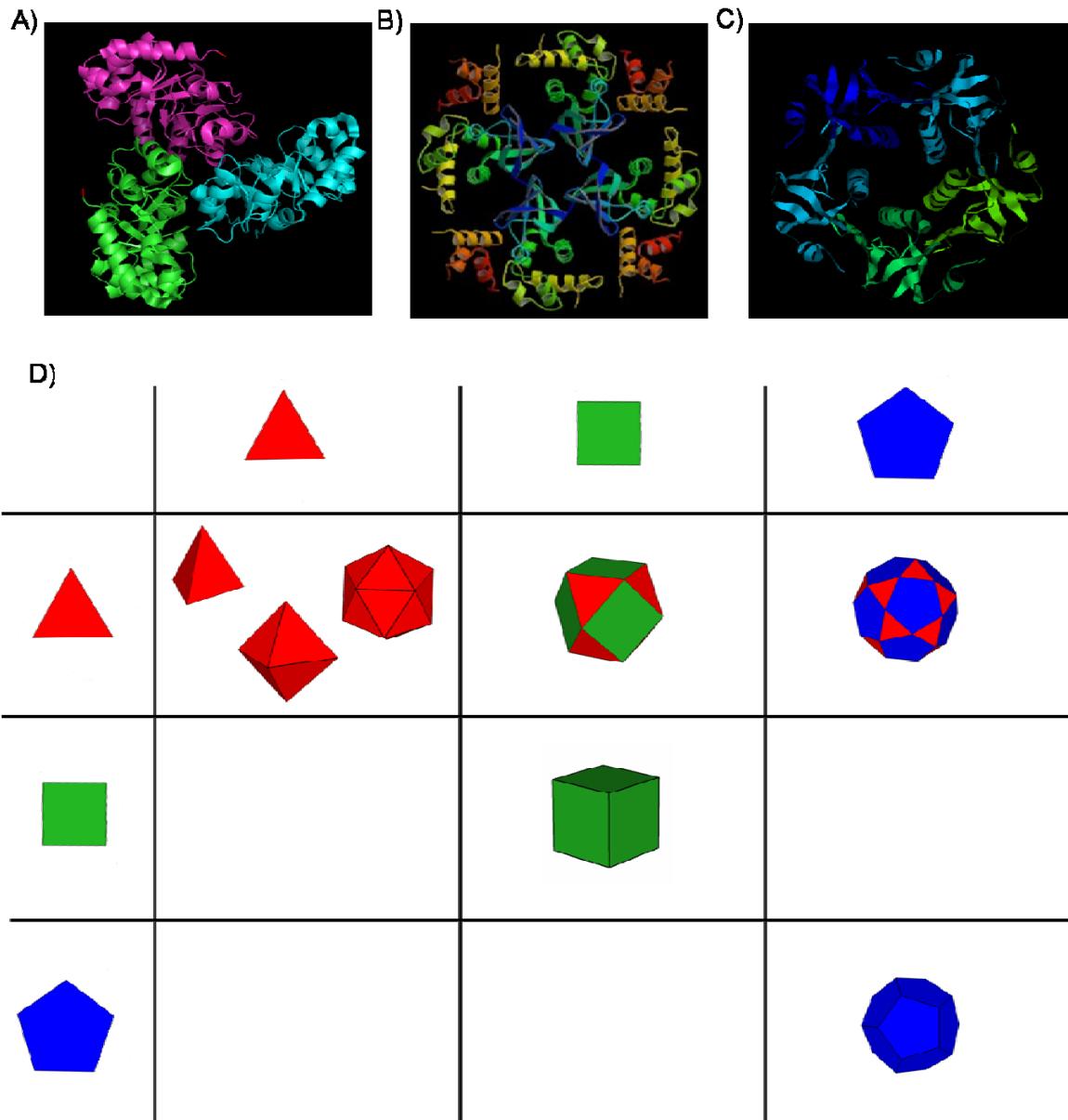
**Figure 2.6.** Peptide directed assembly of proteins. Crosslinking of Ig27 (from human cardiac titan) by a homodimeric coiled coil peptide. Yellow spheres indicate the N-terminal end of the peptides. The figure was adapted from a figure found in reference 31.

### 2.3 - Directing Assembly by Symmetry

A major challenge in designing a strategy for assembling protein cages is determining how to control assembly to yield specific structures, preferably a single, unique structure, that are robustly predictable. A guiding principle is to use symmetry endowed by the protein building block (PBB) used in the construction to direct assembly and determine overall structure.<sup>38</sup> The PBB serves as a scaffold that orients and directs the cross-linking elements, limiting the structures that form upon linkage.

For our strategy we looked to exploit proteins with highly symmetric quaternary structures that would direct assembly towards a limited number of cage structures that are compatible with that symmetry. Trimeric, tetrameric, and pentameric proteins with geometries closely resembling triangles, squares, and pentagons, respectively, are

commonly observed, and examples are illustrated in Figure 2.7. Geometry dictates that few simple polyhedra can form from such shapes as shown in the table in Figure 2.7. By analogy, we hypothesized that the geometry imparted by the quaternary structure would limit the structures that could form and further allow us to predict what cage architectures would form.

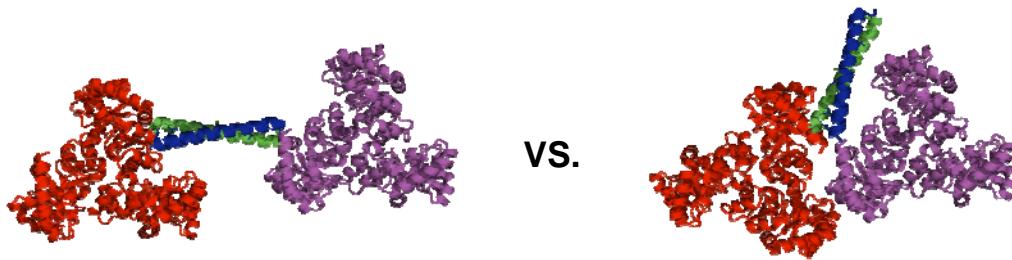


**Figure 2.7.** Protein quaternary structure geometries and structures that could be constructed from their geometric shapes. Protein Crystal Structures: A) The trimer KDPG aldolase (PDB ID: 1WBH), B) the tetramer Ruv A (PDB ID: 1BDX), and C) the pentamer cholera toxin subunit B (PDB ID: 3EFX). D) Table showing the geometric structures that can be constructed from triangles, squares and pentagons. Duplicate structures have been omitted to help with clarity. For structures composed of a mixture of two shapes, only structures arising from the hetero-connection of shapes are provided (other Archimedean solids are possible).

## 2.4 – Incorporating Flexibility Between the Protein Building Block and Peptide Linker

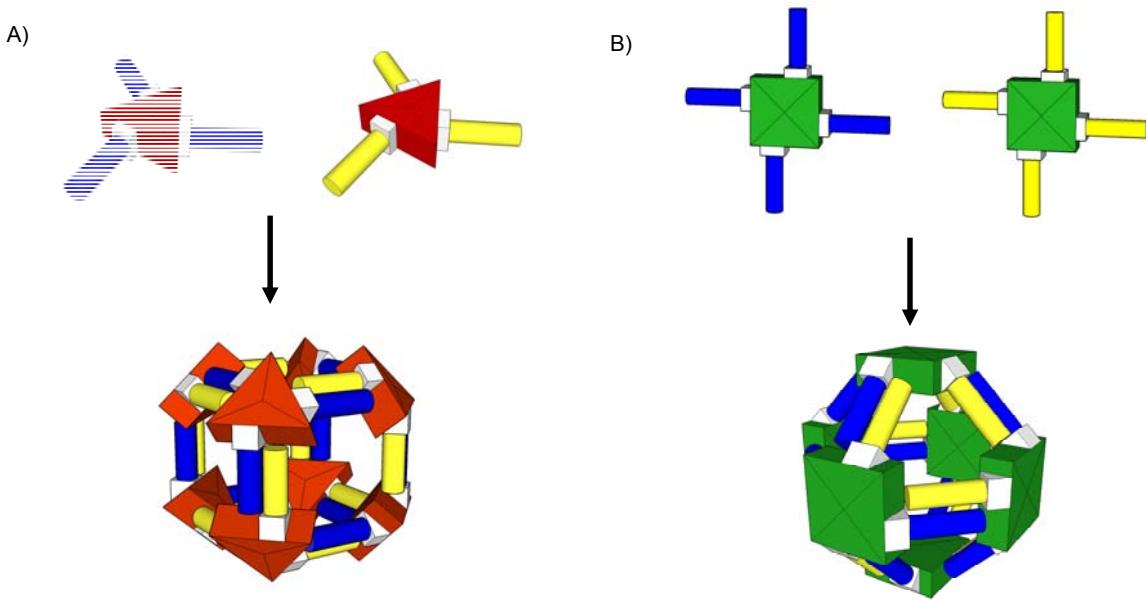
A feature we decided to incorporate into our design strategy is flexibility between the protein building block and coiled coil peptide linker, deviating from other strategies which necessitate construction of rigid building blocks locked in a desired confirmation. Rigidly constructed building blocks are expected to limit the structure that forms upon assembly because all other structures are “locked out”, however this means that the alignment must be precisely controlled, otherwise the structure will not assemble as desired. However the N and C termini of proteins that are used to append linking agents are often unstructured in a large number of proteins. The stringent requirement that proteins have well-defined termini, as proposed by Yeates, greatly limits the pool of proteins which can be used as building blocks.<sup>1</sup> We decided to incorporate a short flexible peptide spacer connecting the protein building block and the peptide linker, allowing a larger number of proteins to be incorporated as potential building blocks. Moreover, flexibility allows the linking subunits to adopt a large range of conformations with respect to the protein, providing the freedom to choose the appropriate geometry. This feature is important because we wanted to determine to what extent assembly could be controlled by simply the symmetry of the protein building block employed. One might expect that incorporating flexibility would yield problems because a large ensemble of structures could, in principle, be formed. However, research in related DNA cage assembly has shown that flexibility can be exploited to construct a broad range of cage structures with a high degree of specificity.<sup>39</sup> We hypothesized that relatively few structures would predominate and that these would be the most highly symmetrical structures.

## 2.5 - The Strategy



**Figure 2.8.** Parallel vs. antiparallel coiled coil association for linking protein building blocks. Antiparallel orientation (left) provides more extended spatial separation than parallel association of coiled coiled peptides placed at the same termini.

Pairs of peptides designed to form antiparallel heterodimeric coiled coils will be fused to the PBBs and facilitate their assembly. Incorporation of heterodimeric coiled coils will provide control over assembly; only when two sets of PBBs containing complementary peptide pairs are mixed, will proteins assemble and be linked together. Formation of coiled coil dimers in the antiparallel orientation is expected to provide spatial separation of the protein scaffolds enabling them to form open cage structures without steric clashes between the PBBs (Figure 2.8). PBBs and coiled coil peptides will be connected through their termini by a short peptide sequence allowing flexibility between the two domains, providing “wiggle room” for PBBs to form cages and broadening the strategy to allow inclusion of more protein as PBBs. Fusion proteins of the PBB, peptide linker, and flexible spacer will be obtained through recombinant gene expression in *E. coli*. A schematic representation of the assembly strategy is shown in Figure 2.9.



**Figure 2.9.** Schematic representation for the assembly strategy of two protein geometries. Protein building blocks (red triangle or green squares) connected to heterodimeric coiled coil peptides (blue or yellow cylinders) by a flexible peptide spacer (white region) assemble upon mixing to form structures that are predicted by the quaternary structure geometry. (Note that subunits must exchange for cubic structure to form from square protein building block).

## 2.6 - Discussion

I have described a general strategy for assembling proteins into cage structures. The strategy limits incorporation of PBBs to proteins with highly symmetric quaternary structures with regular geometries, predicted to guide assembly into a specific cage structure. A large number of proteins exist with quaternary structures of regular geometries. We have also allowed the linkers used (coiled coil peptides) to be flexible, which has not been exploited for making cage structures from proteins. It has been conventional wisdom that linkers must be rigidly oriented on the protein scaffold to limit the structures that form. We specifically wanted to determine if the symmetry of the PBB used is enough to define assembly into specific protein cage structures. Modifications to the flexible peptide region connecting the linker and PBB may allow for greater

specificity to be built in and prescribe assembly of one structure over another. The initial flexible design allows for additional rigidity to be built in as needed in subsequent studies.

The strategy outlined above was designed as a starting point for advancing the construction of protein based biomaterials. Protein cages were specifically targeted due to their importance towards potential applications in nanotechnology. To date, there have been relatively few methods described for assembling protein cages, and those that have been put forward tend to be limited to a subset of proteins that can be used in assembly. We have put forth a new approach incorporating stringent design criteria that could serve as a foundation for advancing research in protein based biomaterials. My thesis seeks to determine whether this design strategy fulfills the criteria that have been laid out here.

## 2.7 - References

- 1 Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc Natl Acad Sci U S A* **98**, 2217-2221 (2001).
- 2 Ni, T. W. & Tezcan, F. A. Structural characterization of a microperoxidase inside a metal-directed protein cage. *Angew Chem Int Ed Engl* **49**, 7014-7018 (2010).
- 3 Hartgerink, J. D. Covalent capture: a natural complement to self-assembly. *Curr Opin Chem Biol* **8**, 604-609 (2004).
- 4 Spencer, D. M., Wandless, T. J., Schreiber, S. L. & Crabtree, G. R. Controlling signal transduction with synthetic ligands. *Science* **262**, 1019-1024 (1993).
- 5 Fegan, A., White, B., Carlson, J. C. & Wagner, C. R. Chemically controlled protein assembly: techniques and applications. *Chem Rev* **110**, 3315-3336 (2010).
- 6 Carlson, J. C. T. *et al.* Chemically Controlled Self-Assembly of Protein Nanorings. *J Am Chem Soc* **128**, 7630-7638 (2006).
- 7 Dotan, N., Arad, D., Frolow, F. & Freeman, A. Self-Assembly of a Tetrahedral Lectin into Predesigned Diamondlike Protein Crystals *Angew Chem Int Ed Engl* **38**, 2363-2366 (1999).
- 8 Kitagishi, H. *et al.* Supramolecular Hemoprotein Linear Assembly by Successive Interprotein Heme-Heme Pocket Interactions. *J Am Chem Soc* **129**, 10326-10327 (2007).
- 9 Corey, D. R. & Schultz, P. G. Generation of a hybrid sequence-specific single-stranded deoxyribonuclease. *Science* **238**, 1401-1403 (1987).

- 10 Howorka, S., Cheley, S. & Bayley, H. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nat Biotechnol* **19**, 636-639 (2001).
- 11 Fong, R. B., Ding, Z., Long, C. J., Hoffman, A. S. & Stayton, P. S. Thermoprecipitation of streptavidin via oligonucleotide-mediated self-assembly with poly(N-isopropylacrylamide). *Bioconjug Chem* **10**, 720-725 (1999).
- 12 Niemeyer, C. M. *et al.* Self-assembly of DNA-streptavidin nanostructures and their use as reagents in immuno-PCR. *Nucleic Acids Res* **27**, 4553-4561 (1999).
- 13 Niemeyer, C. M., Sano, T., Smith, C. L. & Cantor, C. R. Oligonucleotide-directed self-assembly of proteins: semisynthetic DNA--streptavidin hybrid molecules as connectors for the generation of macroscopic arrays and the construction of supramolecular bioconjugates. *Nucleic Acids Res* **22**, 5530-5539 (1994).
- 14 Liu, Y., Lin, C., Li, H. & Yan, H. Aptamer-directed self-assembly of protein arrays on a DNA nanostructure. *Angew Chem Int Ed Engl* **44**, 4333-4338 (2005).
- 15 Chhabra, R. *et al.* Spatially addressable multiprotein nanoarrays templated by aptamer-tagged DNA nanoarchitectures. *J Am Chem Soc* **129**, 10304-10305 (2007).
- 16 Shu Jun Zhen, C. Z. H., Jian Wang, and Yuan Fang Li. End-to-End Assembly of Gold Nanorods on the Basis of Aptamer-Protein Recognition. *J. Phys. Chem. C* **113**, 21543-21547 (2009).
- 17 Miller, R. E., Shelton, E. & Stadtman, E. R. Zinc-induced paracrystalline aggregation of glutamine synthetase. *Arch Biochem Biophys* **163**, 155-171 (1974).
- 18 Dabrowski, M. J. *et al.* Supramolecular Self-Assembly of Glutamine Synthetase: Mutagenesis of a Novel Intermolecular Metal Binding Site Required for Dodecamer Stacking. *Biochemistry* **33**, 14957-14957 (1994).
- 19 Salgado, E. N., Radford, R. J. & Tezcan, F. A. Metal-Directed Protein Self-Assembly. *Accounts of Chemical Research* **43**, 661-672 (2010).
- 20 Radford, R. J., Lawrenz, M., Nguyen, P. C., McCammon, J. A. & Tezcan, F. A. Porous protein frameworks with unsaturated metal centers in sterically encumbered coordination sites. *Chemical Communications* (2011).
- 21 Radford, R. J. & Tezcan, F. A. A Superprotein Triangle Driven by Nickel(II) Coordination: Exploiting Non-Natural Metal Ligands in Protein Self-Assembly. *J Am Chem Soc* **131**, 9136-9137 (2009).
- 22 Salgado, E. N., Lewis, R. A., Faraone-Mennella, J. & Tezcan, F. A. Metal-Mediated Self-Assembly of Protein Superstructures: Influence of Secondary Interactions on Protein Oligomerization and Aggregation. *Journal of the American Chemical Society* **130**, 6082-6084 (2008).
- 23 Kharenko, O. A. & Ogawa, M. Y. Metal-induced folding of a designed metalloprotein. *J Inorg Biochem* **98**, 1971-1974 (2004).
- 24 Dai, Q., Castellino, F. J. & Prorok, M. A single amino acid replacement results in the Ca<sup>2+</sup>-induced self-assembly of a helical conantokin-based peptide. *Biochemistry* **43**, 13225-13232 (2004).
- 25 Dai, Q., Prorok, M. & Castellino, F. J. A New Mechanism for Metal Ion-assisted Interchain Helix Assembly in a Naturally Occurring Peptide Mediated by Optimally Spaced g-Carboxyglutamic Acid Residues. *J. Mol. Biol.* **336**, 731-744 (2004).

- 26 Mason, J. M. & Arndt, K. M. Coil Domains: Stability, Specificity, and Biological Implications. *Chembiochem* **5**, 170-176 (2004).
- 27 Yu, B. Y. Coiled-coils: stability, specificity, and drug delivery potential. . *Advanced Drug Delivery Reviews*. **54**, 1113-1129 (2002).
- 28 Gaidamakov, E. K., Backer, M. V., and Backer, J. M. . Molecular vehicle for target mediated delivery of therapeutics and diagnostics *Journal of Controlled Release* **74**, 341-347 (2001).
- 29 Schweller, R. M., Constantinou, P. E., Frankel, N. W., Narayan, P. & Diehl, M. R. Design of DNA-Conjugated Polypeptide-Based Capture Probes for the Anchoring of Proteins to DNA Matrices. *Bioconjug Chem* **19**, 2304-2307 (2008).
- 30 Veiga, E., Lorenzo, V. d. & Fernandez, L. A. Autotransporters as Scaffolds for Novel Bacterial Adhesins: Surface Properties of Escherichia coli Cells Displaying Jun/Fos Dimerization Domains. *Journal of Bacteriology* **185**, 5585–5590 (2003).
- 31 Dietz, H., Bornschlogl, T., Heym, R., Konig, F. & Rief, M. Programming protein self assembly with coiled coils. *New Journal of Physics* **9**, 424-431 (2007).
- 32 Wu, K., Liu, J., Johnson, R. N., Yang, J. & Kopecek, J. Drug-free macromolecular therapeutics: induction of apoptosis by coiled-coil-mediated cross-linking of antigens on the cell surface. *Angew Chem Int Ed Engl* **49**, 1451-1455 (2010).
- 33 Hillar, A., Culham, D. E., Vernikovska, Y. I., Wood, J. M. & Boggs, J. M. Formation of an antiparallel, intermolecular coiled coil is associated with in vivo dimerization of osmosensor and osmoprotectant transporter ProP in Escherichia coli. *Biochemistry* **44**, 10170-10180 (2005).
- 34 Katz, B. Z. *et al.* Green fluorescent protein labeling of cytoskeletal structures--novel targeting approach based on leucine zippers. *Biotechniques* **25**, 298-302, 304 (1998).
- 35 De Crescenzo, G., Pham, P. L., Durocher, Y., Chao, H. & O'Connor-McCourt, M. D. Enhancement of the antagonistic potency of transforming growth factor-beta receptor extracellular domains by coiled coil-induced homo- and heterodimerization. *J Biol Chem* **279**, 26013-26018 (2004).
- 36 Contarino, M. R. *et al.* Modular, self-assembling peptide linkers for stable and regenerable carbon nanotube biosensor interfaces. *J Mol Recognit* **19**, 363-371 (2006).
- 37 Chao, H., Bautista, D. L., Litowski, J., Irvin, R. T. & Hodges, R. S. Use of a heterodimeric coiled-coil system for biosensor application and affinity purification. *J Chromatogr B Biomed Sci Appl* **715**, 307-329 (1998).
- 38 Yeates, T. O. & Padilla, J. E. Designing supramolecular protein assemblies. *Curr Opin Struct Biol* **12**, 464-470 (2002).
- 39 He, Y. *et al.* Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature* **452**, 198-202 (2008).

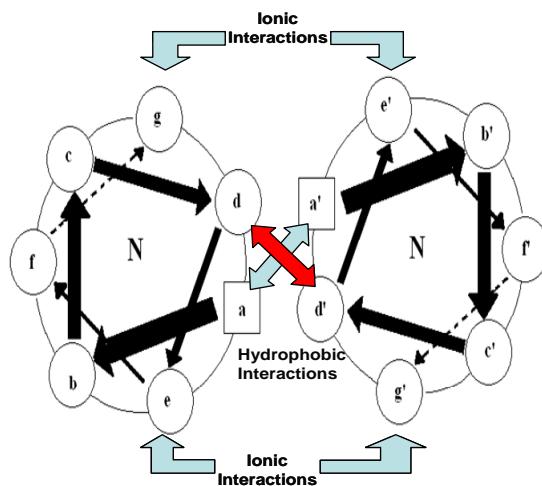
## Chapter 3

### Investigations of a Designed Coiled Coil to be Used for the Controlled Assembly of Protein Building Blocks

#### 3.1 - Introduction

The coiled coil is a structural motif formed by, in the simplest case, two  $\alpha$ -helices that coil around each other in a left handed manner to form a super helix. It is the major structural motif in many fibrous proteins and often mediates protein-protein interactions. Coiled coils are characterized by a heptad repeat of seven amino acids, denoted **a-g** by convention. A helical wheel representation of a parallel dimeric coiled coil is shown in Figure 3.1. Residues in the **a** and **d** positions are predominantly nonpolar and form the inner core of the coiled coil through interhelical hydrophobic interactions. Charged residues in positions **g** and **e** are involved in inter-helical Coulombic interactions that predominantly mediate specificity, i.e. whether homodimeric or heterodimeric, and the directionality, parallel or antiparallel, of the coiled coil formed. For parallel coiled coils, residues at the **e** position on one strand interact with residues at the **g** position on the other strand, whereas for antiparallel coiled coils, the interactions are between the two **g** residues and between the two **e** residues.<sup>1</sup> The packing of residues in the hydrophobic **a** and **d** positions can also be important in specifying parallel/antiparallel orientation. For example, the parallel orientation of the homodimeric leucine zipper region of GCN4, a yeast transcription factor, is determined by asparagine residues buried in the hydrophobic core.<sup>2</sup> In the parallel orientation asparagine residues pack next to one another and can

hydrogen bond, whereas formation of the antiparallel coiled coil would pack the hydrophilic Asp with hydrophobic residues, greatly destabilizing coiled coil formation. Steric interactions between hydrophobic residues in the **a** and **d** positions can also help specify orientation.<sup>3,4</sup>



**Figure 3.1.** Helical wheel representation of a parallel dimeric coiled coil. View is from N-terminus down the helical axis of both peptide strands. Primes are used to distinguish different strands.

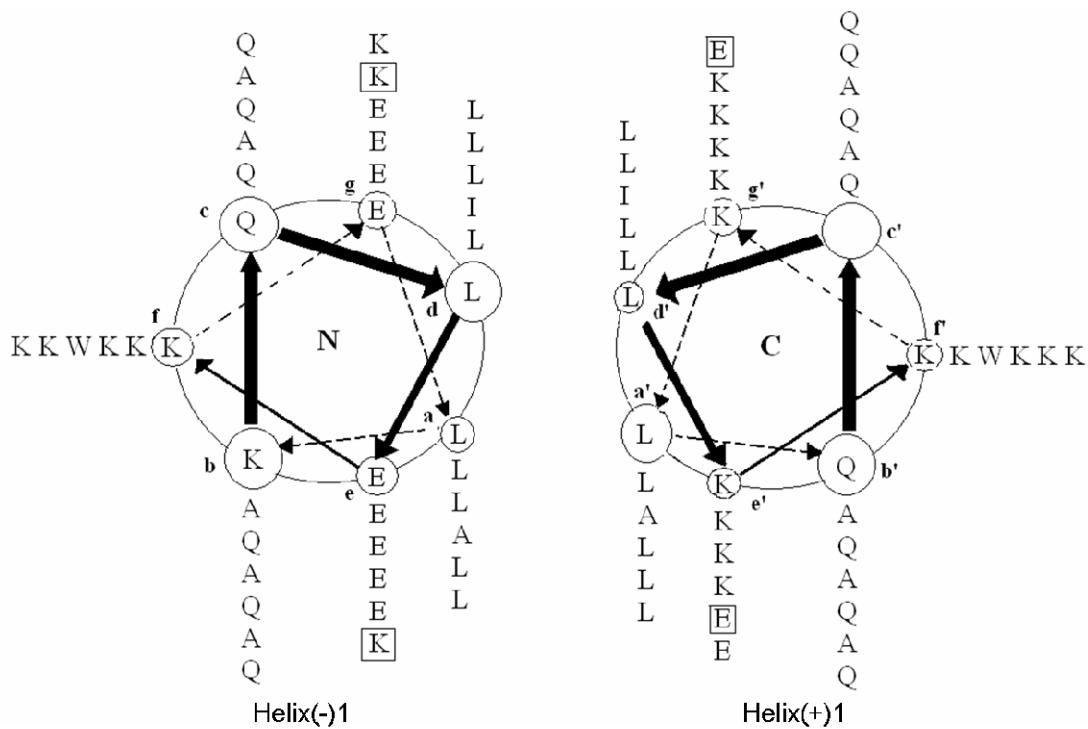
The *de novo* design of coiled coils has been a particularly well-studied area in protein science. Studies have provided great insight into factors that can alter the specificity, stability, oligomerization states, and strand orientation of coiled coils.<sup>1-78</sup> The ability to rationally design coiled coils from first principles has led to their use in applications that include protein design and modeling, hydrogels for use as potential drug delivery agents, affinity purification of proteins, biosensors, drug targeting systems, and other applications.<sup>79-100</sup> Furthermore, dimerization of protein *in vitro* and *in vivo* has been achieved by genetically engineering coiled coil domains into the protein sequence.<sup>96,101-106</sup>

An important component of my strategy for assembling proteins into higher order cage structures is to develop a universal linking agent for assembly of PBBs. In this chapter I describe the design and characterization of peptides designed to associate to form antiparallel heterodimeric coiled coils, ultimately to be used as linkers for assembling proteins into higher order structures. As discussed in the previous chapter, the heterodimerizing coiled coil peptides will allow control over the assembly process and the antiparallel orientation will provide proper spatial arrangement for cage formation.

### 3.2 - Design

The initial design of antiparallel coiled coil peptides was based primarily on work from Martha Oakley and coworkers.<sup>4,47</sup> Our design of two six-heptad peptides, Helix(-)1 and Helix(+)1, is shown in figure 3.2. The name derives from the overall charge of amino acids in the **e** and **g** positions of the peptides, which is expected to impose a heterodimeric association. Association of peptides of the same kind is expected to yield charge-charge repulsion, whereas heterodimerization leads to ion-pairing (i.e. salt bridges). To control the orientation, steric matching of hydrophobic core residues in addition to differential electrostatics was incorporated into the design. Whereas asparagine-asparagine pairing in the hydrophobic core has been shown to control orientation, it comes at the expense of stability. Steric matching of isoleucine and alanine in the hydrophobic **a** and **d** positions of the heptad repeats was chosen to maintain stability, which was thought to be crucial for keeping proteins locked together. The design packs isoleucine next to alanine only in the antiparallel orientation, whereas formation of a parallel dimer would lead to packing of isoleucine with isoleucine

generating a “bulge” and alanine with alanine generating a “hole,” both of which are destabilizing.<sup>3</sup> Complementary electrostatic interactions allow complete charge-charge pairing of residues in the **e** and **g** positions only in the antiparallel orientation, further promoting association in the antiparallel fashion. Lastly, the codon usage of the gene encoding the peptide was optimized for expression in *E. coli* and an N-terminal 6 residue histidine (6xHis) tag was incorporated to allow purification by Ni-affinity chromatography. A thrombin cleavage site was incorporated between the 6xHis tag and coiled coil peptide to allow removal of the tag if required (sequence Figure 3.2). Additional amino acids at the C-termini were included as flexible spacer sequences in the event that the peptides would be appended to the N-terminus a protein.



**Helix(-)1:**  
**cdefgabcdefgab cdefgabcdefgabcdefgab**  
**MGS SHHHHHHSSGLVPRGSHMGSKOLEKELAQALEKELAQALEKELAQALEKELAQALKKKLQGGSGGT**

**Helix(+)1:**  
**cdefgabcdefgab cdefgabcdefgabcdefgab**  
**MGS SHHHHHHSSGLVPRGTHGSSKOLEKELAQALEKKLQAIKKKLAQLKWKKAQALKKKLAQALKKKLQGGSGAS**

**Figure 3.2.** Helical wheel representation and peptide sequences of the de novo designed peptides Helix(-)1 and Helix(+). For simplicity, only amino acids encoding the coiled coil structure are shown in the helical wheel representation. Complementary electrostatic pairs incorporated to control orientation are highlighted by boxes. Heptad repeats are denoted above their position in the peptide sequence. The 6xHis tag is found near the N-terminus and the thrombin cleavage site is underlined.

### 3.3 - Experimental Procedures

#### Materials

DNA modifying enzymes and reagents were purchased from New England Biolabs. DNA primers were purchased from IDT DNA Technologies (Coralville, IA). PBS 10x stock solution was purchased from Invitrogen (Carlsbad, CA). *E. coli* BL21( $\lambda$ DE3) and expression vector pET-28b were purchased from Novagen (Madison,

WI); Pfu turbo DNA polymerase and *E. coli* XL1-Blue were from Stratagene (Cedar Creak, TX). Nickel nitrilotriacetic acid (Ni-NTA) resin, QIAquick gel extraction kit, and QIAprep Spin Miniprep kit were purchased from Qiagen (Valencia, CA). All chemical reagents (IPTG, Sodium Chloride, etc.) were from Fisher (Pittsburgh, PA).

### Peptide Expression Vectors

Genes encoding Helix(-)1 and Helix(+)1 were synthesized by Picoscript (Houston, Texas) and subcloned into the expression vector pET28b(+). The DNA sequence encoding Helix(+)1 as found in pET28b(+), with start and stop sequences underlined and unique restriction sites for genetic manipulation, is given below.

1	ATACC <u>ATGGGCTCCTCTCATCATCATCACACCACAGCTCGGCCTGGTGCCGCCGGCACCCACGGCTCGAGCAAACAG</u>	80
1	TATGGTACCCGAGGAGAGTAGTAGTAGTGAGCTGGTGTGAGGCCGACCACGGCGCCGTGGTGCCGAGCTCGTTGTC	80
	NcoI	
81	TTGGAGAAAGAACTGGCGAACACTCGAAAAGAAA <u>ACTGCAGGCCATTAAAAAAACTGGCGAATTGAAATGGAAGGCCA</u>	160
81	AACCTTTCTTGACCGCCTTGAGCTTTGACGTCCGTAATTTTTTGACCGCGTTAACCTTACCTCCGGGT	160
	PstI	MfeI
161	GGCGCTGAAAAAA <u>ACTGGCACAGCTGAAGAAAAACTCCAAGGTGGCTCCGGAGCTAGCTAATAA</u>	227
161	CCCGCAGTTTTTGACCGTGTCACTCTTTGAGGTTCCACCGAGGCCCTCGATCGATTATT	227
	PvuII	BspEI
	BmtI	
	NheI	

The DNA sequence encoding Helix(-)1 in pET28b(+) as found in pET28b(+), with start and stop sequences underlined and unique restriction sites for genetic manipulation, is given below.

1	ATACC <u>ATGGGAGCAGGCCATCATCATCACAGCAGCGGCCTGGTGCCGCCGGCAGCCATATGGGATCCAAACAG</u>	80
1	TATGGTACCCGCTGCGGTAGTAGTAGTAGTGACGTGTCGCTCGCCGGACCACGGCGCCGTGGTATACCCTAGGTTGTC	80
	NcoI	
	BamHI	
81	TTGGAAAAAGAA <u>CTGGCCCAACTGGAAAAGGAAC<u>CTGCAAGCGATTGAGAAAGAGCTCGCTCAACTGGAATGGGAGGCACA</u></u>	160
81	AACCTTTCTTGACCGGGTTGACCTTTCTTGACGTTCGCTAAC <u>CTTCTCGAGCGAGGTTGACCTTACCCCTCCGTG</u> T	160
	EcoI	
	SacI	
161	AGCTTTGGAAAAAA <u>ACTGGCGCAGCTCAAAAAAAGTTGCAGGGCGGCAGGGCGGTACCTAATAA</u>	227
161	TCGAAAC <u>CTTTTTTGACCGCGTCGAGTTTTTCAACGTCCCGCCGTGCCGCCATGGATTATT</u>	227
	HindIII	
	Acc65I	
	KpnI	

Unique restriction sites were built into the design to allow “cassette” replacement of gene segments. Double digest of plasmids with XhoI/NheI or NdeI/KpnI allows removal of the coiled coil encoding portion of Helix(+)1 and Helix(-)1 respectively.

#### PCR Modification

Genes encoding the peptides were modified by polymerase chain reaction (PCR) using Stratagene Quick Change II site mutagensis kit according to the manufacturers directions. Primers used to introduce a double point mutation into the Helix(-)1 gene to convert two lysines to glutamates were: 5'-GAAAAAAA<sup>\*</sup>ACTGGCGCAGCTCG<sup>\*</sup>AAA AAG<sup>\*</sup>AGTTGCAGGGCGGCAGCGGC-3' and 5'-GCCGCTGCCGCCCTGCAACTC<sup>\*</sup>T TTTTC<sup>\*</sup>GAGCTGCGCCAGTTTTTTC-3'. Primers used to introduce a single mutation of Helix(+)1 to convert a glutamate to lysine were: 5'-GAAC<sup>\*</sup>TGGCGCAACT CA<sup>\*</sup>AAAAGAAACTGCAGGC-3' and 5'-GCCTGCAGTTCTTTGAGTTGCGCCA GTTC-3'. The resulting modified peptides were named Helix(-) and Helix(+). Shortened versions of Helix(-) with only 4 or 5 heptads were made by insertion of a stop codon using the following primers: 5'-CAACTGGAATGGGAGGCATAAGCTTGGA AAAAAAAC-3' and 5'-GTTTTTTCAAAGCTTATGCCTCCCATTCCAGTTG-3' for the 4 heptad and primers 5'-CTTGAAAAAA<sup>\*</sup>ACTGGCGTAAC<sup>\*</sup>TCGAAAAA GAGTTGCAGGG-3' and 5'-CCCTGCAACTCTTTGAGTTACGCCAGTTTTT-3' for the 5 heptad constructs.

#### Peptide Expression and Purification

*E. coli* strains harboring expression constructs were grown on 2xTY medium at 37°C in the presence of kanamycin (50 µg/mL) to maintain selection for the plasmids. Expression of the genes was induced by addition of isopropyl β-D-thiogalactopyranoside

(IPTG) to a final concentration of 1 mM once the cells reached early log phase ( $OD_{600}=0.8$ ). Cultures were grown for 4 hours after addition of IPTG, then the cells were harvested by centrifugation and cell pellets stored at -20° C until needed.

Cell pellets were resuspended in ice-cold lysis buffer (100 mM Tris-HCl, 300 mM sodium chloride, 0.1% Triton-X 100, pH 8.0) containing 1 mM  $\beta$ -mercaptoethanol, 10  $\mu$ M PMSF. Cells were lysed on ice by sonication with a microtip for 6 minutes total run time, 30 second pulse, and 2 minute pause between bursts. Cell debris was removed by centrifugation at 24,000 g for 15 min at 4°C. The supernatant was then heated at 80°C in a water bath for 20 minutes to produce a white cloudy precipitate. Precipitated material was removed by centrifugation at 10,000 g for 15 min at 4°C. The supernatant was loaded onto a 5 mL bench top column of Ni-NTA superflow resin equilibrated in Buffer A (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 7.0). Non-specifically bound proteins were eluted by washing the column with 10 column volumes of buffer A. Finally the 6xHis-tagged peptides were eluted with Buffer A containing 500 mM imidazole. Purified protein solutions were dialyzed twice overnight against 1 L PBS (10 mM sodium phosphate, 150 mM NaCl, pH 7.0) and stored at 4°C. Protein concentrations were determined by UV absorption measured at 280 nm assuming a molar extinction coefficient  $\epsilon_{280} = 5690 M^{-1}cm^{-1}$ , peptides each contain one tryptophan for this purpose.

#### Tris-Tricine SDS-PAGE Analysis

Tris-tricine gels were made at 16.5%T/6.0%C for the separating and 4.0%T/3.0%C for stacking gel. Gels were subjected to electrophoresis overnight at 15 volts and then at 100 volts until the loading dye reached the bottom of the gel. Gels were subsequently stained with Coomasie, then destained and photographed.

### Thrombin Digest

Thrombin digests were performed according to the manufacturer's directions. 0.5 mL digests containing ~0.3 mg peptide and 0.3 units of thrombin in a mixture of PBS and cleavage buffer (10x buffer provided by manufacturer) were incubated overnight at room temperature. Samples were analyzed by HPLC and CD to determine effects of removing the 6xHis tag.

### HPLC Analysis

Chromatography was performed at room temperature with detection at 220 nm using an Alltech reverse phase C-4 column (150 mm x 4.6 mm, 5 $\mu$ ). The column was equilibrated with a solution of 0.15% TFA in deionized water and peptides eluted with an ascending gradient of Acetonitrile (95% Acetonitrile:5% Water and 0.13% TFA) using the time gradient program, shown in Table 1, with a flow rate of 1 mL/min. Peak fractions were collected for analysis by electrospray ionization mass spectrometry (ESI-MS) to confirm identity of peptides.

### Mass Spectrometry

Samples collected from the HPLC analysis was directly analyzed by time-of-flight electrospray mass spectrometry (LCT Micromass) under positive ion mode. ESI-MS samples were introduced into the spectrometer through an in-line HPLC pump in a carrier solvent of 90% methanol/10% water at 0.1 mL/min. The desolvation temperature was 150°C. The spectral data were analyzed using the MassLynx 4.0 software suite. Expected masses were calculated from sequences using ExPASy compute pI/Mw tool.

### Circular Dichroism

Circular dichroism (CD) spectra were recorded on an Aviv model 202 circular dichroism spectrometer. Rectangular quartz cuvettes of 1 mm pathlength were used for all experiments. Experiments were run at 25°C. Typically, CD spectra were the average of three scans between 190 and 250 nm. Molar mean residue ellipticity ( $\theta$ ) was calculated from the equation:

$$\theta = (\theta_{\text{obs}} \times \text{MRW}) / (10 \times l \times C)$$

where  $\theta_{\text{obs}}$  is the ellipticity measured in millidegrees, MRW is the mean residue molecular weight (molecular weight of peptide divided by the number of amino acids), l is the optical pathlength in centimeters, and c is the concentration in milligrams per milliliter. Temperature denaturation was investigated by taking scans of a sample at 25°C and 90°C, and again at 25°C after cooling. Guanidine hydrochloride studies were carried out by monitoring the CD at 222 nm. Peptide samples were prepared from stock solutions of 168  $\mu$ M peptide in buffer (10 mM sodium phosphate, 100 mM sodium chloride, 1 mM EDTA, pH 7.5) and 8 M Guanidine in buffer (same as peptide). Fraction unfolding was determined using the methods described by De Francesco et al. and Litowski et al.<sup>46,55,107</sup> Denaturation data was fit to a dimer association model to determine  $\Delta G_{\text{unfold}}$  and m (m is a constant of proportionality with dimensions of kcal mol<sup>-1</sup> M<sub>GuHCl</sub><sup>-1</sup>) using the program Igor Pro.

