UCLA

UCLA Electronic Theses and Dissertations

Title

Design, Characterization and Applications of Symmetric Protein Scaffolds A

Permalink

https://escholarship.org/uc/item/5z72t53d

Author

Leibly, David

Publication Date

2016

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Design, Characterization and Applications of Symmetric Protein Scaffolds

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Biochemistry and Molecular Biology

By

David James Leibly

2016

© Copyright by

David James Leibly

2016

ABSTRACT OF THE DISSERTATION

Design, Characterization and Applications of Symmetric Protein Scaffolds

by

David James Leibly

Doctor of Philosophy in Biochemistry and Molecular Biology

University of California, Los Angeles, 2016

Professor Todd O. Yeates, Chair

Proteins are essential macromolecules for all living organisms. They provide cellular structure and perform most of the metabolic functions essential for all life. The importance of proteins makes them the most studied and exploited macromolecules. We can exploit the structure of a protein to design a specific therapeutic to treat a disease or we can use proteins as biocatalysts for the efficient creation of molecules. In many instances, these applications of proteins are difficult to achieve. The work in this dissertation focuses on the development and evaluations of novel techniques to aid in the study and use of proteins.

The first part of this dissertation focus on the creation of a series of symmetric oligomers to be used as crystallization scaffolds. Such scaffolds are intended to induce their symmetry onto asymmetric protein crystallization target proteins. The ability to determine the crystal structure can be essential for the creation of new targeted drugs or the better understanding of a biological process. Unfortunately many proteins fail to crystallize for reasons that are not well understood. It is thought that such induction of symmetry and variety of geometrically distinct scaffolds will aid in the crystallization of difficult-to-crystallize proteins. Preliminary results of these novel scaffolds and existing scaffolds are described.

In the second part, applications of symmetric scaffolds for the creation of enzymatic materials are presented. These purely proteinaceous assemblies are designed to replicate previous described enzyme encapsulating materials. These materials typically improve enzyme reaction rates and product extraction. The final part of the dissertation focuses on the shell protein PduA from the 1,2-propanediol-utilization bacterial microcompartment (MCPs). These MCPs encapsulate metabolic pathways and contain volatile or toxic pathway intermediates. Research into turning these MCPs into bioreactors containing non-native enzymes is ongoing in many labs. Full realization of this technology relies on the encapsulation of new metabolic enzymes and transport of novel substrate and products through the shells. These processes are poorly understood, here structural studies of shell protein permutations. These permutations alter the topology of the shell protein allowing the scaffolding of proteins to the exterior surface of the MCPs. Finally, the efforts to elicited the interaction of specific targeting sequences to shell protein by x-ray crystallography are discussed.

This dissertation of David James Leibly is approved.