### Sedimentation Equilibrium Analytical Ultracentrifugation

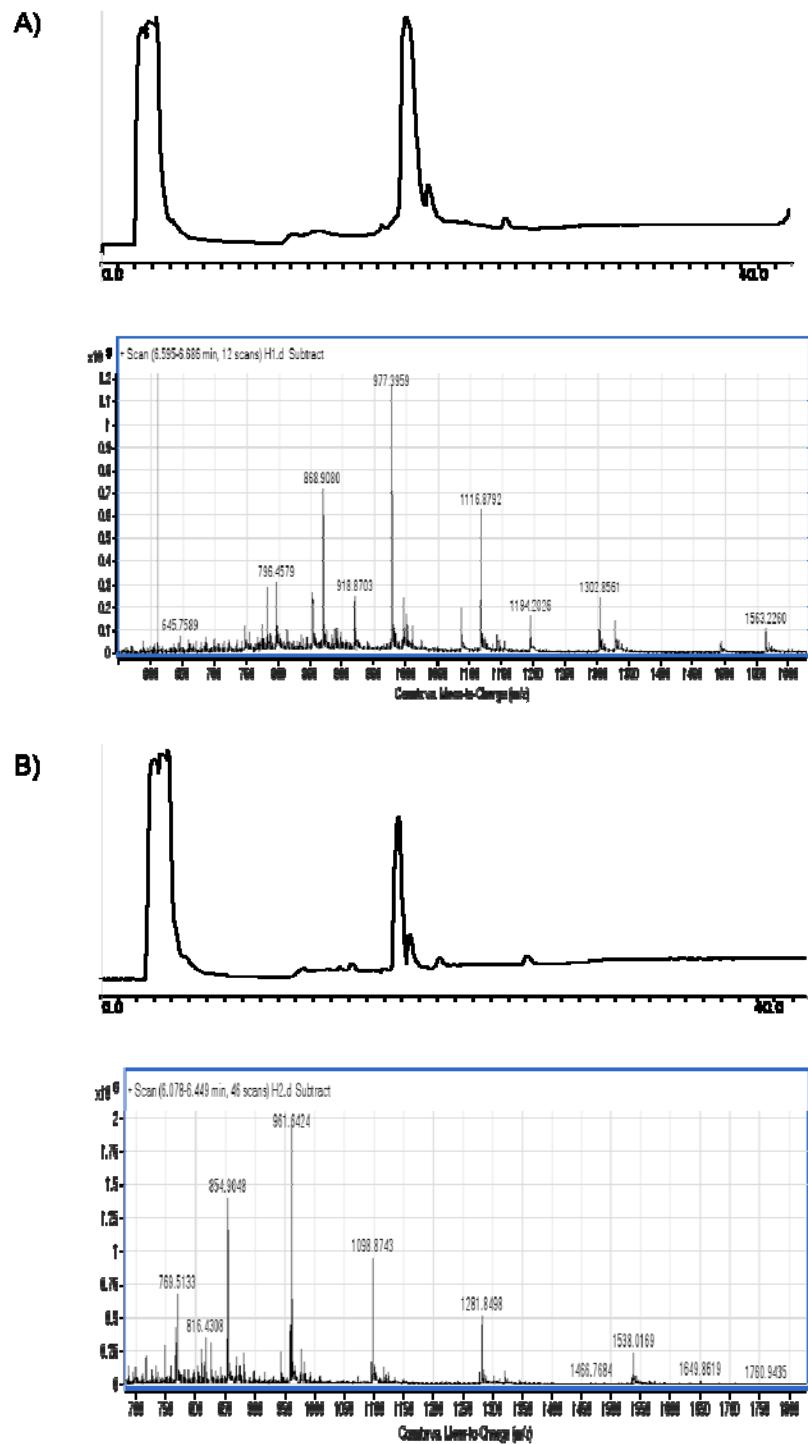
Sedimentation equilibrium analytical ultracentrifugation (SE AUC) experiments were performed on a Beckman XL-I analytical ultracentrifuge. Samples with A<sub>280</sub> of 0.3, 0.5, and 0.7 were loaded into Epon 6 channel centrifuge cells and spun at four speeds (30

krpm, 35 krpm, 40krpm, and 45 krpm). Temperature was set at 20°C. Equilibrium was verified by observing overlap of scans taken several hours apart. Data for scans observed to be at equilibrium were analyzed using the software package Ultrascan using global fit analysis function. Partial specific volumes were calculated from peptide sequences using the program Sednterp.

### 3.4 - Results and Discussion

#### Expression, Purification, and Mass Analysis of Helix(-)1 and Helix(+)1

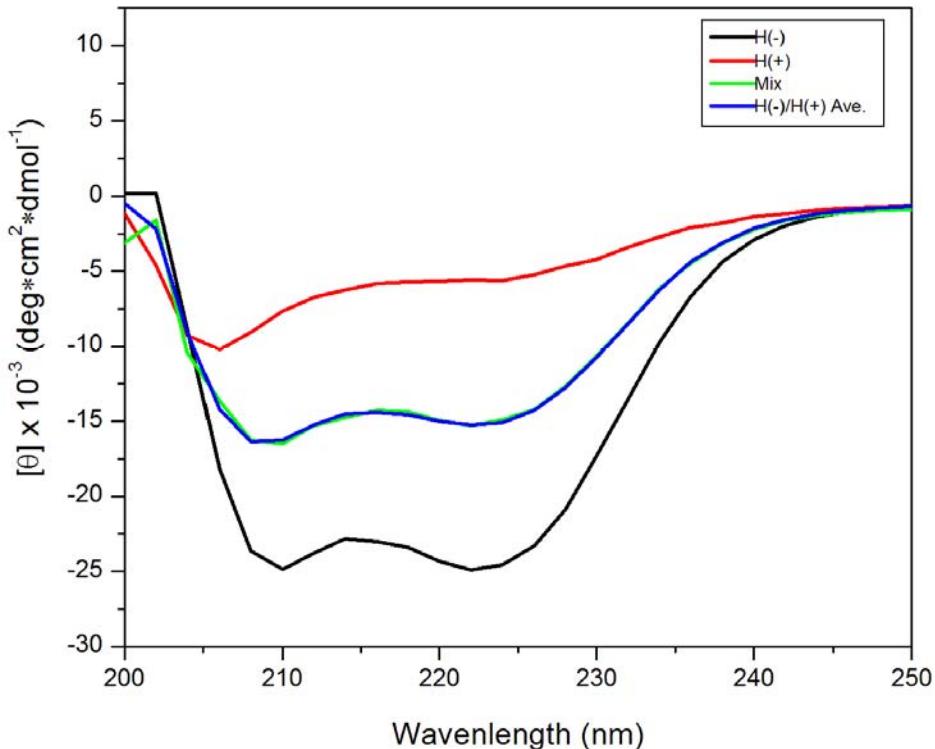
Often, short *de novo* designed peptides can not be obtained through bacterial host overexpression.<sup>108</sup> Small peptides are degraded by bacterial host proteases and can also be toxic to the host. However Helix(-)1 and Helix(+)1, which are 72 residues long, were successfully overexpressed and purified from *E. coli*. Chromatograms from HPLC of nickel affinity-purified samples showed a single major peak during the separation gradient. Mass spectrometry of HPLC fractions gave molecular weights expected for Helix(-)1 and Helix(+)1 with removal of methionine from their N-terminus and serine from the C-terminus of Helix(+)1 (Figure 3.3).



**Figure 3.3.** HPLC chromatograms and ESI-MS spectra of Helix(-)1 and Helix(+)1 peptides. A) Helix(-)1 HPLC chromatogram (top) and the ESI-MS spectrum (bottom) of a fraction taken from the HPLC elution. Deconvolution gives molecular weight of 7811.2 (Calculated 7811.8 Da with loss of Met). B) Helix(+)1 chromatogram (top) and the ESI-MS spectrum (bottom) of a fraction taken from the HPLC elution. Deconvolution gives 7685.13 (Calculated 7688.00 Da with loss of Met and Ser).

### Investigations of Helix(-)1 and Helix(+)1 by CD

Analysis of peptide samples by CD were expected to show spectra consistent with a random coil structure for the individual peptides and the emergence of helical structure upon mixing of Helix(-)1 with Helix(+)1. The spectra obtained for Helix(+)1 shows mostly random coil as expected, although there is a small amount of helicity as judged by a small dip at 222 nm. However, Helix(-)1 exhibited spectra typical for alpha-helix structure, as judged by the absorption minima at 208 and 222 nm. In addition, samples of 1:1 mixtures of Helix(-)1 with Helix(+)1 resulted in CD spectra equivalent to the average of the spectra obtained for the individual peptides, suggesting that the Helix(-)1 self associates and doesn't form a coiled coil with Helix(+)1. A thermal denaturation study of Helix(-)1 showed that it retained most of the helical structure up to the temperature limit of 90°C. Furthermore, addition of guanidine hydrochloride to 4M showed little disruption of the helical secondary structure. Thrombin digests of the peptides was performed to remove the 6xHistidine tag and determine if removal would have any effect on secondary structure. Complete removal of the 6xHistidine tag was verified by HPLC. However analysis by CD showed little change in helical content in comparison to the undigested peptides.

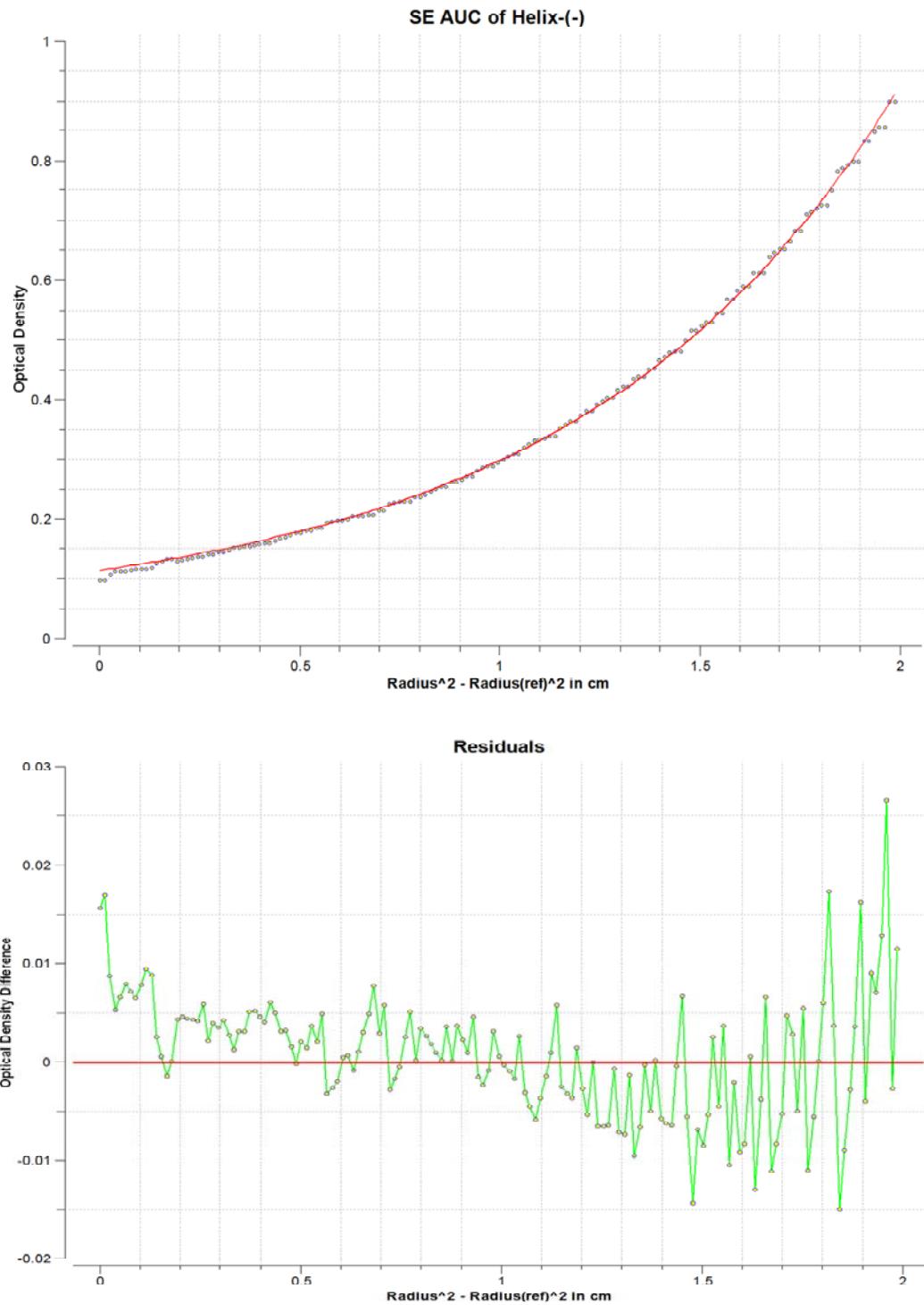


**Figure 3.4.** CD spectra and analysis of the designed peptides Helix(-)1, Helix(+)1, and their 1:1 mixture. The spectra for Helix(-)1 (black trace) indicates the presence of helical structure. The spectra for Helix(+)1 (red trace) is consistent with that of the expected random coil. The 1:1 mixture of the two peptides (green trace) is equivalent to the average of the spectra of the individual peptides (blue trace), indicating no association occurs upon mixing.

#### SE AUC studies of Helix(-)1 and Helix(+)1

SE AUC studies were performed to determine the oligomeric state of the peptides. Helix(-)1 fit well to a single species with a molecular weight of 14,940 Da, close to the calculated weight of 15,616 Da expected for a dimer. The SE AUC results for Helix(-)1 explain the observations that Helix(-)1 is highly helical due to self association to form a dimeric coiled coil. Surprisingly, Helix(+)1 did not fit to a single component species as expected from CD data showing the peptide to be relatively unstructured. Attempts to fit

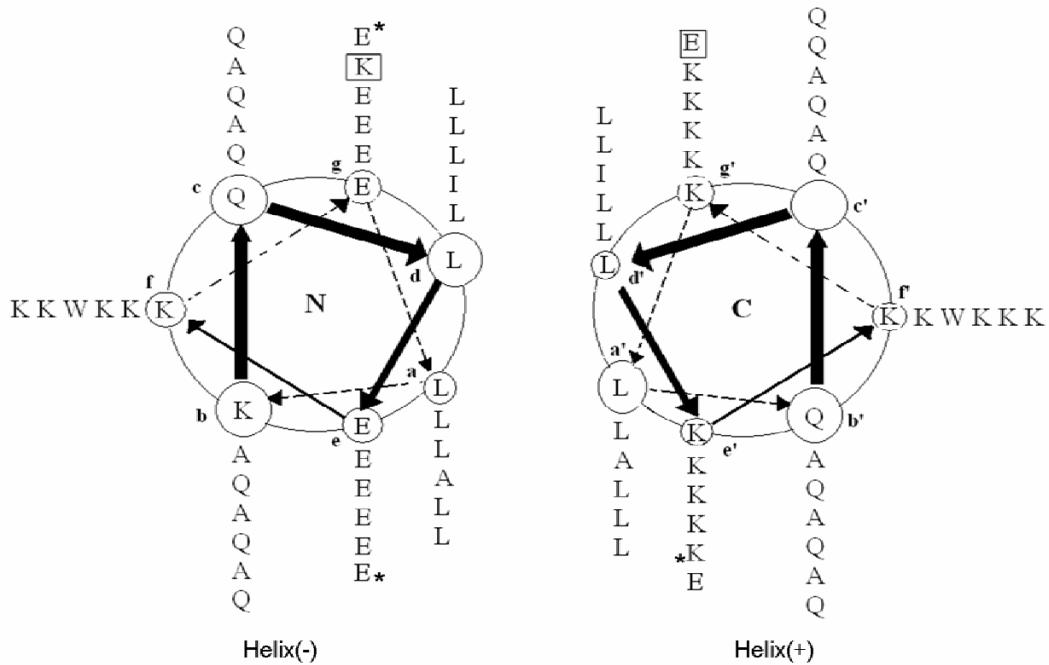
the sedimentation data from Helix(+)1 suggest it self associates. Although simple models (i.e. single component, monomer-dimer, etc.) did not generate good fits to the Helix(+)1 SE AUC data, molecular weight values propagated towards the value of a tetramer, but residuals indicated that other species were present. This is interesting considering that Helix(+)1 is found to be unstructured by CD. However, large concentration gradients are generated during analytical ultracentrifugation, possibly leading to oligomerization of peptides observed to be unstructured by CD.<sup>55</sup>



**Figure 3.5.** Representative SE AUC data scan for Helix(-)1 with data fitted to a single component and residuals for the fit. Helix(-)1 SE AUC data fit to a single species with a molecular weight of 14,940 Da, close to the calculated weight of 15,616 Da for a dimer. Residuals indicate a relatively good fit to the data.

### Peptide Redesign

The data from the SE AUC studies and CD is consistent with self association of Helix(-)1 to form a homodimeric coiled coil. Homodimerization of peptides designed to form heterodimers is not without precedent.<sup>41,52,55</sup> Litowski and Hodges found that peptides with sequences IAAL for **a**, **b**, **c** and **d** positions in the heptad repeat form nearly fully folded homodimeric coiled coils even with 8 Lys-Lys or 8 Glu-Glu residues incorporated to provide electrostatic repulsions. Arndt and coworkers found peptides selected to form highly stable heterodimeric coiled coils also formed homodimers when the complementary peptides were not mixed.<sup>41,52</sup> Mason and Arndt have suggested that the overall electrostatic charge may play a role in peptides displaying non-complementary interactions.<sup>1</sup> Because our design incorporated electrostatic interactions designed to enforce an antiparallel orientation (see Figure 3.2), the overall repulsive charge to minimize self association was likely reduced. To disrupt self dimerization of Helix(-)1, we redesigned the peptide to have only one lysine at the electrostatic interface (Figure 3.6). In addition, Helix(+)1 was altered by changing one glutamate to lysine, which would leave one differential electrostatic pair between the peptides to help control orientation. The new peptides were named Helix(-) and Helix(+).



**Helix(-):**

cdefgab**cdefgab****cdefgab****cdefgab****cdefgab****cdefgab****cdefgab**

MGS SHHHHHHSSGLVPRGHMGSKOLEELAQLEELAQLEEWAALEEKLAQLEKELQGGSGGT

**Helix(+):**

cdefgab**cdefgab****cdefgab****cdefgab****cdefgab****cdefgab****cdefgab**

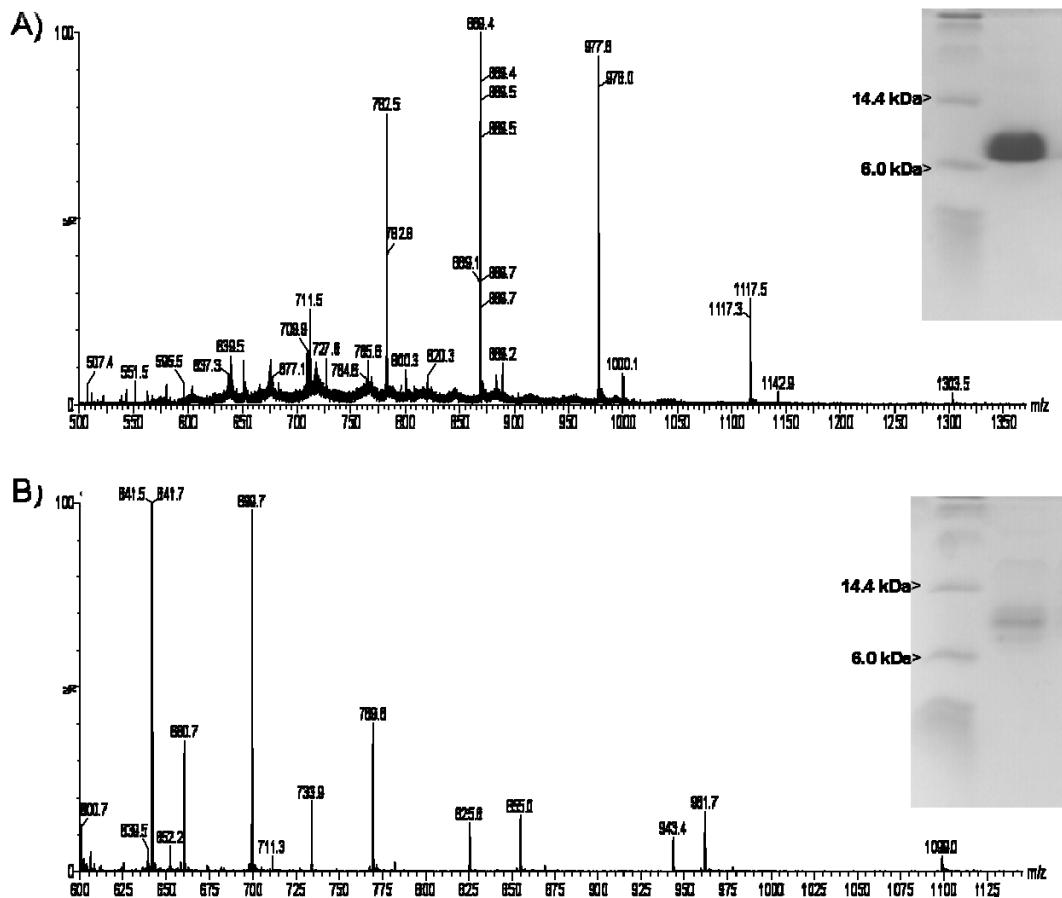
MGS SHHHHHHSSGLVPRGTHGSKOLEELAQLKKKLAQIKKKLAQLKKKLAQLKKKLQGGSGAS

**Figure 3.6.** Helical wheel representation and peptide sequences of the designed modified peptides Helix(-) and Helix(+). For simplicity, only amino acids encoding the coiled coil structure are shown in the helical wheel representation. Complementary electrostatic pairs incorporated to control orientation are outlined by boxes. Heptad repeats are denoted above their position in the peptide sequence. The 6xHis tag is found near the N-terminus and the thrombin cleavage site is underlined. Amino residues changed from Helix(-)1 and Helix(+)1 are denoted by asterisks in the helical wheel representation and are underlined in the peptide sequence.

#### Expression, Purification, and Mass Analysis of Helix(-) and Helix(+)

Over-expression of Helix(-) and Helix(+) in *E. coli* and subsequent purification yielded pure peptides of expected size as observed by tris-tricine SDS-PAGE (Figure 3.7). Mass spectrometry analysis of samples taken from HPLC gave a mass of 7687.8 Da for Helix(+) consistent with the 7688.06 Da expected for removal of Met at the N-

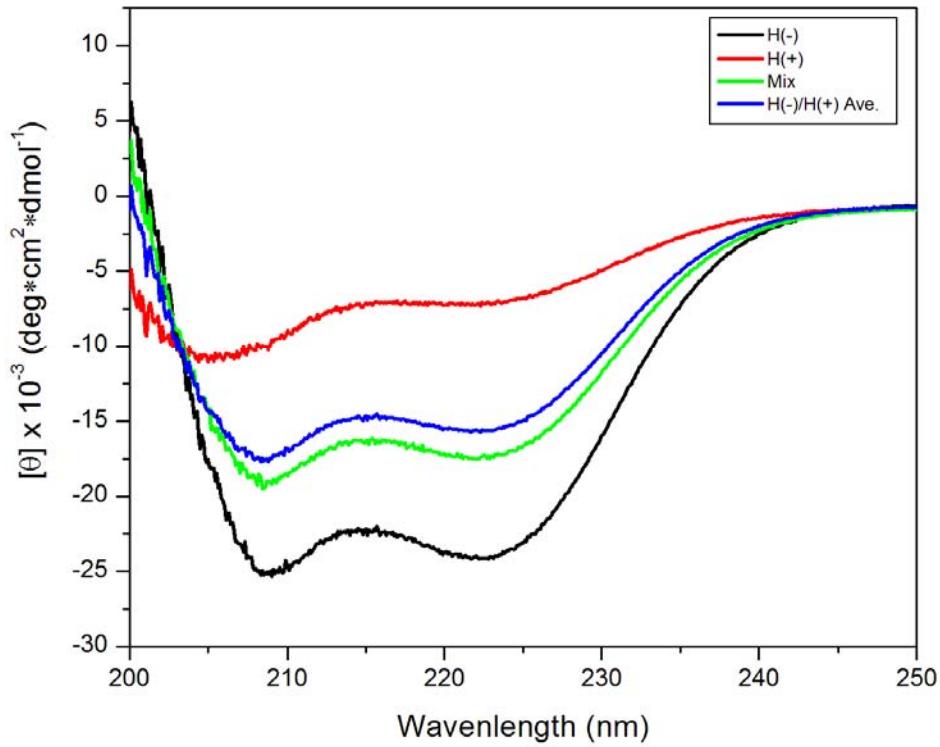
terminus and Ser at the C-terminus, and 7815 Da for Helix(-) consistent with the loss of Met from terminus, calculated to be 7813.7 Da.



**Figure 3.7.** SDS-PAGE and ESI-MS data analysis for Helix(-) and Helix(+). A) ESI-MS spectrum and tris-tricine gel of Helix(-) (inset) purified via Ni-affinity chromatography. B) ESI-MS and Tris-Tricline gel (inset) of Helix(+) purified via Ni-affinity chromatography. Deconvolution of the peaks gives a mass of 7815 Da for Helix(-) and 7687.8 for Helix(+) consistent with masses expected for removal of Met from the N-terminus for both and an additional removal of Ser from the C-terminus of Helix(+), 7813.7 Da and 7688.06 Da respectively.

### Investigations of Helix(-) and Helix(+) by CD

CD of Helix(-) and Helix(+) showed relatively little change from the original design. Helix(-) displayed strong minima signals at 202 and 208 nm indicating the presence of alpha-helical structure. The CD for Helix(+) gave a signal consistent with a mostly random coil structure. Mixtures of the two peptides in a 1:1 ratio gave spectra that were slightly more helical than the average for the CD spectra of the individual peptides, although the increase was negligible in comparison to the helicity observed for Helix(-). In addition, no change in CD of the 1:1 mixture was observed after incubation of the mixture for several days. The results suggest that Helix(-) predominantly self associates with very little, if any, hetero-association with Helix(+).

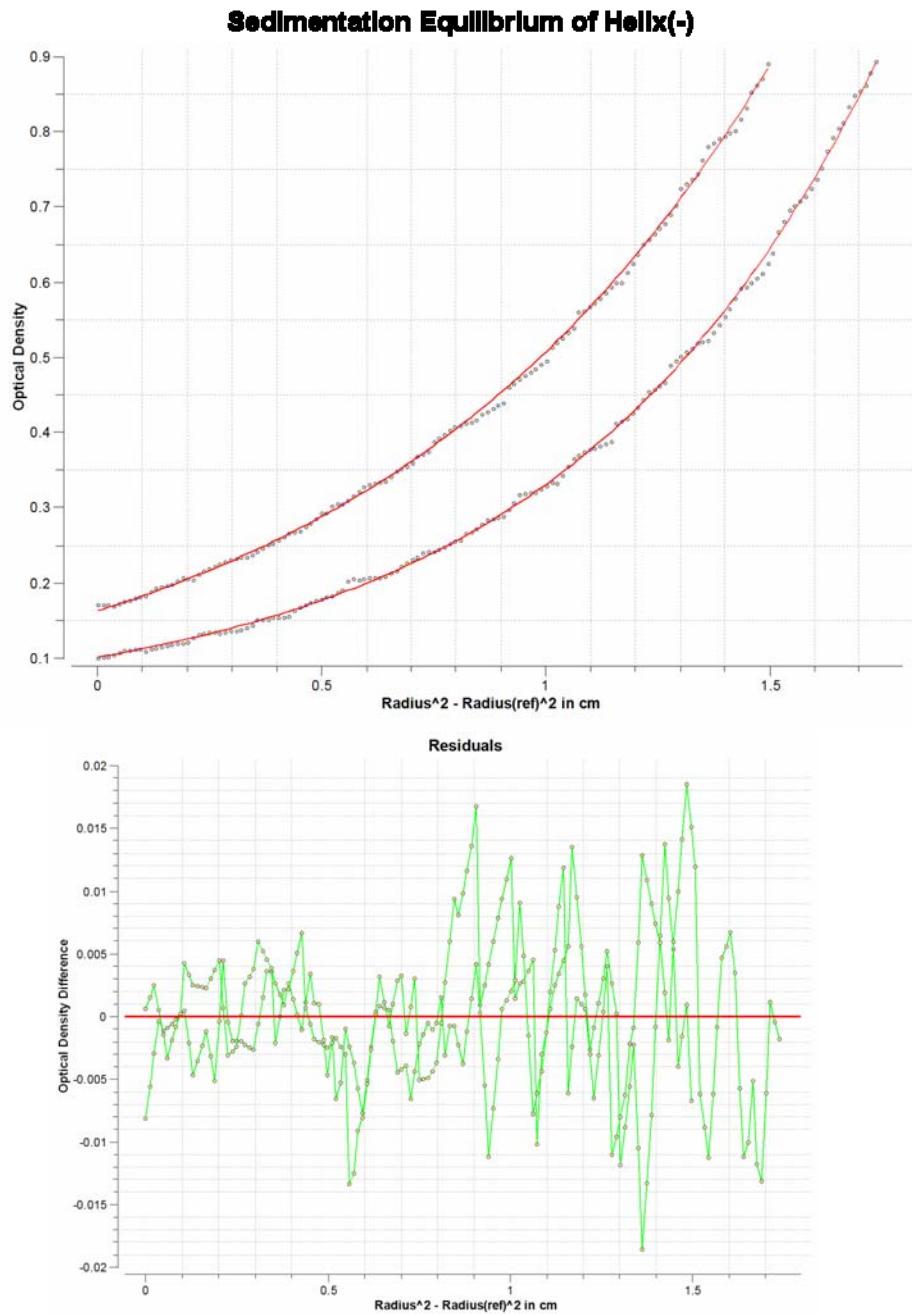


**Figure 3.8.** CD spectra and analysis of the redesigned peptides Helix(-) and Helix(+). Helix(-) (black trace) spectra shows helical structure and Helix(+) (red trace) shows mostly random coil structure. The spectra for the 1:1 mixture of Helix(-) and Helix(+) (green trace) is slightly more helical than the average of the spectra for the two individual peptides (blue trace), suggesting some hetero-association may occur, but Helix(-) is mostly self associating.

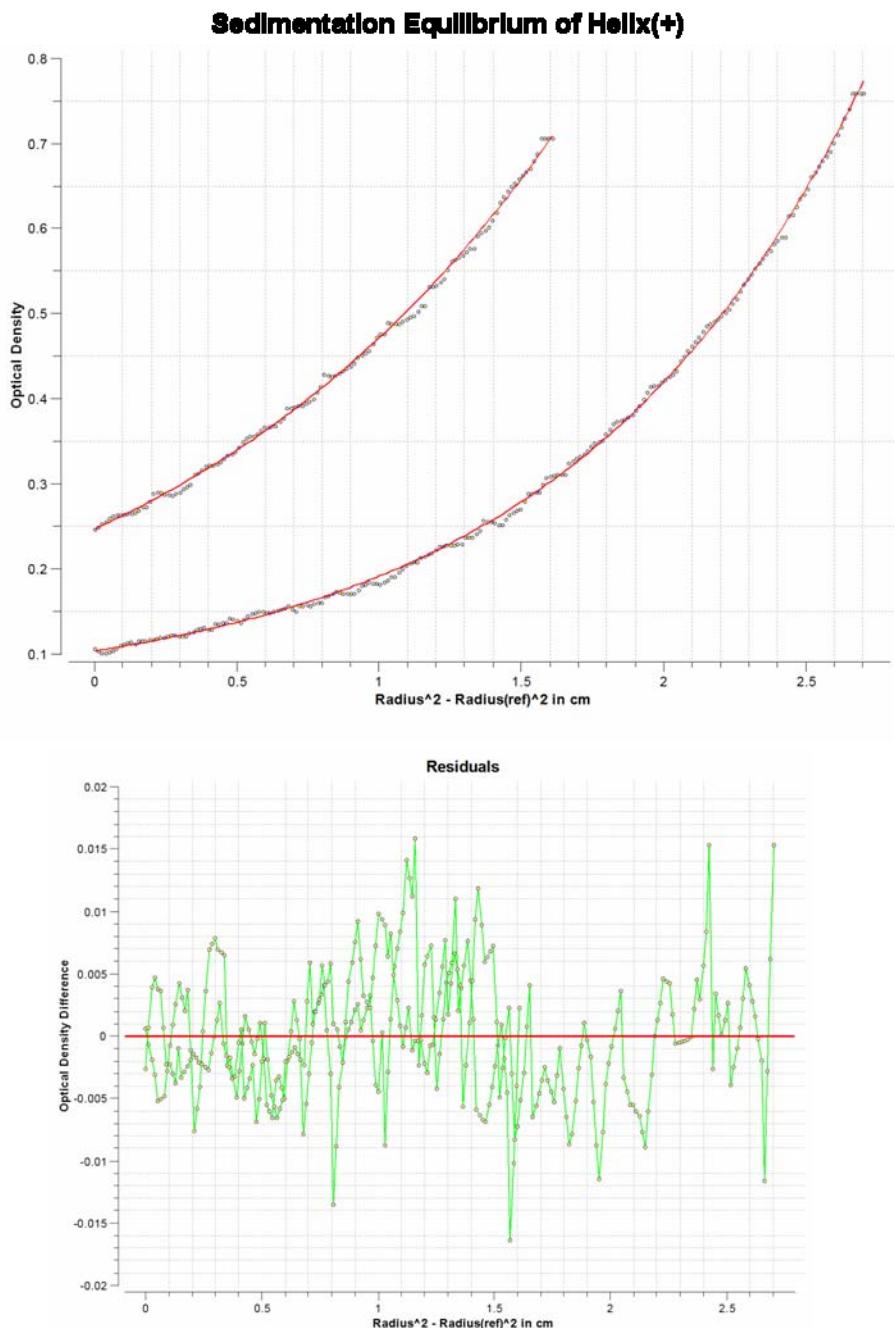
#### SE AUC studies of Helix(-) and Helix(+)

SE AUC results showed that Helix(-) fits well to a monomer-dimer equilibrium model with it existing primarily as a species of 15,180 Da, consistent with the molecular weight of a dimer (calculated dimer=15,634 Da). Helix(+) data fit to single component model with a molecular weight of 14,430 Da, a molecular weight close to the molecular weight calculated for a dimer (calculated dimer=15,376 Da). The results from CD of

Helix(+) suggested that it was not self associating to form a regular oligomer, but was largely monomeric. Oligomerization of peptides occurs, in general, with formation of alpha-helix or beta-sheet structures between peptide, which was not observed by CD. However, the high concentration gradient experienced in the analytical ultracentrifuge likely promoted association, which has been suggested for other peptides.<sup>55</sup> Plots of the fitted data with residuals are given in Figure 3.9 and 3.10 for Helix(-) and Helix(+) respectively.



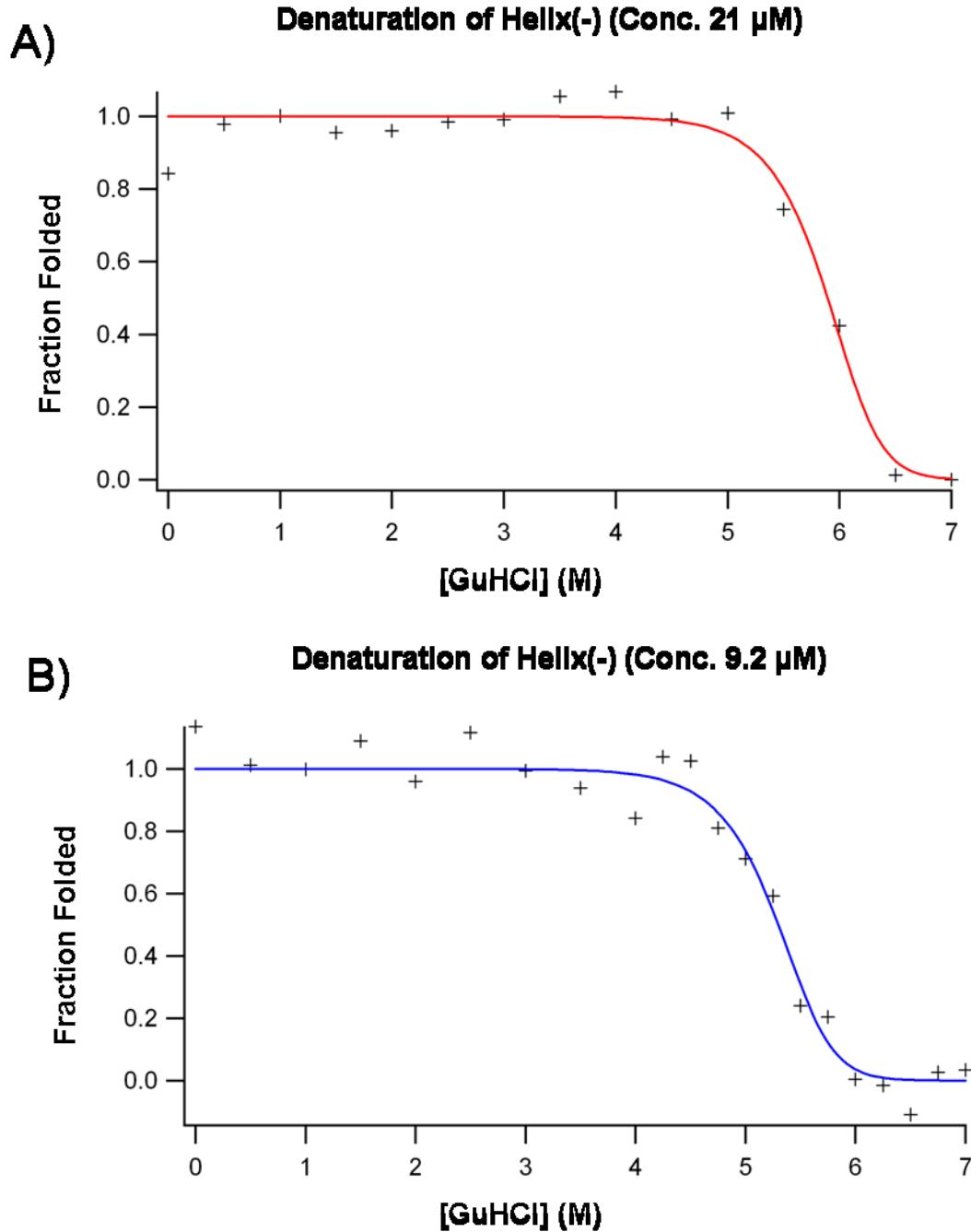
**Figure 3.9.** SE AUC data and fits with residuals for Helix(-). Helix(-) SE AUC data traces (blue circles) are at speeds of 35,000 and 40,000 rpm fit to monomer-dimer equilibrium model (red line). Molecular weight of 15,180 Da determined from fit with an expected molecular weight of 15,634 Da for a dimer. Plots of the residuals for the fits of the scans are given below the sedimentation equilibrium scan data and indicate good fitting to the selected model.



**Figure 3.10.** SE AUC data and fits with residuals for Helix(+). Helix(+) SE AUC data traces (blue circles) at speeds of 30,000 and 35,000 rpm fit to a single component species (red line). Molecular weight of 14,430 Da determined from fit with an expected molecular weight of 15,376 for dimer. Plots of the residuals for the fits of the scans are given below the sedimentation equilibrium scan data and indicate good fitting to the selected model.