James U. Bowie

Pascal Francois Egea

Todd O. Yeates, Committee Chair

University of California, Los Angeles

2016

To those who have inspired and encouraged me to pursue science

TABLE OF CONTENTS

List of Figures xv

List of Tables..xl

Acknowledgements..xl

Vita...xliv

Selected Publications and Presentations xlv

CHAPTER 1. Introduction

1.1 Synthetic Symmetrization as a Crystallization Tool ……………………………………1

1.1.1 Obstacles for Individual Proteins………………………………………..2

1.1.2 Pre-formed Symmetric Scaffolds……………………………………...4

1.2 Higher Order Scaffolds…………………………………………………..………...5

1.3 Alternate Applications of Symmetric Scaffolds……………………………..……10

1.4 Pdu Microcompartment Protein PduA…………………………………..……..10

1.5 Targeting Enzymes to the Microcompartment………………...…………………..12

1.6 Summary of the Dissertation…………………………………………..………..15

CHAPTER 2 - A Suite of Engineered GFP Molecules for Oligomeric Scaffolding…………18

CHAPTER 3 – Split GFP as a Crystallization Scaffold…………………………………..……...41

3.1 Introduction - A Split GFP Mediated Crystallization Scaffold………………..……...41

3.2 Results - Difficulties in Dimer Formation and Purification………………………41

3.2.1 Cysteine Null Constructs………………………………………………..42

3.2.2 Alternate Purification Methods……………………………………..……..45

3.2.2.1 Thiopropyl Sepharose Purification………………………..……..48

3.2.2.2 Oxidizing Cytosolic *E. coli* Expression ……………..………..51

3.2.2.3 Anion Exchange Chromatography…………………………..…...…51

3.2.3 Metal-Mediated Oligomers are Unpredictable ……………………..……...52

3.3 Refolding Attempts of Split GFP Mutants ……………………………………..55

3.3.1 Refolding Attempts of Larger Amounts of Inclusion Bodies ……………..56

3.3.2 IMAC Purification of Refolded Protein……………………………………59

3.3.3 IMAC On Column Refolding Attempts……………………………………59

3.3.4 IEX Purification of Refolded Proteins……………………………………..60

3.3.5 Dynamic Light Scattering Experiments……………………………………63

3.4 Results - Mycobacterium tuberculosis Rv2658c in vitro complementation…………66

3.5 Results- Coexpression Attempts……………………………………….…………66

3.5.1 H3L……………………………………………………………………..…...67

3.5.2 StarD9 Kinesin…………………………………………………………..…74

3.5.3 sfCherry I206K……………………………………………….………..…79

3.6 Discussion…………………………………….……………………………81

3.6.1 Viability of *In Vitro* GFP Complex Formation……………...…………..…81

3.6.2 Purification of GFP 1-9-POI Complexes…………………………………..83

3.6.3 *In Vivo* Complementation of Split GFP for Complex Formation…………84

3.7 Conclusions……………………………………………………………………..84

3.8 Material and Methods…………………………………………………………...85

3.8.1 Cloning of POI constructs………………………………………………85

3.8.2 Protein Expression………………………………………………………86

3.8.3 Protein Purification…………………...…………………………………86

3.8.4 Dynamic Light Scattering………………………………………………88

3.8.5 Crystallization of Complexes……………………………………………88

CHAPTER 4 – Computationally Designed Trimeric GFPs ……………………………..89

4.1 Introduction – Computationally Design GFP Oligomers……………………………89

4.2 Results – Perpendicular Docked c3 Trimers……………………………………………90

4.2.1 Purification and Crystallization of c3 Designs……………………………92

4.2.2 Structure of c3_33 JCSG+ D12 Dimer……………………………………96

4.3 Results – Parallel Docked c3 Trimers…………………………………………………..96

4.3.1 Purification of c3_51 Design……………………………………………...……100

4.3.2 Structure of c3_33 Trimer……………………………………………...…100

4.4 Discussion.………………………………..…………………………...……………103

4.4.1 Issues With GFP Computational Assembly Design ………………...……103

4.4.2 Applications of New GFP Oligomers……………………………..…………104

4.5 Conclusions…………………………………………………………………………105

4.6 Materials and Methods…………………………………………………..…….……106

4.6.1 Cloning………………………………………………………………...…106

4.6.2 Protein Expression……………………………………………...……106

4.6.3 Protein Purification………………………………………………….……106

4.6.4 Protein Crystallization………………………………………..…...……107

4.6.5 X-Ray Data Processing and Structure Solution………………….………107

CHAPTER 5 – Coiled-Coil Fusions to T4-Lysozyme Oligomers…………….……………...…116

5.1 Introduction…………………………………………………..…..………………..116

5.2 Results…………………………………………………………………..………..119

5.2.1 Creation of T4L Coiled Coil Constructs……………………………………119

5.2.2 Selection of Crystallization Targets…………………………………………119

5.2.3 Expression and Purification of constructs………………………………124

5.2.3.1 Co-purification of Complexes……………………………………125

5.2.4 Crystallization Experiments………………………………………………129

5.3 Discussion……………………………………………………………………………129

5.3.1 Coiled-Coil Motifs Significantly Reduce Protein Solubility………………129

5.3.2 Stability Through Co-Purifications…………………………………………130

5.3.3 Rescue Strategies to Validated Coiled-Coil Mediated Oligomers…………130

5.3.3.1 Coiled-Coil Optimization………………………………………130

5.3.3.2 Disulfide Bond Optimization……………………………………131

5.3.4 New Synthetic Symmetry Scaffolds and Fusion Approaches………………132

5.4 Conclusions…………………………………………………………………………133

5.5 Materials and Methods……………………………………………………………133

5.5.1 Cloning………………………………………………………………………133

5.5.2 Protein Expression……………………………………………………………134

5.5.4 Crystallization Experiments…………………………………………………134

CHAPTER 6 – Computationally Designed Cages as Scaffolds.……………………137

6.1 Introduction…………………………………………………………………………137

6.2 Results………………………………………………………………………………140

6.2.1 T3-10 sfGFP Fusions…………………………………………………………140

6.2.1.1 Dynamic Light Scattering of T3-10 sfGFP………………………...……144

6.2.1.2 GFP V206K Constructs…………………………..………….…..……144

6.2.2 T3-10 StarD9 Fusions…………………………..…………………..…145

6.2.2.1 T3-10 StarD9 Crystal Optimization…………………………….…149

6.2.2.2 T3-10 StarD9 Crystals Forms……………………………………..149

6.2.2.3 T3-10 StarD9 Loop Deletion Constructs…………………………..149

6.2.2.4 Molecular Replacement Efforts…………………...………………155

6.3 Discussion……………………………………………………………………155

6.3.1 Viability of Cages as Scaffolds…………………………………………155

6.3.1.1 Internal Cage Fusions………………………………………..……156

6.3.2 Linker Optimization…………………………………………………159

6.5 Conclusions……………………………………………..……………………159

6.6 Materials and Methods…………………………………………………..……159

6.6.1 Cloning…………………………………………………………….…159

6.6.2 Protein Expression…………………………………………………..…160

6.6.3 Protein Purification……………………………………………………160

6.6.4 Crystallization……………………………………………………...…161

6.6.5 X-Ray Data Collection and Processing…………………..…………………161

CHAPTER 7 – Oligomeric Enzyme Based Materials

7.1 Introduction – Sol-Gels for Protein Encapsulation……………………….………162

7.2 Results – Candidates for Enzymatic Gels……………………………………...…162

7.2.1 Bromperoxidase – GFP Dimer Materials………………………………163

7.2.2 Bromperoxidase Activity Assays……………………………………………….168

7.3 Polychlorinated Biphenyl Degradation Pathway BphABCD……………………….………168

7.3.1 Bph Enzyme Constructs……………………………………………..……173

7.3.2 Bph Expression Testing…………………………………………………173

7.4 Discussion……………………………………………………………………………173

7.4.1 Structure Hypothesis for Lack of Bromperoxidase Activity…………………173

7.4.2 Optimization of Bromperoxidase Materials With Multivalent Thiol Cross-linkers…………………………………………………………...…………175

7.4.3 Modifications of Bph Enzymes for Soluble Expression………………………175

7.5 Conclusions……………………………………………………………………..……177

7.6 Materials and Methods………………………………………………………………177

7.6.1 Cloning……………………………………………………………………177

7.6.2 Protein Expression…………………………………………………...……177

7.6.3 Protein Purification……………………………………………..…………178

7.6.4 Bromperoxidase Material Formation………………………………...……178

7.6.5 Bromperoxidase Activity Assays……………………………………………178

CHAPTER 8 – Circular Permutations of PduA

8.1 Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein……………………………………………………181

8.2 Introduction – Design of a More Stable Permuted PduA Cage…………………………..188

8.3 Results- Attempts to Optimize the Permuted PduA Cage……………………………188

8.3.4 IEX Purification of Untagged PduAp……………………………….……188

8.3.2 Rosetta Designed Constructs……………...……………………………………192

8.3.3 BeAtMuSiC Hotspot Mutants………………………………………………..…192

8.3.4 Cysteine Mutants of PduAp……………………………………………………196

8.4 Results - Reconstructing the PduA Hexamer……………………………………..……196

8.5 Redesigned PduA-PduU Permutations………………………………...………………200

8.6 Discussion……………………………………………………………………...……200

8.6.1 Experimental Characterization of Computational Designed Permuted PduA Constructs………………………………………………………………………200

8.6.2 Salt Induced Disruption of Permuted PduA………………………...……….202

8.6.3 TEV Coexpression for In Vivo Affinity Tag Removal……………...…………202

8.6.4 Structure of Redesigned Permuted PduA Hexamer…………………………....204

8.6.5 Interconversion Between Pentamers and Hexamers Has Evolutionary Implications……………………………………………………………………204

8.7 Conclusions………………………………………………………………………..…205

8.8 Materials and Methods……………………………………………………………...…205

8.8.1 Cloning…………………………………………………………………………205

8.8.2 Protein Expression……………………………………………………………...206

8.8.3 Ion Exchange Purification of Untagged Constructs…………………………....206

8.8.4 Purification of Affinity Tagged Proteins……………………………………….206

8.8.5 PduAp7 Crystallization……………………...…………………………………207

8.8.6 PduAp7 X-Ray Data Processing and Structure Solution……………………....208

CHAPTER 9 – Structural Studies of the PduA – PduD Helix Interaction

9.1 Introduction……………………………………………………………………….216

9.1.1 Microcompartment Enzyme Encapsulation………………………….....…….216

9.1.2 Summary of Previous Crystallographic Efforts……………………………..217

9.2 Results……………………………………………………………………………221

9.2.1 PduAD Fusion Model Constructs…..………………………………………..221

9.2.2 PduAD3 Crystal Structures………………………………………………….221

9.3 Discussion………………………………...……………………………………….224

9.3.1 Obscured Electron Density From Six-Fold Symmetry……………...………224

9.3.2 PduAD Crystals..…………………………………………………………….225

9.3.3 Fusion Protein Creates Incorrect Stoichiometry……………………………..…226

9.3.4 Crystallization May Select For Unbound Structures…………………..………226

9.3.5 Ambiguity in the Binding Partners and Orientations of Targeting Tails ………………………………………………………..227

9.3.5.1 Preferential Binding of Tails to Specific Shells……………………..227

9.3.6 Crystallographic Rescue Through Broken Symmetry Scaffolds………………228

9.4 Conclusions………………………………………………………………..………228

9.5 Materials and Methods……………………………………………………………228

9.5.1 Cloning……………………………………………...………………..228

9.5.2 Protein Expression…………………………………...…………………229

9.5.3 Protein Purification…………………………………………………………..229

9.5.4 PduAD Crystallization……………………………………………………….229

9.5.4.1 PduAD3 Optimization……………………………………………….229

9.5.5 X-ray Data Processing and Structure Solution……………………………………230

9.5.5.1 PduAD3 P1 Crystal Form………………………………………………230

9.5.5.2 PduAD3 $P4_322$ Crystal Form………………………………………..230

References……………………………………………………………………..……………………232

LIST OF FIGURES

CHAPTER 1. Introduction

Figure 1.1 **Concept of synthetic symmetrization.** Adapted with permission from Banatao *et al*, 2006 (DOI 10.1073/pnas.0607674103). An asymmetric molecule such as a '5' can be made symmetric by introduction of a point of attachment (yellow circle). This can be translated to an asymmetric protein or protein complex and the introduction of a point mutation such as a cysteine or metal binding site. As the position of this mutation is changed on the surface of the molecule different oligomers are formed, each with unique opportunities to form a crystal lattice. Although the overall geometry of oligomer changes the tertiary structure of the asymmetric subunit remains unchanged……………………………………………………………...……….3

Figure 1.2 **Scaffold Induced Synthetic Symmetrization.** Instead of introducing symmetrizing mutations onto a protein of interest, the mutations can be made on a core protein, in this case GFP 1-9. The protein of interest can then be fused to the GFP 10/11 hairpin segment. Multiple GFP 1-9 cores with unique oligomer mutations can be made, once complemented the symmetry of the core will be induced onto the protein of interest…………………………………………………………………….………….6

Figure 1.3 **Presence of Multiple Cysteines Results in Unpredictable Oligomerization.** Although it was thought that cysteine 70 was buried in the core of split GFP based upon the sfGFP structure, it was reactive and formed mulitmers. These are likely hetero-oligomers between the C70 and the introduced mutation. The unexpected reactivity of C70 indicates that it is essential that only one cysteine is present during the cysteine oxidation step. This would require the GFP dimer cores to be oxidized to dimers prior to complementation with a protein containing native cysteines…………………………………………………………………...8

Figure 1.4 **Metal Chelation is Buffer Specific.** Examples of unintended metal chelation events dues to the crystallization conditions. In each case the histidine or cysteine side chains where point mutations to bind the metal ion. It was found that acidic side chains or lysine side chain contributed to one half of the metal contact depending on the buffer conditions. Small molecule chelators like citrate can also alter the metal –chelation when present in crystallization screens………………..9

Figure 1.5. **The Pdu MCP.** Adapted with permission from Jorda *et al* 2006 (DOI 10.1002/pro.2196). The 1,2-propandiol utilization microcompartment encapsulates the enzymes for 1,2-proapandiol catabolism to propionate. Part of this pathway involves the intermediate propionaldehyde (red box) this DNA modifying cytotoxic intermediate is sequestered in the shell and away from the bacterial cytoplasm…………………………………………………………13

Figure 1.6. **The Pdu MCP operon and shell proteins.** The BMC assemble into hexagonal tiles that are the building blocks of the facets of the microcompartment while the pentameric BMV are supposed to close the vertices. BMC are typically hexameric and perforated by a rather narrow pore that is believed to facilitate the transport of 1,2-proapndiol across the shell. The BMC shell proteins come in different versions, PduA/and its highly homologous twin, PduJ : the canonical representation of a shell protein most abundant shell proteins, assemble into hexamer and central pore………………………………………………..……………………...14

CHAPTER 3 – Split GFP as a Crystallization Scaffold

Figure 3.1. **Structure of split GFP in complex with sfCherry.** Adapted from Nguyen *et al.* 2013, Reproduced with permission of the International Union of Crystallography (http://journals.iucr.org/). (A) Crystal structure of split GFP 1-9 in complex with the sfCherry revealed the presence of a crystal artifact dimer interface along strands 10/11 of the GFP barrel. (B) The linkage between sfCherry and the GFP 1-9 core is a rigid linkage as seen but the backbone electron density (right), this is a result of a double pass linkage from strands 10/11 being interested in a loop of sfCherry. A terminal fusion would result in a single pass linkage……………………………43

Figure 3.2. **Multiple Oligomers Formed When Two Cysteines Are Present.** The presence of cysteine 70 in addition to the introduced cysteine mutants resulted in a predominant dimer species for all except the D117C mutants. In all cases higher order oligomers were also observed. It is likely that he dimeric species are contaminated heterodimers of the introduced mutant and Cys70 and Cys70 homodimers……………………………………………………………………44

Figure 3.3. **Size exclusion purification of GFP disulfide dimers.** (A) Size exclusion chromatography failed to fully resolve GFP disulfide dimers from monomers. (B) The major peak that corresponded to the molecular weight of a dimer (~50 kDa) also contained GFP monomers when fractions were analyzed by non-reducing SDS-PAGE. The absorbance at 280nm trace from the D190C SEC purification represents the typical resolution between dimer and monomer peaks when run on a s200 16/60 sizing column…………………………………………………....46

Figure 3.4. **The presence of cysteine 70 leads to heterogeneous oligomers.** (A) Cysteine 70 (red residue) was thought to be sufficiently buried in the core of the GFP barrel to be nonreactive. (B) Non-reducing SDS-PAGE gels indicated that while Cys70 was present, high-order oligomers were formed. Upon mutation to alanine the only bands present on non-reducing SDS-PAGE corresponded to monomers and dimers……………………………………………………………………….47

Figure 3.5. **Native PAGE of GFP metal binding mutants indicating native oligomer presence.** Native PAGE gels from assays of metal chelating mutants of GFP showed residual dimers in solution at higher concentrations (protein was at 10mg/mL when loaded onto gels). The faint high molecular weight bands in the "mono" lanes and the high molecular weight bands for the D21H/K26H mutant……………………………………………………………………….49

Figure 3.6. **Thiopropyl sepharose purification of GFP D190C.** (A) Thiopropyl sepharose was used in an attempt to remove unreacted cysteine monomers from the dimers instead of SEC. (B) When this resin was used on a mixture of D190C dimers and monomers, the monomers did not bind the column (FT lane). A minor band from unbound dimers was present in the wash and only monomers were present in the elution. When the dimers were mixed with the resin, the resin reacted with the disulfide bond, reducing the dimers to monomers that were bound to the resin……………………………………………………………………...50

Figure 3.7. **Expression test and non-reducing IMAX purification of GFP disulfide dimers in SHuffle and BL21 cell lines.** SHuffle and BL21 cell lines were used to compare expression levels of monomers and dimers. Both the D117C and D190C mutants were screened in both cell lines and gels with the most prominent overexpression are presented. For the SHuffle cell lines low protein expression resulted in poor purity of GFP (Elute & Reduced lanes) and no apparent over expressed bands in the crude or soluble fractions. The BL21 had clear overexpressed GFP monomers at 25 kDa in the soluble fraction. Upon elution from the Ni^{2+} resin in non-reducing conditions GFP dimers were formed (50 kDa band). The fraction of dimers was increased when treated with $CuSO_4$………..53

Figure 3.8. **Anion exchange chromatography separation of GFP D190C monomers and dimers.** GFP dimers and monomers could be resolved with anion exchange at pH 9.5. Two peaks were resolved when a gradient from 0M-1M NaCl was run. The first peak corresponded to monomers that were too dilute to be observed by SDS-PAGE. The second peak corresponded to pure dimers with no contaminating monomers. This was reproducible for all five disulfide dimers…………………54

Figure 3.9. **Expression Test of Disulfide Dimers Mutations on Split GFP 1-9 backbone.** (A) All five cysteine mutants were well expressed (T, total cell fraction lanes) but were completely insoluble (no 25 kDa band present in the S, soluble fraction lanes). (B) The original constructs to express GFP 1-9 with cysteine mutants contained a N-terminal hexa-histidine affinity tag and TEV cleavage site……...57

Figure 3.10. **Rapid-dilution Refolding of Split GFP 1-9 K26C Inclusion Bodies.** Rapid dilution refolding of was performed with varying starting amounts of inclusion bodies dissolved in 1mL of TNG buffer + 9M urea (D lanes). After refolding in to 25mL of TNG buffer and subsequent filtration to remove aggregates the majority of protein appeared soluble (R lanes). In each lane (D and R) the same amount of protein was loaded for direct comparison of protein remaining in solution after refolding. Concentration of refolded protein resulted in total precipitation of protein……………………………………………………………………58