### Determination of the Stability of Helix(-) homo-dimer

Guanidinium hydrochloride denaturation studies were performed to determine the overall stability of the Helix(-) dimer. Denaturation curves are consistent with cooperative two state folding by Helix(-) (Figure 3.11). The Helix(-) dimer is very stable with denaturation midpoints above 5 M for the two different concentrations (9.2  $\mu$ M and 21  $\mu$ M). Fitting the denaturation curves to a dimeric model gave  $\Delta G_{\text{unfold}}$  of  $24.5 \pm 2.6$  kcal/mol and  $26.8 \pm 3.9$  kcal/mol at concentrations of 9.2  $\mu$ M and 21  $\mu$ M, respectively. Highly cooperative folding is evident from the m values of  $3.3 \pm 0.$  kcal  $\text{mol}^{-1}$   $\text{M}_{\text{GuHCl}}^{-1}$  for 9.2  $\mu$ M and  $3.5 \pm 0.65$  kcal  $\text{mol}^{-1}$   $\text{M}_{\text{GuHCl}}^{-1}$  for 21  $\mu$ M determined from the fits.



**Figure 3.11.** Denaturation of Helix(-) and determination of  $\Delta G_{\text{unfold}}$  and  $m$ . A) Fraction folded data for Helix(-) at a concentration of 21  $\mu$ M and fit to a dimeric species. B) Fraction folded data for the Helix(-) at a concentration of 9.2  $\mu$ M and fit to a dimeric species. The fits gave  $\Delta G_{\text{unfold}}$  of  $24.5 \pm 2.6$  kcal/mol and  $26.8 \pm 3.9$  kcal/mol and  $m$  values of  $3.3 \pm 0.45$  kcal  $\text{mol}^{-1} \text{M}_{\text{GuHCl}}^{-1}$  and  $3.5 \pm 0.65$  kcal  $\text{mol}^{-1} \text{M}_{\text{GuHCl}}^{-1}$  for the data at peptide concentrations of 9.2  $\mu$ M and 22  $\mu$ M, respectively.

### Reduced Heptad Helix(-) Constructs

Our finding that increased electrostatic charge does not deter self association of our peptides agrees with the findings of Litowski and Hodges.<sup>55</sup> In their studies, increasing the stability of the sequence and number of heptad repeats caused peptides designed to form heterodimeric coiled coils to self associate. In addition, the most stable heterodimeric peptide pair discovered by Arndt and coworkers contained two g/e' pairs predicted to be repulsive, suggesting that favorable packing of the core was in part responsible.<sup>1</sup> The most stable peptides in these studies contained 4 heptad repeats, whereas our design incorporated 6 heptad repeats. Furthermore, it was found that the peptides in all these studies formed heterodimers upon mixing of complementary peptides, whereas the peptides in our study were never found to heterodimerize. Co-expression of the two peptides was unsuccessful; only Helix(-) expressed whereas Helix(+) was not observed to express. It was thought that the additional heptad repeats in our peptide system provided more stability to homodimers of Helix(-). To investigate our system further, plasmid constructs encoding Helix(-) with 4 and 5 heptad repeats were developed and attempts were made to express the peptides in *E.coli*. Although these peptides were still of relatively long length, even trace amounts of the peptide could not be found upon purification. This is consistent with loss of structure, as unstructured peptides are more amenable to degradation when expressed in *E. coli*.

#### 3.5 - Conclusions

I described the design of two heterodimeric antiparallel coiled coil peptides. These peptides can be overexpressed in *E.coli* and purified to high purity. However, the peptides exhibit self association with Helix(-)1 and Helix(-) forming a very stable dimeric

coiled coil. Homodimerization of peptides designed to form heterodimers is not uncommon and has been observed in peptides with sequences encoding greater structural stability. Results from our investigations provide further evidence for the loss of specificity when incorporating sequences for greater stability. One explanation for this phenomenon is that the overall destabilization of homodimerization from charge-charge repulsion is not enough to overcome the stability provided by hydrophobic packing of the amino acids in the **a** and **d** positions. Sequence modifications to increase electrostatic repulsive forces were unable to overcome the self-dimerization of the peptide constructs described herein. However, the modifications made to Helix(-)1 to make Helix(-) changed the overall charge of the peptide slightly, from overall charge of the coiled coil region of -1 to -5. Helix(+1) and Helix(+) were not observed to form coiled coils under normal conditions and contained the greatest overall electrostatic charge for the coiled coil sequence, +12 and +13 respectively. Our findings suggest that overall charge plays a significant role in determining specificity when sequences encoding highly stable structures are used. Trying to avoid self-dimerization by reducing the number of heptad repeats in Helix(-), thus lowering the stability of association, led to the inability to harvest the peptide from *E. coli*. This finding is consistent with loss of structure, which often enables peptides to be degraded.<sup>4</sup> This study illustrates the importance of balancing elements that dictate specificity with those that enhance stability when designing coiled coil peptides. Furthermore, this study illustrates that design principles are far from complete, especially for larger peptides.

The goal of this research was to develop a pair of heterodimer-forming peptides for linking protein building blocks together to form supramolecular protein architectures.

While these studies don't show association of the peptides to form heterodimers, attachment to larger protein subunits could alter the peptide properties and eliminate self association. The high stability of the coiled coil interaction is desired for robust linking of PBBs together; therefore, we chose to use the final 6-heptad peptide designs, Helix(-) and Helix(+), in our design strategy for constructing protein cages.

### 3.6 – References

- 1 Mason, J. M. & Arndt, K. M. Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem* **5**, 170-176 (2004).
- 2 Gonzalez, L., Jr., Woolfson, D. N. & Alber, T. Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat Struct Biol* **3**, 1011-1018 (1996).
- 3 Monera, O. D., Zhou, N. E., Lavigne, P., Kay, C. M. & Hodges, R. S. Formation of parallel and antiparallel coiled-coils controlled by the relative positions of alanine residues in the hydrophobic core. *J Biol Chem* **271**, 3995-4001 (1996).
- 4 Gurnon, D. G., Whitaker, J. A. & Oakley, M. G. Design and characterization of a homodimeric antiparallel coiled coil. *J Am Chem Soc* **125**, 7518-7519, doi:10.1021/ja0357590 (2003).
- 5 Zhu, B. Y., Zhou, N. E., Semchuk, P. D., Kay, C. M. & Hodges, R. S. Design, synthesis and structural characterization of model heterodimeric coiled-coil proteins. *Int J Pept Protein Res* **40**, 171-179 (1992).
- 6 Lovejoy, B. et al. Crystal structure of a synthetic triple-stranded alpha-helical bundle. *Science* **259**, 1288-1293 (1993).
- 7 Monera, O. D., Zhou, N. E., Kay, C. M. & Hodges, R. S. Comparison of antiparallel and parallel two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization. *J Biol Chem* **268**, 19218-19227 (1993).
- 8 O'Shea, E. K., Lumb, K. J. & Kim, P. S. Peptide 'Velcro': design of a heterodimeric coiled coil. *Curr Biol* **3**, 658-667 (1993).
- 9 Harbury, P. B., Kim, P. S. & Alber, T. Crystal structure of an isoleucine-zipper trimer. *Nature* **371**, 80-83 (1994).
- 10 Krylov, D., Mikhailenko, I. & Vinson, C. A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions. *EMBO J* **13**, 2849-2861 (1994).
- 11 Monera, O. D., Kay, C. M. & Hodges, R. S. Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci* **3**, 1984-1991 (1994).
- 12 Monera, O. D., Kay, C. M. & Hodges, R. S. Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils. *Biochemistry* **33**, 3862-3871 (1994).

- 13 Su, J. Y., Hodges, R. S. & Kay, C. M. Effect of chain length on the formation and stability of synthetic alpha-helical coiled coils. *Biochemistry* **33**, 15501-15510 (1994).
- 14 Zhou, N. E., Kay, C. M. & Hodges, R. S. The net energetic contribution of interhelical electrostatic attractions to coiled-coil stability. *Protein Eng* **7**, 1365-1372 (1994).
- 15 Zhou, N. E., Kay, C. M. & Hodges, R. S. The role of interhelical ionic interactions in controlling protein folding and stability. De novo designed synthetic two-stranded alpha-helical coiled-coils. *J Mol Biol* **237**, 500-512 (1994).
- 16 Betz, S., Fairman, R., O'Neil, K., Lear, J. & DeGrado, W. Design of two-stranded and three-stranded coiled-coil peptides. *Philos Trans R Soc Lond B Biol Sci* **348**, 81-88 (1995).
- 17 Kohn, W. D., Kay, C. M. & Hodges, R. S. Protein destabilization by electrostatic repulsions in the two-stranded alpha-helical coiled-coil/leucine zipper. *Protein Sci* **4**, 237-250 (1995).
- 18 Kohn, W. D., Monera, O. D., Kay, C. M. & Hodges, R. S. The effects of interhelical electrostatic repulsions between glutamic acid residues in controlling the dimerization and stability of two-stranded alpha-helical coiled-coils. *J Biol Chem* **270**, 25495-25506 (1995).
- 19 Lavigne, P. *et al.* Preferential heterodimeric parallel coiled-coil formation by synthetic Max and c-Myc leucine zippers: a description of putative electrostatic interactions responsible for the specificity of heterodimerization. *J Mol Biol* **254**, 505-520 (1995).
- 20 Lumb, K. J. & Kim, P. S. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* **34**, 8642-8648 (1995).
- 21 Nautiyal, S., Woolfson, D. N., King, D. S. & Alber, T. A designed heterotrimeric coiled coil. *Biochemistry* **34**, 11645-11651 (1995).
- 22 Woolfson, D. N. & Alber, T. Predicting oligomerization states of coiled coils. *Protein Sci* **4**, 1596-1607 (1995).
- 23 Boice, J. A., Dieckmann, G. R., DeGrado, W. F. & Fairman, R. Thermodynamic analysis of a designed three-stranded coiled coil. *Biochemistry* **35**, 14480-14485 (1996).
- 24 Lavigne, P., Sonnichsen, F. D., Kay, C. M. & Hodges, R. S. Interhelical salt bridges, coiled-coil stability, and specificity of dimerization. *Science* **271**, 1136-1138 (1996).
- 25 Lupas, A. Coiled coils: new structures and new functions. *Trends Biochem Sci* **21**, 375-382 (1996).
- 26 Monera, O. D., Sonnichsen, F. D., Hicks, L., Kay, C. M. & Hodges, R. S. The relative positions of alanine residues in the hydrophobic core control the formation of two-stranded or four-stranded alpha-helical coiled-coils. *Protein Eng* **9**, 353-363 (1996).
- 27 Sosnick, T. R., Jackson, S., Wilk, R. R., Englander, S. W. & DeGrado, W. F. The role of helix formation in the folding of a fully alpha-helical coiled coil. *Proteins* **24**, 427-432 (1996).

- 28 Yu, Y., Monera, O. D., Hodges, R. S. & Privalov, P. L. Investigation of electrostatic interactions in two-stranded coiled-coils through residue shuffling. *Biophys Chem* **59**, 299-314 (1996).
- 29 Yu, Y., Monera, O. D., Hodges, R. S. & Privalov, P. L. Ion pairs significantly stabilize coiled-coils in the absence of electrolyte. *J Mol Biol* **255**, 367-372 (1996).
- 30 Kohn, W. D., Kay, C. M. & Hodges, R. S. Positional dependence of the effects of negatively charged Glu side chains on the stability of two-stranded alpha-helical coiled-coils. *J Pept Sci* **3**, 209-223 (1997).
- 31 Moitra, J., Szilak, L., Krylov, D. & Vinson, C. Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. *Biochemistry* **36**, 12567-12573 (1997).
- 32 Ogiara, N. L., Weiss, M. S., Degrado, W. F. & Eisenberg, D. The crystal structure of the designed trimeric coiled coil coil-VaLd: implications for engineering crystals and supramolecular assemblies. *Protein Sci* **6**, 80-88 (1997).
- 33 Szilak, L., Moitra, J. & Vinson, C. Design of a leucine zipper coiled coil stabilized 1.4 kcal mol-1 by phosphorylation of a serine in the e position. *Protein Sci* **6**, 1273-1283 (1997).
- 34 Wolf, E., Kim, P. S. & Berger, B. MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci* **6**, 1179-1189 (1997).
- 35 Kohn, W. D., Kay, C. M. & Hodges, R. S. Orientation, positional, additivity, and oligomerization-state effects of interhelical ion pairs in alpha-helical coiled-coils. *J Mol Biol* **283**, 993-1012 (1998).
- 36 Krylov, D., Barchi, J. & Vinson, C. Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *J Mol Biol* **279**, 959-972 (1998).
- 37 Lumb, K. J. & Kim, P. S. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* **37**, 13042 (1998).
- 38 Oakley, M. G. & Kim, P. S. A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry* **37**, 12603-12610 (1998).
- 39 Wagschal, K., Tripet, B. & Hodges, R. S. De novo design of a model peptide sequence to examine the effects of single amino acid substitutions in the hydrophobic core on both stability and oligomerization state of coiled-coils. *J Mol Biol* **285**, 785-803 (1999).
- 40 Wagschal, K., Tripet, B., Lavigne, P., Mant, C. & Hodges, R. S. The role of position a in determining the stability and oligomerization state of alpha-helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins. *Protein Sci* **8**, 2312-2329 (1999).
- 41 Arndt, K. M. et al. A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble. *J Mol Biol* **295**, 627-639 (2000).
- 42 Tripet, B., Wagschal, K., Lavigne, P., Mant, C. T. & Hodges, R. S. Effects of side-chain characteristics on stability and oligomerization state of a de novo-designed model coiled-coil: 20 amino acid substitutions in position "d". *J Mol Biol* **300**, 377-402 (2000).
- 43 Akey, D. L., Malashkevich, V. N. & Kim, P. S. Buried polar residues in coiled-coil interfaces. *Biochemistry* **40**, 6352-6360 (2001).

- 44 Keating, A. E., Malashkevich, V. N., Tidor, B. & Kim, P. S. Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. *Proc Natl Acad Sci U S A* **98**, 14825-14830 (2001).
- 45 Lee, D. L., Lavigne, P. & Hodges, R. S. Are trigger sequences essential in the folding of two-stranded alpha-helical coiled-coils? *J Mol Biol* **306**, 539-553 (2001).
- 46 Litowski, J. R. & Hodges, R. S. Designing heterodimeric two-stranded alpha-helical coiled-coils: the effect of chain length on protein folding, stability and specificity. *J Pept Res* **58**, 477-492 (2001).
- 47 McClain, D. L., Binet, J. P. & Oakley, M. G. Evaluation of the energetic contribution of interhelical Coulombic interactions for coiled coil helix orientation specificity. *J Mol Biol* **313**, 371-383 (2001).
- 48 McClain, D. L., Woods, H. L. & Oakley, M. G. Design and characterization of a heterodimeric coiled coil that forms exclusively with an antiparallel relative helix orientation. *J Am Chem Soc* **123**, 3151-3152 (2001).
- 49 Moll, J. R., Ruvinov, S. B., Pastan, I. & Vinson, C. Designed heterodimerizing leucine zippers with a range of pIs and stabilities up to 10(-15) M. *Protein Sci* **10**, 649-655 (2001).
- 50 Oakley, M. G. & Hollenbeck, J. J. The design of antiparallel coiled coils. *Curr Opin Struct Biol* **11**, 450-457 (2001).
- 51 Acharya, A., Ruvinov, S. B., Gal, J., Moll, J. R. & Vinson, C. A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: amino acids I, V, L, N, A, and K. *Biochemistry* **41**, 14122-14131 (2002).
- 52 Arndt, K. M., Pelletier, J. N., Muller, K. M., Pluckthun, A. & Alber, T. Comparison of in vivo selection and rational design of heterodimeric coiled coils. *Structure* **10**, 1235-1248 (2002).
- 53 Hicks, M. R., Walshaw, J. & Woolfson, D. N. Investigating the tolerance of coiled-coil peptides to nonheptad sequence inserts. *J Struct Biol* **137**, 73-81 (2002).
- 54 Kim, B. M. & Oakley, M. G. A general method for selection and screening of coiled coils on the basis of relative helix orientation. *J Am Chem Soc* **124**, 8237-8244 (2002).
- 55 Litowski, J. R. & Hodges, R. S. Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity. *J Biol Chem* **277**, 37272-37279 (2002).
- 56 McClain, D. L., Gurnon, D. G. & Oakley, M. G. Importance of potential interhelical salt-bridges involving interior residues for coiled-coil stability and quaternary structure. *J Mol Biol* **324**, 257-270 (2002).
- 57 De Crescenzo, G., Litowski, J. R., Hodges, R. S. & O'Connor-McCourt, M. D. Real-time monitoring of the interactions of two-stranded de novo designed coiled-coils: effect of chain length on the kinetic and thermodynamic constants of binding. *Biochemistry* **42**, 1754-1763 (2003).
- 58 Kwok, S. C. & Hodges, R. S. Clustering of large hydrophobes in the hydrophobic core of two-stranded alpha-helical coiled-coils controls protein folding and stability. *J Biol Chem* **278**, 35248-35254 (2003).

- 59 Lee, D. L., Ivaninskii, S., Burkhard, P. & Hodges, R. S. Unique stabilizing interactions identified in the two-stranded alpha-helical coiled-coil: crystal structure of a cortexillin I/GCN4 hybrid coiled-coil peptide. *Protein Sci* **12**, 1395-1405 (2003).
- 60 Walshaw, J. & Woolfson, D. N. Extended knobs-into-holes packing in classical and complex coiled-coil assemblies. *J Struct Biol* **144**, 349-361 (2003).
- 61 Kwok, S. C. & Hodges, R. S. Effect of chain length on coiled-coil stability: decreasing stability with increasing chain length. *Biopolymers* **76**, 378-390 (2004).
- 62 Kwok, S. C. & Hodges, R. S. Stabilizing and destabilizing clusters in the hydrophobic core of long two-stranded alpha-helical coiled-coils. *J Biol Chem* **279**, 21576-21588 (2004).
- 63 Lai, J. R., Fisk, J. D., Weisblum, B. & Gellman, S. H. Hydrophobic core repacking in a coiled-coil dimer via phage display: insights into plasticity and specificity at a protein-protein interface. *J Am Chem Soc* **126**, 10514-10515 (2004).
- 64 Lu, S. M. & Hodges, R. S. Defining the minimum size of a hydrophobic cluster in two-stranded alpha-helical coiled-coils: effects on protein stability. *Protein Sci* **13**, 714-726 (2004).
- 65 Pandya, M. J. *et al.* Sequence and structural duality: designing peptides to adopt two stable conformations. *J Am Chem Soc* **126**, 17016-17024 (2004).
- 66 Lupas, A. N. & Gruber, M. The structure of alpha-helical coiled coils. *Adv Protein Chem* **70**, 37-78 (2005).
- 67 Slovic, A. M., Lear, J. D. & DeGrado, W. F. De novo design of a pentameric coiled-coil: decoding the motif for tetramer versus pentamer formation in water-soluble phospholamban. *J Pept Res* **65**, 312-321 (2005).
- 68 Acharya, A., Rishi, V. & Vinson, C. Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). *Biochemistry* **45**, 11324-11332 (2006).
- 69 Hadley, E. B. & Gellman, S. H. An antiparallel alpha-helical coiled-coil model system for rapid assessment of side-chain recognition at the hydrophobic interface. *J Am Chem Soc* **128**, 16444-16445 (2006).
- 70 Iranzo, O., Ghosh, D. & Pecoraro, V. L. Assessing the integrity of designed homomeric parallel three-stranded coiled coils in the presence of metal ions. *Inorg Chem* **45**, 9959-9973 (2006).
- 71 Mason, J. M., Schmitz, M. A., Muller, K. M. & Arndt, K. M. Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. *Proc Natl Acad Sci U S A* **103**, 8989-8994 (2006).
- 72 Mason, J. M., Muller, K. M. & Arndt, K. M. Positive aspects of negative design: simultaneous selection of specificity and interaction stability. *Biochemistry* **46**, 4804-4814 (2007).
- 73 Mason, J. M., Muller, K. M. & Arndt, K. M. Considerations in the design and optimization of coiled coil structures. *Methods Mol Biol* **352**, 35-70 (2007).
- 74 Straussman, R., Ben-Ya'acov, A., Woolfson, D. N. & Ravid, S. Kinking the coiled coil--negatively charged residues at the coiled-coil interface. *J Mol Biol* **366**, 1232-1242 (2007).

- 75 Hadley, E. B., Testa, O. D., Woolfson, D. N. & Gellman, S. H. Preferred side-chain constellations at antiparallel coiled-coil interfaces. *Proc Natl Acad Sci U S A* **105**, 530-535 (2008).
- 76 Tsatskis, Y. *et al.* Core residue replacements cause coiled-coil orientation switching in vitro and in vivo: structure-function correlations for osmosensory transporter ProP. *Biochemistry* **47**, 60-72 (2008).
- 77 Mason, J. M., Hagemann, U. B. & Arndt, K. M. Role of hydrophobic and electrostatic interactions in coiled coil stability and specificity. *Biochemistry* **48**, 10380-10388 (2009).
- 78 Steinkruger, J. D., Woolfson, D. N. & Gellman, S. H. Side-chain pairing preferences in the parallel coiled-coil dimer motif: insight on ion pairing between core and flanking sites. *J Am Chem Soc* **132**, 7586-7588 (2010).
- 79 Beasley, J. R. a. H., M. H. . Protein Design: The Choice of de Novo Sequences. . *J. Biol. Chem.* **272**, 2031-2034 (1997).
- 80 Bryson, J. W. *et al.* Protein design: a hierachic approach. *Science* **270**, 935-941 (1995).
- 81 Dieckmann, G. R. & DeGrado, W. F. Modeling transmembrane helical oligomers. . *Curr. Opin. Struct. Biol.* **7**, 486-494 (1997).
- 82 Jing, P., Rudra, J. S., Herr, A. B. & Collier, J. H. Self-assembling peptide-polymer hydrogels designed from the coiled coil region of fibrin. *Biomacromolecules* **9**, 2438-2446 (2008).
- 83 Krishna, O. D. & Kiick, K. L. Protein- and peptide-modified synthetic polymeric biomaterials. *Biopolymers* **94**, 32-48 (2010).
- 84 Nakaji-Hirabayashi, T., Kato, K. & Iwata, H. Self-assembling chimeric protein for the construction of biodegradable hydrogels capable of interaction with integrins expressed on neural stem/progenitor cells. *Biomacromolecules* **9**, 1411-1416 (2008).
- 85 Tang, A., Wang, C., Stewart, R. J. & Kopecek, J. The coiled coils in the design of protein-based constructs: hybrid hydrogels and epitope displays. *J Control Release* **72**, 57-70 (2001).
- 86 Wang, C., Kopecek, J. & Stewart, R. J. Hybrid hydrogels cross-linked by genetically engineered coiled-coil block proteins. *Biomacromolecules* **2**, 912-920 (2001).
- 87 Wang, C., Stewart, R. J. & Kopecek, J. Hybrid hydrogels assembled from synthetic polymers and coiled-coil protein domains. *Nature* **397**, 417-420 (1999).
- 88 Xu, C., Breedveld, V. & Kopecek, J. Reversible hydrogels from self-assembling genetically engineered protein block copolymers. *Biomacromolecules* **6**, 1739-1749 (2005).
- 89 Xu, C. & Kopecek, J. Genetically engineered block copolymers: influence of the length and structure of the coiled-coil blocks on hydrogel self-assembly. *Pharm Res* **25**, 674-682 (2008).
- 90 Yang, J., Xu, C., Kopeckova, P. & Kopecek, J. Hybrid hydrogels self-assembled from HPMA copolymers containing peptide grafts. *Macromol Biosci* **6**, 201-209 (2006).

- 91 Yang, J., Xu, C., Wang, C. & Kopecek, J. Refolding hydrogels self-assembled from N-(2-hydroxypropyl)methacrylamide graft copolymers by antiparallel coiled-coil formation. *Biomacromolecules* **7**, 1187-1195 (2006).
- 92 Zhang, L., Furst, E. M. & Kiick, K. L. Manipulation of hydrogel assembly and growth factor delivery via the use of peptide-polysaccharide interactions. *J Control Release* **114**, 130-142 (2006).
- 93 Wu, K., Liu, J., Johnson, R. N., Yang, J. & Kopecek, J. Drug-free macromolecular therapeutics: induction of apoptosis by coiled-coil-mediated cross-linking of antigens on the cell surface. *Angew Chem Int Ed Engl* **49**, 1451-1455 (2010).
- 94 Schweller, R. M., Constantinou, P. E., Frankel, N. W., Narayan, P. & Diehl, M. R. Design of DNA-Conjugated Polypeptide-Based Capture Probes for the Anchoring of Proteins to DNA Matrices. *Bioconjug Chem* **19**, 2304-2307 (2008).
- 95 Gaidamakov, E. K., Backer, M. V., and Backer, J. M. Molecular vehicle for target mediated delivery of therapeutics and diagnostics *Journal of Controlled Release* **74**, 341-347 (2001).
- 96 Shekhawat, S. S., Porter, J. R., Srivatsa, A. & Ghosh, I. An autoinhibited coiled-coil design strategy for split-protein protease sensors. *J Am Chem Soc* **131**, 15284-15290 (2009).
- 97 Yu, B. Y. Coiled-coils: stability, specificity, and drug delivery potential. *Advanced Drug Delivery Reviews*. **54**, 1113-1129 (2002).
- 98 Chao, H., Bautista, D. L., Litowski, J., Irvin, R. T. & Hodges, R. S. Use of a heterodimeric coiled-coil system for biosensor application and affinity purification. *J Chromatogr B Biomed Sci Appl* **715**, 307-329 (1998).
- 99 Contarino, M. R. *et al.* Modular, self-assembling peptide linkers for stable and regenerable carbon nanotube biosensor interfaces. *J Mol Recognit* **19**, 363-371 (2006).
- 100 Tripet, B. *et al.* Engineering a de novo-designed coiled-coil heterodimerization domain for the rapid detection, purification and characterization of recombinantly expressed peptides and proteins. *Protein Eng* **9**, 1029-1042 (1996).
- 101 De Crescenzo, G., Pham, P. L., Durocher, Y., Chao, H. & O'Connor-McCourt, M. D. Enhancement of the antagonistic potency of transforming growth factor-beta receptor extracellular domains by coiled coil-induced homo- and heterodimerization. *J Biol Chem* **279**, 26013-26018 (2004).
- 102 Hillar, A., Culham, D. E., Vernikovska, Y. I., Wood, J. M. & Boggs, J. M. Formation of an antiparallel, intermolecular coiled coil is associated with in vivo dimerization of osmosensor and osmoprotectant transporter ProP in Escherichia coli. *Biochemistry* **44**, 10170-10180 (2005).
- 103 Katz, B. Z. *et al.* Green fluorescent protein labeling of cytoskeletal structures--novel targeting approach based on leucine zippers. *Biotechniques* **25**, 298-302, 304 (1998).
- 104 Minten, I. J., Hendriks, L. J., Nolte, R. J. & Cornelissen, J. J. Controlled encapsulation of multiple proteins in virus capsids. *J Am Chem Soc* **131**, 17771-17773 (2009).
- 105 Veiga, E., Lorenzo, V. d. & Fernandez, L. A. Autotransporters as Scaffolds for Novel Bacterial Adhesins: Surface Properties of Escherichia coli Cells

- Displaying Jun/Fos Dimerization Domains. *Journal of Bacteriology* **185**, 5585–5590 (2003).
- 106 Dietz, H., Bornschlogl, T., Heym, R., Konig, F. & Rief, M. Programming protein self assembly with coiled coils. *New Journal of Physics* **9**, 424-431 (2007).
- 107 De Francesco, R., Pastore, A., Vecchio, G. & Cortese, R. Circular dichroism study on the conformational stability of the dimerization domain of transcription factor LFB1. *Biochemistry* **30**, 143-147 (1991).
- 108 Muller, K. M., Arndt, K. M. & Alber, T. Protein fusions to coiled-coil domains. *Methods Enzymol* **328**, 261-282 (2000).

## Chapter 4

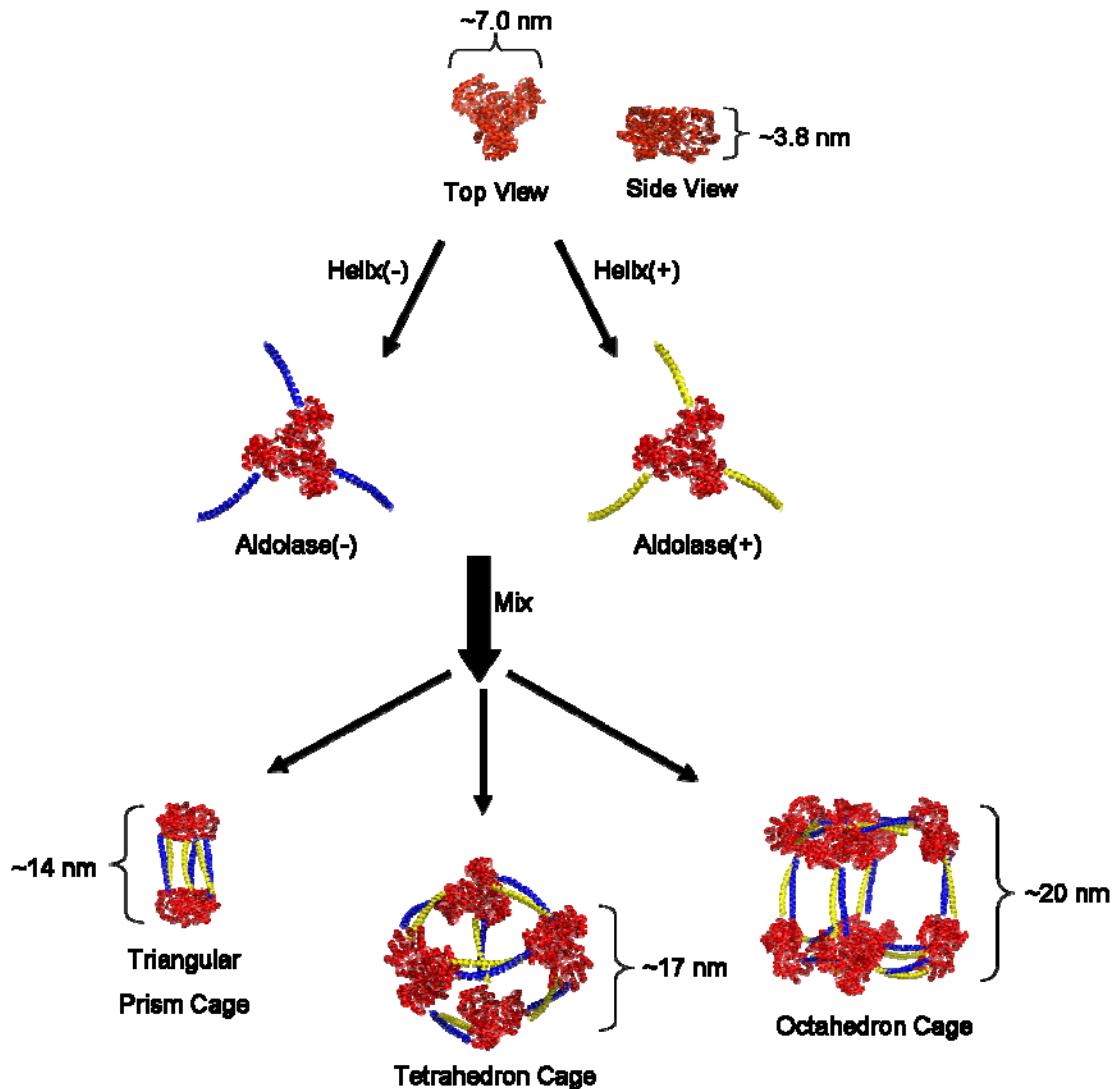
### Proof of Concept: Investigations of Symmetry Directed Protein Cage Assembly of the Trimeric KDPG Aldolase

#### 4.1 - Introduction

Strategies for assembling proteins into higher order cage structures to date have been, in general, limited to only specific proteins and not broadly applicable.<sup>1,2</sup> We devised a new strategy for assembling protein cages that is potentially applicable to many proteins, as discussed in detail in Chapter 2 of this dissertation. Briefly, this approach utilizes proteins with symmetrical quaternary structures as building blocks. The symmetry acts as a guide to assembly, restricting the possible cage structures that can form. Assembly will be driven by *de novo* designed heterodimeric coiled coils, described in detail in Chapter 3 of this dissertation, and will provide control over assembly. Although the *de novo* designed peptides were not observed to form heterodimeric coiled coils, it was thought that attachment of the peptides to a large protein may inhibit self-dimerization of the Helix(-) and allow association with the complementary peptide Helix(+). The PBB subunits and the coiled coil peptides are connected by a short flexible peptide segment allowing the proteins a degree of freedom to form cage structures compatible with the symmetry.

This chapter describes an experimental test of our strategy for assembling protein cages. For this study, fusion proteins of the trimeric protein KDPG aldolase and Helix(-) and Helix(+) peptides were constructed.<sup>3</sup> According to our design hypothesis KDPG

aldolase would be preferred to form four structures compatible with its  $C_3$  symmetry; tetrahedral, octahedral, icosahedral, or a simple dimeric or triangular prism (dimer of trimers) cage structure. In addition, valency requirements further limit the structures to either the prism or octahedral structures if there is no exchange of subunits between PBBs. However, in the event that subunit exchange occurs, assembly of the tetrahedron and icosahedron are also predicted to form, although the entropic penalty for forming an icosahedron may be too great to allow formation. A model depicting our strategy for assembly based on the crystal structure of KDPG aldolase and the structures predicted to form are provided in Figure 4.1.



**Figure 4.1.** KDPG aldolase crystal structure assembly model. The coiled coil peptides (blue and yellow helix) grafted to KDPG aldolase (red) cause assembly into cage structures upon mixing. Size descriptions of the KDPG aldolase and the predicted structures are provided for future reference.

#### 4.2 - Experimental Procedures

##### Materials

DNA modifying enzymes and reagents were purchased from New England Biolabs. DNA primers were purchased from IDT DNA Technologies (Coralville, IA). PBS 10x stock solution was purchased from Invitrogen (Carlsbad, CA). *E. coli* BL21( $\lambda$ DE3) and expression vector pET-28b were purchased from Novagen (Madison,

WI); Pfu turbo DNA polymerase and *E. coli* XL1-Blue were from Stratagene (Cedar Creak, TX). Nickel nitrilotriacetic acid (Ni-NTA) resin, QIAquick gel extraction kit, and QIAprep Spin Miniprep kit were purchased from Qiagen (Valencia, CA). DNP-10 Non-Condensed Silicon Nitride tips were purchased from Veeco Probes (Camarillo, CA). Gold nanoparticles (Product #15702-20) were from Ted Pella, Inc. (Redding, CA). Vinlyec (Polyvinyl Formal) Film grids, stabilized with Carbon were purchased from Ernest F. Fullam, Inc. (Clifton Park, NY). All chemical reagents (IPTG, Sodium Chloride, etc.) were from Fisher (Pittsburgh, PA).

#### Gene Splicing and Subcloning

Genes encoding *T. maritima* KDPG Aldolase in pUC19 were kindly provided by Carol Fierke and Manoj Cherian (Univ. of Michigan). Genes encoding the coiled-coil linkers Helix(-) and Helix(+) were obtained as described previously in Chapter 3. The plasmids were digested with Nco I and Xho I restriction enzymes and resulting KDPG Aldolase gene and linearized plasmids containing Helix(-) and Helix(+) constructs were purified using a 2% agarose gel. The DNA fragments were extracted from the gel using QIAquick Gel Extraction Kit and the DNA fragments ligated together with T4 DNA Ligase. Ligation products were transformed into electrocompetent XL1-Blue *E. coli* cells and cells containing plasmids were selected for by growing transformed cells on LB agar plates with kanamycin (50 µg/mL). Single colonies were used to inoculate 5 mL 2x YT broth, allowed to incubate overnight, and resulting cells were pelleted by centrifugation. DNA was extracted from cell pellets using QIAprep Spin Miniprep Kit (Qiagen, MD) and samples of the resulting DNA extracts were sent for sequencing (U of M Sequencing Core, MI) to confirm their sequence identity. Expression vectors containing the desired

gene fusions were further modified to encode a 6xHis tag at the C-terminus to improve purification of expressed fusion proteins. Nucleotide sequences encoding 6xHis tags were already present in pET28b(+) allowing incorporation through PCR mutagenesis and restriction digest-ligation steps. Nucleotide sequences encoding the proteins in pET28b(+) are given in the appendix.

Vectors containing the recombinant gene fusions were transformed separately into electrocompetent BL21(λDE3) cells for protein expression. Transformed cells containing plasmids were selected for by growing cells on LB agar plates containing kanamycin (50 µg/mL). Single colonies were grown up in 5 mL 2x YT broth with kanamycin for selection overnight and deeps were made from the cultures. Deeps were stored at -20 °C and -80 °C.

#### Protein Sequences and Physical Constants

The fusion proteins were named Aldolase(-) and Aldolase(+), with the (-) and (+) being derived from the coiled coil peptide linker, Helix(-) or Helix(+). Sequences for the final protein constructs containing a 6xHis tag with the flexible spacer peptide sequences containing a thrombin cleavage site underlined are as follows:

#### Aldolase(-)

MAMEELFKKH KIVAVLRLANS VEEAKEKALA VFEGGVHLIE ITFTVPDADT VIKELSFLKE	60
KGAIIGAGTV TSVEQCRKAV ESGAEFIVSP HLDEEISQFC KEKGVFYMPG VMTPTELVKA	120
MKLGHHTILKL FPGEVVGPQF VKAMKGPFPN VKFVPTGGVN LDNVCEWFKA GVLAVVGSA	180
LVKGTPDEV R EKAKAFVEKI RGCTELE <u>PRG SHMGSKQLEK ELAQLEKELQ AIEKELAQLE</u>	240
WEAQALEKKL AQLEKELOQGG SGGTSHHHHH HS	272

### Aldolase(+)

MAMEELFKKH KIVAVLRANS VEEAKEKALA VFEGGVHLIE ITFTVPDADT VIKELSFLKE	60
KGAIIGAGTV TSVEQCRKAV ESGAEFIVSP HLDEEISQFC KEKGVFYMPG VMTPTELVKA	120
MKLGHТИLKL FPGEVVGPQF VKAMKGPFPN VKFVPTGGVN LDNVCEWFKA GVLAVGVGSA	180
LVKGTPDEVN EKAKAFVEKI RGCTELEPRG <u>THGSSKQLEK</u> ELAQLKKKLQ AIKKKLAQLK	240
WKAQALKKKL AQLKKKLQGG SGASSQLVHH HHH	273

Molecular weight and pI were calculated from sequences using Protein Calculator v3.3 (Chris Putnam, Scripps). Aldolase(-) had a calculated molecular weight of 29,716 Da and pI of 6.60. Aldolase(+) had a calculated molecular weight of 29,931 Da and pI of 9.44. The calculated molecular weight of the parent KDPG aldolase was 23,244 with a pI of 6.56.

### Protein Expression

Cultures of 5 mL 2x YT broth with kanamycin (final conc. 50 µg/mL) were inoculated from bacteria stocks stored at -20 °C and allowed to incubate overnight in a Lab-Line incubator-shaker at 37 °C and 250 rpm shaker speed. Overnight cultures were transferred to 1 L 2x YT broth with kanamycin in an Erlenmeyer flask and incubated at 37 °C and 250 rpm shaker speed. Once the cells reached early log phase ( $OD_{600}=0.8$ ), expression of the genes was induced by addition of isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Cultures were incubated for 4 hours after addition of IPTG, then cells were harvested by centrifugation on a Beckman Avanti J-20I centrifuge at 4000xg for 15 minutes at 4 °C. Cell pellets were separated from supernatant by decanting and the pellets were left in the centrifuge tubes and placed in freezer at -20 °C overnight.