Figure 3.11. **Rapid Dilution Refolding of split GFP K26C and Subsequent IMAC Purification.** IMAC was attempted on refolded GFP 1-9 K26C, this was used as a way to concentrate and purify the refolded protein. The majority of the protein failed to bind the Ni^{2+} resin (IMAC FT and wash lanes). The eluted protein was pure but predominately monomeric. The protein could be dimerized with $CuSO_4$. However, the protein failed to complement proteins containing the 10 -11 strands and could not be concentrated without precipitation…………………………….61

Figure 3.12. **SDS-PAGE On Column refolding of Split GFP K26C.** Denatured protein was bound the $Ni2^+$ IMAC resin in TNG + 9M urea buffer, protein was refolded by running a gradient to TMG + 0M urea. Initial runs contained 30mM imidazole in the buffers to prevent non-specific protein binding to the resin. When it was observed that the protein was eluted during the wash the experiment was repeated with no imidazole present in the binding or wash buffers. The GFP continues to leach off the column during the refolding gradient. The top gel was performed in reducing conditions while the bottom gel was non-reducing……………………62

Figure 3.13. **IEX Purification of Refolded Split GFP 1-9 K26C.** Refolded protein with no hexa-histidine tag could be purified and partially concentrated with anion exchange chromatography at pH 9.5. This resulted in more efficient protein binding and higher yields then IMAC purification but the protein failed to efficiently complement control proteins *in vitro*……………………………...64

Figure 3.14. **Dynamic Light Scattering Results from Freshly Refolded Split GFP 1-9 Mutants.** DLS was performed on freshly refolded and filtered mutants of GFP 1-9. All five mutants existed primarily as large soluble aggregates. Only K26C had any protein with a radius consistent with monomeric GFP in solution (red circle)………………………………………………………………………65

Figure 3.15. **XtalPred Results for *M. tuberculosis* Rv 2658c.** The sequence for the MtbCon protein used as a complementation control is predicted to be 22% disordered (underlined residues). Although this protein efficiently complemented split GFP 1-9, the percentage of disorder and low complexity make it a poor crystallization target……………………………………………………………………..68

Figure 3.16. **Complementation Efficiency of MtbCon with GFP 1-9 Mutants.** MtbCon was able to complement all five mutants of GFP 1-9. The D102C mutant was the least efficient with only 40% of the fluorescent intensity of GFP 1-9 with no mutations. Introduction of the C48A/C70A backbone mutations reduced the efficiency to 80% of GFP 1-9 alone. This is roughly equivalent to 10% of the total refolded protein being viable for complementation (~8mg from 75mg of denatured protein)……………………………………………………………..69

Figure 3.17. **GFP 1-9 and Crystallization Target Constructs.** Original constructs were intended to be expressed and purified separately then mixed *in vitro* once the GFP core was dimerized. Once it was concluded that the N-terminal hexa-histidine tag and TEV cleavage site could interfere with complementation, new constructs for coexpression and purification were made. The GFP 1-9 has no affinity tags and contains the C48A/C70A backbone and applicable oligomer mutations. The crystallization construct features the N-terminal hexa-histidine tag and TEV cleavage site for complex purification and the internal 10/11 strand sequence for complementation. Successful complementation would result in a green color of the cell pellet. To allow for coexpression the crystallization target were cloned into a pET vector and GFP 1-9 was cloned into a pACYC vector………………70

Figure 3.18. **SEC Purification of Dimerized GFP-H3L complexes.** Three peaks observed for all GFP-H3L SEC runs. The first minor peak corresponded to the void volume of the column and contained aggregated protein. The second peak was green in color (indication of complex formation) and the third major peak was H3L alone. The non-reducing SDS-PAGE gels are presented showing the prominent band from H3L overexpression contaminating the complex after IMAC. The fractions from the complex peak showed the presence of a ~ 50 kDa band corresponding to the GFP 1-9 dimer core and a ~25 kDa band for the H3L…………………………...72

Figure 3.19. **Native PAGE Gel from SEC Purified GFP-H3L Complex.** Native gels analyzed the fractions pooled from the SEC runs for complex purity. For both mutants the complex was contaminated with monomeric H3L and a proteolysis production from H3L. The complexes were from thawed protein that was flash frozen immediately after purification. The H3L lanes is protein that was left at 4°C for one week and was significantly degraded by a contaminating protease……………………………………………………………………..73

Figure 3.20. **Anion Exchange Chromatograms and Native PAGE Gels of GFP 1-9 – H3L Complexes.** IEX of the GFP-H3L complexes resulted in a minor peak for the complex and a major peak for H3L monomers. The complex peak contained contaminating H3L and monomeric complex……………………………………75

Figure 3.21. **SEC After Anion Exchange of GFP 1-9 –H3L Dimer Complexes.** SEC of the complex peak from IEX was able to fully resolve the GFP-H3L dimer complex from contaminating monomeric complex and monomeric H3L. Native PAGE gels showed smeared band for the SEC complex peak, smearing of the band is likely due to overloading the gel. With the addition of β-mercaptoethanol the dimer complex was partially reduced to the monomeric state…………………………..76

Figure 3.22. **Crystals of GFP D117C – H3L Dimers.** After approximately seven months crystals of GFP-H3L dimers from a D117C mutation were obtained in a single condition. These this triangle plate crystals of 50-75µm diffracted weakly to ~10Å at the Advanced Photon Source synchrotron……………………………………77

Figure 3.23. **Crystals of GFP D21H/K26C - StarD9 With Cu^{2+}.** Crystals grew in approximately three months. Cu^{2+} ions were added immediately prior to crystallization experiments in order to induce oligomer formation. These small rod crystals (~2µm x20 µm) did not diffract……………………………………..78

Figure 3.24. **Size Exclusion Chromatography Fractions of sfCherry I206K – GFP 1-9 K26C Complex.** Fractions that corresponded the complex peak indicated bands for intact sfCherry/10/11 I206K (black box), GFP 1-9 K26C (green box) and light induced cleavage products from sfCherry (red box). Presence of the cleavage products made it difficult to obtain intact complexes. For each mutant the fractions containing the least contamination were pooled for crystallization experiments……………………………………………………………………80

CHAPTER 4 – Computationally Designed Trimeric GFPs

Figure 4.1. **Docking of GFP Monomers to a Three-Fold Symmetry Axis.** Initial design strategy focused on docking the GFP barrels perpendicular to the symmetry axis (triangle). The GFP barrels start away from the axis and are allowed to dock together towards the axis. If the barrels dock in an energetically favorable manner, interfaces are designed. The starting GFP barrels are then rotated 5° and docked again in the new orientation…………………………………………………...91

Figure 4.2. **The Four Orientations of Perpendicular c3 Trimers.** From the ten individual computational docking runs, each run generated one trimer design. Four general classes of trimer designs were found to be energetically favorable in from these 10 runs. The design numbered by occurrence (Design 1 most common, Design 4 the least from the 10 runs). Five sequences from each design when then selected for experimental characterization……………………………………………..93

Figure 4.3. **Expression Tests and Purification of Designs.** (A) SDS-PAGE gels of the soluble from expression of the four designs. The 25 kDa bands correspond to the molecular weight of the GFP barrels. Designs that resulted in soluble, properly folded protein. The final two lanes were point mutations that emerged during cloning and did not result in properly folded protein. (B) All four SEC elution profiles from the purifications featured a single peak that eluted at the volume consistent with monomers. (C) Native-PAGE gel of the four SEC purified proteins compared to known monomers and dimers of GFP. Sequence variations caused the proteins to run at different rates, but all are putatively monomeric. Design c3_14 (1-4) was partially aggregated, as the protein did not enter the gel (red circle)…………………………………………………………………..94

Figure 4.4. **Crystals of c3 Designs.** The crystals observed from the c3_31 (3-1) and c3_33 (3-3) designs, the only two that crystallized. These represent all the crystal morphologies observed………………………………………………………..95

Figure 4.5. **The c3_33 JCSG D12 Dimer Compared to Design.** The dimeric structures varied significantly form the intended design with none of the designed contacts preserved……………………………………………………………………97

Figure 4.6. **Observed Dimer is a New GFP Crystallographic Artifact Dimer.** The observed dimer is different from the traditionally observed "native" GFP crystallographic dimer. The previously observed dimer is a c2 symmetric dimer with the interface primarily along strand 11. Strand 11 has been colored in red for both dimers to highlight the differences in the structures………………………..98

Figure 4.7. **General Design Orientation of Parallel Trimers.** With the failure of the perpendicular dimers, new designs were based on trimers with the GFP barrels parallel to the symmetry axis. With these designs, the interface would be larger and no rely on any interactions from the loops on the end of the barrels. However, all the termini would be on the same end of the bundled trimer. All the termini in the same location have the potential to create steric hindrance for fusion proteins………………………………………………………………………...99

Figure 4.8. **C3_52 SEC Elution Profile and Crystal.** (A) The protein was run on an analytical sizing column resulting in poor resolution between the void peak and trimer peak. Only fractions of the trimer peak were used for crystallization. (B) The looped crystal used for data collection had cubic morphology. The crystal is colorless due to the lighting when the picture was taken. The crystal was faintly green in the crystallization drop due the acid conditions (pH 5.5)……………..101

Figure 4.9. **Comparison of the c3_51 Design and Structure.** The Rosetta designed structure features all the barrels parallel to each other and the symmetry axis. In the crystal structure, the barrels were rotated approximately 17° from the design. The residues that were mutated for the design are highlighted in yellow. These residues formed the trimer interface for the crystal structure, but none of the specifically designed interactions were retained………………………………..102

CHAPTER 5 – Coiled-Coil Fusions to T4-Lysozyme Oligomers

Figure 5.1. **Model of the Coiled-Coil Pair Fused to T4L.** The x-ray crystal structure of a highly stable version of T4L (PDB ID 3FA0, cyan) was aligned with the c-Fos (magenta)/c-Jun (purple) structure. The model that indicated no steric hindrance is presented. Here, the c-Jun sequence is to be fused to the N-terminus of T4L, the crystallization target would be fused to the N-terminal end of the c-Fos coil (red box). This coiled-coil model is from the native c-Fos/c-Jun sequence; the engineered heterodimeric pair for expression constructs replaces these sequences for expression constructs. Mutations are highlighted for disulfide dimers (yellow) and metal chelating sites (orange)……………………………………………118

Figure 5.2. **Homology Model and Secondary Sequence Prediction of IDP01024.** The homology model of IDP01024, *Salmonella enterica* anti-sigma28 factor FlgM used for construct construction is presented; the predicted C-terminal helix where the coiled-coil is fused is highlighted in red……………………………….………...121

Figure 5.3. **Homology Model and Secondary Sequence Prediction of IDP01204.** The homology model of IDP01204, *Bacillus anthracis str. Ames* glycosyl transferase, group 2 family protein used for construct construction is presented; the predicted C-terminal helix where the coiled-coil is fused is highlighted in red. This protein was previously found to crystallize through CSGID efforts but failed to diffract…………………………………………………………………..122

Figure 5.4. **Homology Model and Secondary Sequence Prediction of IDP90101.** The homology model of IDP90101, *Salmonella enterica* sseL deubiquitinase used for construct construction is presented; the predicted C-terminal helix where the coiled-coil is fused is highlighted in red……………………………………..…123

Figure 5.5. **Size Exclusion Elution of IDP01204 Co-purified with T4L 40H/44H.** SEC was performed on the co-purified proteins after TEV cleavage to remove the HexaHis affinity tag. The two prominent peaks correspond to the T4L-IDP01204 complex and un-complexed T4L monomers. Fractions from the complex peak with equimolar amounts of the two proteins were pooled and used fro crystallization experiments…………………………………………………………………..127

Figure 5.6. **Size Exclusion Elution of IDP01204 Co-purified with T4L 125C/128C.** SEC was performed on the co-purified proteins after TEV cleavage to remove the HexaHis affinity tag. The three prominent peaks correspond to aggregated protein, T4L-IDP01204 complex peak and un-complexed T4L monomers. Fractions from the complex peak with equimolar amounts of the two proteins were pooled and used fro crystallization experiments. Due to the significant aggregation of this construct, care was taken to choose fractions of the complex furthest from the void…………………………………….…..…………128

CHAPTER 6 – Computationally Designed Cages as Scaffolds

Figure 6.1. **Structure of T3-10 and Fusion Construct Designs.** (A) The crystal structure of the tetrahedral cage T3-10 (PDB 4EGG). T3-10 originally crystallized in space group C2 with the tetrahedral 2-fold symmetry axis replicated with the 2-fold crystallographic symmetry axis. (B) General design scheme for T3-10 fusions, linker sequences could be varied between the T3-10 and target genes. A C-terminal hexa-histidine tag would be used for purification………………...…..139

Figure 6.2. **Experimental Characterization of T3G Fusion in Solution.** (A) SDS-PAGE gel of IMAC purified T3G resulted in highly pure protein (red circle) subsequent SEC purification further purified the protein and gave initial indications the tetrahedral cage was still formed in solution. (B) SEC-MALS was used to validate the size of the cage in solution. The cage was predicted to have a mass of 577 kDa; the single peak from SEC had a calculated mass of 565±4 kDa indicating a fully formed cage in solution. A proposed model of the T3G cage is also presented. (C) DLS was used to determine the changes in particle size in a concentration dependent manner. As the protein concentration increased the radius of the particles also increased, an indication of aggregation……………141

Figure 6.3. **Optimization of the T3G Crystallization Construct, T3G-CTD.** (A) The crystallographic dimer of sfGFP (PDB 2B3P). This dimer is mediated by a specific interaction between Val206 from each chain. The mutation, V206K is intended to prevent this dimer from forming in new T3G constructs. (B) The structure of T3-10 (PDB 4EGG) found no electron density for the C-terminal residues highlighted in red. Lack of density indicates disorder, resulting in a highly flexible linkage between the two proteins. These disordered residues were eliminated from new constructs resulting in T3-10 C-terminal deletion (CTD). (C) Native PAGE gel of both T3G Full Length (FL) and T3G CTD. T3G CTD ran as a single band while T3G FL had multiple bands present……………………..142

Figure 6.4. **Crystals of T3G CTD.** Crystals of T3G CTD rapidly grew as small rods in one condition from the screen JCSG+ (0.1M Na Cacodylate pH 6.5, 40% MPD, 5% Peg 8000). Crystals diffracted to approximately 18-20Å at the Advanced Photon Source and could not be optimized……………………………………………..143

Figure 6.5. **Expression Tests of T3-10 CTD StarD9.** (A) T3-10 CTD was genetically fused to StarD9 with a GSGTGSG linker and a C-terminal histidine affinity tag from IMAC purification. (B) IPTG induction for 4 hours at 37°C resulted in a poorly expressed band after IMAC purification with a 70kDa protein contaminating the pure protein that could not be further purified due to protein loss during SEC. (C) Auto-induction for 72 hours at 20°C produced protein that could be purified by SEC with some fractions contaminated by a smaller molecular weight protein. The most homogenous fractions (red oval) were pooled for crystallization experiments………………………………………………………………..146