### Purification of Non-His Tagged Fusion Proteins

Frozen cell pellets were thawed at 4 °C and then resuspended in Lysis Buffer (50 mM Tris-HCl, 150 mM NaCl, 10% glycerol, pH 8.0) by stirring with a stir rod. Upon resuspension of the cell pellet, benzonase, β-mercaptoethanol (1 mM) and PMSF (0.1 mM) were added, then the cell suspension was sonicated on ice to lyse the cells. Cell lysate was centrifuged at 23,000 x g for 15 minutes on a Sorvall Super T 21 centrifuge at 4 °C. The supernatant was decanted into a clean centrifuge tube and heated at 80 °C for 30 minute yielding a suspension of dense white precipitate. The precipitate was removed by centrifugation at 23,000 x g for 15 minutes. The supernatant was then dialyzed twice against 20 mM Tris, 50 mM NaCl, pH 7.5 in preparation for ion exchange chromatography. Ion exchange chromatography (IEC) was performed on a Mono Q anion exchange column equilibrated with IEC buffer A (20 mM Tris, 50 mM NaCl, pH 7.5) using a Biologic HR FPLC (Biorad, CA). Separation was carried out using an increasing gradient of IEC buffer B (20 mM Tris, 1 M NaCl, pH 7.5) with IEC buffer A (20 mL total volume, flow rate 1.0 mL/min). Alternatively, proteins were subjected to size exclusion chromatography on a sephacryl S-200 packed column also using a Biologic HR FPLC (Biorad, CA) and elution buffer of 25 mM Tris, 150 mM NaCl, pH 7.5. Neither method was able to completely purify fusion proteins from DNA contamination without significant loss of materials.

### Purification of Aldolase(-)

Frozen cell pellets were thawed at 4 °C and then resuspended in Lysis Buffer (50 mM Tris-HCl, 150 mM NaCl, 10% glycerol, pH 8.0) by stirring with a stir rod. Upon resuspension of the cell pellet, benzonase, β-mercaptoethanol (1 mM) and PMSF (0.1

mM) were added, then the cell suspension was sonicated on ice to lyse the cells. Cell lysate was centrifuged at 23,000 x g for 15 minutes on a Sorvall Super T 21 centrifuge at 4 °C. Supernatant was decanted into a clean centrifuge tube and heated at 80 °C for 30 minute yielding a suspension of dense white precipitate. Precipitate was removed by centrifugation at 23,000 x g for 15 minutes. The supernatant was loaded onto a 5 mL column of Ni-NTA superflow resin equilibrated in Buffer A (50 mM Tris-HCl, 300 mM NaCl, 2.5 mM imidazole, 1 mM β-mercaptoethanol, 10% glycerol, pH 8.0) at a flow rate of 0.5 mL/min using a Biologic HR FPLC (Biorad, CA) and the column washed with 30 mL of buffer A. Non-specifically bound proteins were eluted with a step gradient (20 mL at flow rate of 2 mL/min for each step) of increasing concentrations (7.5 mM, 27.5 mM and 52.5 mM) of imidazole in Buffer A. Finally the 6xHis-tagged fusion proteins were eluted with Buffer A containing 500 mM imidazole. This yielded protein that was greater than 95% pure as judged by SDS-PAGE. Purified protein solutions were dialyzed twice overnight against 1.5 L Storage Buffer (50 mM Tris-HCl, 150 mM NaCl, 0.5 mM EDTA, 10% glycerol, pH 8.0) and stored at -20 °C. Protein concentrations were determined by UV absorption measured at 280 nm assuming a molar extinction coefficient  $\epsilon_{280} = 12,900 \text{ M}^{-1}\text{cm}^{-1}$ .

#### Purification of Aldolase(+)

Aldolase(+) was found to strongly bind DNA, which could not be removed by the purification procedure previously described. For Aldolase(+), cell pellets were resuspended in denaturation buffer (50 mM Tris HCl, 150 mM NaCl, 6 M guanidine hydrochloride, 5% glycerol, pH 8.0) then sonicated on ice. Cell lysate was then centrifuged at 23,000 x g for 15 minutes at 4 °C. Supernatant was applied to a column

packed with Ni-NTA superflow resin as described previously for Aldolase(-) and then washed with denaturing wash buffer (50 mM Tris, 6 M guanidium-HCl, 300 mM NaCl, 5 mM imidazole, 1 mM  $\beta$ -mercaptoethanol, 10% glycerol, pH 8.0) until the absorbance detector showed a stable flat baseline. The protein was refolded on the column using a decreasing linear gradient of denaturing buffer with Buffer A, (60 mL total volume, flow rate 0.5 mL/min) before being eluted with 500 mM imidazole as described previously. Alternatively, the protein can be purified by first lysing cells by sonication in Lysis Buffer, heating the cell lysate at 80 °C for 30 minutes, removing precipitates by centrifugation, applying the supernatant a nickel column, then washing with denaturation buffer, and subsequently performing the refolding procedure described above. However, the yields using this method were lower than the denatured purification described previously. All subsequent preparation and analyses were done as described for Aldolase(-).

#### Catalytic Activity Assay

Catalytic activity was assayed using a lactate dehydrogenase-coupled assay.<sup>4</sup> HEPES (10 mM, pH 8.0, 100  $\mu$ L), NADH (230-280  $\mu$ M), lactate dehydrogenase (0.8 Units), KDPG (1 mM) were mixed in a quartz cuvette at 34 °C. The assay was started by addition of KDPG aldolase (final concentration 1  $\mu$ M) or the fusion proteins (final concentrations 0.5 - 0.6  $\mu$ M) and activity followed by monitoring the decrease in absorbance at 340 nm.

#### High Performance Size Exclusion Chromatography

Protein samples were dialyzed against phosphate buffer (1 mM, pH 8.0) in order to remove salts. Dialyzed protein solutions were then flash frozen with liquid nitrogen

and lyophilized. Prior to injection, the protein samples were reconstituted in PBS buffer. The samples were then subjected to centrifugation using a bench-top centrifuge at 10,000xg for 10 minutes to pellet any precipitate present. Concentrations of protein samples were determined at 280 nm using UV/Vis spectroscopy. 100 µL samples of 1-2 mg/mL protein were injected onto size exclusion chromatography (SEC) columns. Protein mixture samples were allowed to incubate at 4 °C for 2 or 24 hours before injection. In addition, commercial stocks of bovine serum albumin (BSA) and carbonic anhydrase (CA) were chromatographed as controls. Samples were submitted to Ankur Desai of MNIMBS, who performed all runs and raw data analysis. All SEC experiments were performed on an Alliance Waters 2695 separation module equipped with a 2487 dual wavelength UV absorbance detector (Waters Corporation), a Wyatt HELEOS Multi Angle Laser Light Scattering (MALLS) detector, and an Optilab rEX differential refractometer (Wyatt Technology Corporation). Columns employed were TosoHaas TSK-Gel Guard PHW 06762 (75 mm × 7.5 mm, 12 mm), G 2000 PW 05761 (300 mm × 7.5 mm, 10mm), G 3000 PW 05762 (300 mm × 7.5 mm, 10 mm), and G 4000 PW (300 mm ×7.5 mm, 17 mm). Column temperature was maintained at  $25 \pm 0.1$  °C with a Waters temperature control module. The isocratic mobile phase was 0.1 M phosphate buffered saline pH 7.4, at a flow rate of 1 mL/min. The number average molecular weight,  $M_n$ , was calculated with Astra5.3.14 software (Wyatt Technology Corporation) based on the molecular weight distribution. Although  $d\eta/dc$  for KDPG aldolase has not been previously determined, most researchers in the field use a fixed  $d\eta/dc$  value of  $0.18 \pm 0.03$  cm<sup>3</sup>/g regardless of the protein and buffer solution used.<sup>5</sup> The  $d\eta/dc$  value used for KDPG aldolase and the fusion proteins was set to 0.186 cm<sup>3</sup>/g.

### Sedimentation Velocity Analytical Ultracentrifugation

All protein samples were dialyzed against PBS with 5% glycerol, pH 8.0. Samples were filtered with 0.1 micron anatop filters from Whatmann prior to loading samples. A series of individual proteins, Aldolase(-) or Aldolase(+), were run with absorptions of 0.2, 0.4, 0.6, 0.8, and 1 at 280 nm. Mixture samples made of 1:1 ratio of Aldolase(-) and Aldolase(+) were incubated overnight at 4 °C. Samples were loaded into sector-shaped double channel centerpieces, temperature-equilibrated for 2 hours prior to sedimentation in the non-spinning rotor, and samples spun at 16,000 and 30,000 rpm at 24 °C using an AN50TI rotor. Absorbance scans were collected continuously at a wavelength of 280 nm. Data analysis was with the enhanced van Holde-Weischet analysis module followed by 2-dimensional sedimentation spectrum analysis (2-DSA) using the finite element modeling module provided with the Ultrascan software (<http://www.ultrascan.uthscsa.edu>). Sedimentation profiles were analyzed at a grid resolution of 62500 using 25 grid repetitions. Confidence levels for statistics were derived from 2-DSA data refinement using a genetic algorithm followed by 30 Monte Carlo simulations. Calculations to analyze analytical ultracentrifugation data were performed on the UltraScan LIMS cluster at the Bioinformatics Core Facility at the University of Texas Health Science Center at San Antonio, the Lonestar cluster at the Texas Advanced Computing Center (supported by NSF Teragrid Grant #MCB070038 to Borries Demeler), and the National Supercomputer HLRB-II at the Leibnitz-Rechenzentrum, Munich, Germany (supported by Project pr28ci to Johannes Buchner and T. M. F.).

### Negative Stain Transmission Electron Microscopy

Samples (10  $\mu$ L, 6  $\mu$ M protein) were applied to formvar coated grids and incubated for 1-2 minutes and excess liquid was wicked away with filter paper. Grids were then washed twice with 10  $\mu$ L of distilled water, wicking away liquid after each addition with filter paper, and then stained with 1  $\mu$ L 2% uranyl acetate for 1 minute and excess stain was wicked away with filter paper. Images were taken on a Philips CM12 scanning tunneling electron microscope at an accelerating voltage of 60 kV. Higher resolution images were taken by Justin Schilling in the Skinoitis lab at room temperature at an accelerating voltage of 100 kV on a Morgagni 268(D) transmission electron microscope equipped with a mounted Orius SC200W CCD camera. Samples imaged by Mr. Schilling were stained with 1% uranyl formate on glow discharged formvar carbon coated grids by a procedure similar to the one described above.

### Atomic Force Microscopy

Mica sheets, 1x1 cm, attached to metal pucks were incubated with 40  $\mu$ L HEPES buffer (10 mM, pH 7.0) and incubated at room temperature under a beaker for 30 minutes. 0.5-1  $\mu$ L of protein solutions (6  $\mu$ M for fusion proteins samples and 100  $\mu$ M KDPG Aldolase in Storage Buffer) were applied to the center of the mica sheets and allowed to incubate for 30 minutes at room temperature under a beaker. 250  $\mu$ L of HEPES buffer was then applied to the mica surface to stop protein absorption. The mica was then washed with 4 mL of HEPES buffer and then the sample was imaged. For gold nanoparticle samples, 0.1 % poly-L-lysine (20  $\mu$ L) was applied to freshly cleaved mica sheets and allowed to incubate at room temperature for 30-60 seconds. The mica was then rinsed with 1 mL deionized water and dried with nitrogen gas. A gold colloid

solution was added to the mica and incubated for 5 minutes, then the surface rinsed with 1 mL deionized water and dried with nitrogen gas. The mica was then incubated at 60 °C in an oven for 1 hour. Imaging was performed in tapping mode on a Nanoscope IIIa Multimode AFM from Digital Instruments (DI, Veeco Metrology, Santa Barbra, CA) equipped with an “E” scanner. Veeco DNP-10 non-condensed silicon nitride tips (force constant=0.06 N/m) were used for imaging under HEPES Buffer (10 mM, pH 7.0) filtered with a 0.2 µM filter. Typical scan sizes were 5 µm x 5 µm or 2 µm x 2 µm and were taken at 0.5-1 Hz scanning speeds. Images were processed and height distribution determined using the program Gwyddion. Thresholds for particle height counting was set to 2.5 nm.

#### Cryo-Electron Microscopy

Protein solutions of >10 mg/mL of Aldolase(-) and Aldolase(+) were mixed 1:1 and incubated overnight. Precipitates were removed by centrifugation at 10,000 x g at 4 °C for 10 minutes. Precipitate cleared samples were applied to a Sephadryl S-400 packed size exclusion column equilibrated with elution buffer (50 mM Tris, 100 mM NaCl, pH 8.0) and separations were run at a flow rate of 0.5 mL/min on a Biologic HR FPLC (Biorad, CA). A fraction collector was set to collect 1 mL fractions over the course of the separation. Fractions were analyzed by static light scattering on a Zetasizer Nano ZS (Malvern Instruments, UK) and the instrument software was used to determine the hydrodynamic radius of the species present in collected fractions. Fractions with sizes consistent with octamers or dimers were combined and concentrated to concentrations greater than 2 mg/mL, as judged by the absorbance of protein solutions at 280 nm. Samples were immediately taken to our collaborators to be prepared for Cryo-EM. After

images were acquired, particles were picked and uploaded for reconstruction analysis performed by Dr. Min Su.

#### 4.3 - Results

##### Design

In formulating the strategy for assembling proteins into cage structures the following criteria were identified as important for incorporation into the design:

- 1) Assembly should be genetically encoded to eliminate the need for additional cross-linking steps and allow *in vivo* assembly if desired.
- 2) Assembly should only occur under specified conditions and be reversible
- 3) The strategy should allow a broad range of proteins to be incorporated as building components
- 4) Assembly should lead to the formation of a limited number of predictable cage architectures
- 5) Assembled proteins should retain their structure and biological activity.

To attempt to meet these criteria we designed, *de novo*, a heterodimeric antiparallel coiled coil peptide pair, Helix(-) and Helix(+). Control over assembly can be assumed by constructing two sets of protein building blocks, each one containing one of the peptide constructs. Only upon mixing should assembly of the protein building blocks occur. Furthermore, coiled coils can be designed to be highly stable or can be modified to be stimuli responsive. The antiparallel orientation is expected to allow protein building blocks to form cage structures without steric interference. The peptides can be genetically encoded for recombinant expression at either the N or C termini of many proteins to attach the peptides to the desired protein building block. Generally expression

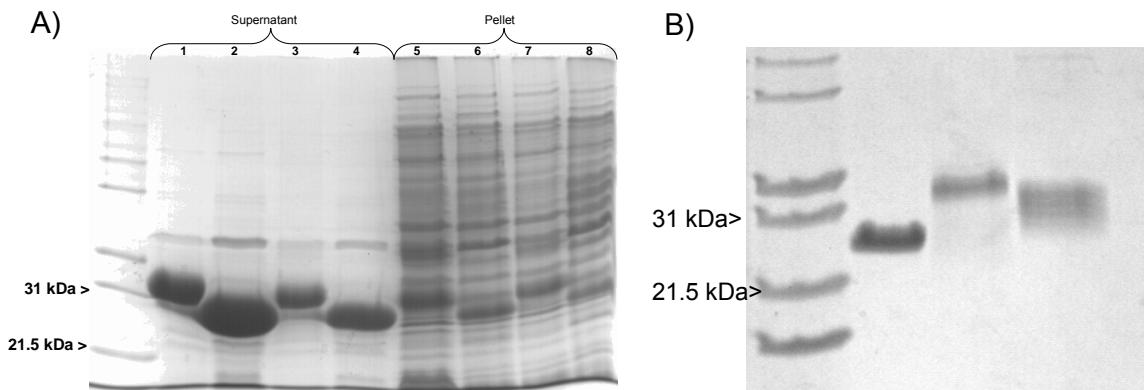
of fusion proteins does not disrupt normal folding of the “tagged” protein and it retains its biological activity.

To fulfill the criterion that a limited number of predictable cage structures should be formed, our strategy was to employ proteins with highly symmetrical quaternary structures as protein building blocks. Depending on the geometry of the protein building blocks, only a limited set of structures are predicted to form. We specifically incorporated flexibility between the protein building block and dimerization peptides by joining the two with a short flexible peptide sequence, expanding the number of proteins that could potentially be used to construct cages.

For our study we selected KDPG aldolase, from the thermophilic bacteria *Thermotoga maritima*, a trimeric protein with a quaternary structure exhibiting C<sub>3</sub> symmetry, as the protein building block (Figure 4.1). KDPG aldolase was chosen because it is highly stable; the final residues of the C-terminus were unstructured and not suspected to be important for folding; it expressed well in *E. coli* and gene constructs were readily available from Dr. Manoj Cheriyon and Prof. Carol Fierke. We predicted that the triangular shape would limit assembly to a triangular prism cage (dimeric cage) or octahedral cage structure when considering the valency requirements of the heterodimeric linkers. In addition, tetrahedron and icosahedron cages were also predicted in the case that subunit exchange occurred. However, the entropic penalty of assembling 12 subunits together to form an icosahedron was thought to be too great for formation to occur.

## Protein Expression and Purification

Initial fusion protein constructs did not contain 6xHis tags. The un-tagged constructs expressed well in *E.coli* as soluble proteins and heating of cell lysates and removal of precipitates by centrifugation appeared to produce >90% pure protein as judged by SDS-PAGE (Figure 4.2A). However, it was found that DNA contaminated the heat purified proteins and it could not be readily removed by treatment with nuclease and dialysis or by ion exchange chromatography or size exclusion chromatography. Plasmids were modified to encode the expression of a C-terminal 6xHis tag to allow purification by nickel-affinity chromatography and the resulting proteins were named Aldolase(-) and Aldolase(+), abbreviated A(-) and A(+). A(-) and A(+) were purified by nickel affinity chromatography and highly pure protein samples devoid of DNA contamination were obtained (Figure 4.2B). Yields of ~25 mg protein per liter of media were typical for purifications.



**Figure 4.2.** SDS-PAGE analysis of KDPG aldolase and fusion protein constructs. A) Purification by heat treatment. Lanes 1,3, 5, 7 are for fusion protein samples and 2, 4, 6, and 8 are wild type KDPG aldolase. Supernatant (left lanes) and Pellet (right lanes) fractions are labeled. B) 6xHis tagged protein constructs purified by nickel affinity chromatography. Samples left to right: Molecular weight marker, KDPG aldolase, A(-), and A(+). Molecular weights of 23.2 kDa, 29.7 kDa, and 29.9 kDa are expected for the monomeric subunits of KDPG aldolase, A(-), and A(+) respectively.

### Catalytic Activity of A(-) and A(+)

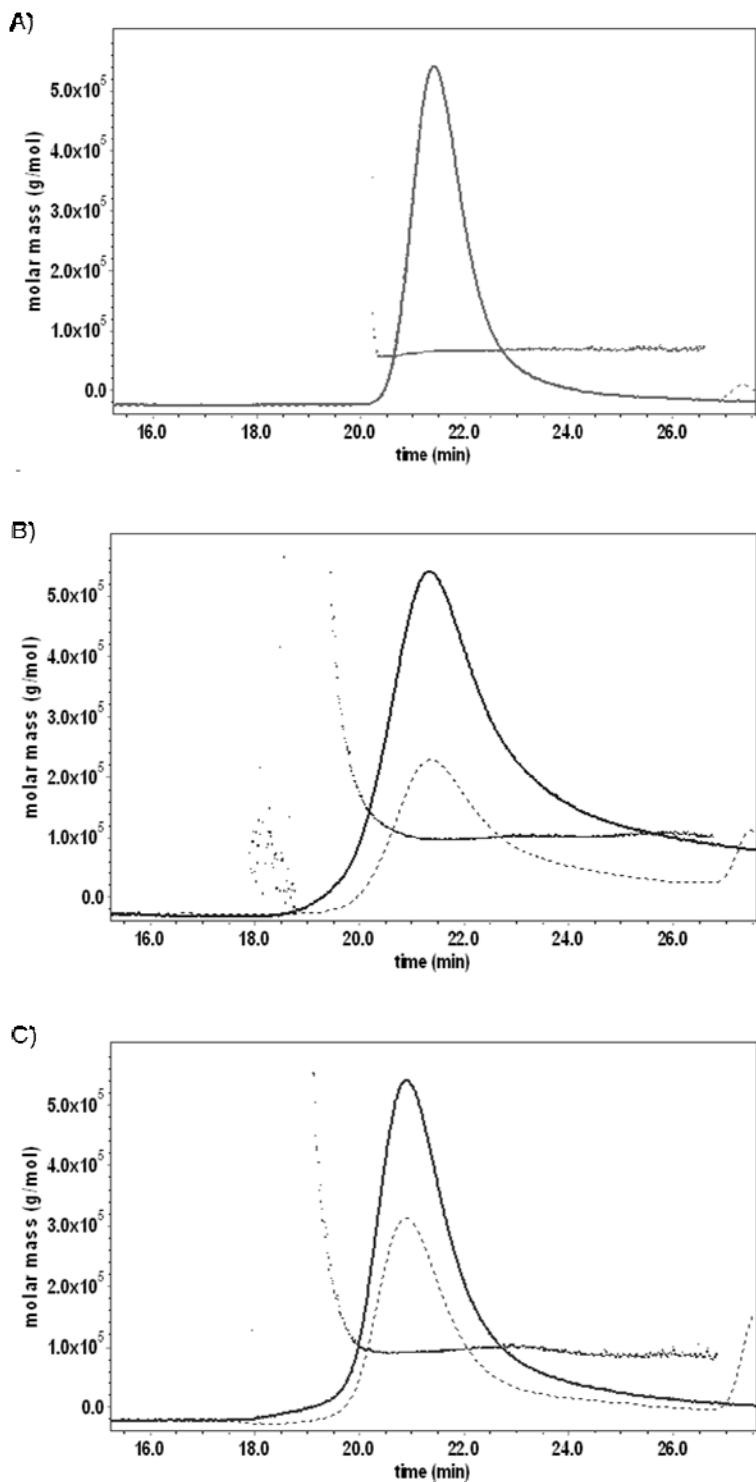
The catalytic activity of the fusion proteins was measured and compared with the activity of wild type aldolase to determine if the additional peptide sequence had any deleterious effects on folding and quaternary structure formation. It is known that disruption of the quaternary structure of aldolase causes almost complete loss of catalytic activity (C. A. Fierke, personal communication). Enzyme assays gave  $k_{cat}$  values of 3.4 sec<sup>-1</sup>, 3.6 sec<sup>-1</sup> and 3.0 sec<sup>-1</sup> for aldolase, A(-) and A(+) respectively. In addition, a 1:1 mixture of A(-) and A(+) was found to have a  $k_{cat}$  of 3.4 sec<sup>-1</sup>. These results provide evidence that appendage of peptides to aldolase has no significant effect on the folding and maintains the original quaternary structure.

### High Performance Size Exclusion Chromatography Investigations

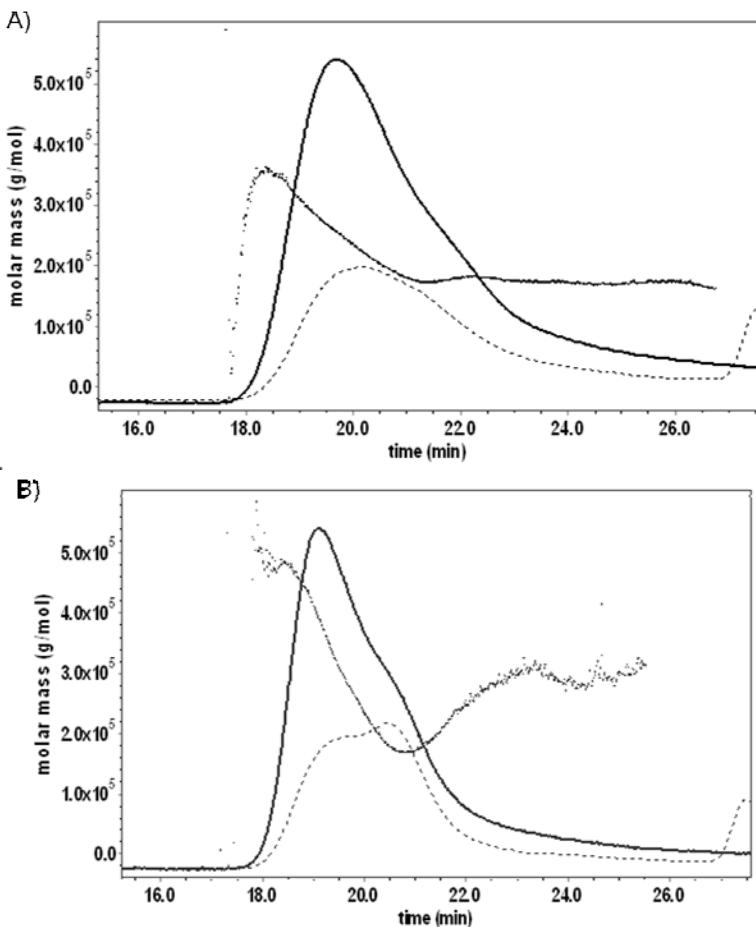
The oligomeric states of the fusion proteins were investigated by high performance size exclusion chromatography (HP-SEC) incorporating dual MALLS and refractive index (RI) detectors. Molecular weights can be directly calculated from MALLS and RI data. This method is superior to the way in which molecular weights are typically determined, because SEC using calibration curves for a set of molecular weight standards detected by U.V. absorbance can often lead to errors in molecular weight calculations. A further advantage is that molecular weights for specific fractions or “slices” of the elution profile can be determined from dual MALLS/RI data.

Molecular weights determined for the control proteins BSA and CA agreed with molecular weights expected for these well characterized proteins (Appendix A). Aldolase analyzed by HP-SEC produced a single peak with a number average molecular weight ( $M_n$ ) of 66 kDa (Figure 4.3A), consistent with the molecular weight of 69.7 kDa

expected for a trimer, and polydispersity index (PDI) of 1.00. The PDI is a measure of homogeneity, with 1.0 being homogenous and higher values indicating heterogeneous species are present. Fusion proteins A(-) and A(+) also exhibited single elution peaks with an  $M_n$  of 102 kDa and 93.9 kDa respectively (Figure 4.3B and C), consistent with the molecular weight of ~90 kDa expected for trimeric fusion proteins. PDIs of 1.02 and 1.01 for A(-) and A(+) showed that they were relatively monodisperse. Interestingly, the results show that attachment of Helix(-) to KDPG aldolase disrupts Helix(-) ability to homodimerize, as discussed previously (see Chapter 3). Appending Helix(-) to KDPG aldolase potentially inhibits homodimerization through sterics, disallowing Helix(-) to associate as it does as an isolated peptide. Effects on the overall electrostatics upon attachment of Helix(-) to KDPG aldolase may also play a role.



**Figure 4.3.** Elution profiles and molecular weight analysis of KDPG aldolase and fusion proteins. A) KDPG aldolase, B) A(-) and C) A(+). MALLS and refractive index detector outputs are given as solid and dashed traces respectively. Molecular weight analysis of the elution profile is shown as a dotted trace (appearing solid in parts) that does not encompass the entire chromatogram. In A the intensities of MALLS and RI are identical.



**Figure 4.4.** HP-SEC elution profiles and molecular weight analysis of 1:1 mixtures of A(-) and A(+). A) Sample incubated for 2 hours. B) Sample incubated for 24 hours. Light scattering detector and refractive index detector outputs are given as solid and dashed traces respectively. Molecular weight analysis of the elution profile is shown as a dotted trace (appearing solid in parts) that does not encompass the entire chromatogram.

HP-SEC was further used to analyze oligomerization states upon mixing of A(-) with A(+). Elution profiles with molecular weight analyses for samples incubated at 2 hours and 24 hours are shown in Figure 4.4. Close inspection of the molecular weight distribution across the elution profile of the 2 hour incubation period revealed a  $M_n$  of ~360 kDa at the start of the elution followed by a linear decline to give a  $M_n$  of ~180 kDa (Figure 4.4A). The molecular weights and profile are consistent with overlapping peaks

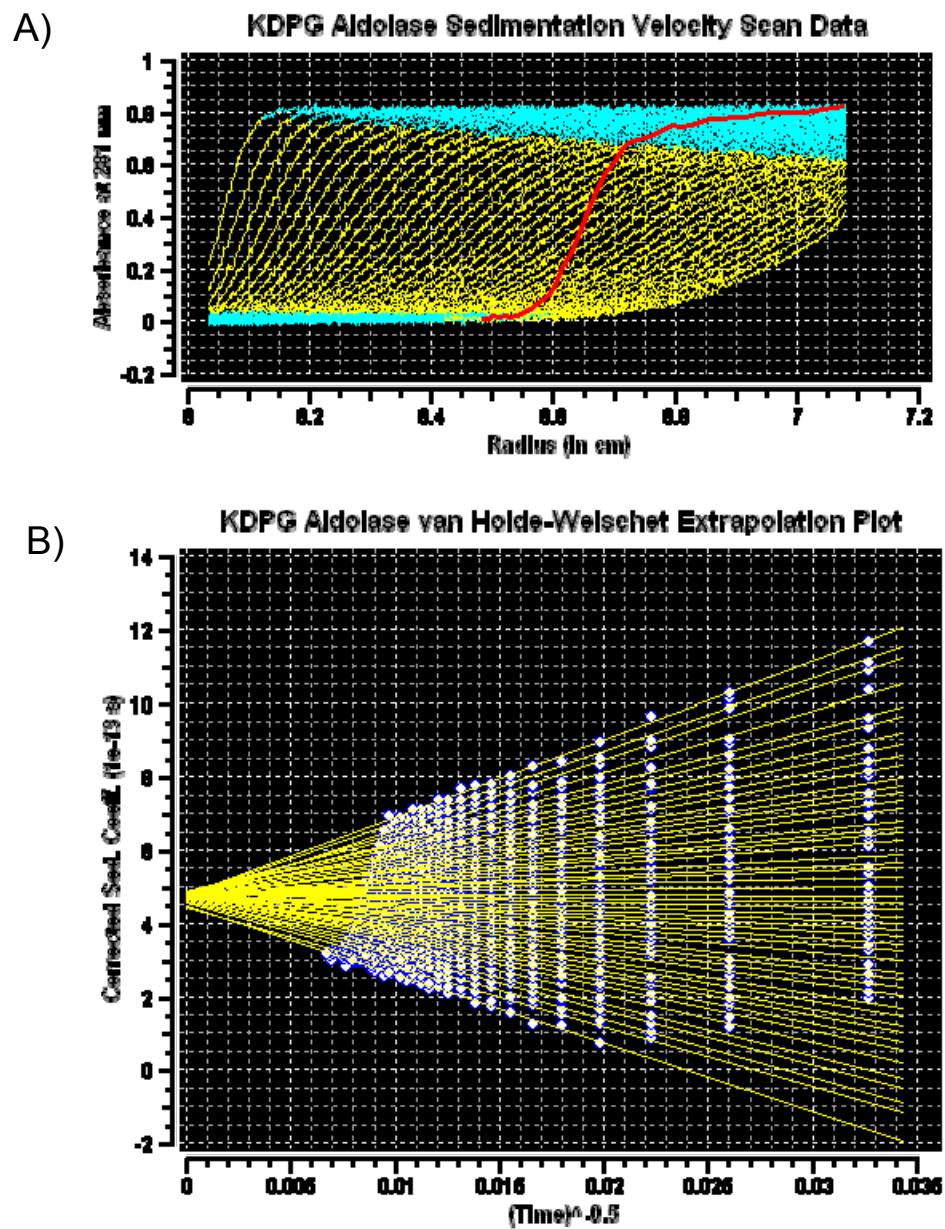
for a tetramer of trimers and a dimer of trimers, which are the oligomeric states expected for the formation of tetrahedron and dimeric cage structures. Although one can not discount formation of trimer of trimers, the peak shape better represents the formation of two species. The elution profile for the sample incubated for a longer time shows formation of higher molecular weight aggregates.

Interpretation of the 24 hour data was less straight-forward than the 2 hour incubation data. A species of ~190 kDa consistent with a dimer of trimers was again observed, however the molecular weights for species consistent with a pentamer of trimers (species of ~470 kDa) and a trimer of trimers (species of ~280 kDa) are seen at the start and end of the elution profile (Figure 4.4B). Cage structures are not predicted for a pentamer or trimer species, but could be remnants of an octamer of trimers expected to form octahedron cages. Dissociation of protein quaternary structures on SEC due to dilution is a well known phenomenon. However, the formation of octamers of trimers remains speculative and must be substantiated by other techniques.

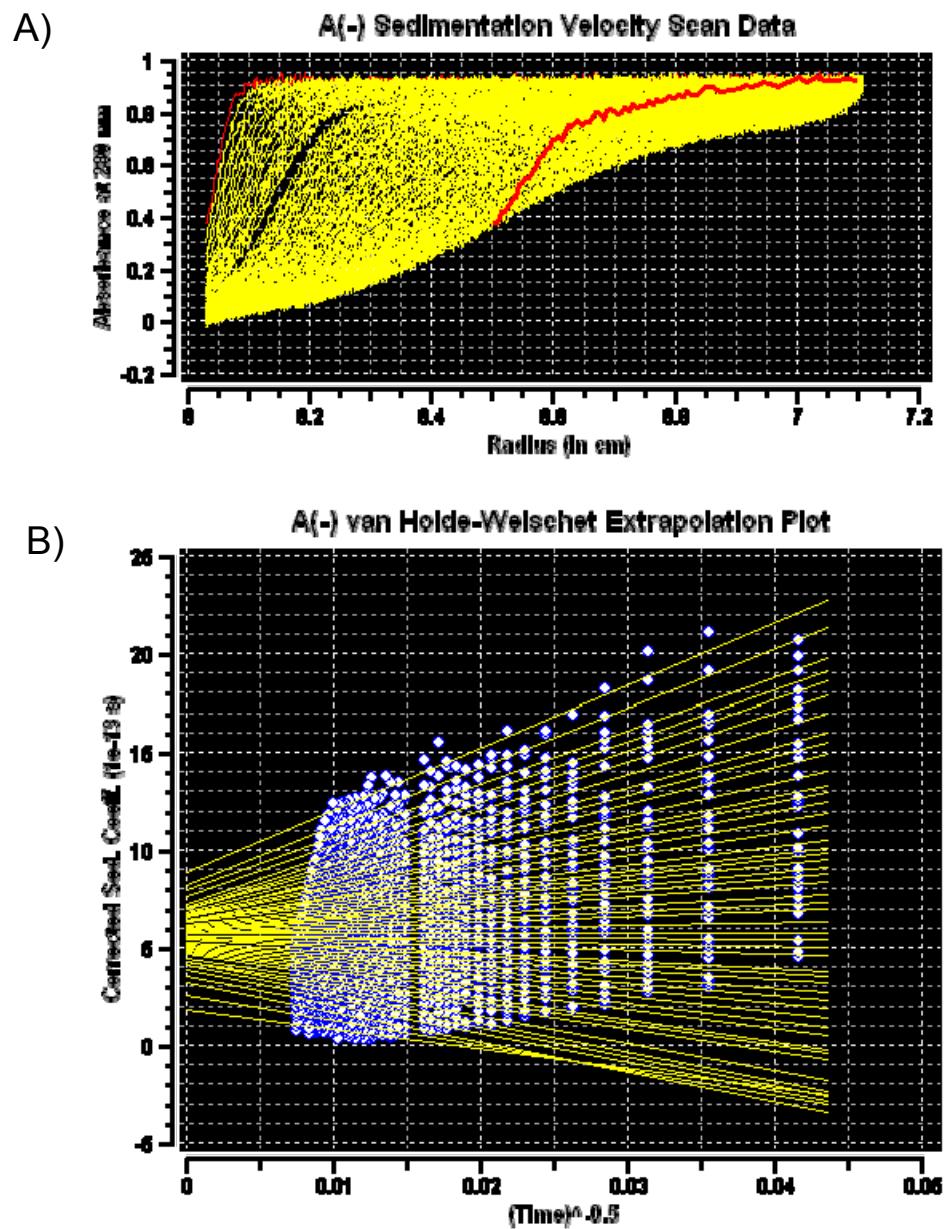
#### Analysis by Sedimentation Velocity Analytical Ultracentrifugation

More conclusive evidence that A(-) and A(+) assemble into protein cages came from sedimentation velocity analytical ultracentrifugation (SV AUC) studies. SV AUC data was first analyzed using the van Holde-Weischet method, which is generally used to determine the heterogeneity of a sample.<sup>6</sup> The method plots sedimentation coefficients for boundary fractions of each scan vs. the reciprocal square root of the scan time and then the data is extrapolated. Experimental van Holde-Weischet plots are given in Figure 4.5, 4.6, 4.7, and 4.8 for KDPG aldolase, A(-), A(+), and mix of A(-) and A(+), respectively. To explain the plots further, sedimentation coefficients are calculated for

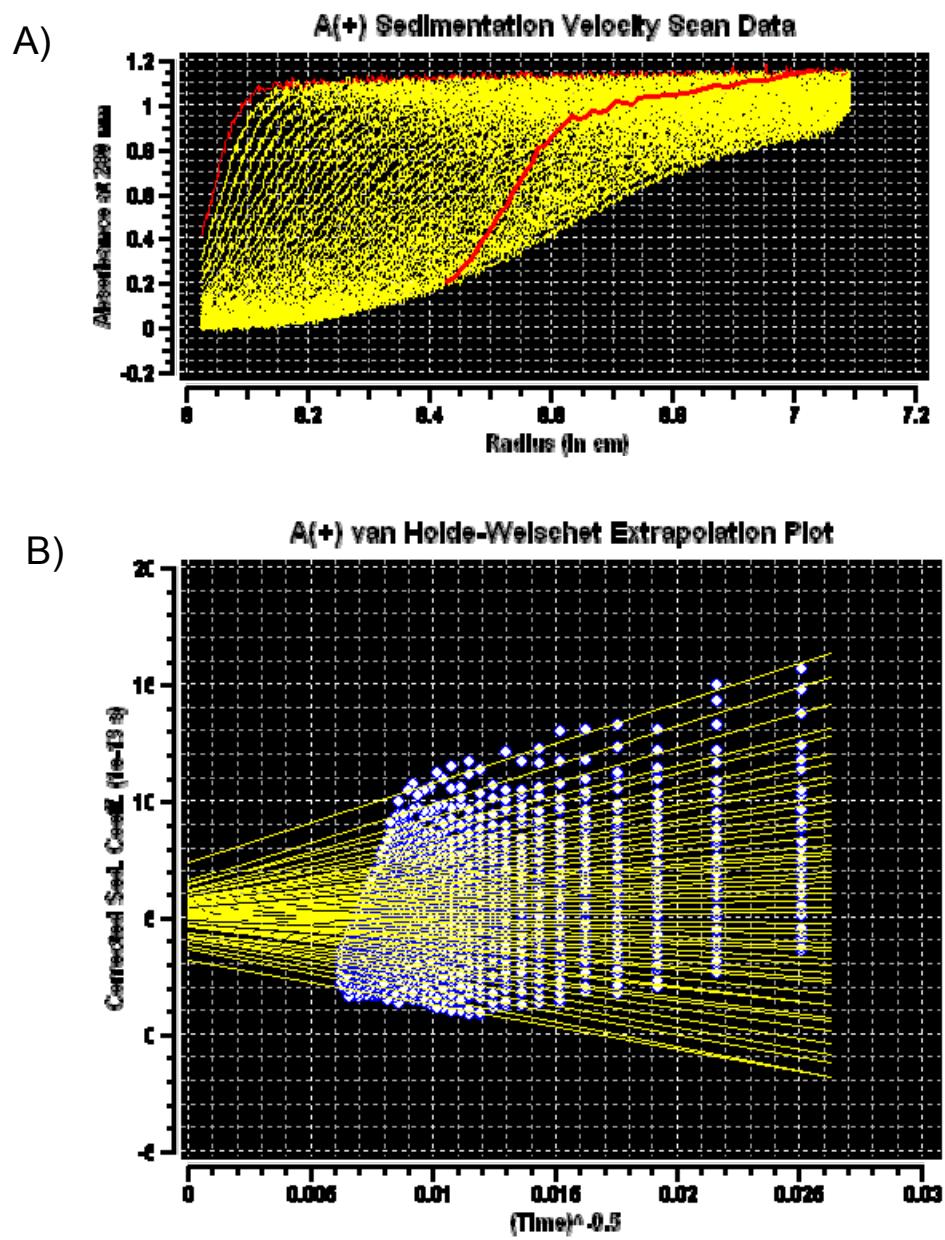
different radial positions along the sedimentation boundary of a single scan, which are plotted against the reciprocal square root of the scan time. A single scan produces a set of vertically aligned data points, a result of the scan time being the same for all sedimentation coefficients in a single scan. Multiple scans are analyzed in the manner described, excluding later scans affected by back diffusion of proteins sedimented near the centrifuge cell wall. Extrapolation to the sedimentation coefficient data is made to a reciprocal time of zero. Intersection of all the extrapolations to one point on the y-axis indicates that one species is present and the point of intersection is the sedimentation coefficient,  $s_{20w}$ , of that species. Multiple intersection points indicates that multiple species were present in the sample (i.e. the sample is heterogeneous).



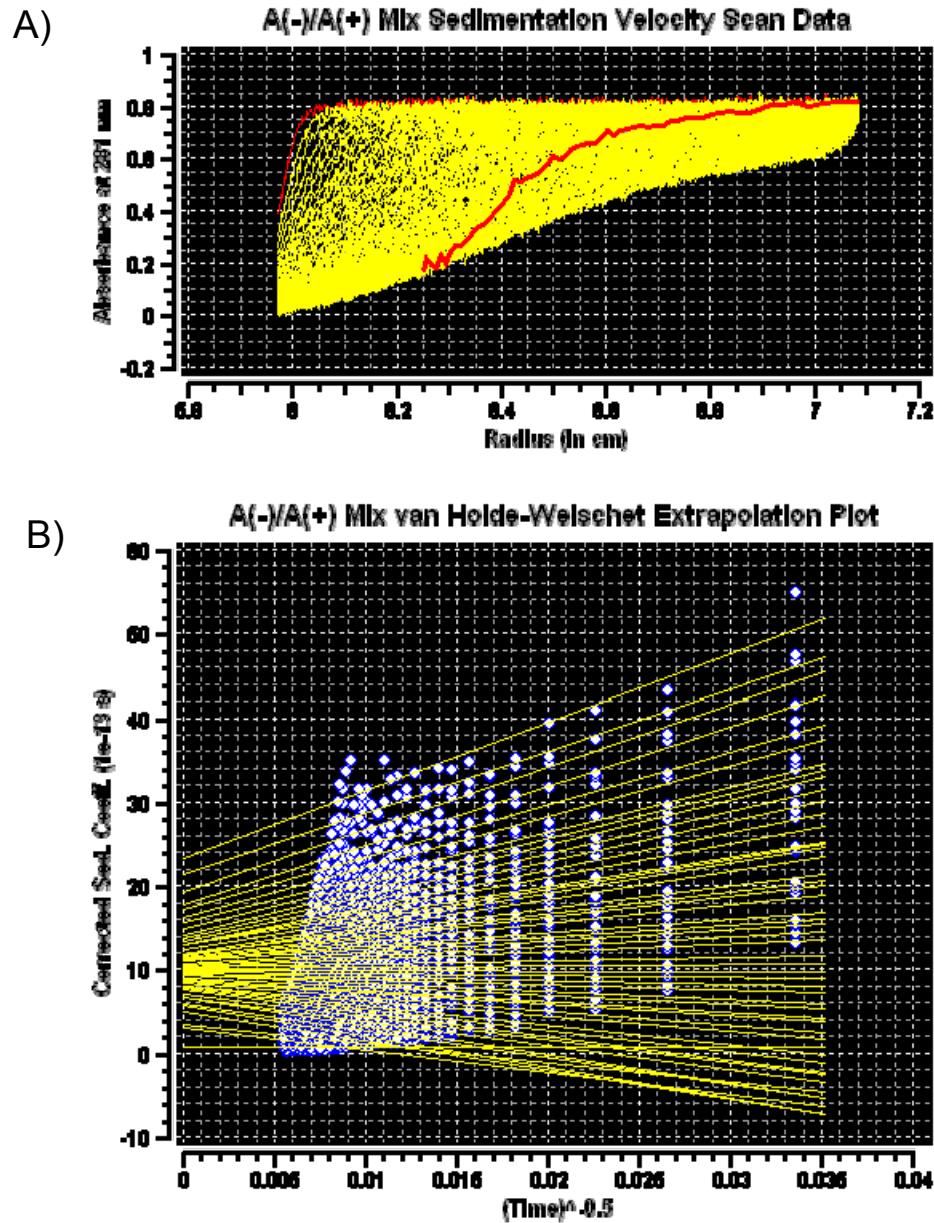
**Figure 4.5.** Sedimentation velocity analytical ultracentrifugation data and van Holde-Weischet analysis for KDPG aldolase. A) Sedimentation velocity data scans for KDPG aldolase. The red line indicates where back diffusion occurs and data after that line is omitted from van Holde-Weischet analysis. B) van Holde-Weischet extrapolation plot for scan data in A. Intersection by all lines to one point indicates that the sample contains a single species with  $s_{20w}$  of ~4.8 S.



**Figure 4.6.** Sedimentation velocity analytical ultracentrifugation data and van Holde-Weischet analysis for A(-). A) Sedimentation velocity scans for A(-). The red line indicates where back diffusion occurs and data after that line is omitted from van Holde-Weischet analysis. B) van Holde-Weischet extrapolation plot for scan data in A. Extrapolations intersect at varying points showing that more than one species was present in the sample.

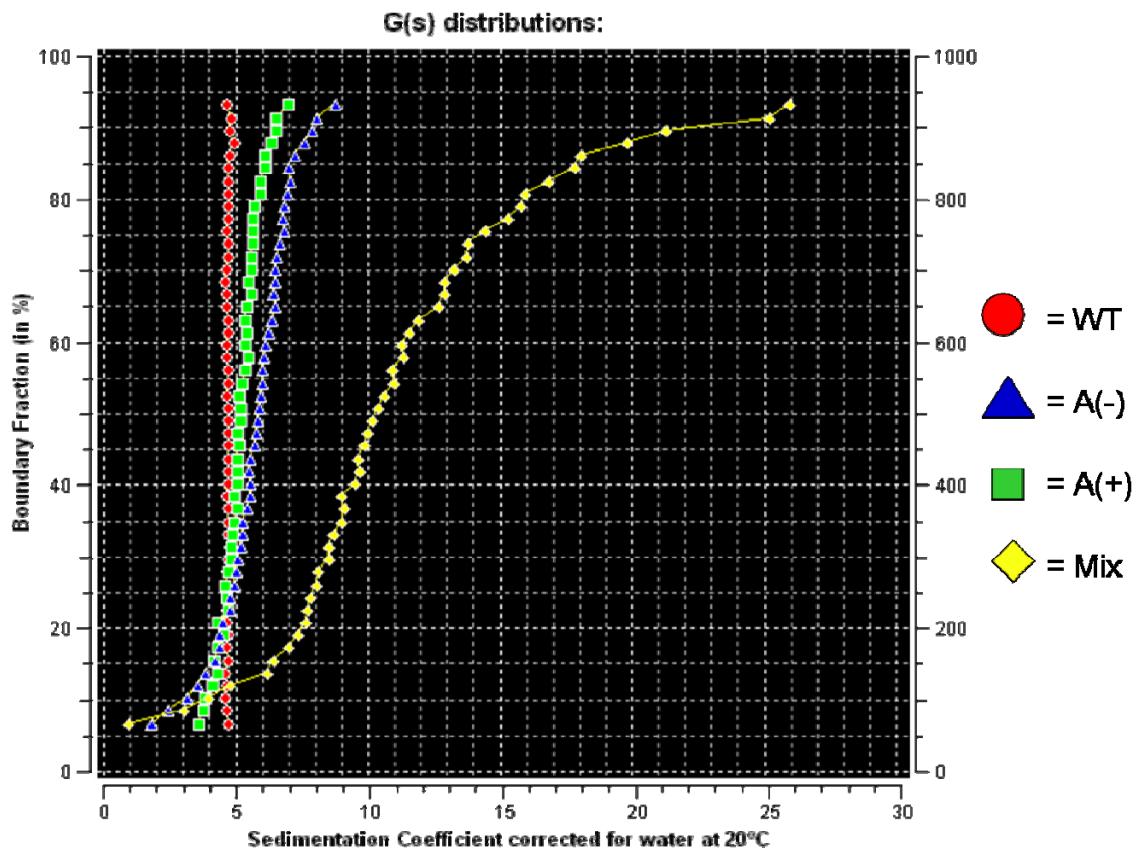


**Figure 4.7.** Sedimentation velocity analytical ultracentrifugation data and van Holde-Weischet analysis for A(+). A) Sedimentation velocity scans for A(+). The red line indicates where back diffusion occurs and data after that line is omitted from van Holde-Weischet analysis. B) van Holde-Weischet extrapolation plot for scan data in A. Extrapolations intersect at varying points showing that more than one species was present in the sample.



**Figure 4.8.** Sedimentation velocity analytical ultracentrifugation data and van Holde-Weischet analysis for 1:1 A(-)/A(+) mix. A) Sedimentation velocity scans for 1:1 A(-)/A(+) mixture. The red line indicates where back diffusion occurs and data after that line is omitted from van Holde-Weischet analysis. B) van Holde-Weischet extrapolation plot for scan data in A. Extrapolations intersect at multiple points with sedimentation coefficients much larger than seen for A(-) and A(+). These results indicate that upon mixing of A(-) and A(+) a heterogeneous mixture of aggregates much larger than the individual A(-) and A(+).

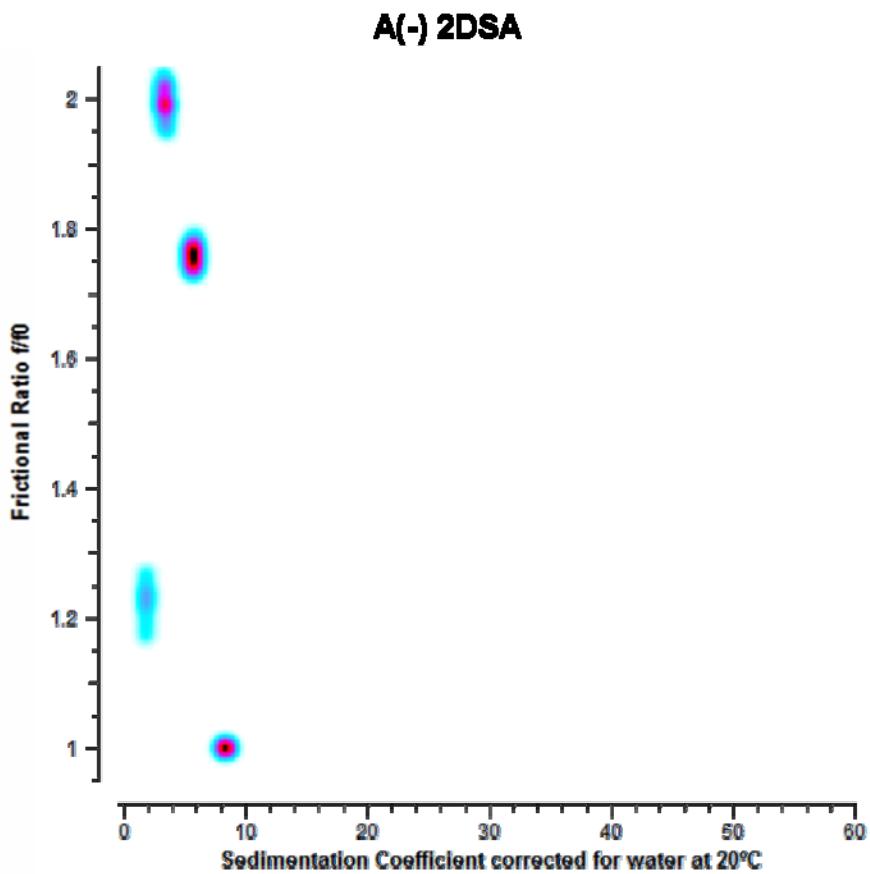
Alternatively data from van Holde-Weischet can be converted to a distribution plot, which plots the  $s_{20w}$  vs. the boundary fraction as a percentage. This allows for the calculation of the percentage of a species with a certain  $s_{20w}$ . Single species are represented by completely vertical segments with the same  $s_{20w}$  values. Tilting of the  $s_{20w}$  points indicates heterogeneity within the sample. A van Holde-Weischet distribution plot ( $G(s)$  distributions) displaying the data is given in Figure 4.9. The van Holde-Weischet distribution plot shows that wild type KDPG aldolase exists as a single species with a sedimentation coefficient,  $s_{20w}$ , of 4.79, in close agreement with the value determined for a closely related homolog.<sup>7</sup> A(-) and A(+), by contrast, were slightly heterogeneous as the  $s_{20w}$  do not align in a straight vertical arrangement, suggesting that appendage of the peptides likely cause some self-association. Mixing A(-) and A(+) showed an increase in sedimentation coefficient values, in agreement with the assembly of the proteins into higher order aggregates upon mixing.



**Figure 4.9.** SV AUC van Holde-Weischet distributions plot. Plot shows that KDPG aldolase (WT) is a single homogenous species, while A(-) and A(+) show some heterogeneity. The A(-)/A(+) mixture (Mix) shows increased S values compared to A(-) and A(+), with a larger degree of heterogeneity.

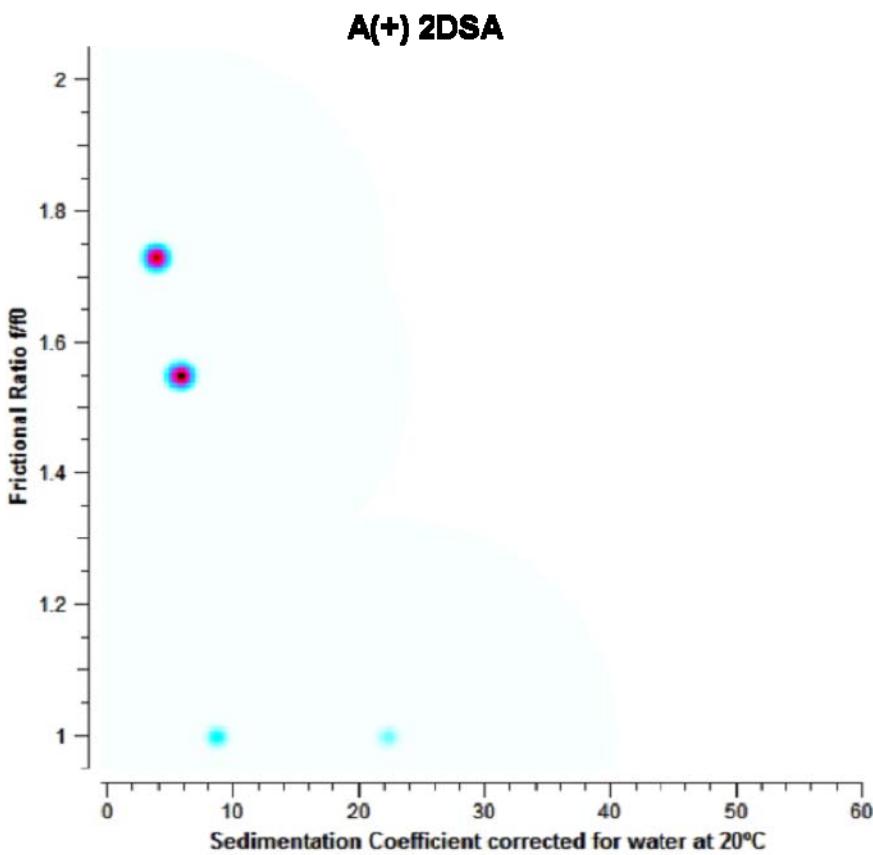
SV AUC data were further analyzed by Titus Franzmann by 2-dimensional sedimentation spectrum analysis (2DSA) to determine the number of species present, their  $s_{20w}$  values, the percentage of each present, and ultimately the molecular weights for each species from each data set. Plots and corresponding data tables of the 2DSA data generated for A(-), A(+), and the 1:1 mixture of A(-) and A(+) are given in Figures 4.10, 4.11, and 4.12, respectively. The frictional ratio ( $f/f_0$ ) describes the shape of the species, with  $f/f_0$  of 1 implying spherical structures and much higher values ( $f/f_0 > 1.3$ ) for more extended structures. The 2DSA plots and tables indicate that A(-) and A(+) are found mainly as two species of extended structure with molecular weights consistent with the

expected trimer and an unexpected dimer of trimers. Whereas the analysis shows some self-association of the protein building blocks, it is apparent that these multimers are transient and are displaced by complementary hetero-association upon mixing. This is evident by the assembly of much higher order structures and the absence of these species upon mixing. The 2DSA of the mixture shows the emergence of three major species that make up ~50% of all species present with  $s_{20w}$  of 9.2, 17.6, and 26.9 S and frictional ratios ( $f/f_0$ ) close to 1. Molecular weights for these species correspond to those predicted for a dimer, tetramer, and octamer of trimers. In addition, the frictional ratios of close to 1 imply spherically symmetric structures, consistent with the formation of closed cage structure. Two other species are present in relatively large abundance with  $s_{20w}$  of 6.8 S and 12.6 S displaying frictional ratios indicative of elongated structures. Whereas the exact identity cannot be conclusively determined, the molecular weights of 216 kDa and 414 kDa for the two species indicate oligomers are composed of approximately 7 and 14 aldolase fusion monomers or approximately 2.4 and 4.6 trimers. Exchange of subunits upon mixing could be a possible explanation for the observed states, with the smaller species retaining a single subunit and the larger species retaining two subunits. Both species are found in approximately equal abundance. Alternatively and more probable, these species are extended dimer of trimers and tetramer of trimers unable to form a cage or trimer of trimers and pentamer of trimers, which were observed in the HP-SEC experiment.



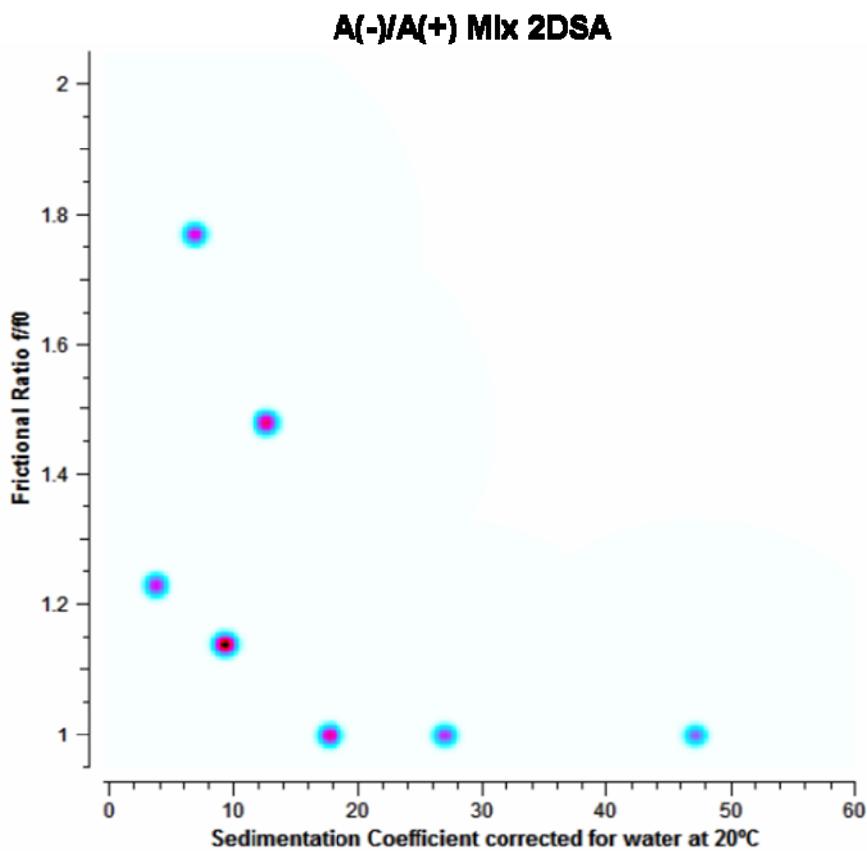
	$s_{20w}$ (Svedberg)	f/f₀	Mol. W. (kDa)	Cone. (%)
1	1.8	1.22	16.7	12
2	3.27	1.99	84.9	31
3	5.64	1.77	163	39
4	8.3	1.00	123	18

**Figure 4.10.** Results of the 2DSA for A(-) samples. Each spot on the 2DSA plot indicates a unique species. Data from the 2DSA is also given in tabulated form below the 2DSA plot with additional information about the molecular and total concentration, as a percentage, of each species. The major species are consistent with a trimer (expected ~89 kDa) and dimer of trimers (expected ~178 kDa). The species of 8.3 S likely is the result of dissociation of trimers forming the dimer of trimers. The smaller 1.8 S species is a potential degradation product that was not removed during dialysis (Mol. W. cutoff of 10 kDa for tubing).



	$s_{20w}$ (Svedberg)	$f/f_0$	Mol. W. (kDa)	Conc. (%)
1	<b>3.8</b>	<b>1.73</b>	<b>91</b>	<b>44</b>
2	<b>5.7</b>	<b>1.55</b>	<b>143</b>	<b>47</b>
3	<b>8.7</b>	<b>1</b>	<b>130</b>	<b>7</b>
4	<b>22.2</b>	<b>1</b>	<b>585</b>	<b>2</b>

**Figure 4.11.** Results of the 2DSA for A(+) samples. Each spot on the plot indicates a unique species. Data is also given in tabulated form below the data plot. The major species with an  $s_{20w}$  of 3.8 S has molecular weight consistent with the trimeric protein (89.8 kDa). The second major species with an  $s_{20w}$  of 5.7 S is suspected to have resulted from the dissociation of a dimer of trimers, yielding the molecular weight between a trimer and dimer of trimers. Minor amounts of other aggregates are also observed, possibly formed from misfolded protein.

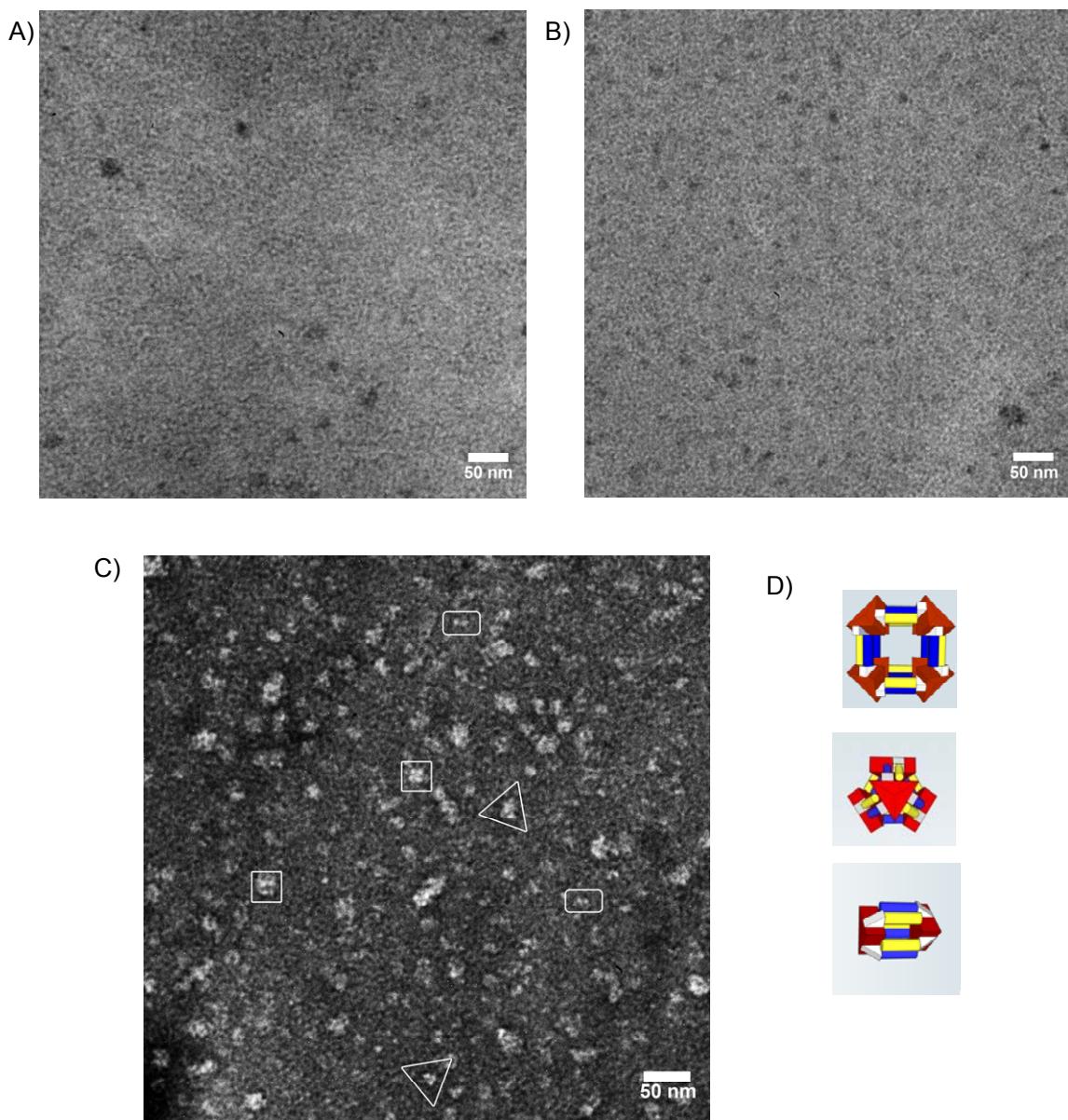


	$s_{20w}$ (Svedberg)	$f/f_0$	Mol. W. (kDa)	Conc. (%)
1	3.65	1.24	49	10
2	6.83	1.76	217	13
3	9.22	1.13	176	25
4	17.6	1.00	380	16
5	12.6	1.46	414	17
6	26.9	1.00	721	10
7	47.1	1.00	1,667	7

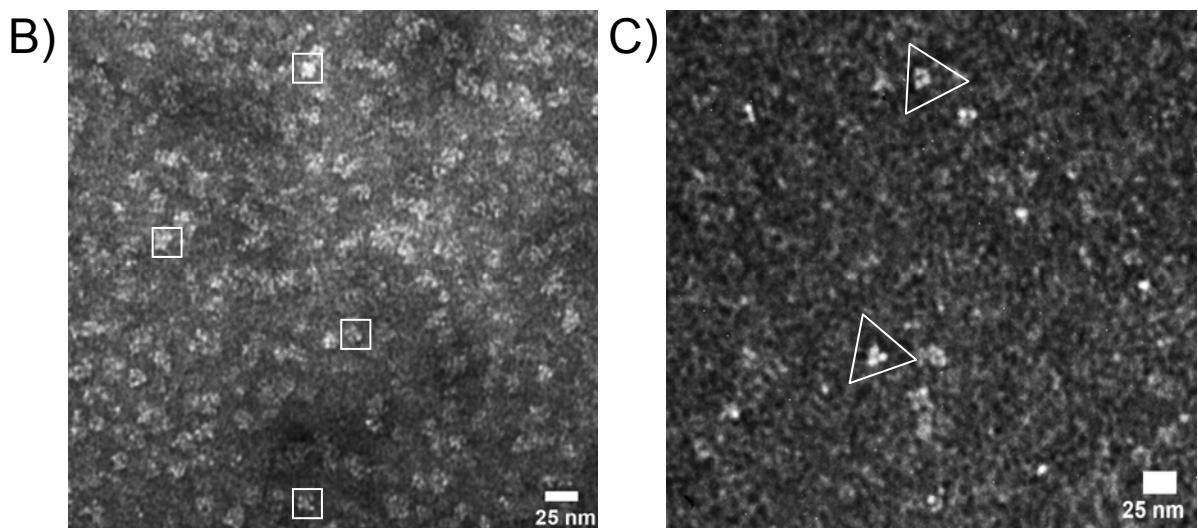
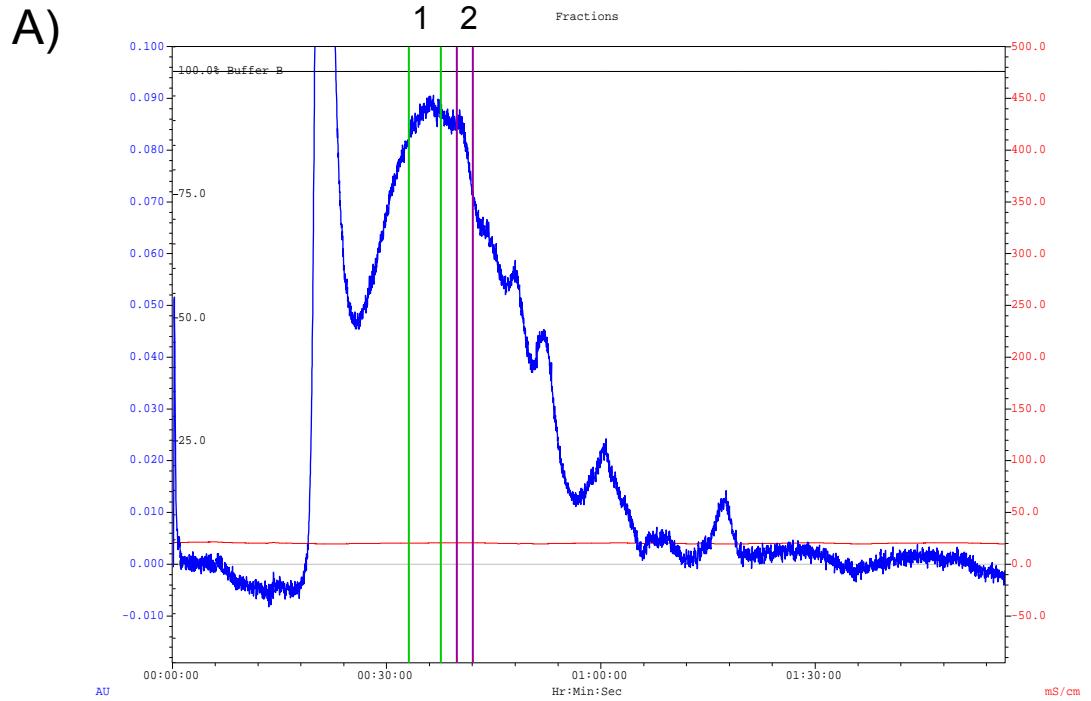
**Figure 4.12.** Results of the 2DSA for A(-)/A(+) mixture samples. Each spot on the plot indicates a unique species. Data is also given in tabulated form below the 2DSA data plot. Species 3, 4, and 6 give molecular weights and  $f/f_0$  consistent with those predicted for “sandwich” (~180 kDa), tetrahedral (~360 kDa), and octahedral (~720 kDa) cage structures. Other species are discussed in more detail within the text.

### Investigations of Protein Assembly by Transmission Electron Microscopy

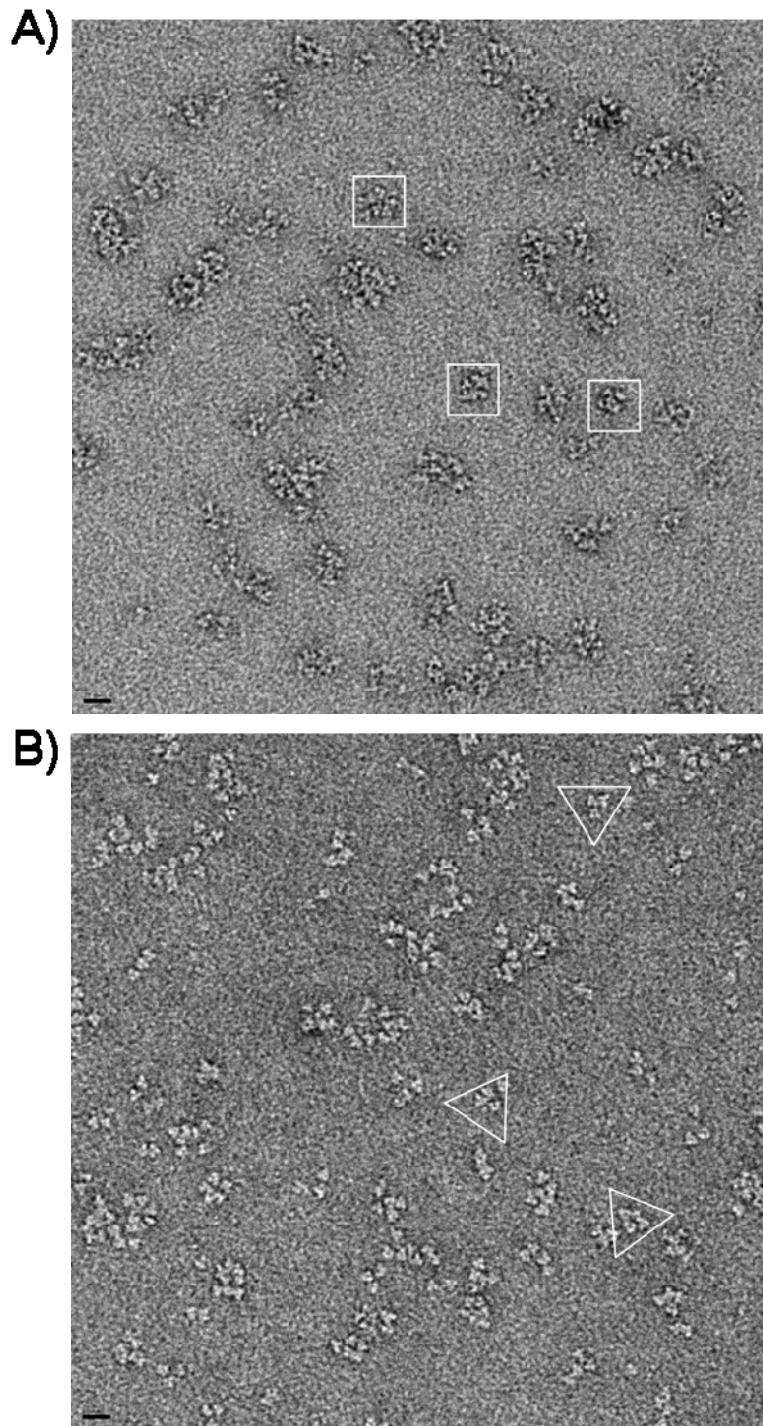
Negative stain transmission electron microscopy (TEM) of samples of A(-) and A(+) gave images that were similar to control samples containing buffer(Figure 14.3A and B, buffer not shown). TEM images of 1:1 mixtures revealed discrete particles with sizes of 10-20 nm consistent with sizes expected for the cage structures (Figure 4.13C). Particles exhibiting morphologies consistent with that of octahedron, tetrahedron and dimeric cages were observed (Figure 14.3C and schematics in 14.3D). SEC was performed on samples to determine if the different protein species could be separated to get images containing only one species. Protein standards were run first to determine what elution volume octahedron (octamer of trimers) or tetrahedron (tetramer of trimers) would likely elute. Elution profiles for mixed sample did not show separation of discrete species, so fractions were taken from volumes expected to contain the octahedron and tetrahedron cage species of interest (Figure 4.14A). SEC fractions contained particles with expected morphologies for octahedron and tetrahedron, although other morphologies were also observed (Figure 4.13B and C). Investigation of SEC fractions under higher resolution, performed by Mr. Justin Schilling in the Skiniotis lab, suggested that most of the protein particles collapse under the harsh staining conditions. However, the higher resolution images allowed the observation of triangular particles of sizes consistent with KDPG aldolase and clusters that corresponded well with expected octahedron and tetrahedron structures (Figure 4.15).



**Figure 4.13.** TEM images of A(-), A(+), and 1:1 A(-)/A(+) mix with schematic visual aides. A) TEM image of A(-). B) TEM image of A(+). C) TEM image of 1:1 mixture of A(-) and A(+). Squares, triangles, and rectangles are used to outline particles with morphologies consistent with those of octahedral, tetrahedral, and “sandwich” cage structures respectively. D) Schematic representations of octahedron, tetrahedron, and “sandwich” cage structures to serve as visual aides.



**Figure 4.14.** Size exclusion chromatography separation of A(-)/A(+) mixture aggregates and TEM images of selected fractions. A) Size exclusion chromatogram for elution 1:1 mixture of A(-) and A(+) with fractions used for TEM imaging outlined with green and purple lines. B) TEM image of sample from fraction 1 shown in the chromatogram in B. C) TEM image of sample from fraction 2 shown in the chromatogram in B. Squares and triangles outline particles showing octahedron and tetrahedron morphologies respectively.



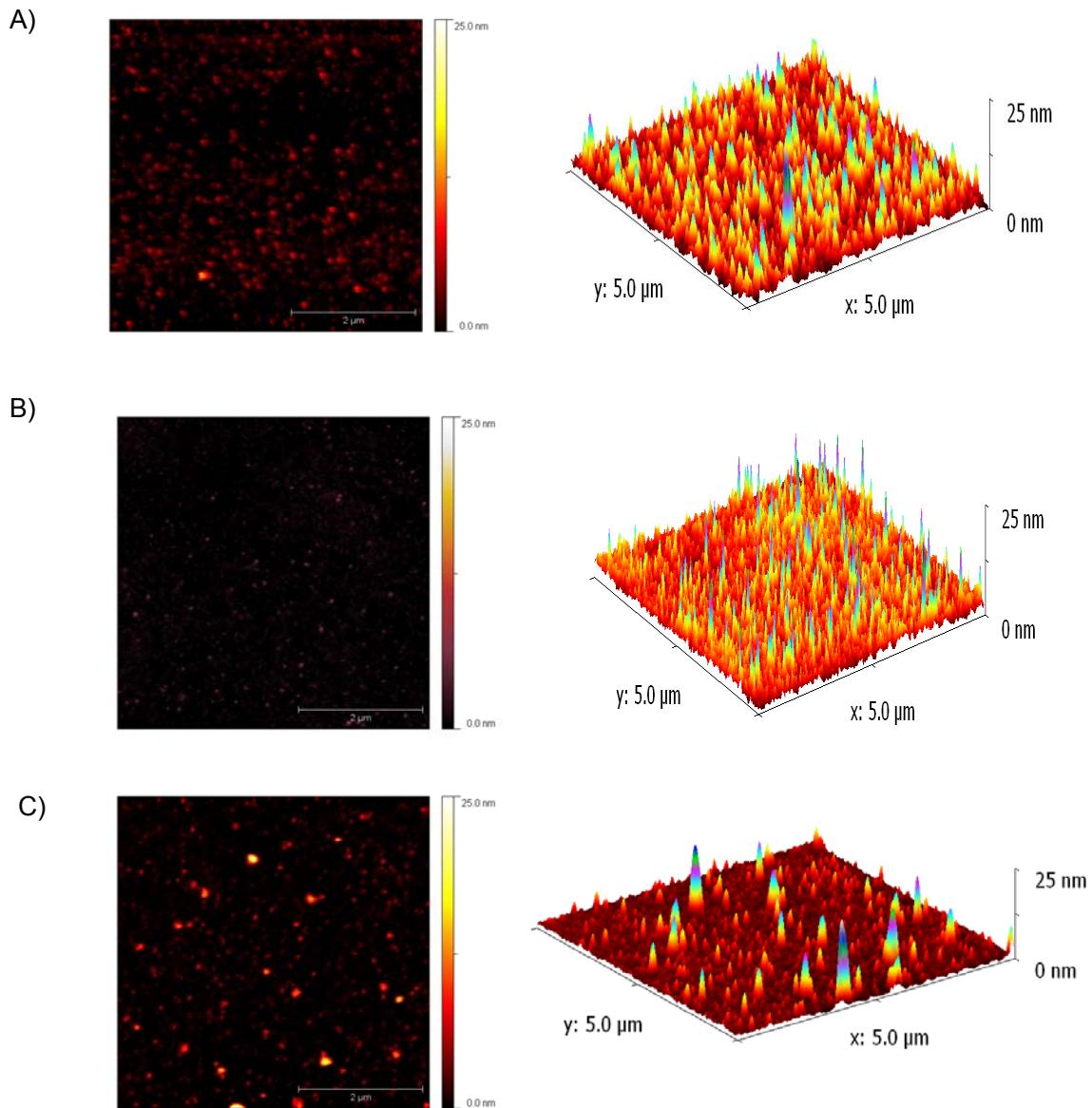
**Figure 4.15.** High resolution TEM images of SEC purified A(-)/A(+) mixture protein fractions. A) Sample taken from fraction of similar elution volume to that of fraction 1 shown in Figure 4.14A. B) Sample taken from fraction of similar elution volume to that of fraction 2 shown in figure 4.14B. The scale bar represents a length of 20 nm. Squares and triangles outline example protein clusters displaying octahedral and tetrahedral “like” arrangements respectively.