Figure 6.6. **Figure 6.6. Successive SEC Purifications of T3-10 CTD StarD9.** (A) SEC immediately after IMAC resulted in significant protein contamination of the T3-10-StarD9 cage (blue trace). A second SEC purification of the peak containing the cage fusion resulted in a single SEC peak of pure protein (red trace). (B) Native-PAGE of the fraction pooled from the second SEC purification. The protein ran as one distinct band in the resolving layer of the gel. The upper band is protein that remained in the stacking layer of the gel; this had also been observed with the T3G constructs on native gels……………………………………………………147

Figure 6.7. **Crystal of T3-10 CTD Full-Length StarD9 Fusions.** (A) Thin plate crystals grew in conditions based on screening condition JCSG+ G2 (0.02M $MgCl_2$, 0.1M Hepes pH 7.3-7.5, 20-24% polyacrylic acid 5100). Diffraction quality crystals were not obtained in any of the conditions and the crystals could not be improved through optimization. (B) Rod like crystals grew in conditions based upon the hits from the Pact screen in condition E9 (0.15M K/Na tartrate, 18-22% Peg 3350, 0.1M hepes pH 7.2-7.6). These crystals weakly diffracted at low resolution and were extensively optimized…………………………………………………..148

Figure 6.8. **T3-10 CTD StarD9 Cage Fusion Purified with AMP-PNP.** (A) In an attempt the improve crystal diffraction the fusion cage was purified in the presence of AMP-PNP instead of ATP. SEC fractions highlighted by the red box were pooled for crystallization experiments. (B) In addition to the previously observed rod crystals, new trigonal crystals that only grew with AMP-PNP were observed…..150

Figure 6.9. **Oil Overlay Optimized Crystals of T3-10 CTD StarD9.** The rod crystal form from Pact E9 conditions was only grew to larger sizes when vapor diffusion was slowed. To slow diffusion, an overly of 1:1 silicone:paraffin oil was applied to the top of the well solution……………………………………………………………151

Figure 6.10. **The Three Distinct Crystal Forms of T3-10 CTD StarD9.** Three unique crystal forms were found, all grew in conditions based on Pact E9 with only variations to K/Na tartrate, Peg 3350 or pH changes. None of the crystals resulted in data sets that could be solved by molecular replacement……………………………...152

Figure 6.11. **XtalPred Results for the Kinesin Domain of StarD9.** The kinesin domain of StarD9 features a large disordered loop (red box). This loop is absent from other kinesin domains including the closest homolog of StarD9, Kif1A. This loop was removed and replaced with a 'GSG' linker. This construct crystallized in the same space group as the full-length protein but diffracted to a higher resolution………………………………………………………………..153

Figure 6.12. **Diffraction Image from T3-10 CTD StarD9 Loop Deletion Construct.** Diffraction from the optimized construct with optimized crystal growth conditions diffracted with strong, well-defined reflections to 6.6Å. Full data sets were obtained from several isomorphous crystals to similar resolutions…………….154

Figure 6.13. **Designed Cages Have Internal Cavities Amenable to Internal Fusions.** The structure of the two-component cage T33-28 features an internal cavity with a diameter of 70Å, sufficient to accommodate a small crystallization target. Since the cage is composed of two separate protein components, not every subunit requires a fusion to the crystallization target reducing the possibility of steric clashes. Since the crystallization target is internal to the cage the crystal contacts on the exterior of the cage remain unchanged allowing the fusion version of the cage to crystallize in the same manner as the apo protein……………………...157

CHAPTER 7 – Oligomeric Enzyme Based Materials

Figure 7.1. **Amorphous Proteins Gels Formed Through Oligomer Linkages.** (A) Connections of an oligomeric enzyme (grey) through a flexible protein linker (green) would create an amorphous gel-like material. In this example the linkage would be through GFP disulfide dimers. Flexibility comes from both a flexible sequence connecting the two proteins and the rotation of the disulfide bond allowing multiple conformations. (B) The crystal structure of the trimeric bromperoxidase 2A (PDB 1BRO). This protein lacks native cysteines and contains a permissive loop for insertion of GFP strands 10 & 11, making this an ideal candidate for material formation via split GFP dimers…………………...164

Figure 7.2. **Formation of Large Bromperoxidase – GFP Oligomers.** (A) Non-reducing SDS-PAGE gels of purified bromperoxidase-GFP complexes that have been reduced and oxidized. When oxidized the ~25 kDa monomeric GFP band is converted to the higher molecular weight (~45 – 50 kDa) band indicating dimer formation. The major band in each lane at 30 kDa corresponds to the bromperoxidase. For each complex there was incomplete complementation as seen by the drastic difference in band intensity between the two proteins. (B) Native-PAGE gel of the bromperoxidase-GFP complexes that have been reduced and oxidized. The three bands in the reduced (+DTT) lanes correspond to bromperoxidase trimers with differing numbers of complexed GFP. When the disulfide bonds are oxidized, large complexes are formed that fail to enter the resolving layer of the pre-cast gel……………………………………………165

Figure 7.3. **Monochlorodimedone Bromperoxidase Assay and pH Optimization.** (A) Bromperoxidase activity can be assayed by monitoring the conversion of monochlorodimedone to the brominated form with a corresponding loss of absorbance at 290nm. (B) Screening of pH variations indicated that the bromperoxidase-GFP fusions were most active at pH 4.75 or 5.75. At pH 4.75 protein precipitation was observed leading to pH 5.75 to be used for all future assays…………………………………………………………………...166

Figure 7.4. **Bromperoxidase-GFP Activity Assays.** Assays for each complex was performed at pH 5.75 in triplicate and compared to a control with no enzyme present as monochlorodimedone can be brominated when bromine ion are present. Each assay was performed with protein from SEC purified protein was reduced during the purification. Additional DTT was added to fully reduce any disulfides that may have inadvertently formed; Cu^{2+} was added to oxidize the disulfides. For each complex the oxidized versions had reduced activity after four hours. After a full 24 hours every protein sample fully brominated the monochlorodimedone. Error bars represent the standard deviation of the three experiments…………..167

Figure 7.5. **Crystal Structures of the Four Bph Enzymes.** (A) Structure of BphA from *Burkholderia* (PDB 2XSH) a hexamer with C3 symmetry. (B) Structure of BphB from *Pseudomonas* (PDB IBDB) a tetramer with D2 symmetry. (C) Structure of BphC from *Burkholderia* (PDB 1LGT) an octamers with D4 symmetry. (D) Structure of BphD from *Burkholderia* (PDB 2OG1) a tetramer with D2 symmetry…………………………………………………………………..169

Figure 7.6. **Model of the BphA-BphB Coiled-Coil Pair.** Model of the coiled-coil linkage between monomers of BphA (green) and BphB (purple). The N-terminal coiled-coil helix is indicated in red and the corresponding C-terminal helix is modeled in blue……………………………………………………………………...…170

Figure 7.7. **Model of the BphC-BphD Coiled-Coil Pair.** Model of the coiled-coil linkage between monomers of BphC (orange) and BphD (green). The N-terminal coiled-coil helix is indicated in red and the corresponding C-terminal helix is modeled in blue……………………………………………………………………...171

Figure 7.8. **SDS-PAGE Gel from Bph Coiled-Coil Fusion Expression Testing.** Coiled-coil fusions of BphA, BphB and BphC were overexpressed in high levels (red box) in the total cell lysate (T lanes) but totally absent from the soluble fraction (S lanes). Lack of an overexpression band for BphD suggests that the construct failed to express……………………………………………………………………..172