### Atomic Force Microscopy for Investigating Protein Heights

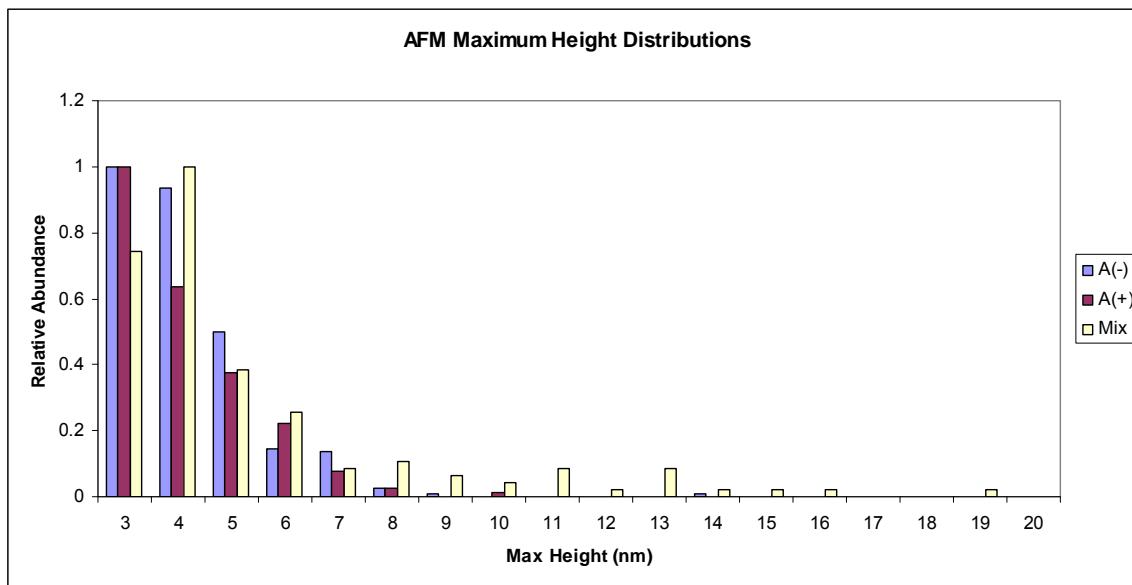
Analysis of protein samples by atomic force microscopy (Figure 4.16) was limited to the determination of heights because the tips used for scanning have much larger diameters than size of the proteins being examined. However, information obtained from AFM was expected to complement the findings from TEM, which were limited to x-y dimensional space. Images taken of KDPG aldolase, A(-) and A(+) in aqueous buffer produced similar results with most particles having maximum heights measuring 3.0-4.0 nm (Figure 4.17 and Figure 4.18B), consistent with the ~3.8 nm maximum height expected for KDPG aldolase lying with its triangular face, the greatest surface area available, against the mica. Taller height values were also observed, likely due to proteins lying in alternative fashions, such as on edge. The longest distance measured across the trimeric structure of KDPG aldolase is less than 10 nm, consistent with this, almost no particles above 10 nm were observed by AFM for the non-mixed proteins. Images taken of samples comprising a 1:1 mixture of A(+) and A(-) showed populations of particles with taller heights of 12-20 nm, consistent with dimensions expected for cages (Figure 4.17). Far fewer particles with sizes of 12-20 nm were seen in the samples comprising mixtures of A(-) and A(+) than would have been expected based on the other methods utilized. Most particles comprising A(-) and A(+) mixtures had heights similar to individual A(-) and A(+) samples.

It was observed that A(-) and A(+) attached to mica particularly well and required vigorous washing to obtain images of separated protein particles. However, wild type KDPG aldolase did not attach well to mica and concentrations in excess of 10 times that of A(-) and A(+) had to be applied without rinsing to obtain images, which still

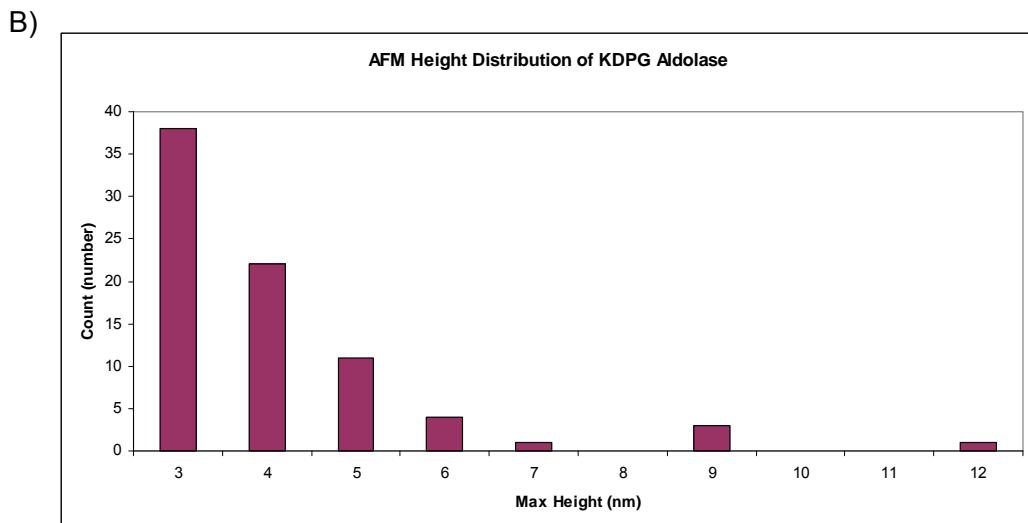
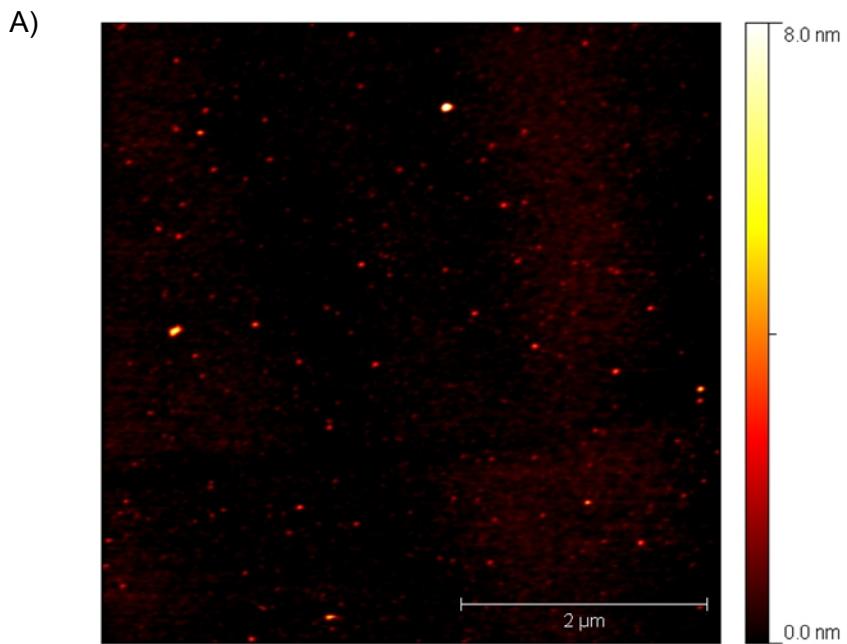
showed relatively few particles (Figure 4.18). These findings suggest that the charged peptides of A(-) and A(+) have an affinity towards the mica surface used for imaging. Cage assemblies formed from A(-) and A(+) are potentially washed away because coiled coil peptides are charge-paired and have less affinity towards the mica and/or the cages break apart. Alternatively, the large dilution upon washing could cause dissociation of the coiled coil, allowing peptides to interact and bind to the mica surface. Although the relative population of species with sizes predicted for the protein cages is small and statistically minute, the observation of them by AFM provides further evidence supporting the formation of cages upon mixing of A(-) and A(+). Checks were made to investigate whether proteins were picked up by the tip during scanning, which would be a concern for proteins that don't appear to adhere well to the surface, but this did not appear to be an observed problem.



**Figure 4.16.** AFM images of protein samples. A) Two image perspectives from AFM imaging of A(-). B) Two image perspectives from AFM imaging A(+). and C) Two image perspectives from AFM imaging a 1:1 mixture of A(-) and A(+). Images shown are for the same scans. The scale bar in the AFM images represents 2  $\mu\text{m}$  for all images. Differences in the x-y dimensions of particles are due to differences in the tip used for imaging (the tip is larger than the particles imaged and thus the x-y from the AFM images are images of the tips used).



**Figure 4.17.** AFM height distributions for images taken of A(-), A(+), and an A(-)/A(+) mixture (data for images in Figure 4.15). The data was normalized for comparison. The images contained the following total number of counted particles: A(-) 304 particles, A(+) 181 particles, and A(-)/A(+) mix 140 particles. Only the A(-)/A(+) mix sample shows a population of particles with sizes expected for the predicted cage structures (12-20 nm).

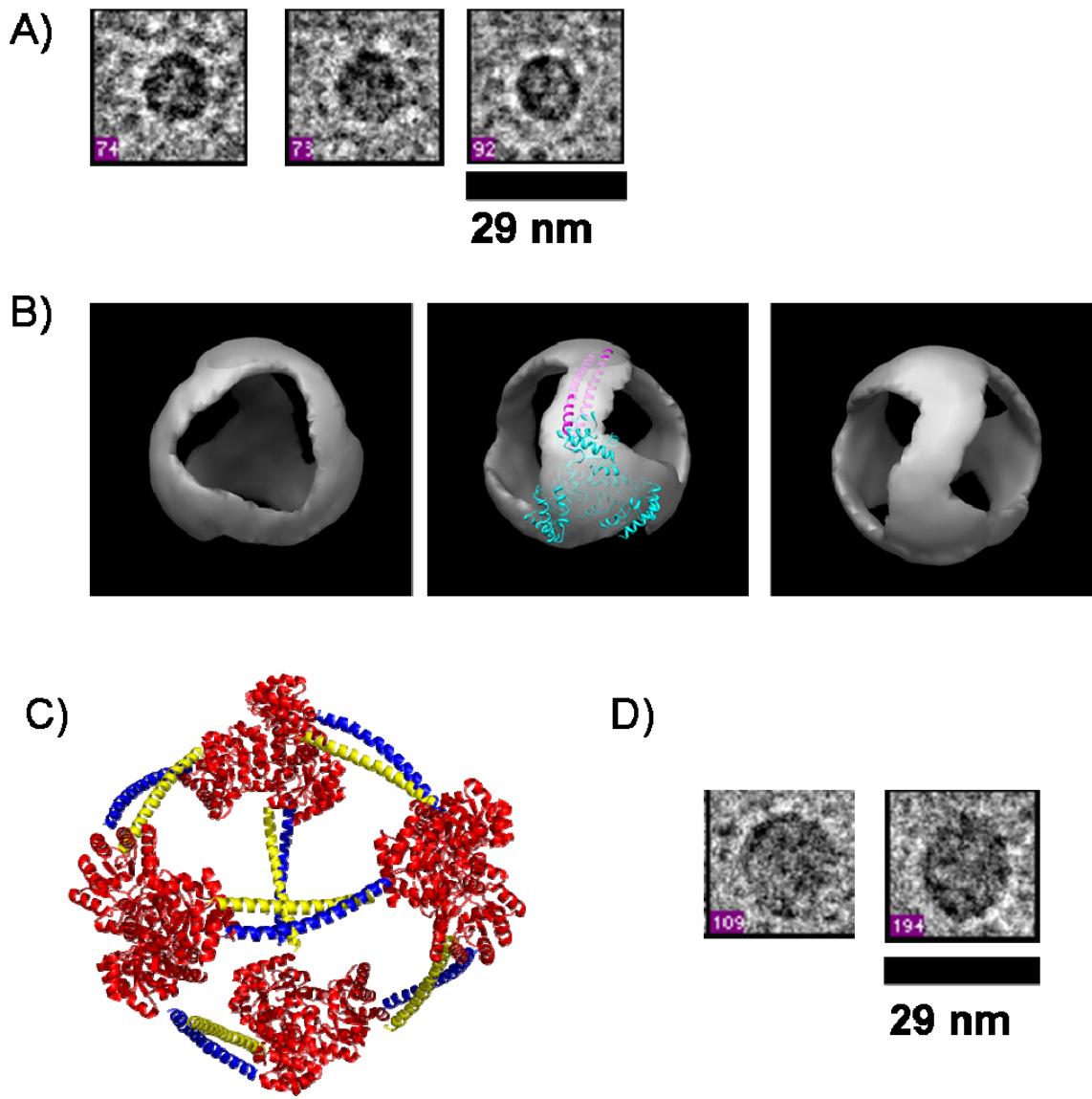


**Figure 4.18.** Atomic force microscopy imaging of KDPG aldolase and analysis. A) AFM image of KDPG aldolase. B) Maximum height distribution graph of particles from image shown in A. Only 80 total particles above 2.25 nm were counted even though over 10 times more protein was applied than for the fusion protein samples.

#### New Insights from Cryo-Electron Microscopy

Whereas experiments in cryo-electron microscopy and subsequent image reconstruction are far from complete at the time of writing, promising results have been obtained after only a few imaging attempts. Fractions separated by SEC have been

imaged and discrete particles have been observed. Sizes for the particles imaged were consistent with values expected for tetrahedron cage structures (~16 nm) (Figure 4.19A). Dr. Min Su has collected a set of 383 particle images and has generated a low resolution reconstruction model of a tetrahedron from these. The crystal structure of KDPG aldolase fits well into the triangular portions of the reconstruction map, as does a model for a coiled coil containing 6-heptad repeats excised from a protein crystal structure (Figure 4.19B). However, the reconstruction is relatively low resolution and the electron density of the model is thin relative to the dimensions of the component structures. This is likely due to fluctuations of the cage structure which would smear out electron density. This may be caused by the flexible spacer connecting the peptide to the aldolase subunits. By obtaining images of many more particles we hope to boost the resolution to give a more defined reconstruction. A protein crystal structure model based on the cryo-EM reconstruction is shown in Figure 4.19C. Detection of particles that may correspond to the octahedron has proved more difficult. A few particles with sizes expected for an octahedron (~20 nm) have been observed (Figure 4.19D). However, images suggest that the structure is breaking down, possibly during the freezing process. Chemical cross-linking studies to stabilize the structure are underway to try to overcome this problem.



**Figure 4.19.** Results from cryo-electron microscopy of A(-)/A(+) mix samples separated by SEC. A) Cryo-images of particles with sizes and shapes consistent with the tetrahedron structure. B) Particle reconstruction model built from larger library of particles that include those in A. The center image shows the fit of crystal structure of KDPG aldolase and a 6 heptad coiled coil. C) Model constructed from the crystal structure of KDPG aldolase and a 6 heptad coiled coil based on the reconstruction model. D) Images taken of larger particles that have sizes and shape consistent with octahedron.

#### 4.4 - Discussion

I described experiments to investigate a new and general approach for constructing protein cages. Fusion proteins of KDPG aldolase and coiled coil peptides

with 6xHis tags were successfully expressed and purified from *E. coli*. Grafting the peptides to KDPG aldolase resulted in no deleterious effects to the function of the enzyme and mixtures of the two proteins showed no reduction in activity. HP-SEC and SV AUC demonstrated that neither the oligomeric state nor quaternary structure of the KDPG aldolase protein was affected by incorporation of the peptides. Although some transient self-dimerization of A(+) and A(-) was observed by AUC, assembly into higher order structures was only observed upon mixing of the two protein constructs.

HP-SEC data showed species with molecular weights consistent with a tetrahedron and triangular prism cage formed after two hours, but much larger structures were observed for later incubation periods. Definitive evidence of cage formation and the identities of the structures formed came from SV AUC studies. Analysis of SV AUC data indicated the formation of three dominant species with molecular weights consistent with a prism, tetrahedron, and octahedron cage. The frictional ratios for these species showed they were spherical ( $f/f_0=1$ ) as one would expect for a closed cage structure. Closer inspection of the smallest species (~175 kDa) shows its frictional ratio deviates slightly from spherical, consistent with formation of the slightly more elongated structure expected for a prism cage. Furthermore, negative stain TEM and AFM provide more evidence for cage formation as the morphologies and sizes observed from these imaging techniques agree with prism, tetrahedron and octahedron formation. In addition, data being obtained from cryo-EM is providing even more definitive evidence; a low resolution reconstruction of a tetrahedron has been generated.

Formation of tetrahedron would be unexpected if A(-) and A(+) subunits persist as homotrimers. Tetrahedron formation could only result from the exchange of subunits

between A(-) and A(+). Such exchange, although hard to detect, is not unreasonable to suspect since the interactions between subunits is non-covalent and monomeric aldolase mutants have been constructed by only a double point mutation at the association interface.

While the observed structures fit well to those predicted from the symmetry, other structures also appear to be formed. Additional non-cage structures are also observed to assemble. However, one must keep in mind that our design invokes minimal restrictions, practically none, on the types and number of structures that could potentially form. The symmetry of the PBB, to our knowledge, provides the only major control over the final assembly. Connection of the peptide linkers to the PBB by the flexible peptide sequence is expected to allow the peptide linkers a high degree of structural freedom and not preclude it to a single defined orientation. This was done deliberately so that we could test what structures do form with only the symmetry of the protein to specify what is made. One might predict that giving such a large degree of conformational freedom would allow many structures to form. However, results from SV AUC show that from the infinitely large number of possible assemblages that can form, only six major species are observed to form (disregarding the degradation species). That only six major species form is quite astounding; more importantly three of the six species have molecular weights and shapes consistent with the prism, tetrahedral, and octahedral cages predicted to form, with all additional spectroscopic evidence supporting their formation. The method proposed by Yeates and co-workers, which required rigidly defined protein building components, also produced alternative protein structures that were not predicted to form in their design, although they were not exclusively identified.<sup>1</sup> The findings

presented here suggest that the symmetry of the protein building blocks dictates, to a high degree, the structure that forms and imposing rigid restrictions as prescribed by Yeates are not necessarily required to construct desired cage structures. Our strategy has the potential to allow a greater number of proteins to be incorporated as protein building blocks. The results presented here should serve as a foundation for future research in assembling protein cage biomaterials. Further design modifications may allow a higher degree of selectivity for limiting the cage structures that form when multiple architectures are possible.

#### 4.5 – References

- 1 Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments *Proc Natl Acad Sci U S A* **98**, 2217-2221 (2001).
- 2 Ni, T. W. & Tezcan, F. A. Structural characterization of a microperoxidase inside a metal-directed protein cage. *Angew Chem Int Ed Engl* **49**, 7014-7018 (2010).
- 3 Fullerton, S. W. *et al.* Mechanism of the Class I KDPG aldolase. *Bioorg Med Chem* **14**, 3002-3010 (2006).
- 4 Griffiths, J. S. *et al.* Cloning, isolation and characterization of the Thermotoga maritima KDPG aldolase. *Bioorg Med Chem* **10**, 545-550 (2002).
- 5 Ball, V. & Ramsden, J. J. Buffer Dependence of Refractive Index Increments of Protein Solutions. *Biopolymers* **46**, 489-492 (1998).
- 6 Demeler, B. & van Holde, K. E. Sedimentation velocity analysis of highly heterogeneous systems. *Anal Biochem* **335**, 279-288 (2004).
- 7 Meloche, H. P. & Wood, W. A. Crystallization and Characteristics of 2-Keto-3-Deoxy-6-Phosphogluconic Aldolase. *J Biol Chem* **239**, 3515-3518 (1964).

## Chapter 5

### Conclusions and Future Directions

This thesis evaluates a method for assembling proteins into cage structures. The strategy is markedly different from those previously proposed by others. Most notably, the PBBs and protein dimerization domains are not rigidly linked to achieve a fixed geometry, but instead are connected through a flexible linker.<sup>1,2</sup> We predicted that the topology of the assembled protein cage could be controlled by the symmetry associated with the PBB's quaternary structure, which would favor cage structures that are compatible with that symmetry. As an experimental test of this strategy, KDPG aldolase was selected as a trimeric PBB and a set of *de novo* designed heterodimeric coiled-coil peptides were designed to act as linkers. Overall the experimental results support our hypothesis, and the strategy appears promising.

Mixing fusion proteins A(-) and A(+) produced species consistent with a dimeric prism, a tetrameric tetrahedron, and an octameric octahedron, which were observed by multiple complementary spectroscopic techniques. Preliminary results from cryo-EM provide convincing evidence for assembly of the tetrahedron cage. Further investigations are underway to bolster confidence in the cryo-EM reconstruction of the tetrahedral cage and also to try to obtain an image reconstruction of the octrahedral cage. As discussed in previous chapters, a limited number of non-cage species were also observed to form. However, this has also been the case for other strategies that utilized more stringent requirements to control assembly of PBBs with the aim of achieving unique structure.<sup>1</sup>

The method developed in this dissertation provides greater potential for incorporating a larger set of proteins as potential building blocks than previous methods.

In formulating our design we tried to incorporate a set of five criteria (listed in Chapter 2 and Chapter 4) that we felt were important for making the method robust and more universal. The results from my investigations show that we were able to meet many of these criteria. The coiled coil peptides Helix(-) and Helix(+) provided a genetically encoded linker (criterion 1) that allowed the controlled assembly of proteins (criterion 2); assembly only occurred upon mixing of proteins containing each peptide. Assembly led to the formation of a limited number of predictable cage structures (criteria 4), although a few non-cage structures formed as mentioned. Assembly of the proteins into the higher order structures also did not adversely affect the biological activity *in vitro* and would not be expected to do so *in vivo* (criterion 5).

Design criterion 1 was to allow assembly of structures *in vivo*, however this was not explicitly shown (by for example expression of both A(-) and A(+) in the same cell) and monitoring assembly *in vivo* would likely be difficult. However examples of using coiled coils to link proteins *in vivo* have been described.<sup>3-6</sup> Zang et al. showed supramolecular self assembly of eCFP and eYFP can be mediated by lithocholic acid and  $\beta$ -cyclodextrin to give a FRET response *in vivo*.<sup>7</sup> In future work, the ability of Helix(-) and Helix(+) to dimerize proteins *in vivo* could be determined by performing a FRET assay of enhanced cyan fluorescent protein (eCFP) and enhanced yellow fluorescent protein (eYFP) tagged with Helix(-) and Helix(+). Assembly (dimerization) of the fluorescent proteins mediated by the heterodimeric coiled coil formation would be expected to show a FRET response. To evaluate *in vivo* assembly three cell lines of *E.*

*coli* would be made, one containing a plasmid for expression of fusion of eCFP with one peptide linker, one containing a plasmid for expression of the eYFP with the other peptide linker, and finally a cell line expressing both. Cell lines containing both proteins would be expected to show a FRET response if Helix(-) and Helix(+) link eCFP and eYFP.

In the near future, additional work is needed to prove that our method can be generalized to a broad range of proteins (criterion 3). In addition, the method needs to be investigated further to determine if the strategy yields a limited number of predictable structures for other protein symmetries. I showed that KDPG aldolase can be assembled into cages using the strategy and yields results consistent with those hypothesized. However, the triangular shape of KDPG aldolase allows the assembly of several cage structures based on the symmetry. Other symmetries by contrast have more limitations on the structures they can form (Figure 2.7). To show that our strategy can incorporate more proteins and can be used to form fewer structures, I propose that an additional study be performed using KDPG aldolase as one building block and a tetrameric protein with a “square” quaternary structure, such as RuvA (Figure 2.7), as the other. Based on symmetry, a triangle and square are expected to form one unique cage structure, a cuboctahedron (Figure 2.7). Construction of the tetrameric fusion protein will only require removal of the KDPG aldolase gene from one of my plasmid constructs, replacement with the gene encoding the desired tetrameric protein subunit, and subsequent expression and purification by nickel affinity chromatography. Analysis of mixture samples by analytical ultracentrifugation should allow a quick assessment to see that fewer structures are made, ideally one species consistent with a cuboctahedral cage.

One aspect of the design strategy that could use further improvement is providing more control over the structures that form. Modifications to the design that can further bias assembly toward the formation to a single structure, even when a PBB structure allows for more, would be desirable. Two elements of the design, the coiled coil linker and the flexible spacer region, are areas where further refinement could provide more specificity in cage assembly. A relatively simple modification would involve shortening the coiled coil linker domains to encode fewer heptad repeats. Reducing the number of heptad repeats from 6 to 5 or 4 repeats is expected to lower the overall stability of the coiled coil. The reduced stability may help limit assembly to smaller cage structures, which have lower entropic cost to assembly. I have already constructed plasmids encoding 4 and 5 heptad repeats for future analysis. Another modification to the linking element would be to incorporate peptides designed to form coiled coils upon metal binding. Assembly could be controlled by titrating metal ions into proteins containing the metal binding peptide linkers. By controlling the rate of the titration, the assembly process might be controlled. Titration rates could be used to specify assembly under thermodynamic or kinetic control, giving preference to the assembly of one cage structure over another. The amino acid sequence in the flexible spacer region is another area that could be modified, for example, by shortening the sequence or otherwise designing in decreased flexibility. The sequence of the flexible spacer was already encoded by pET28b(+) and was not specifically designed. Further investigations could be performed to optimize the spacer sequence to provide more control over assembly.

The results presented in this dissertation provide a foundation for future research towards constructing protein based biomaterials. *De-novo* construction of protein cage

architectures has been very limited and provides an approach for overcoming some of the obstacles that have hampered further development. Whereas I have focused on the assembly of protein cage structures, modifications to the strategy presented here could be used to also construct extended protein structures. Adapting the flexible region to be more rigid to enforce extended linker geometries, as was done for studies for constructing extended DNA structures, could allow construction of protein fibers. The peptide sequence between the PBB and peptide linker could be replaced with a sequence encoding the formation of a short alpha-helix, such as the sequence used by Padilla et al.,<sup>1</sup> providing a more rigid connection to project the coiled coil peptides for assembly in a more linear arrangement. Additionally, shortening the peptide sequence connecting the PBB and peptide linker could help limit assembly into linear extended structures by restricting the overall flexibility between the PBB and peptide linker. Further investigations will be required to determine if the method I have outlined can be used as a general strategy for constructing higher order protein structures with useful properties, i.e. protein biomaterials.

## References

- 1 Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc Natl Acad Sci U S A* **98**, 2217-2221 (2001).
- 2 Ni, T. W. & Tezcan, F. A. Structural characterization of a microperoxidase inside a metal-directed protein cage. *Angew Chem Int Ed Engl* **49**, 7014-7018 (2010).
- 3 De Crescenzo, G., Pham, P. L., Durocher, Y., Chao, H. & O'Connor-McCourt, M. D. Enhancement of the antagonistic potency of transforming growth factor-beta receptor extracellular domains by coiled coil-induced homo- and heterodimerization. *J Biol Chem* **279**, 26013-26018 (2004).
- 4 Hillar, A., Culham, D. E., Vernikovska, Y. I., Wood, J. M. & Boggs, J. M. Formation of an antiparallel, intermolecular coiled coil is associated with in vivo dimerization of osmosensor and osmoprotectant transporter ProP in Escherichia coli. *Biochemistry* **44**, 10170-10180 (2005).

- 5 Katz, B. Z. *et al.* Green fluorescent protein labeling of cytoskeletal structures--novel targeting approach based on leucine zippers. *Biotechniques* **25**, 298-302, 304 (1998).
- 6 Veiga, E., Lorenzo, V. d. & Fernandez, L. A. Autotransporters as Scaffolds for Novel Bacterial Adhesins: Surface Properties of Escherichia coli Cells Displaying Jun/Fos Dimerization Domains. *Journal of Bacteriology* **185**, 5585–5590 (2003).
- 7 Zhang, L., Wu, Y. & Brunsveld, L. A synthetic supramolecular construct modulating protein assembly in cells. *Angew Chem Int Ed Engl* **46**, 1798-1802 (2007).

## Appendix A

### DNA and protein sequences of KDPG aldolase and fusion proteins.

#### A.1 Description

The following appendix provides DNA and protein sequences for the major proteins investigated. DNA sequences show unique restriction sites used to insert the respective genes into the expression vectors. In addition, other unique restriction sites are included, which might be useful for further future modifications. KDPG aldolase is found in the pUC18 vector and the fusion proteins are all found in pET28b(+) expression vectors. Calculated molecular weight and pI are given for each protein sequence, unless noted.

#### A.2 Nucleotide and protein sequence for the KDPG aldolase construct.

##### KDPG Aldolase DNA Sequence

1	CGAATTCGAGCTTACCATGGCAATGGAAGAACTGTTCAAAAAACACAAGATTGTAGCCGTGCTGAGGGCAAACAGTGTG	80
1	GCTTAAGCTCGAGATGGTACCGTTACCTTCTTGACAAGTTTTGTGTTAACATCGGCACGACTCCGTTGTCACAC	80
	EcoRI	EcoICRI
	SacI	
	NcoI	
81	GAAGAAGCGAAAAGAAAAGCGTTGGCTGTTTGAAAGGAGGAGTTCACCTCATCGAAATCACCTTCACTGTTCCAGACGC	160
81	CTTCTTCGCTTCTTCCGCAACCGACAAAAACTCCTCCTCAAGTGGAGTAGCTTGTGAAAGTACAAGGCTGCG	160
161	TGACACAGTCATCAAAGAAACTCTCGTTCCTCAAGGAAAAGGTGCCATAATAGGTGCAGGTACAGTGACGAGTGTGCAAC	240
161	ACTGTGTCAGTAGTTCTTGAGAGCAAGGAGTTCCCTTCCACGGTATTATCCACGTCCATGTCACTGCTCACAGCTTG	240
241	AGTGCAGAAAAGCTGTAGAAAGTGGAGCAGAGTTCATCGTCAGTCCACACCTTGACGAGGAAATCTCTCAATTCTGCAA	320
241	TCACGTCTTCGACATTTCACCTCGTCTCAAGTAGCAGTCAGGTGTGAACTGCTCCTTAGAGAGTTAACAGCTT	320
321	GAAAAAGGTGTCTTCTACATGCCGGTGTGATGACACCCCACCGAACTTGTAAAAGCCATGAAACTCGGTACACGATTT	400
321	CTTTTCCACAGAAGATGTACGGGCCACACTACTGTGGGTGGCTGAAACATTCGGTACTTTGAGCCAGTGTGCTAAAA	400
401	GAAACTCTCCCTGGAGAAAGTGGTGGGACCTCAGTTGTAAAAGCGATGAAAGGACCGTTCCCCAATGTGAAATTGTC	480
401	CTTGAGAAGGGACCTCTCACCAACCTGGAGTCACACATTTCGCTACTTCTGGCAAGGGTTACACTTTAACGACG	480
481	CCACTGGAGGCCTGAATCTGGACAACGTGTGAGTGGTCAAGGCCGGAGTCCTCGCTGTTGGTGGAAAGTGCCTG	560
481	GGTGACCTCCGACTTAGACCTGTTGCACACACTCACCAAGTTCCGGCTCAGGAGCGACAACCACACCTCACCGC	560
561	GTGAAAGGAACACCAGACGAGGTGAGAGAAAAAGCAGAAAGCAGTCTGAGAAAAGATCAGGGGTGCACAGAACTCGAGCA	640
561	CACTTCCCTGTGGTCTGCTTCACTCTCTTCTGCAAGCATTTCTAGTCCCCCAGTGTCTTGAGCTCGT	640
	ApalI	XbaI

641 CCACCACCACCACCACTGAGATCCGGCATGCAAGC	675
641 GGTGGTGGTGGTGGTGA <del>C</del> TAGGCCGTACGTTCG	675
	SphI

### KDPG Aldolase Protein Sequence

MAMEELFKKH KIVAVLRANS VEEAKEKALA VFEGGVHLIE ITFTVPDADT VIKELSFLKE	60
KGAIIGAGTV TSVEQCRKAV ESGAEFIVSP HLDEEISQFC KEKGVFYMPG VMTPTELVKA	120
MKLGHITLKL FPGEVVGPQF VKAMKGPFPN VKFVPTGGVN LDNVCEWFKA GVLAVVGSA	180
LVKGTPDEV <del>R</del> EKAKAFVEKI RGCTELEHHH HHH	213

The calculated molecular weight and pI for the sequence was calculated to be 23,244 Da and 7.05, respectively.

### A.3 - Nucleotide and protein sequence for the Non-6xHis tagged Aldolase(+).

#### Non-6xHis tagged Aldolase(+) DNA Sequence

1 CCATGGCAATGGAAGA <del>CT</del> TTCAAAAAACACAAGATTGTAGCCGTGCTGAGGGCAAACAGTGTGGAAGAACGAAAGAA	80
1 GGTACCGTTACCTTCTTGACAAGTTTTGTGTTCAACATCGGCACGACTCCC <del>G</del> TTGTACACCTTCTTCGCTTCTT	80
	NcoI
81 AAGGC <del>G</del> TTGGCTGTTTGAGGAGGAGTTCACCTCATCGAA <del>T</del> CACCTTCACTGTTCCAGACGCTGACACAGTCATAA	160
81 TTCCGCAACCGACAAA <del>A</del> CTTCCTCAAGTGGAGTAGCTTAGGAAAGT <del>G</del> ACAAGGTCTGCGACTGTGTCAGTAGTT	160
161 AGAACTCTCGTTCTCAAGGAAAAGGTGCCATAATAGGTGCAGGTACAGT <del>G</del> ACGAGTGTGAA <del>C</del> AGTCAGTCAGAAAAGCTG	240
161 TCTTGAGAGCAAGGAGT <del>T</del> CC <del>T</del> TTTCCACGGTATTATCCACGTCATGTC <del>A</del> CTGCTCACAGCTTGT <del>C</del> ACGTCTTCGAC	240
241 TAGAAAGTGGAGCAGAGTTCATCGTCAGTCCACACCTGACGAGGAA <del>A</del> ATCTCTCAATTCTGCAAAGAAAAGGTGTCTTC	320
241 ATCTTTCACCTCGTCTCAAGTAGCAGT <del>G</del> AGGTGTGAA <del>C</del> TC <del>G</del> TCTTTAGAGAGTTAAGACGTTCTTTCCACAGAAC	320
321 TACATGCCCGGTGTGATGACACCCACCGAACTTGTAAAAGCCATGAA <del>A</del> ACTCGGT <del>C</del> ACACGATTTGAA <del>A</del> ACTCTCCCTGG	400
321 ATGTACGGGCCACACTACTGTGGGTGC <del>T</del> GAACATTTCGGTACTTTGAGCCAGTGTGCTAAA <del>A</del> CTTGAGAAGGGACC	400
401 AGAA <del>G</del> GTGGGGACCTCAGTTGTA <del>A</del> AGCGATGAAAGGACCGTTCCCA <del>A</del> TGTGAA <del>A</del> TCGTGCCACTGGAGGCGTGA	480
401 TCTTCACCACCC <del>T</del> GGAGTC <del>A</del> ACATTTCGCTACTTTCTGGCAAGGGTTACACTTTAAGC <del>A</del> GGGTGACCTCCGCACT	480
481 ATCTGGACACAGTGTGAGTGGTTCAAGGCCGGAGTCCTCGCTGGTGGAGTGC <del>G</del> CCTGTGAAAGGAACACCCA	560
481 TAGACCTGTTGACACACTCACCAAGTT <del>C</del> GGCCTCAGGAGCGACAACCACACCTTCAGCGAACACTTCTTG <del>G</del> GT	560
561 GACGAGGTGAGAGAAAAGC <del>A</del> AGCGTTCTGAGAAAAGATCAGGGGTGCACAGAA <del>A</del> CTCGAGGCCGCGGGCACCCACGG	640
561 CTGCTCCACTCTCTTT <del>C</del> GGTT <del>C</del> GAAGCATTTCTAGTCCCCACGTGTCTTGAGCTGGCGC <del>G</del> CCGTGGGTGCC	640
	ApaLI
	XbaI
641 CTCTAGCAAACAGTGGAGAAA <del>G</del> ACTGGCGCAACTC <del>A</del> AAAAGAAA <del>A</del> ACT <del>G</del> CAGGCCATTAAAAAA <del>A</del> ACTGGCGCAATTGA	720
641 GAGATCGTTGTCAACCTTTCTTGACCGCGTTGAGTTTCTTGACGT <del>C</del> GGTAATT <del>T</del> TTTGACCG <del>G</del> T <del>T</del> A <del>C</del> CT	720
	PstI
	MfeI
721 AATGGAAGGCC <del>C</del> AGGC <del>G</del> CTGAAAAAA <del>A</del> CTGGCACAGCTGAAGAAAAACT <del>C</del> CAAGGTGGCTCCGGAGCTAGCTAATAA	800
721 TTACCTCCGGTCCG <del>G</del> ACTTTTTTGACCG <del>T</del> GT <del>C</del> GACTTCTTTTGAGGTTCCACCGAGGC <del>T</del> CGATTATT	800
	PvuII
	BspEI
	BmtI
	NheI
801 CTCGTGCACCACCACCACCACTGA	827
801 GAGCACGTGGTGGTGGTGGT <del>G</del> ACT	827
	BssSI
	ApaLI

### Non-6xHis Tagged Aldolase(+) Protein Sequence

MAMEELFKKH KIVAVLRLANS VEEAKEKALA VFEGGVHLIE ITFTVPDADT VIKELSFLKE	60
KGAIIGAGTV TSVEQCRKAV ESGAEFIVSP HLDEEISQFC KEKGVFYMPG VMTPTELVKA	120
MKLGHITILKL FPGEVVGPQF VKAMKGPFPN VKFVPTGGVN LDNVCEWFKA GVLAVGVGSA	180
LVKGTPDEVR EKAKAFVEKI RGCTELEPRG THGSSKQLEK ELAQLKKLQ AIKKKLAQLK	240
WKAQALKKKL AQLKKKLQGG SGAS	264

The calculated molecular weight and pI for the sequence was calculated to be 28,681 Da and 9.44, respectively.

### A.4 – Nucleotide and protein sequence for the Non-6xHis Tagged Aldolase(-)

#### Non-His Tagged Aldolase(-) DNA Sequence

1 CCATGGCAATGGAAGAACTGTTAAAAAACACAAGATTGTAGCCGTGCTGAGGGCAAACAGTGTGGAAGAAGCGAAAGAA	80	
1 GGTACCGTTACCTTCTTGACAAGTTTTGTGTTCTAACATCGGCACGACTCCGTTGTCACACCTTCTCGCTTCTT	80	
NcoI		
81 AAGGC GTTGGCTGTTGAAGGAGGGAGTTCACCTCATCGAAATCACCTC ACTGTTCCAGACGCTGACACAGTCATCAA	160	
81 TTCCGCAACCGACAAAACCTCCTCTCAAGTGGAGTAGCTTAGTGGAAAGTGACAAGGTCTGCACTGTGTCAGTAGTT	160	
161 AGAACTCTCGTCCCTCAAGGAAAAGGTGCCATAATAGGTGCAGGTACAGTGACGAGTGTGCAACAGTGCAGAAAAGCTG	240	
161 TCTTGAGAGCAAGGAGTTCCCTTCCACGGTATTATCCACGTCATGTCACTGCTCACAGCTGTACGTCTTCGAC	240	
241 TAGAAAGTGGAGCAGAGTCATCGTCAGTCCACACCTTGACGAGGAAATCTCTCAATTCTGCAAAGAAAAGGTGTCTTC	320	
241 ATCTTTCACCTCGTCTCAAGTAGCAGTCAGGTGTGGAACTGCTCCTTAGAGAGTTAACAGCTTCTTTCCACAGAAC	320	
321 TACATGCCCGGTGTGATGACACCCACCGAACTTGAAAAGCCATGAAACTCGGTACACGATTGAAACTCTCCCTGG	400	
321 ATGTACGGGCCACACTACTGTGGGTGCGTTGAACATTTCGGTACTTTGAGCCAGTGTGCTAAAACCTTGTGAGAAGGGACC	400	
401 AGAAAGTGGGGACCTCAGTTGTAAGCGATGAAAGGACCGTCCCCATGTGAAATTCTGCCCCACTGGAGGCGTGA	480	
401 TCTTCACCAACCTGGAGTCAAACATTTCGCTACTTTCTGGCAAGGGTTACACTTTAACGACGGGTGACCTCCGCACT	480	
481 ATCTGGACAACGTGTGAGTGGTTCAAAGCCGGAGTCCTCGCTGTTGGTGTGAAAGTGCCTGTGAAAGGAACACCA	560	
481 TAGACCTGTTGCACACACTCACCAAGTTCCGGCCTCAGGAGCGACAACCACACCTTCACGCGAACACTTCCGTGGT	560	
561 GACGAGGTGAGAGAAAAGCAAAGCGTTCTGTAGAAAAGATCAGGGGGTGCACAGAACCTCGAGCCGGCGCAGCCATAT	640	
561 CTGCTCCACTCTCTTTCGTTGCAAGCATCTTCTAGTCCCCACGTGTCTTGAGCTGGCGCGCCGTCGGTATA	640	
ApaII	XbaI	NdeI
641 GGGATCCAACAGTTGGAAAAGAACTGGCCAAGTGGAAAGGAACGTGCAAGCGATTGAGAAAAGAGCTCGCTCAACTGG	720	
641 CCCTAGGTTGTCAACCTTTCTTGACCGGGTTGACCTTCCCTTGACGTTGCTAATCTTCTCGAGCGAGTTGACC	720	
BamHI	EcoICRI	
	SacI	
721 AATGGGAGGCACAAGCTTGAAAAAAACTGGCGCAGCTCGAAAAGAGTTGCAGGGCGGCAGCGCGGTACCTAATAA	800	
721 TTACCCCTCCGTGTTGAAACCTTTTGACCGCGTCAGCTTTCTCAACGTCCC CGCGTCGCCGCGCATGGATTATT	800	
HindIII	Acc65I	
	KpnI	
801 GAATT CAGAAGGGAGATATACTTGGGCTCTCATCATCACACCACAGCTCG	857	
801 CTTAAGTCTCCTCTATATGGAACCCGAGGAGAGTAGTAGTGTGAGTCAG	857	
EcoRI		

### Non-6xHis tagged Aldolase(-) Protein Sequence

MAMEELFKKH KIVAVLRLANS VEEAKEKALA VFEGGVHLIE ITFTVPDADT VIKELSFLKE	60
KGAIIGAGTV TSVEQCRKAV ESGAEFIVSP HLDEEISQFC KEKGVFYMPG VMTPTELVKA	120
MKLGHITLKL FPGEVVGPQF VKAMKGPFPM VKFVPTGGVN LDNVCEWFKA GVLAVGVGSA	180
LVKGTPDEVR EKAKAFVEKI RGCTELEPRG SHMGSKQLEK ELAQLEKELQ AIEKELAQLE	240
WEAQALEKKL AQLEKELQGG SGGT	264

The molecular weight and pI of the non-6xHis tagged Aldolase(-) was calculated to be 28,719 kDa and 5.95, respectively.