Figure 7.9. **Structure of Bromoperoxidase Active Site and Substrate Channel.** The crystal structure of one subunit from the bromperoxidase trimer used in the protein gel experiments (PDB 1BRO). The active site is highlighted in red; the substrate channel is indicated with the black arrow. The loop where the 10/11 GFP strands were inserted to mediate the linkage to split GFP is highlighted in blue. This loop is connected to a helix (green arrow) perturbation of this helix could partially occlude the substrate channel, reducing the apparent activity of the enzyme. Insertion of the GFP strands into the bromperoxidase sequence may have moved the helix but the effect was most apparent when the GFP cysteines were oxidized to form dimers, changing the oligomeric state of the complex in solution and potentially perturbing the helix……………………………………………………174

CHAPTER 8 – Circular Permutations of PduA

Figure 8.1. **Slight Deviation of The PduA Terminal Helix Drastically Alters Geometry.** (A) PduAp4 forms a pentamer instead of the intended hexamer. (B) A slight deviation in the designed loop to allow the circular permutation resulted in a shift (blue) of the PduA wild type terminal helix (green). This shift results in a more compact monomer that can pack as a pentamer……………………………......189

Figure 8.2. **Size Exclusion Chromatography Elution Profile of PduAp4.** The SEC elution profile of IMAC purified PduAp4 resulted in to peaks. The first peak corresponded to the void volume of the s200 superdex column. It was first thought this peak contained only protein aggregates. After the cage structure was solved it was determined that this peak contains the icosahedral cage. The main peak corresponded to the pentamer subunits of the cage. Dilution of the protein during SEC run caused disruption of the cage………………………………….190

Figure 8.3. **Dynamic Light Scattering Data of PduAp4 Buffer Screening.** DLS was utilized to rapidly screen for conditions that readily formed the cage. In this experiment, it was determined that the cage was most abundant at pH 9.0 and 250mM NaCl (particles with a radius of ~10nm). The cage was also present as a minor species in a buffer of pH 8.0 and 50mM NaCl. It was later determined that the presence of aggregates in samples obscured the true distribution of particles in solution. Subsequent experiments concluded the cage was the predominant species at pH 8.0 and 50mM NaCl (chapter 8.1). All other buffer conditions resulted in dissociation of the cage into smaller subunits (pentamers or monomers)…………………………………………………………...191

Figure 8.4. **Energy Scores From Asp43/Lys76 PduAp4 Mutants.** The global (blue) and interface (red) energy scores of each mutant were compared to the scores from the PduAp4 cage (dashed lines). Mutants that had lower energies then the PduAp4 sequence were thought to form a more stable assembly………………………..193

Figure 8.5. **Ion Exchange Chromatography Purification of Two PduAp4 Mutants.** IEX was employed for the purification of the PduAp constructs with no affinity tag. Cells were lysed in a buffer at pH 9.0 with only 50mM NaCl. After the cells were lysed and clarified the soluble fraction was diluted 2x in a buffer lacking NaCl. This was then applied to a Q Sepharose cation exchange to remove contaminating *E. coli* proteins (Q Elute). Flow through from this column was then immediately applied to S Sepharose anion column and washed to remove unbound protein. PduAp mutants were eluted off with NaCl (S Elute 1, 0.5M NaCl; S Elute 2 1.0M NaCl)……………………………………………………………………194

Figure 8.6. **PduAp4 Mutants Are Purified as Monomers.** IEX purified PduAp4 mutants were further purified on a Superos6 SEC column intended for the purification of large complexes. All PduAp4 mutants (red trace) eluted as a significantly smaller protein then the known cage T3G (blue trace). Subsequent analysis of the mutants on an s200 SEC column indicated the protein was monomeric, eluting at the volume expected of a 10 kDa protein…………………………………………..195

Figure 8.7. **Crystal Structure of PduAp7.** (A) The experimental crystal structure of PduAp7 (blue) formed the designed structure (green), with a backbone RMSD of only 0.9Å. (B) The PduAp7 structure is nearly identical to wild type PduA (magenta) when two structures are aligned. This suggests PduAp7 can be used to replace PduA in the Pdu shell. Once this is achieved the pore may be fully occluded to study the role of the other pores in the shell or allow external fusion of proteins to the external surface of the shell via the external N-terminus of PduAp7……………………………………………………………………197

Figure 8.8. **Soluble Expression Testing of PduAp7 – PduU Barrel Fusions.** Soluble fractions from the first 18 of 19 PduAp7 – PduU Barrel Fusions, construct 19 was screened at a later date and was insoluble. Only one design, P7U-7 (red arrow) was soluble. The 14.5 kDa band present in all lanes is chicken egg white lysozyme added during cell lysis…………………………………………….198

Figure 8.9. **Purification of P7U-7.** (A) S200 SEC elution of IMAC purified and TEV cleaved P7U-7. The major peak of the purification corresponded to a predicted molecular weight of 65 kDa. The leading fractions of the peak were pooled for crystallization experiments. (B) SDS-PAGE gel of the P7U-7 purification. A double band was apparent in the IMAC purified protein, the lower band increased and the upper band decreased upon TEV cleavage. The IMAC column used for this purification was previously used for another attempted purification and it is likely the TEV protease from that purification was not fully eluted from the resin prior to this experiment. Incomplete TEV cleave and removal of the affinity tag suggests a buried tag or TEV site for a fraction of the hexamer subunits……...199

Figure 8.10. **Crystals of P7U-7.** Crystals were obtained after one round of streak seeding (original drop: seeds crushed and diluted 4x with reservoir solution). Reservoir solution contains 0.15M Ca(OAc)2, 18% PEG-1000 and 0.1M Immidazole, pH 8.0. Drops set up with 10mg/mL PduA-P7, 2.25µL total volume drops, 2:1 (sample:crystallant). Well conditions were the same as those from the original well/drop used to seed this 24-well plate. These crystals will be screened for diffraction in August 2016…………………………..201

CHAPTER 9 – Structural Studies of the PduA – PduD Helix Interaction

Figure 9.1. **Rosetta Opimized Fusion Model of a PduA-PduD N-terminal Fusion.** (A) The monomer of the PduAD model has the PduD tail located above the PduA C-terminal tail (green box). The structure of the PduA BMC domain is not altered. (B) When the PduA hexamer forms the PduD tail is predicted to sit on the interface of two monomers (red arrow)…………………..…………………….218

Figure 9.2. **Hexametric Model of the PduAD Rosetta Design.** (A) In the computational model sixe PduAD fusions could pack as a hexamer with the PduD tails pack tightly against the PduA C-terminal helices (green box). (B) Viewed down the six-fold axis the PduAD fusions do not disrupt the hexamer or central pore of PduA………………………………………………………………….......219

Figure 9.3. **PduAD Construct Designs.** (A) PduAD featured the first 90 residues, a variable linker and PduD residues 2-16. (B) A turn (green box) is required to set the correct geometry to allow the tail fold on to the PduA putative binding site. (C) Four versions of the loop were attempted, PduAD1 is the Rosetta optimized sequence, PduAD2 is a long flexible linker, PduAD3 and PduAD4 are sequences with a high propensity to form beta turns……………………………………...220

Figure 9.4. **Crystal Structure of the PduAD3 P1 Crystal Form.** (A) Crystals in the P1 crystal form grew as well-defined hexagons. (B) The crystal structure was found to have layers of the PduA hexamers with a large space between the layers. Although the PpduAD model would full this space, no density was observed besides that of the PduA BMC core…………………………………………….222