### A.5 – Nucleotide sequence for the KDPG Aldolase(+).

#### Aldolase(+)

1 CCATGGCAATGGAAGAACGTGTTCAAAAAACACAAGATTGAGCCGTGCTGAGGGCAAACAGTGAGAAGAAGCGAAAGAA	80
1 GGTACCGTTACCTTCTTGACAAGTTTTGTGTTCTAACATCGGCACGACTCCCGTTGTCACACCTTCTTCGCTTCTT	80
NcoI	
81 AAGGCCTGGCTGTTTGAAGGAGGAGTCACCTCATCGAAATCACCTCAGTGTCCAGACGCTGACACAGTCATAA	160
81 TTCCGCAACCGACAAAACCTCCTCTCAAGTGGAGTAGCTTAGTGGAAAGTGACAAGGTCTGCGACTGTGTCAGTAGTT	160
161 AGAACTCTCGTCTCAAGGAAAAGGTGCCATAATAGGTGCAGGTACAGTGACGAGTGTGCAACAGTGCAAGAAAAGCTG	240
161 TCTTGAGAGCAAGGAGTCCCTTTCCACGGTATTATCCACGTCATGTCAGTGCACAGCTGTCACGCTTGTACCGTCTTCGAC	240
241 TAGAAAGTGGAGCAGAGTCATCGTCAGTCCACACCTTGACGAGGAAATCTCTCAATTCTGCAAAGAAAAGGTGTCTTC	320
241 ATCTTCACCTCGTCTCAAGTAGCAGTGTGGAACTGCTCTTGTAGAGAGTTAACAGCTTTCCACAGAAG	320
321 TACATGCCCGGTGTGATGACACCCCACCGAACCTTGAAAGGCCATGAAACTCGGTACACGATTGAAACTCTCCCTGG	400
321 ATGTACGGGCCACACTACTGTGGGTGGCTAACATTTCGGTACTTGAGCCAGTGTGCTAAAACATTGAGAAGGGACC	400
401 AGAAAGTGGTGGGACCTCAGTTGAAAGCGATGAAAGGACCGTTCCCAATGTGAAATTGCCCCACTGGAGGCGTGA	480
401 TCTTCACCACCCCTGGAGTCAAACATTTCGCTACTTCCCTGGCAAGGGTTACACTTAAGCACGGTGACCTCCGCACT	480
481 ATCTGGACAACGTGTGAGTGGTTCAAAGCCGGAGTCCTCGTGTGGTGGAAAGTGCCTGTGAAAGGAACACCA	560
481 TAGACCTGTTGCACACACTCACCAAGTTCGGCTCAGGAGCGACAACCACACCTTCACCGAACACTTCCGTGGT	560
561 GACGAGGTGAGAGAAAAGCAAAGCGTTCTGTAGAAAAGATCAGGGGTGCACAGAACTCGAGCCGCGCACCACGG	640
561 CTGCTCCACTCTCTTTCTGGCAAGCATTTCTAGTCCCCCACGTGTCTTGAGCTCGCGCCTGGTGCC	640
ApaLI XbaI	
641 CTCTAGCAAAACAGTTGGAGAAAGAACGTGGCGCAACTCAAAAGAAACTGCAGGCCATTAAAAAAACTGGCGCAATTGA	720
641 GAGATCGTTGTCAACCTCTTCTTGACCGCGTTGAGTTTTCTTGACGTCCGGTAATTGACCGCTTAACT	720
PstI MfeI	
721 AATGGAAGGCCAGGGCTGAAAAAAAAACTGGCACAGCTGAAGAAAAAACTCCAAGGTGGCTCCGGAGCTAGCTCACAA	800
721 TTACCTTCCGGTCCGGACTTTTTTGACCGTGTGACTTCTTTGAGGTTCCACCGAGGCCTCGATCGAGTGT	800
PvuII BspEI BmtI	
801 CTCGTGCACCAACCACCAACTAA 827	
801 GAGCACGTGGTGGTGGTGGTGGTGGTGGATT 827	
BssSI	
ApalI	

Calculated molecular weight and pI for Aldolase(+) can be found in Chapter 4.

#### A.6 – Nucleotide sequence encoding Aldolase(-) DNA.

##### Aldolase(-)

1 CCATGGCAATGGAAGAACTGTTCAAAAAACACAAGATTGTAGCCGTGCTGAGGGCAAACAGTGTGGAAGAACGCAGAAAGAA 80  
1 GGTACCGTTACCTTCTTGACAAGTTTTGTGTTCTAACATCGCACGACTCCCCTTGTACACCTTCTCGCTTC 80  
NcoI

81 AAGGCCTGGCTGTTTGAAGGAGGAGTTCACCTCATCGAAATCACCTCACTGTTCCAGACGCTGACACAGTCATCAA 160  
81 TTCCGCAACCGACAAAACCTCCTCAAGTGGAGTAGCTTAGTGGAAAGTACAAGGTCTGCGACTGTGTCAGTAGTT 160

161 AGAACTCTCGTCTCAAGGAAAAGGTGCCATAATAGGTGCAGGTACAGTGACGAGTGTGCAACAGTCAGCAAAGAGCTG 240  
161 TCTTGAGAGCAAGGAGTTCCCTTTCCACGGTATTATCCACGTCCATGTCAGTCACAGCTTGTACGCTTTGAC 240

241 TAGAAAGTGGAGCAGACTTCATCGCAGTCCACACCTTGACGAGGAAATCTCTCAATTCTGAAAGAAAAGGTGTCTTC 320  
241 ATCTTTCACCTCGTCTCAAGTAGCAGTCAGGTGTGAACTGCTCTTAGAGAGTTAACAGCTTCTTTCCACAGAAG 320

321 TACATGCCCGGTGTGATGACACCCACCGAACTTGTAAAAGCCATGAAACTCGGTACACGATTTGAAACTCTCCCTGG 400  
321 ATGTACGGGCCACACTACTGTGGGTGGCTGAACATTTCGGTACTTGAGCCAGTGTGCTAAAACCTTGAGAAGGGACC 400

401 AGAAGTGGGGGACCTCAGTTGTAAAAGCGATGAAAGGACCGTCCCCATGTGAAATTCTGCCCCACTGGAGGCGTGA 480  
401 TCTTCACCAACCTGGAGTCAAACATTTCGCTACTTCTGGCAAGGGTTACACTTTAACGACGGGTGACCTCCGCACT 480

481 ATCTGGACAACGTGTGAGTGGTTCAAAGCCGGAGTCCTCGCTGGTGGAAAGTGCCTGTGAAAGGAACACCA 560  
481 TAGACCTGTTGCACACACTCACCAAGTTCGGCCTCAGGAGCGACAACCACACCTTCACGCGAACACTTCCTGTGGT 560

561 GACGAGGTGAGAGAAAAGCAAAAGCCTCGTAGAAAAGATCAGGGGGTGCACAGAACTCGAGCCGCGCCAGCCATAT 640  
561 CTGCTCCACTCTCTTTCTGTTGCAAGCATCTTCTAGTCCCCACGTGTCTTGAGCTCGCTGGCGCCGCGTATA 640  
ApaLI XhoI NdeI

641 GGGATCCAACAGTTGGAAAAAGAACTGGCCAACCTGGAAAAGGAACGTGCAAGCGATTGAGAAAAGAGCTCGCTCAACTGG 720  
641 CCCTAGGTTGTCAACCTTTCTTGACCGGGTTGACCTTCTGACGTTGCTAACTCTTCTGAGCGAGTTGACC 720  
BamHI EcoI CRI SacI

721 AATGGGAGGCACAAGCTTGGAAAAAAACTGGCGCAGCTCGAAAAGAGTTGCAGGGGGCAGCGCGGTACCTCACAA 800  
721 TTACCCCTCGTGTGAAACCTTTTGACCGCGTCAGCTTTCTCAACGTCCCCGGTGCAGCCGCGTGGAGTGT 800  
HindIII Acc65I KpnI

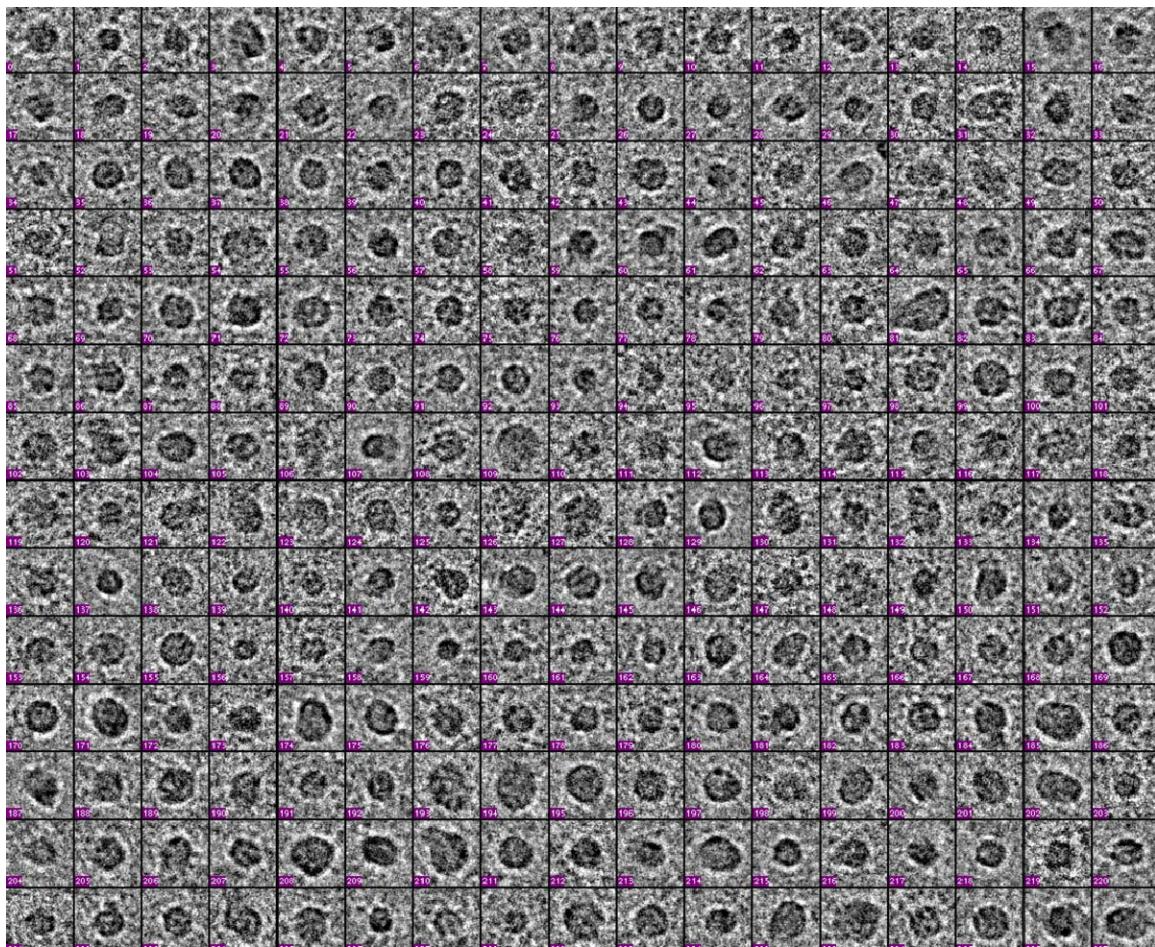
801 GAATTCAAGAGGAGATACCTTGGCTCTCATCATCACCAACCGCTAA 857  
801 CTTAAGTCTCCCTATATGGAACCCGAGGAGAGTAGTAGTGTGGTGGCGATT 857  
EcoRI

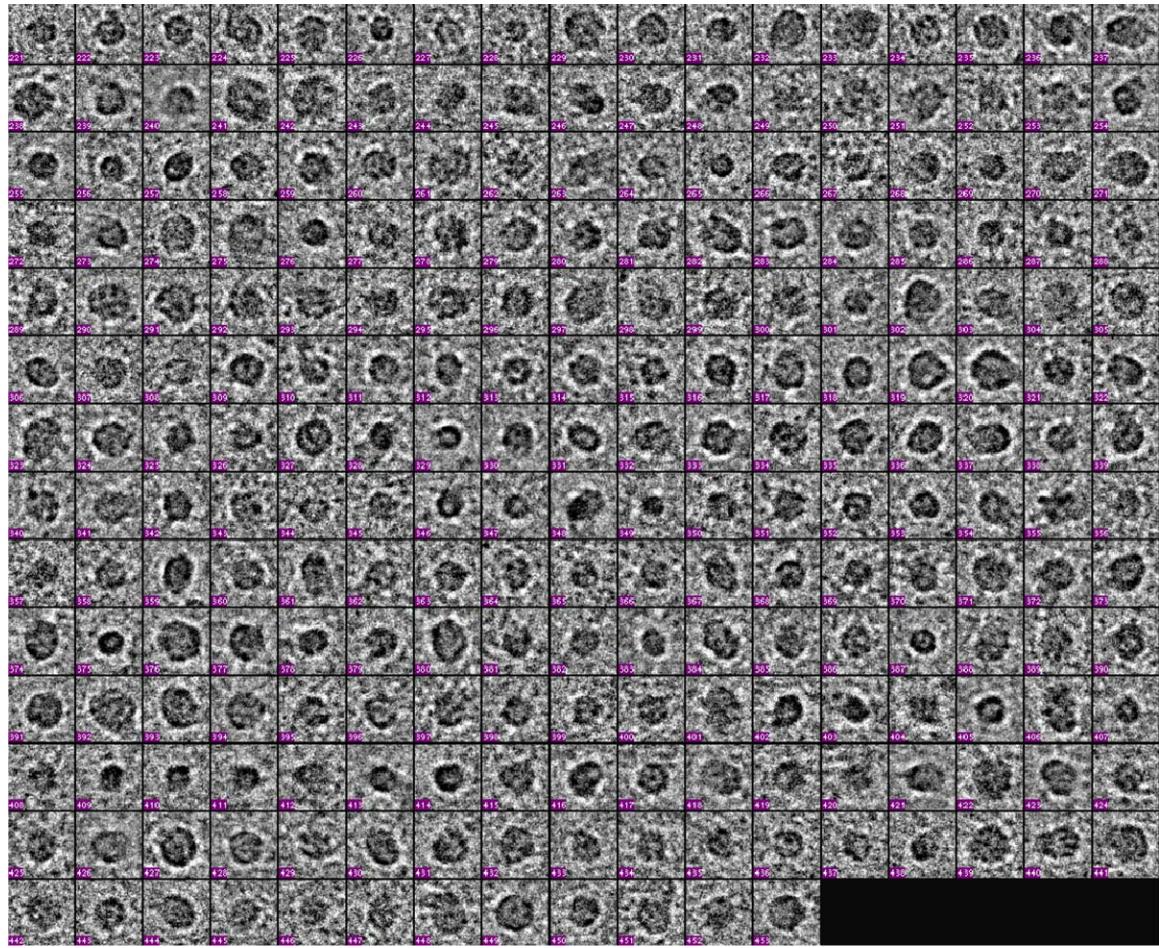
Calculated molecular weight and pI for Aldolase(-) can be found in Chapter 4

## Appendix B

### Cryo-EM image gallery

Provided below is the photo gallery of particles obtained from initial studies in imaging the protein cages by Cryo-EM. The data was collected by Dr. Min Su. A population of the particles taken from the images below were used in making the reconstruction model in Figure 4.18B. Each box is ~29 nm x 29 nm.





## Appendix C

### The Subunit Structure of Benzylsuccinate Synthase<sup>†</sup>

Lei Li<sup>#</sup>, Dustin P. Patterson<sup>#</sup>, Christel C. Fox<sup>#</sup>, Brian Lin<sup>#</sup>, Peter W.

Coschigano<sup>§</sup> and E. Neil G. Marsh<sup>#\*</sup>

<sup>#</sup>Department of Chemistry, University of Michigan, Ann Arbor, MI 48109-1055, USA

and <sup>§</sup>Department of Biomedical Sciences Ohio University, Athens OH 45701-2979, USA

Biochemistry **2009**, 48, 1284-1292.

\*Much of the work contained below was performed by Lei Li. I performed EPR experiments and was involved in some aspects of the molecular biology and protein purification.

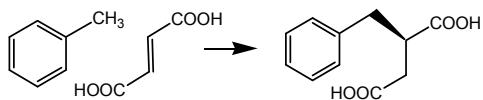
## Abstract

Benzylsuccinate synthase is a member of the glycyl radical family of enzymes. It catalyzes the addition of toluene to fumarate to form benzylsuccinate as the first step in the anaerobic pathway of toluene fermentation. The enzyme comprises three subunits  $\alpha$ ,  $\beta$  and  $\gamma$  that in *Thauera Aromatica* T1 strain are encoded by the *tutD*, *tutG* and *tutF* genes, respectively. The large  $\alpha$ -subunit contains the essential glycine and cysteine residues that are conserved in all glycyl radical enzymes. However, the function of the small  $\beta$ - and  $\gamma$ -subunits has remained unclear. We have over-expressed all three subunits of benzylsuccinate synthase in *E. coli*, both individually and in combination. Co-expression of the  $\gamma$ -subunit (but not the  $\beta$ -subunit) is essential for efficient expression of the  $\alpha$ -subunit. The benzylsuccinate synthase complex lacking the glycyl radical could be purified as an  $\alpha_2\beta_2\gamma_2$  hexamer by nickel-affinity chromatography through a 'His<sub>6</sub>' affinity tag engineered onto the C-terminus of the  $\alpha$ -subunit. Unexpectedly, BSS was found to contain two iron-sulfur clusters, one associated with the  $\beta$ -subunit and the other with the  $\gamma$ -subunit that appear to be necessary for the structural integrity of the complex. The spectroscopic properties of these clusters suggest that they are most likely [4Fe-4S] clusters. Removal of iron with chelating agents results in dissociation of the complex; similarly a mutant  $\gamma$ -subunit lacking the [4Fe-4S] cluster is unable to stabilize the  $\alpha$ -subunit when the proteins are co-expressed.

<sup>1</sup> The abbreviations used are: BSS, benzylsuccinate synthase; NTA, nitrilotriacetic acid; OBP, *o*-bathophenanthroline; HPA, 4-hydroxyphenylacetic acid

## Introduction

Benzylsuccinate synthase<sup>1</sup> (BSS) catalyzes a most unusual chemical transformation in which toluene is added across the double bond of fumarate to produce (*R*)-benzylsuccinate (scheme 1). This is the first step in the anaerobic pathway of toluene metabolism that allows various denitrifying and sulfate-reducing bacteria such as *Thauera aromatica* and *Desulfobacula toluolica* to live on toluene as their sole source of carbon and energy under anaerobic conditions (1, 2). The metabolism of toluene and related aromatic compounds is of particular interest both because these compounds represent an important class of pollutants that are long-lived in the environment and have relatively high water solubility, and because they involve novel approaches to the activation of unreactive C-H bonds (3, 4).



**Scheme 1:** Reaction catalyzed by benzylsuccinate synthase

Sequence similarities initially identified BSS as a member of the glycyl radical-containing group of enzymes that includes pyruvate formate-lyase and anaerobic ribonucleotide reductase (5-10). Later EPR studies show that the resting enzyme harbors an organic radical similar to that observed in these enzymes (11-13). Like other glycyl radical enzymes, BSS is extremely oxygen sensitive, and exposure to air results in oxidative cleavage of the  $\alpha$ -subunit at the site of the presumed glycyl radical (5). The glycyl radicals in these enzymes are generated from the cognate glycine residue by specific activases that are members of the family of S-adenosylmethionine radical enzymes (9, 14, 15).

The proposed mechanism for the BSS-catalyzed reaction (3,16) involves transfer of the glycyl radical to an active site cysteine residue, which then abstracts hydrogen from toluene to generate a benzylic radical. The benzylic radical subsequently undergoes addition to fumarate to generate the C-3 radical of (R)-benzylsuccinate. The abstracted hydrogen is transferred to the product to generate benzylsuccinate and the cysteinyl radical, and finally, the radical is transferred back to glycine.

Mechanistic studies in our laboratory have focused on providing evidence of putative radical intermediates in the reaction and probing the energetics of the reaction. Evidence for the formation of the C-3 radical of benzylsuccinate as an intermediate comes from experiments to investigate the stereochemistry of hydrogen transfer from toluene to fumarate and to the alternate co-substrate maleate (17). When fumarate was the co-substrate, *syn* addition of toluene to the double bond of fumarate was observed. However, when maleate was the co-substrate, the addition of toluene occurred in an *anti* fashion. These observations are explained by the formation of the C-3 radical of benzylsuccinate in which rotation about the C-2, C-3 bond can occur to relieve the sterically unfavorable *cis* conformation of the carboxylate groups when maleate is the co-substrate.

Using deuterated toluene, we have undertaken steady-state deuterium kinetic isotope effect measurements to probe the free energy profile of the BSS reaction (18). By comparing the relative magnitudes of the  $DV$  and  $D(V/K)$  isotope effects, measured experimentally, with those predicted by a density functional theoretical study (19), we obtained evidence that the enzyme altered the free energy profile such that the starting glycyl radical and the intermediate cysteinyl and benzylic radicals appear to be much

closer in energy than calculated. This suggests that the enzyme plays a role in stabilizing the cysteinyl and benzylic radicals relative to the glycyl radical, which may be an important and general mechanism for controlling the reactivity of radicals in this class of enzymes.

Further insight into the unusual energetics of the BSS reaction comes examining the exchange reaction of *p*-cresol with the benzyl portion of benzylsuccinate to form (4-hydroxybenzyl)-succinate (20). This allowed us to investigate the kinetics of the reverse reaction: disproportionation of benzylsuccinate to toluene and fumarate. Even though the equilibrium constant for the reverse reaction is extremely unfavorable ( $K_{eq} \sim 8 \times 10^{-11}$  M at 4 °C), the enzyme catalyzes the reverse reaction at rate only 250-fold slower than the forward reaction. During the reverse reaction, partial exchange of the migrating hydrogen with the solvent is observed. This provides the first direct evidence that the hydrogen is transferred to a labile site on the protein during catalysis, which is consistent with the participation of the proposed active site cysteine residue in the mechanism of BSS.

The rapid loss of activity encountered during attempts to purify the enzyme and its extreme oxygen sensitivity (5) has severely hindered efforts to study the enzyme. The above experiments could all be conducted with cell-free extracts of BSS because of the lack of alternate reaction pathways for toluene. However, for many studies it is necessary to have large amounts of pure enzyme and for some purposes the ability to generate BSS radical-free would be extremely useful. Here we report the over-expression of the BSS subunits in *E. coli* and characterization of the properties of the radical-free enzyme.

Unexpectedly, the  $\beta$ - and  $\gamma$ -subunits are revealed to be iron-sulfur proteins and the iron sulfur clusters appear to play a key role in the assembly of the holo-enzyme.

## Materials and Methods

### Materials

DNA-modifying enzymes and reagents were purchased from New England Biolabs (Beverly, MA). Oligonucleotide primers were obtained either from IDT Integrated DNA Technologies (Coralville, IA) or from Invitrogen (Carlsbad, CA). *E. coli* BL21( $\lambda$ DE3) and expression vectors pET-28b, pACYC-Duet, pET-Duet, pRSF-Duet were purchased from Novagen (Madison, WI); Pfu turbo DNA polymerase and *E. coli* XL1-Blue were from Stratagene (Cedar Creek, TX).. Bovine gamma globulin standard and Bradford reagent were purchased from Bio-Rad Laboratories, Inc (Hercules, CA) and used for routine protein concentration determination. Nickel nitrilotriacetic acid (Ni-NTA) resin and QIAquick PCR purification kit were purchased from Qiagen (Valencia, CA). Sephadex G-25 resin and Superose 6 pre-packed FPLC column were purchased from GE Biosciences (Piscataway, NJ). All other chemicals used were of the highest grade commercially available.

### Subcloning of tut Genes into Expression Vectors

The BSS proteins were overexpressed in *E. coli* using the commercially available pET series of vectors to place the genes under control of the T7 promoter and, where desired, introduce a His<sub>6</sub> tag at either the N- or C-terminus. Standard PCR and site-directed mutagenesis techniques were used to introduce appropriate restriction sites into the genes of interest to facilitate their subcloning into the desired expression vectors. A summary description of the plasmids generated in this study is given in Table 1; a

detailed description of the construction of these plasmids is included as supporting material.

<b>Plasmid name</b>	<b>Vector used</b>	<b>Gene inserted</b>	<b>Cloning sites</b>	<b>Description</b>
pET28_F	pET28b	<i>tutF</i>	<i>NdeI/EcoRI</i>	Expression of BSS- $\gamma$ subunit with a N-terminus His <sub>6</sub> tag
pET28_G	pET28b	<i>tutG</i>	<i>NdeI/HindIII</i>	Expression of BSS- $\beta$ subunit with a N-terminus His <sub>6</sub> tag
pET28_D	pET28b	<i>tutD</i>	<i>NdeI/HindIII</i>	Expression of BSS- $\alpha$ subunit with a N-terminus His <sub>6</sub> tag
pET-Duet_F	pET-Duet	<i>tutF</i>	<i>NcoI/HindIII</i>	Expression of BSS- $\gamma$ subunit with no tag (negative control for monitoring BSS $\alpha$ subunit expression in cell-free extract)
pET-Duet_D_F	pET-Duet	<i>tutD</i> <i>tutF</i>	<i>NdeI/KpnI</i> <i>NcoI/HindIII</i>	Expression of BSS- $\alpha$ and $\gamma$ subunits with a C-terminus His <sub>6</sub> tag on the $\alpha$ subunit
pRSF-RSF_G	pRSF-Duet	<i>tutG</i>	<i>NdeI/XhoI</i>	Expression of BSS- $\beta$ subunit with no tag
pET-Duet_D_F_stop	pET-Duet	<i>tutD</i>	<i>NdeI/KpnI</i>	Expression of BSS- $\alpha$ subunit with a C-terminus His <sub>6</sub> tag
pET-Duet_D_F_C9S	pET-Duet	<i>tutD</i> <i>tutF_C9S</i>	<i>NdeI/KpnI</i> <i>NcoI/HindIII</i>	Expression of BSS- $\alpha$ and $\gamma$ subunits with $\alpha$ subunit carrying a C-terminus His <sub>6</sub> tag and $\gamma$ subunit a C9S mutation
pRSF-Duet_G_C29S	pRSF-Duet	<i>tutG_C29S</i>	<i>NdeI/XhoI</i>	Expression of BSS- $\beta$ subunit with a C29S mutation
pACYC-Duet_F_C9S	pACYC-Duet	<i>tutF_C9S</i>	<i>BamHI/XhoI</i>	Expression of BSS- $\gamma$ subunit with a N-terminus His <sub>6</sub> tag and a C9S mutation
pET28_G_C29S	pET28b	<i>tutG_C29S</i>	<i>NdeI/HindIII</i>	Expression of BSS- $\beta$ subunit with a N-terminus His <sub>6</sub> tag and a C29S mutation

**Table 1.** Summary of BSS expression constructs used in this study

### Expression of BSS Proteins

The expression constructs containing the various *tut* genes were transformed into *E. coli* BL21(λDE3) cells. A single colony of transformed cells was used to inoculate 5 mL of LB medium containing the appropriate antibiotic to maintain selection for the plasmid. The cultures were grown to saturation at 37°C with vigorous shaking and then used to inoculate 1 L of antibiotic-containing LB medium. Upon reaching early log phase ( $OD_{600} = 0.8$ ), expression of the genes was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The cultures were grown for 4 h before harvesting by centrifugation (6000 rpm, 15 min at 4 °C). The medium was decanted and the cell pellet stored at -20 °C.

For the expression of the BSS holo-enzyme complex it was found advantageous to modify the expression protocol as follows. The medium was supplemented with 150 mg  $Fe(NH_4)_2(SO_4)_2$  per liter of culture. Upon reaching  $OD_{600} = 0.8$ , the culture was allowed to cool to ambient temperature before induced by addition IPTG to a final concentration of 0.5 mM. The culture was grown for 3 more hours at room temperature with gentle stirring, to minimize oxygenation, before harvesting by centrifugation.

### Purification of BSS Proteins

To purify BSS-β and BSS-γ, the cell pellet from 1 L of culture was re-suspended in 30 mL of lysis buffer containing 50 mM Tris, 300 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol, 1 mM PMSF at pH 8.0. The cells were lysed by sonication on ice and centrifuged at 17,000 rpm for 20 min at 4°C. The supernatant was loaded onto a 1.5 mL Ni-NTA-agarose column and moved into a Coy anaerobic glove box. The rest of the protein purification steps were carried out under an inert atmosphere. The column was

washed with 20 volumes of the buffer containing 50 mM Tris, 300 mM NaCl, 20 mM imidazole, at pH 8.0. The tagged BSS proteins were eluted with the same buffer but containing 300 mM imidazole. The eluted proteins were yellow-brown colored and were determined to be essentially homogenous by SDS-PAGE on a 12% tricine gel stained with Coomassie blue. BSS was eluted with the same buffer but containing 100 mM imidazole. Proteins were dialyzed three times against the buffer containing 25 mM Tris and 300 mM NaCl at pH 8.0 and stored at -20 °C. A typical purification yielded 30 - 40 mg protein/L of cell culture. To purify the complex of BSS- $\alpha$  with BSS- $\beta$  and BSS- $\gamma$  the protocol was modified such that 100 mM imidazole was used to elute the proteins and the protein purity was assessed by SDS-PAGE using 10 % tricine gels.

#### Purification and Refolding of BSS- $\alpha$ from Inclusion Body

Despite surveying a number of conditions, BSS- $\alpha$  protein, when expressed alone, was always produced as inclusion bodies and could only be purified by Ni-NTA chromatography in 8M urea. Briefly, inclusion bodies from 1L media were solubilized in 100 mM NaH<sub>2</sub>PO<sub>4</sub> buffer, pH 8.0, containing 10 mM Tris/Cl and 8.0 M urea. The resulting protein was loaded onto 1.5 mL Ni-NTA-agarose column and non-specifically bound proteins removed by washing with the same buffer adjusted to pH 6.3. The His-tagged  $\alpha$  subunit was eluted by lowering the pH of the buffer further to pH 4.5. The protein obtained by this procedure was >95% homogeneous as judged by SDS-PAGE. Typically, the purification yielded 50 mg of unfolded protein per liter of LB media.

BSS- $\alpha$  protein was refolded by dialysis of the purified, urea-denatured protein against progressively lower concentrations of urea in a buffer containing 10% glycerol

and 5 mM fumarate. Importantly, it was found that the presence of fumarate was essential to obtain good yields of refolded protein.

#### Determination of Protein Concentrations

Protein concentrations were calculated based on the absorption of aromatic residues at 280 nm in the presence of 6 M guanidine hydrochloride using the method of Gill and von Hippel (21).

#### Iron Determination

Iron content was determined using *o*-bathophenanthroline (OBP) under reductive conditions after digestion of the protein in 0.8% KMnO<sub>4</sub>/1.2 N HCl as described by Fish (22). Iron standards were prepared from commercially available ferric chloride and ferrous ammonium sulfate.

#### Reconstitution of Iron-Sulfur Clusters

Iron-sulfur clusters were reconstituted using a modified protocol described by Jarrett (23). The reactions were set up in an anaerobic glove box and contained DTT (5 mM), FeCl<sub>3</sub> (500 μM), and Na<sub>2</sub>S (500 μM) in 1 mL of protein solution. The solutions were incubated for 20 hrs at 4 °C. Excess reagents were removed by dialysis three times against a buffer containing 25 mM Tris-Cl, 300 mM NaCl at pH 8.0. The extent of cluster formation was assessed by UV-visible spectroscopy (spectra taken using an anaerobic cuvette) and from the iron content determined as described above.

#### Size exclusion chromatography

A 50 μM solution of BSS (250 μL) was loaded onto a Superose 6 FPLC column pre-equilibrated in 25 mM Tris/Cl, 300 mM NaCl, 10 mM β-mercaptoethanol (pH 8.0) at 4 °C. The column was eluted at 0.1 mL/min and the eluant monitored at 280 nm. The

column was calibrated using standard proteins for molecular weight determination (BioRad).

#### Removal of Iron-Sulfur Clusters from BSS

Iron was removed from the protein by treating it with the chelating agent OBP at a final concentration of 200 mM (24). The reaction was typically allowed to proceed at room temperature for 3 h in air or at 4 °C for 16 hrs in an anaerobic chamber. Occasionally some precipitation occurred during the reaction which was removed by centrifugation at 13000 rpm for 15 min at 4 °C.

#### EPR Spectroscopy

Samples were prepared in an anaerobic glovebox. Proteins were dialyzed twice against 50 mM Tris-HCl, 200 mM NaCl, 15% glycerol, and 0.03% sarkosyl (pH 8.0). Dialyzed protein samples were concentrated to ~30 µM using an Amicon (Bedford, MA) Microcon centrifugal ultrafiltration device according to the manufacturer's instructions. Samples were reduced by addition of sodium dithionite, at a final concentration of 2 mM, and immediately transferred to an EPR tube, fitted with an airtight septum and frozen and stored in liquid nitrogen. All EPR spectra were obtained on a Bruker EMX electron spin resonance spectrometer equipped with a Bruker 4102-ST general purpose cavity (rectangular TE102) and an Oxford liquid helium cryostat. Conditions for the acquisition of EPR spectra were as follows: temperature, 10 K; microwave frequency, 9.41 GHz; microwave power, 2.05 mW; modulation frequency, 100 kHz; modulation amplitude, 10.00 G; conversion time and time constant, 40.96 ms; 16-scan average. The data were analyzed using the Bruker Win-EPR data manipulation program.

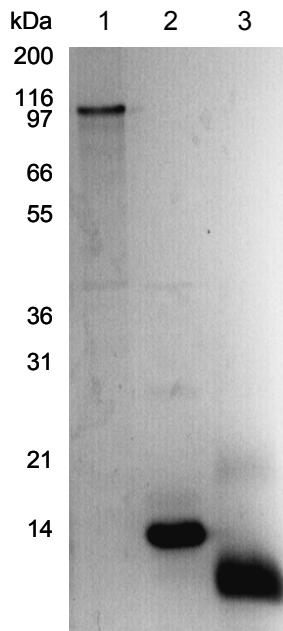
## Results

Although BSS is expressed at quite high levels in *T. aromatica*, the protein has proven extremely hard to purify from this organism. In part, this is because the wild-type enzyme contains a reactive glycyl radical which renders it unstable to air, but significant losses in activity were encountered even during purification under rigorously anaerobic conditions in a glove box (ref 16 and unpublished results of M.S. Huhta and E.N.G. Marsh). To circumvent this problem, we sought to over-express and purify BSS from *E. coli* where the protein could be produced without the catalytic glycyl radical. This would render the protein air stable and thereby greatly facilitate purification and physical characterization of the BSS complex, as well as allowing the individual protein subunits to be purified and characterized. In particular, we wished to investigate the function of the small  $\beta$ - and  $\gamma$ -subunits that have been shown to be essential for enzyme activity (25, 26).

### Expression of the Tut Genes

Initially, the *tutD*, *tutF*, and *tutG* genes (encoding the BSS- $\alpha$ ,  $\gamma$  and  $\beta$  subunits respectively) were each subcloned into the expression vector pET28b (constructs pET28\_D, pET28\_F, pET28\_G, Table 1) which placed them under control of the T7 promoter and also introduced a His<sub>6</sub> affinity tagging sequence at the N-terminus of each protein. Each gene was strongly over-expressed from this vector in *E. coli*. The BSS- $\beta$  and  $\gamma$  subunits were produced as soluble proteins, and could be purified straightforwardly by chromatography on a Ni-NTA column (Figure 1). However, the  $\alpha$ -subunit was always produced as insoluble inclusion bodies. Attempts to express BSS- $\alpha$  in soluble form by varying expression conditions, by, for example, lowering growth temperature,

reducing the concentration of IPTG or inducing expression under anaerobic conditions, met with no success. Therefore, the protein was solubilized with urea and purified from inclusion bodies by Ni-NTA chromatography under denaturing conditions; the protein was then allowed to refold by removing the urea by dialysis to give homogenous, soluble protein (Figure 1).

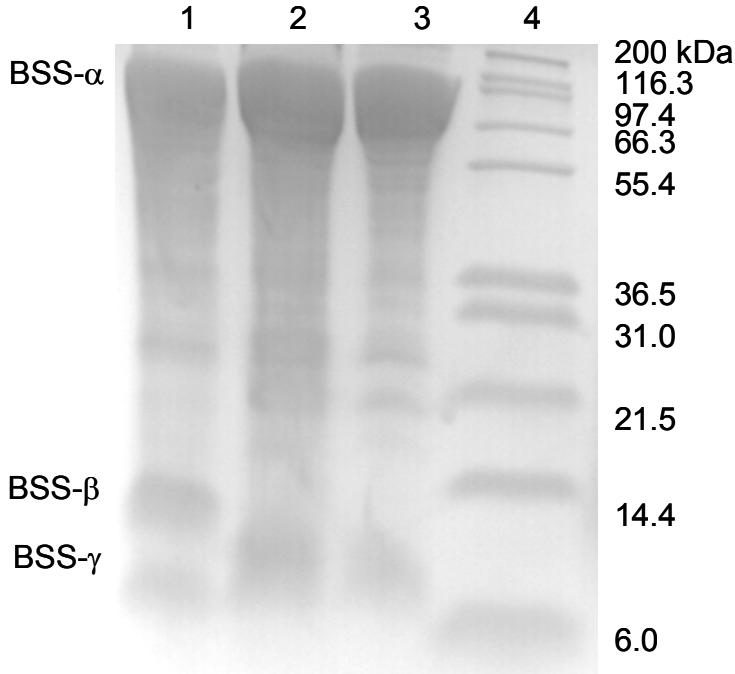


**Figure 1.** Over-expression and purification of individual BSS subunits. *Lane 1*: BSS- $\alpha$ , purified from inclusion bodies and refolded; *lane 2*: BSS- $\beta$ , purified as soluble protein by Ni-NTA chromatography; *lane 3*: BSS- $\gamma$ , purified as soluble protein by Ni-NTA chromatography.

Interestingly, it was found that inclusion of 5 mM fumarate in the refolding buffer significantly improved the yield of soluble BSS- $\alpha$ , suggesting that this substrate may be stabilizing the refolded protein, or possibly nucleating the protein folding. However, surprisingly, the refolded protein did not bind to a Ni-NTA column, suggesting that the N-terminal His<sub>6</sub> tag may be sequestered within the refolded protein. This raised the possibility that the His tag might interfere with the correct folding of the protein.

We therefore considered the possibility that the  $\beta$ - and  $\gamma$ -subunits might be required for the correct folding of the  $\alpha$ -subunit. To investigate whether this was the case we used the “Duet” expression system developed by Novagen to express each of the genes from its own T7 promoter. The *tutD* and *tutF* genes were subcloned into the two multiple cloning sites of pETDuet<sup>TM</sup> (plasmid pET-Duet\_D\_F), which is under the control of the *ColE1* replicon and selected by ampicillin resistance. The *tutG* gene was subcloned into the second multiple cloning site of pRSFDuet<sup>TM</sup> (plasmid pRSF-Duet\_G), which is under the control of the *RSF1030* replicon and selected by kanamycin resistance. As part of this cloning strategy a sequence encoding His<sub>6</sub> affinity tag was introduced at the 3'-terminus of the *tutD* gene.

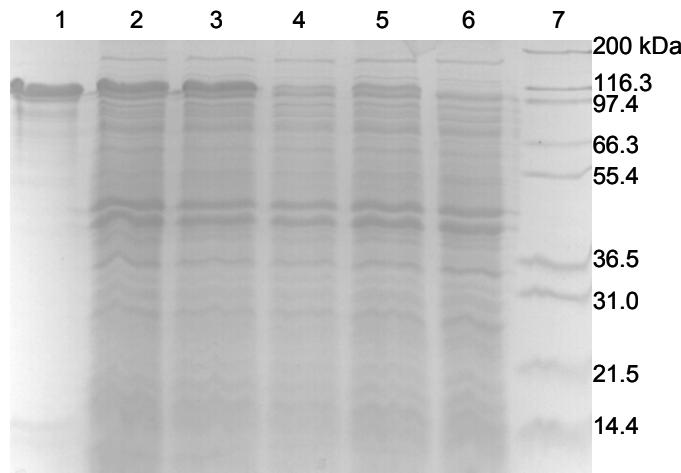
Induction of *E. coli* BL21 harboring the plasmids described above with IPTG at room temperature with limited aeration (gentle stirring) resulted in all three genes being expressed as soluble proteins. Furthermore, the intact BSS complex could easily be purified through the His<sub>6</sub> tag on the  $\alpha$ -subunit by chromatography on a Ni-NTA column, as shown in Figure 2. This result implies that the BSS- $\beta$  and  $\gamma$  subunits are tightly bound to the  $\alpha$  subunit since the small subunits were not themselves tagged and further suggested that one or both of them might be important for the correct folding of the  $\alpha$ -subunit.



**Figure 2.** Over-expression and purification of BSS- $\alpha\beta\gamma$  and BSS- $\alpha\gamma$  complexes. *Lane 1:* BSS- $\alpha_2\beta_2\gamma_2$  complex, purified by Ni-NTA chromatography; *lane 2:* BSS- $\alpha_2\gamma_2$ , complex purified by Ni-NTA chromatography; *lane 3:* BSS- $\alpha_2\gamma_2$ , complex after gel filtration; *lane 4:* molecular weight standards. Note: the lanes have been heavily overloaded so that the weakly staining  $\beta$  and  $\gamma$  subunits can be seen.

To investigate the effects of BSS- $\beta$  and BSS- $\gamma$  on the folding of BSS- $\alpha$ , *E. coli* strains were constructed in which either BSS- $\beta$  or BSS- $\gamma$  were absent. For BSS- $\gamma$  this was achieved by introducing a stop codon after the Phe13 codon in the *tutF* gene (plasmid pET-Duet\_D\_F\_stop, Table 1); for BSS- $\beta$  this was simply achieved using cells only transformed with plasmid pET-Duet\_D\_F. The genes were induced by addition of 0.5 mM IPTG to cultures, which were gently stirred at room temperature for 3 hours. Cell free extracts were prepared and the amount of soluble BSS- $\alpha$  produced evaluated by SDS PAGE. As shown in Figure 3, the absence of the  $\beta$ -subunit appears to have little effect on the level of soluble BSS- $\alpha$ . However, disruption of BSS- $\gamma$  resulted in a significant reduction in the level at which BSS- $\alpha$  accumulated in the cells. Consistent

with this, the complex of BSS- $\alpha$  with BSS- $\gamma$  could be purified from the supernatant by Ni-NTA chromatography (Figure 2), whereas attempts to purify BSS- $\alpha$  by Ni-NTA chromatography from cells lacking BSS- $\gamma$  were unsuccessful.



**Figure 3.** Effect of the BSS- $\beta$  and BSS- $\gamma$  proteins on the expression of BSS- $\alpha$  in *E. coli*. *Lane 1*: BSS- $\alpha$  standard; *lane 2*: co-expression of *tutD*, *tutF* and *tutG* genes (BSS- $\alpha$  strongly expressed); *lane 3*: co-expression of *tutD* and *tutF* genes (BSS- $\alpha$  strongly expressed); *lane 4*: co-expression of *tutD* and *tutG* genes (BSS- $\alpha$  weakly expressed); *lane 5*: expression of only *tutD* gene (BSS- $\alpha$  weakly expressed); *lane 6*: expression of only *tutF* gene (negative control for BSS- $\alpha$  expression); *lane 7*: molecular weight standards.

#### Characterization of Iron-Sulfur Clusters in BSS

The purified BSS protein complex was yellow-brown in color and the u.v.-visible spectrum of the enzyme exhibited a maximum at 420 nm, suggesting the presence of one or more iron sulfur clusters in the protein (Figure 4). Analysis of the iron content of the protein (Table 2) gave a value of  $7.1 \pm 0.3$  iron atoms per  $\alpha\beta\gamma$  trimer, which, together with the characteristic absorption at 420 nm, pointed to the complex most likely containing two  $[4\text{Fe}-4\text{S}]^{2+}$  clusters. Sequence analysis of the BSS proteins identified two cysteine-rich ferridoxin-like motifs, shown below, that could serve as binding sites for  $[4\text{Fe}-4\text{S}]$  clusters in the sequences of BSS- $\beta$  and BSS- $\gamma$ : BSS- $\beta$ ,  $^{26}\text{C}-\text{X}-\text{X}-^{29}\text{C}-\text{X}_{14}-^{44}\text{C}-$

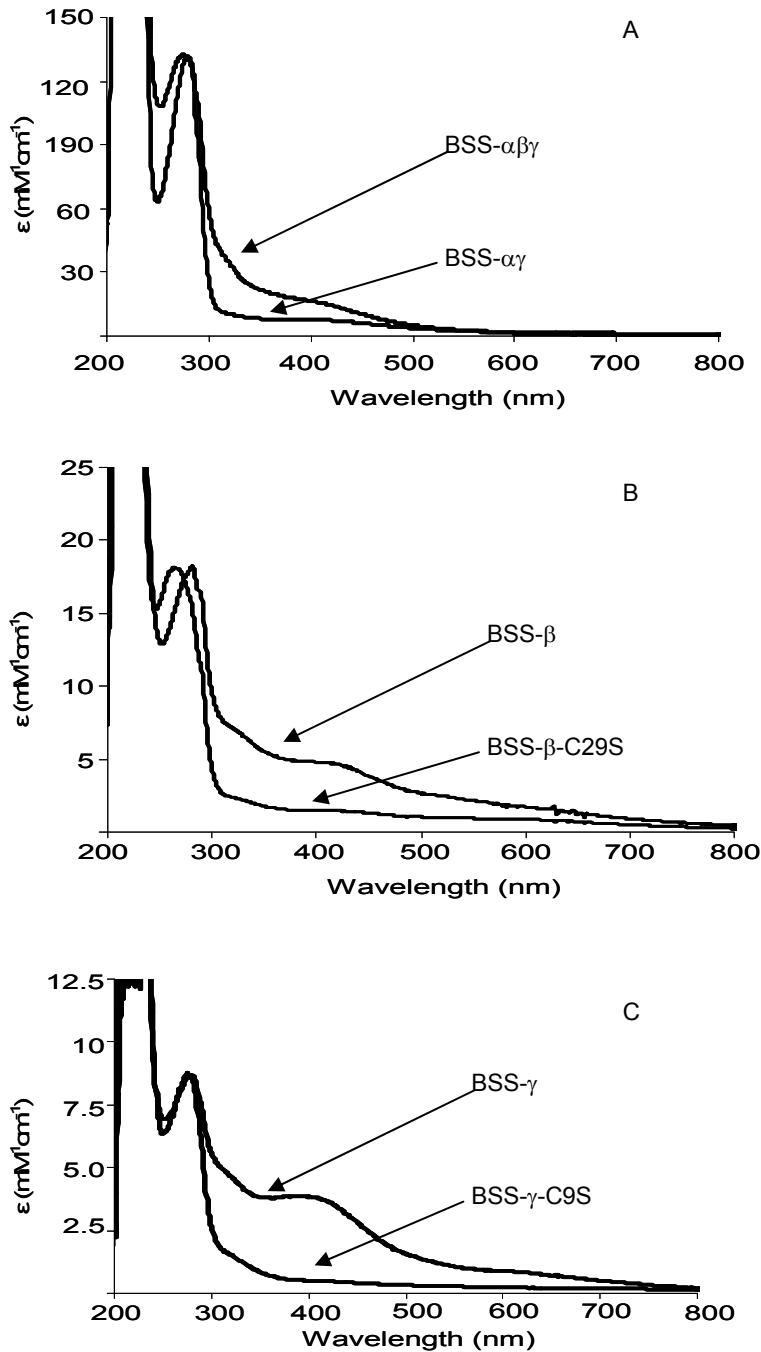
$X_{22}$ - $^{67}\text{C}$ ; and BSS- $\gamma$ ,  $^6\text{C}$ -X-X- $^9\text{C}$ -X<sub>11</sub>- $^{29}\text{C}$ -X<sub>23</sub>- $^{53}\text{C}$ . Consistent with this, the BSS- $\alpha$ :BSS- $\gamma$  complex exhibited a lower  $\epsilon_{420}$  (Table 2) and contained only  $2.9 \pm 0.3$  iron atoms per  $\alpha\gamma$  heterodimer, indicating that the remaining iron-sulfur was associated with the  $\beta$ -subunit.

Protein or protein complex	Fe content	$\epsilon_{420}$ ( $\text{mM}^{-1}\text{cm}^{-1}$ )
BSS- $\alpha\beta\gamma$	$7.1 \pm 0.3$ <sup>a</sup>	18500 <sup>a</sup>
BSS- $\alpha\gamma$	$2.9 \pm 0.3$ <sup>a</sup>	8100 <sup>a</sup>
BSS- $\gamma$	$1.5 \pm 0.2$ <sup>b</sup> $0.5 \pm 0.2$ <sup>a</sup>	3800 <sup>b</sup>
BSS- $\gamma$ -C9S	$0.2 \pm 0.12$ <sup>b</sup> $< 0.11$ <sup>a</sup>	7000 <sup>b</sup>
BSS- $\beta$	$2.0 \pm 0.2$ <sup>b</sup> $0.7 \pm 0.2$ <sup>a</sup>	6200 <sup>b</sup>
BSS- $\beta$ -C29S2	$0.4 \pm 0.2$ <sup>b</sup> $\sim 0.1$ <sup>a</sup>	1800 <sup>b</sup>

**Table 2.** Iron content and extinction coefficients of BSS subunits. <sup>a</sup>Iron content of protein as isolated . <sup>b</sup>Iron content after reconstitution of iron-sulfur cluster under anaerobic conditions.

The UV-visible spectra of the individually expressed  $\beta$ - and  $\gamma$ -subunits, purified by Ni-NTA chromatography from recombinant *E. coli*, each exhibited characteristic maxima at 420 nm suggesting that they contained iron-sulfur clusters (Figure 4). However, the extinction coefficients for the clusters were rather small and iron analysis indicated that less than one iron atom was bound per protein molecule (Table 2), suggesting that most of the iron may have been lost during purification. The iron-sulfur clusters could be partially reconstituted by incubating the proteins with Fe(II) and sulfide ions under reducing conditions, following established protocols described in the

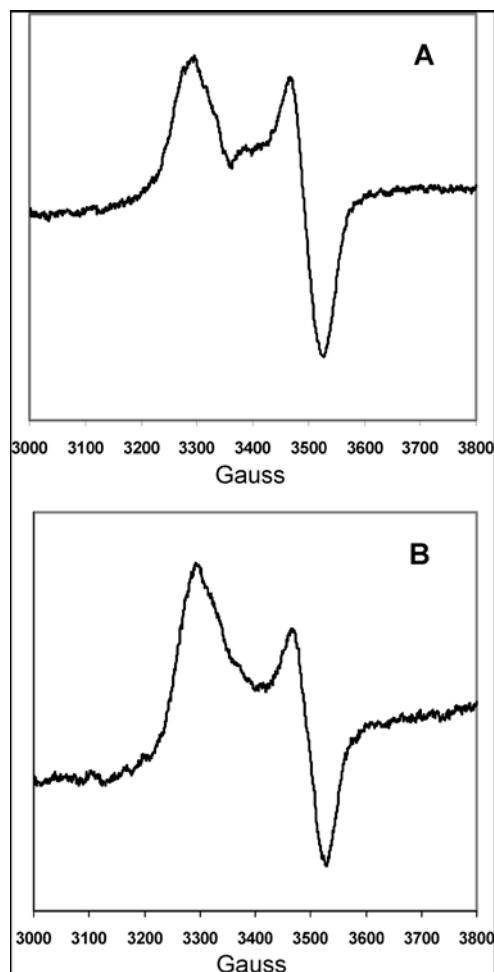
Materials and Methods section. After reconstitution and purification the iron content increased to 1.5 – 2.0 iron atoms per protein molecule, with a corresponding increase in  $\epsilon_{420}$ ; this is still significantly lower than the iron content of the holo-enzyme and suggests that in the absence of BSS- $\alpha$  the iron-sulfur clusters are very labile. Possibly, when BSS- $\beta$ , BSS- $\gamma$  and BSS- $\alpha$  associate together the small proteins become more structured, and thus the binding site for the iron-sulfur cluster better pre-organized, which would stabilize the metal cluster.



**Figure 4.** U.V.-visible spectra of BSS proteins. **A:** Spectra of the BSS- $\alpha\beta\gamma$  complex and the BSS- $\alpha\gamma$  complex as isolated from *E. coli*; **B:** Spectra of BSS- $\beta$  protein and the BSS- $\beta$ -C29S mutant after reconstitution of iron-sulfur clusters; **C:** Spectra of BSS- $\gamma$  protein and the BSS- $\gamma$ -C9S mutant after reconstitution of iron-sulfur clusters

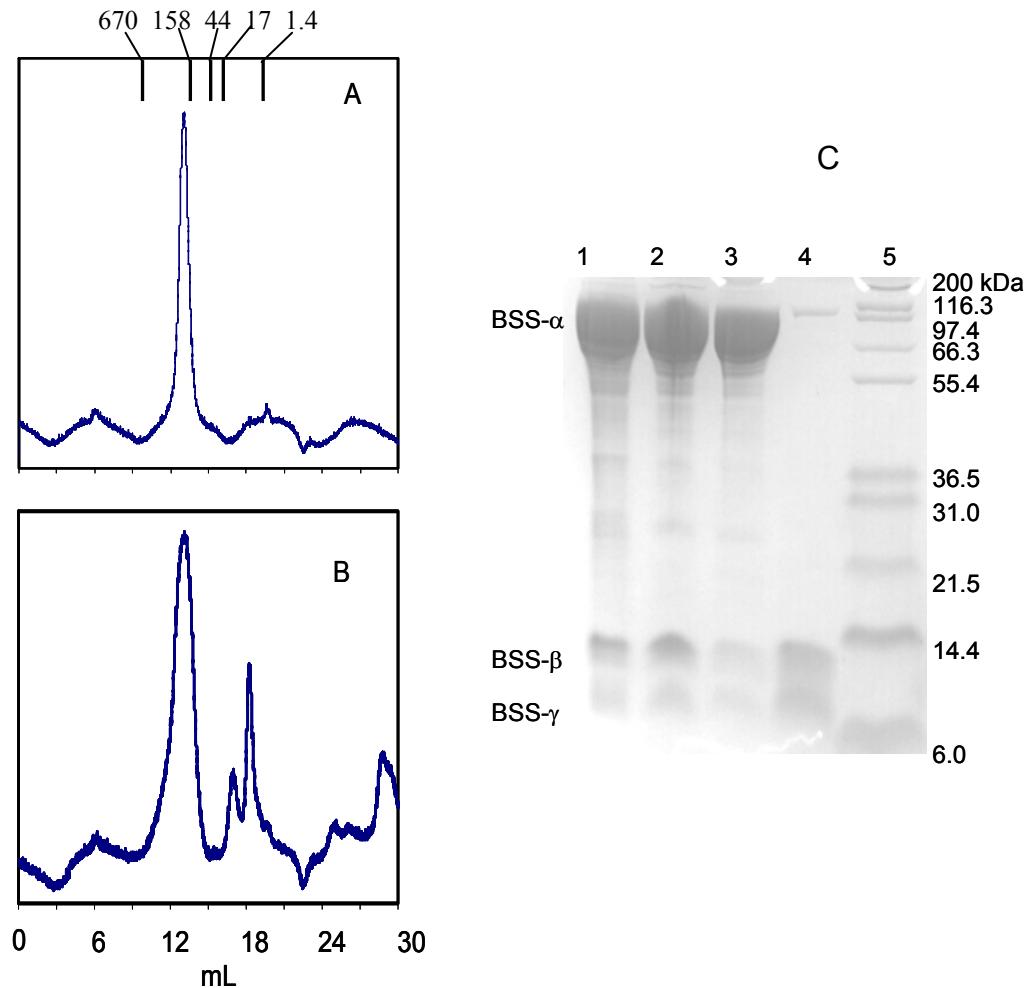
Further evidence of the identity of the iron-sulfur clusters comes from the EPR spectra of the sodium dithionite-reduced BSS- $\alpha\beta\gamma$  complex and the BSS- $\alpha\gamma$  complex,

shown in Figure 5. These spectra are characteristic of  $[4\text{Fe}-4\text{S}]^+$  iron-sulfur clusters, with  $g_{\parallel}$  a of 2.04 and a  $g_{\perp}$  of 1.94. However, the isolated BSS- $\beta$  and BSS- $\gamma$  proteins when reduced gave no EPR signals. Isolated BSS- $\beta$  and BSS- $\gamma$  were reduced immediately upon addition of sodium dithionite, whereas the BSS-  $\alpha\beta\gamma$  complex and the BSS-  $\alpha\gamma$  complex were reduced slowly over the course of  $\sim$ 30 min, as determined by bleaching of the sample color. This suggests that the clusters in the isolated BSS-  $\beta$  and BSS- $\gamma$  proteins are much more exposed to solvent and that reduction results in the loss of the cluster from the protein, consistent with the lability of the unreduced clusters noted above.



**Figure 5.** EPR spectra of sodium dithionite-reduced BSS complexes as isolated from *E. coli*: (A) spectrum of the BSS-R $_{\gamma}$  complex and (B) spectrum of the BSS-R $_{\beta\gamma}$  complex.

The finding that BSS- $\beta$  and  $\gamma$  subunits contained iron-sulfur clusters was unexpected because there is no obvious requirement for a redox cofactor in the mechanism of the radical reaction catalyzed by BSS. We therefore considered whether the clusters may play a structural role. The molecular weight of the recombinant BSS was estimated by gel filtration chromatography using a Superose-6 FPLC column. The protein migrated in a single symmetrical peak (Figure 6) with apparent  $M_r$  of 230000, which is consistent with an  $\alpha_2\beta_2\gamma_2$  subunit structure for the enzyme, as reported for the enzyme purified from *T. aromatica* (16). Analysis of the protein peak after gel filtration confirmed that all three subunits were present (Figure 5).



**Figure 6.** Dissociation of BSS upon treatment with iron chelators. **A:** Chromatography of BSS on Superose-6 gel filtration column. The intact BSS complex elutes with apparent  $M_r = 230000$  (elution volumes of molecular weight standards are indicated above chromatogram). **B:** elution profile of BSS after treatment with iron chelator. The first peak contains primarily BSS- $\alpha$ ; the second BSS- $\beta$  and BSS- $\gamma$ ; the third peak contains no proteins and is probably chelating agent. **C:** analysis of column fractions by SDS-PAGE; *lane 1*: BSS purified by chromatography on Ni-NTA column; *lane 2*: BSS after gel filtration chromatography - material from peak in chromatograph A; *lane 3*: BSS after treatment with iron chelator and gel filtration chromatography - material from first peak (BSS- $\alpha$ ) in chromatograph B; *lane 4*: BSS after treatment with iron chelator and gel filtration chromatography - material from second peak (BSS- $\beta$  and  $\gamma$ ) in chromatograph B. *lane 5*: molecular weight standards.

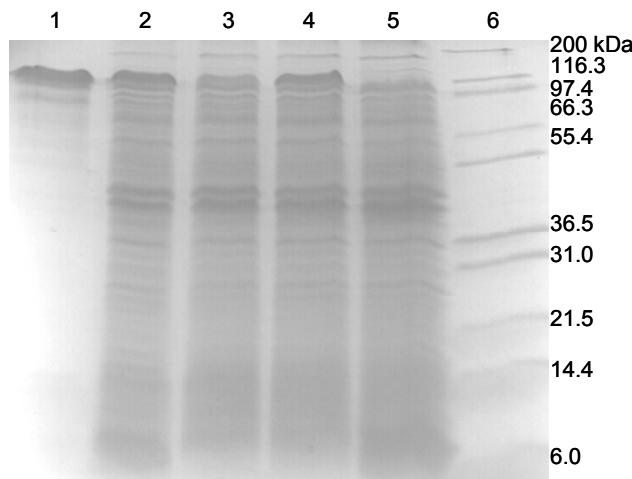
To investigate the effect of removing the iron-sulfur clusters on the quaternary structure of BSS, the protein was incubated with the iron chelating agent OBP overnight under anaerobic conditions, followed by analysis by gel filtration on a Superose-6

column. The OBP-treated protein eluted as two peaks (Figure 6) from the column. SDS-PAGE analysis showed that the first peak is mainly BSS- $\alpha$ , with traces of BSS- $\beta$  and  $\gamma$  remaining; elution volume of this peak indicates that the isolated  $\alpha$ -subunit remains dimeric. The second peak contained BSS- $\beta$  and  $\gamma$  subunits. Thus removal of the iron-sulfur clusters from the small subunits causes them to dissociate from the large  $\alpha$  subunit.

To further investigate the role of the iron-sulfur clusters in the folding and stability of BSS, mutations were introduced to disrupt the binding sites for the clusters. The putative cysteine ligands, Cys29 in BSS- $\beta$  and Cys9 in BSS- $\gamma$  were mutated to serine (plasmids pACYC-Duet\_F\_C9S and pET28\_G\_C29S) and the mutant proteins expressed and purified from *E. coli*. Consistent with these cysteine residues providing the ligands to iron, the mutant proteins contained almost no iron [ $\sim 0.1$  Fe/protein (Table 2)], and attempts to reconstitute the iron-sulfur clusters in these proteins resulted in less than 0.5 Fe/protein being incorporated (Table 2 and Figure 4). Such low levels of iron incorporation may well be due to non-specific binding of the metal to the protein.

We then examined the effect of co-expressing the mutant BSS- $\beta$  and  $\gamma$  proteins on the stability of BSS- $\alpha$  expressed in *E. coli*. This was achieved by co-transforming *E. coli* BL12 with either plasmids pET-Duet\_D\_F and pRSF-Duet\_G\_C29S or plasmids pET-Duet\_D\_F\_C9S and pRSF-Duet\_G. Similar to the experiments described above, the expression of the *tut* genes was induced with IPTG for 3 hours, cell-free extracts prepared and the amount of soluble BSS- $\alpha$  protein produced examined by SDS-PAGE. The results (Figure 7) mirrored those obtained from the gene deletion experiments. The Cys9Ser mutation in BSS- $\gamma$  resulted in a noticeable reduction the amount of BSS- $\alpha$  produced; in contrast the Cys29Ser in BSS- $\beta$  did not significantly alter BSS- $\alpha$

expression. The result indicates that the iron-sulfur cluster in BSS- $\gamma$  is necessary for this subunit to bind to BSS- $\alpha$ , which is consistent with the gel filtration data that shows that removal of the clusters results in dissociation of the subunits.



**Figure 7.** Effect of mutations that remove the iron-sulfur clusters in BSS- $\beta$  and BSS- $\gamma$  on expression of BSS- $\alpha$  in *E. coli*. *Lane 1:* BSS- $\alpha$  standard; *lane 2:* co-expression of *tutD*, *tutF* and *tutG* genes (BSS- $\alpha$  strongly expressed); *lane 3:* co-expression of *tutD*, *tutG* and *tutF-C9S* genes (BSS- $\alpha$  weakly expressed); *lane 4:* co-expression of *tutD*, *tutF* and *tutG-C29S* genes (BSS- $\alpha$  strongly expressed); *lane 5:* expression of only *tutF* gene (negative control for BSS- $\alpha$  expression); *lane 6:* molecular weight standards.

## Discussion

Here we report the expression and purification from *E. coli* and initial physical characterization of BSS from *T. aromatica*. The recombinant enzyme is produced without the catalytic glycyl radical, resulting in a stable protein that facilitates its handling and characterization. In this form the protein is, of course, inactive so that its catalytic properties cannot be examined, however that fact that recombinant BSS can be isolated as the iron-sulfur-containing hetero-hexameric complex strongly suggests that it is correctly folded. Generation of the radical-containing form of BSS will require the requisite activase enzyme be co-expressed in *E. coli*; experiments to accomplish this are

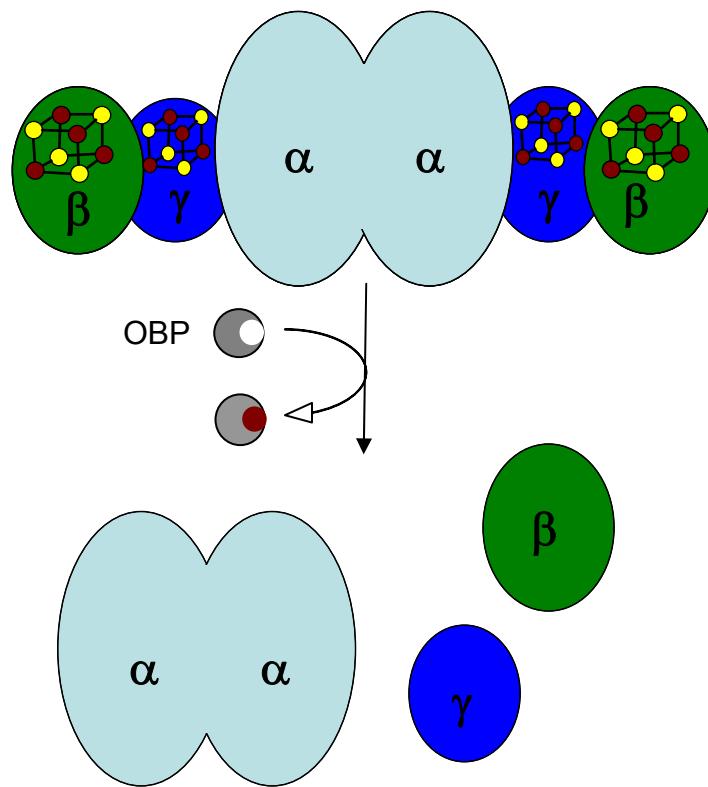
currently on-going in our laboratory. Interestingly, it appears that the endogenous *E. coli* activase enzymes for pyruvate formate lyase and anaerobic ribonucleotide reductase, which are expressed when *E. coli* is grown anaerobically, are unable to activate BSS (L. Li and D. P. Patterson, unpublished results).

BSS has a more complex structure than other better-characterized glycyl radical enzymes: pyruvate formate lyase, anaerobic ribonucleotide reductase and the B<sub>12</sub>-independent glycerol dehydratase are all homodimers and contain no metal clusters (27-29). In contrast, BSS comprises three subunits, α, β and γ, and we have shown here that both the β- and γ-subunits contain [4Fe-4S] clusters. Genetic experiments in Coschigano's laboratory have demonstrated that both small subunits are essential for *T. aromatica* to grow on toluene, indicating that they are required for the large subunit to be active (26). In this study we sought to gain insight into their biochemical function.

One function of BSS-γ appears to be to stabilize the structure of the active site-containing BSS-α subunit. Thus in the absence of the γ subunit BSS-α was expressed poorly, if at all, in *E. coli* and could not be purified from induced cells by Ni-NTA affinity chromatography. The iron-sulfur cluster appears to be essential for BSS-γ to bind BSS-α as chelation of the iron results in dissociation of the subunits. Furthermore, mutation of the putative iron ligand Cys9 to alanine abolished both formation of the cluster and expression of BSS-α in cells containing the mutant BSS-γ protein. This mutation has also been shown to abolish the ability of *T. aromatica* to grow on toluene (25).

The biochemical function of BSS-β is less clear. The iron-sulfur cluster appears to be important for BSS activity because *T. aromatica* carrying the BSS-β C29S mutation

cannot grow on toluene (25). However, neither deletion of this subunit unit, nor mutation of the iron-sulfur cluster away from the protein appears to affect the stability of the  $\alpha$  subunit as judged by the levels of expression of BSS- $\alpha$  in *E. coli*. The BSS- $\alpha\gamma$  complex can purified as a stable protein in the absence of BSS- $\beta$ , whereas in the absence of BSS- $\gamma$  the BSS- $\alpha\beta$  complex cannot. Taken together, these results tentatively suggest that the  $\gamma$ -subunit may serve as a linker between the  $\alpha$  and  $\beta$ -subunits, as illustrated in Figure 8.



**Figure 8.** Structural model for BSS, illustrating the location of the [4Fe-4S] clusters and the proposed quaternary structure of the enzyme. Chelation of iron by OBP results in dissociation to the dimeric  $\alpha$ -subunit and monomeric  $\beta$ - and  $\gamma$ -subunits.

Structurally, the most closely related glycyl radical enzyme to BSS appears to be 4-hydroxyphenylacetate (HPA) decarboxylase which catalyzes the formation of *p*-cresol

in various Clostridia (10). This enzyme comprises two subunits and is active as an  $\alpha_4\beta_4$  hetero-octamer. The large ( $\sim 100$  kDa) subunit contains the catalytic glycyl radical, whereas the small ( $\sim 9.5$  kDa) subunit was recently shown to contain a [4Fe-4S] cluster by Selmer and co-workers (30), who perceptively speculated the BSS- $\beta$  and  $\gamma$  subunits might similarly be iron-sulfur proteins. It also was found that co-expression of the HPA decarboxylase  $\beta$ -subunit was essential to obtain soluble  $\alpha$ -subunit in *E. coli* (31), suggesting that HPA decarboxylase- $\beta$  and BSS- $\gamma$  subunits have similar functions.

Whereas the [4Fe-4S] clusters may simply play a structural role in BSS, as the  $\gamma$ -subunit appears to, this seems somewhat unlikely, especially for the cluster associated with the  $\beta$ -subunit. It has been suggested that in HPA decarboxylase the  $\beta$ -subunit may serve to regulate the enzyme by reducing the catalytic glycyl radical and thereby deactivating the enzyme, as the glycyl radical in this enzyme has a fairly short lifetime (30). It is plausible that the small subunits in BSS play a similar role and that perhaps BSS- $\beta$  interacts with a further regulatory protein involved in sensing toluene concentrations.

Lastly, we note also that serine phosphorylation has been implicated in the regulation of HPA decarboxylase activity (30, 31) suggesting a complicated mechanism for the control of *p*-cresol production. This raises the intriguing question of whether phosphorylation may play a role in the activation of BSS. It appears that BSS and HPA decarboxylase form a subclass of glycyl radical enzymes that is distinct both structurally and mechanistically from the better understood paradigm represented by pyruvate formate lyase.

## Acknowledgement

We thank Prof. Stephen Ragsdale and Dr. Gunes Bender for assistance with obtaining EPR spectra and Andrew Markham for valuable technical assistance. This research was supported by grants from the National Institutes of Health, GM 59227 to E.N.G.M. and 1K99ES017177-01 to L.L.; D.P.P. acknowledges the support of NIH funded Chemistry Biology Interface training grant T32 GM008597

## Supporting Material

A detailed description of the construction of the plasmids used in this study is available as supporting material with the web version of this manuscript.

## References

- 1 Biegert, T., Fuchs, G., and Heider, F. Evidence that anaerobic oxidation of toluene in the denitrifying bacterium *Thauera aromatica* is initiated by formation of benzylsuccinate from toluene and fumarate, *Eur. J. Biochem.* 238, 661-668 (1996).
- 2 Rabus, R., and Heider, J. Initial reactions of anaerobic metabolism of alkylbenzenes in denitrifying and sulfate reducing bacteria, *Arch. Microbiol.* 170, 377-384 (1998).
- 3 Spormann, A. M., and Widdel, F. Metabolism of alkylbenzenes, alkanes, and other hydrocarbons in anaerobic bacteria, *Biodegradation* 11, 85-105 (2000).
- 4 Heider, J., Spormann, A. M., Beller, H. R., and Widdel, F. Anaerobic bacterial metabolism of hydrocarbons, *FEMS Microbiol. Rev.* 22, 459-473 (1998).
- 5 Leuthner, B., and Heider, J. A two-component system involved in regulation of anaerobic toluene metabolism in *Thauera aromatica*, *FEMS Microbiol. Lett.* 166, 35-41 (1998).
- 6 Coschigano, P. W. Transcriptional analysis of the *tutE tutFDGH* gene cluster from *Thauera aromatica* strain T1, *Appl. Environ. Microbiol.* 66, 1147-1151 (2000).
- 7 Achong, G. R., Rodriguez, A. M., and Spormann, A. M. Benzylsuccinate synthase of *Azoarcus* sp strain T: Cloning, sequencing, transcriptional organization, and its role in anaerobic toluene and m-xylene mineralization, *J. Bacteriol.* 183, 6763-6770 (2001).
- 8 Frey, P. A. Radical mechanisms of enzymatic catalysis, *Ann. Rev. Biochem.* 70, 121-148 (2001).
- 9 Marsh, E. N. G., Huhta, M. S., and Patwardhan, A. S-adenosylmethionine radical enzymes, *Bioorganic Chem.* 32, 326-340 (2004).

- 10 Selmer, T., Pierik, A. J., and Heider, J. New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria, *Biol. Chem.* 386, 981-988 (2005).
- 11 Duboc-Toia, C., Hassan, A. K., Mulliez, E., Ollagnier-de Choudens, S., Fontecave, M., Leutwein, C., and Heider, J. Very high-field EPR study of glycyl radical enzymes, *J. Am. Chem. Soc.* 125, 38-39 (2003).
- 12 Verfurth, K., Pierik, A. J., Leutwein, C., Zorn, S., and Heider, J. Substrate specificities and electron paramagnetic resonance properties of benzylsuccinate synthases in anaerobic toluene and m-xylene metabolism, *Arch. Microbiol.* 181, 155-162 (2004).
- 13 Krieger, C. J., Roseboom, W., Albracht, S. P. J., and Spormann, A. M. A stable organic free radical in anaerobic benzylsuccinate synthase of *Azoarcus* sp strain T, *J. Biol. Chem.* 276, 12924-12927 (2001).
- 14 Frey, P. A, Hegeman, A. D., and Reed, G. H. Free radical mechanisms in enzymology. *Chem. Rev.* 106, 3302-3316 (2006).
- 15 Walsby, C. J., Ortillo, D., Yang, J., Nnyepi, M. R., Broderick, W. E., Hoffman, B. M., and Broderick, J. B. Spectroscopic approaches to elucidating novel iron-sulfur chemistry in the "Radical-SAM" protein superfamily. *Inorg. Chem.* 44, 727-741 (2005).
- 16 Leuthner, B., Leutwein, C., Schulz, H., Horth, P., Haehnel, W., Schiltz, E., Schagger, H., and Heider, J. Biochemical and genetic characterization of benzylsuccinate synthase from *Thauera aromatica*: A new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol. Microbiol.* 28, 615-628 (1998).
- 17 Qiao, C. H., and Marsh, E. N. G. Mechanism of benzylsuccinate synthase: Stereochemistry of toluene addition to fumarate and maleate, *J. Am. Chem. Soc.* 127, 8608-8609 (2005).
- 18 Li, L., and Marsh, E. N. Deuterium isotope effects in the unusual addition of toluene to fumarate catalyzed by benzylsuccinate synthase. *Biochemistry* 45, 13932-13938 (2006).
- 19 Himo, F. Catalytic mechanism of benzylsuccinate synthase, a theoretical study, *J. Phys. Chem. B* 106, 7688-7692 (2002).
- 20 Li, L., and Marsh, E. N. G. Mechanism of Benzylsuccinate Synthase Probed by Substrate Exchange. *J. Am. Chem. Soc.* 128, 16056-16057 (2006).
- 21 Gill, S. C., and Vonhippel, P. H. Calculation of Protein Extinction Coefficients from Amino-Acid Sequence Data, *Anal. Biochem.* 182, 319-326 (1989).
- 22 Fish, W. W. Rapid colorimetric micromethod for the quantitation of complexed iron in biological samples, *Methods Enzymol.* 158, 357-364 (1988).
- 23 Ugulava, N. B., Gibney, B. R., and Jarrett, J. T. Biotin synthase contains two distinct iron-sulfur cluster binding sites: chemical and spectroelectrochemical analysis of iron-sulfur cluster interconversions, *Biochemistry* 40, 8343-8351 (2001).
- 24 Carroll, K. S., Gao, H., Chen, H., Leary, J. A., and Bertozzi, C. R. Investigation of the iron-sulfur cluster in *Mycobacterium tuberculosis* APS reductase: implications for substrate binding and catalysis, *Biochemistry* 44, 14647-14657 (2005).

- 25 Bhandare, R., Calabro, M., and Coschigano, P. W. Site-directed mutagenesis of the *Thauera aromatica* strain T1 *tutE tutFDGH* gene cluster, *Biochem. Biophys. Res. Comm.* 346, 992-998 (2006).
- 26 Coschigano, P. W. Construction and characterization of insertion/deletion mutations of the *tutF*, *tutD*, and *tutG* genes of *Thauera aromatica* strain T1, *FEMS Microbiol. Lett.* 217, 37-42 (2002).
- 27 Becker, A., Fritz-Wolf, K., Kabsch, W., Knappe, J., Schultz, S., and Wagner, A. F. V. Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase, *Nat. Struct. Biol.* 6, 969-975 (1999).
- 28 Logan, D., Andersson, J., Sjoberg, B.-M., and Nordlund, P. A glycyl radical site in the crystal structure of a class III ribonucleotide reductase, *Science* 283, 1499-1504 (1999).
- 29 O'Brien, J. R., Raynaud, C., Croux, C., Girbal, L., Soucaille, P., and Lanzilotta, W. N. Insight into the mechanism of the B<sub>12</sub>-independent glycerol dehydratase from *Clostridium butyricum*: Preliminary biochemical and structural characterization, *Biochemistry* 43, 4635-4645 (2004).
- 30 Yu, L., Blaser, M., Andrei, P. I., Pierik, A. J., and Selmer, T. 4-hydroxyphenylacetate decarboxylases: Properties of a novel subclass of glycyl radical enzyme systems, *Biochemistry* 45, 9584-9592 (2006).
- 31 Andrei, P. I., Pierik, A. J., Zauner, S., Andrei-Selmer, L. C., and Selmer, T. Subunit composition of the glycyl radical enzyme p-hydroxyphenylacetate decarboxylase - A small subunit, HpdC, is essential for catalytic activity, *Eur. J. Biochem.* 271, 2225-2230 (2004).