Figure 9.5. **Crystal Structure of the PduAD3 $P4_322$ Crystal Form.** (A) Crystals in the $P4_322$ were found to be large rods. (B) A single hexamer was in the asymmetric unit (central hexamer) with six hexamers forming crystal contacts, one on each edge of the central hexamer. (C) PduAD model is placed into the density of the structure. No electron density is present for the PduD tail helix……………………………………223

LIST OF TABLES

CHAPTER 4 – Computationally Designed Trimeric GFPs

Table 4.1 Expressed Sequences of GFP c3 Perpendicular Designs………….………..108

Table 4.2 Rosetta Values Used to Select Table 4.1 Sequences………………...……….111

Table 4.3 Sequences of GFP c3 Parallel Designs………………………,…..………112

Table 4.4 Data and Refinement Statistic for c3_51………………..………...………115

CHAPTER 5 – Coiled-Coil Fusions to T4-Lysozyme Oligomers

Table 5.1. Sequences of T4-Lysozyme Coiled-Coil Oligomer Constructs……………..135

Table 5.2. Crystallization Targets Derived From CSGID Database……………..……..136

CHAPTER 7 – Oligomeric Enzyme Based Materials

Table 7.1. Bph Enzyme Coiled-Coil Expression Constructs………………………………179

CHAPTER 8 – Circular Permutations of PduA

Table 8.1. Sequences of Rosetta Designed Permuted PduA Icosahedral Constructs…………………………………………………………..…..…208

Table 8.2. Sequences of PduAp4 Icosahedral Constructs From BeAtMuSiC Predictions…………………………………………………………………..209

Table 8.3. Sequences of Rosetta Designed PduAp4 Derived Hexamers……………….212

Table 8.4. Data Collection and Refinement Statistics for the PduAp7 Structure…….213

Table 8.5. Sequences of PduAp7-PduU Barrel Fusions…………………………….....214

CHAPTER 9 – Structural Studies of the PduA – PduD Helix Interaction

Table 9.1. Sequences of PduAD Fusion Constructs………………………………………231

ACKNOWLEDGMENTS

Graduate school is a long, stressful, trying and rewarding journey, one that cannot be accomplished alone. Only with the support and encouragements of many others can one successfully obtain a PhD. Foremost I would like to thank Rachel So, the endless support and encouragement of my partner has allowed me to get this far. Support from my parents; Judy and Jim, Rachel's family; Sue, Sarah, Treasa and Victor, has been invaluable the past five years.

My professional and scientific success is a direct result of my many mentors at both UCLA and University of Washington. I cannot thank my advisor, Todd Yeates, enough for his mentorship and support. Prof. Yeates has made me a better scientist and person through his mentorship. His support of the independent pursuits of his students is something that is rare in many labs. It immensely helps turn graduate students into scientists. I would like to thank my committee members, Prof. Joe Loo, Prof. Jim Bowie, Prof. Pascal Egea and Prof. Kent Hill for their support and suggestions. I would also like to extend a special tank you to Prof. David Eisenberg, his suggests and questions during joint group meetings and journal clubs have challenged me and in turn made me a much better critical thinker. All students in Yeates' group have the unique opportunity for Prof. Eisenberg to serve as a quasi-advisor to them.

All the members of the Yeates' group have helped to create an intellectually stimulating and enjoyable environment to work in. Specifically I would thank Dr. Dan McNamara, Dr. Julien Jorda, Dr. Michael Thompson, Josh Laniado and Diana Garnica for all of their help throughout the past five years. Finally, I cannot thank Dr. Duilio Cascio and Dr. Michael Sawaya enough for all of the knowledge of X-ray crystallography they have passed along to me, and allowing me the opportunity to frequently travel to the Advanced Photon Source for data collection on numerous occasions.

The decision to pursue a career in science and attend graduate school can only come from the influence and encouragement from numerous mentors throughout one's life. For myself my high science teachers Linda Sorenson and Jeff Dineen were the most influential for my career as a scientist. The decision to attend graduate school is a direct result from my time at the University of Washington; and I would like to thank all of my co-workers from my time there. If it weren't for the influence and support of Prof Wesley Van Voorhis, Dr. Alberto Napuli, Dr. Eric Larson, Lynn Barrett and Steve Nakazawa Hewitt there is no way I would have ever considered attending graduate school. Without the help of my undergraduate assistant, Mary Nguyen I would not have accomplished all that I did in Seattle. Finally I would like to that Dr. Eric Larson, a crystallography whom I work closely with and encouraged me to attend UCLA and work with Prof. Yeates.

For three years at UCLA, I received funding from the UCLA--NIH Cellular and Molecular Biology Training Program (Ruth L. Kirschstein National Research Service AwardGM007185). I would like to thank Prof. Steve Clarke, Christine Briganti, and Jon Lowenson who were responsible for the program while I was a fellow.

Previous Publications and Contributions of Co-Authors:

Figures 1.1 and 1.5 have been adapted with permission from Banatao *et al*, 2006 (DOI 10.1073/pnas.0607674103, United States National Academy of Sciences, no license required) and Jorda *et al* 2006 (DOI 10.1002/pro.2196, John Wiley and Sons, license number 3877860137538) respectively.

Chapter 2 is a manuscript that has been previously published (DOI 10.1016/j.str.2015.07.008) and has been reprinted with permission from Elsevier (license 3877321093784). Dr. Mark Arbing performed the construct cloning; Inna Pashkov and Dr.

Michael Sawaya determined the structure of PBD ID 4W6B. Natasha Devore evaluated the mutants for their ability to complement, Dr. Geoffrey Waldo developed the original split GFP sequence. Dr. Thomas Terwilliger oversaw the work as part of a larger crystallization technology collaboration.

Chapter 8.1 is a manuscript previously published by the Royal Society of Chemistry (DOI 10.1039/C6CC00851H) and has been reproduced with permission (no license required). Dr. Julien Jorda computationally designed the initial constructs; Dr. Michael Thompson performed initial cloning efforts for the constructs.

Professor Todd Yeates directed all research presented in this dissertation.

VITA

2009	B.S. in Microbiology, University of Washington, Seattle
2009 – 2011	Research Assistant, Van Voorhis Group, University of Washington
2011 – 2013	University Fellowship, Graduate Division, UCLA
2011 – 2012	Teaching Assistant, Department of Chemistry and Biochemistry, UCLA
2011–12, 2014 –15	Edwin G. Pauley Graduate Student Fellowship, UCLA
2013 – 2016	UCLA - NIH Cellular and Molecular Biology Training Fellowship
2016	UCLA Biochemistry Dissertation Award

SELECT PUBLICATIONS AND PRESENTATIONS

J Jorda, **D Leibly**, M Thompson, T Yeates. Structure of a novel 13nm dodecahedral nanocage assembled from a redesigned bacterial micrcompartment shell protein. *Chem Comm* 52, 5041-44 (2016)

DJ Leibly, MA Arbing, I Pashkov, N DeVore, GS Waldo, TC Terwilliger and TO Yeates. A suite of engineered GFP molecules for oligomeric scaffolding. *Structure* 23(9). 1754-1768. (2015)

DE McNamara, JK Laniado, **DJ Leibly**, NP King, JB Bale, W Sheffler, TP Korman, JU Bowie, D Baker, TO Yeates. (2015). Self-assembly of multi-component symmetric nanomaterial based on fusion strategies.11th Annual Protein Engineering Summit, Boston, MA, USA May 2015

MC Thompson, D Cascio **DJ Leibly** and TO Yeates. An allosteric model for control of pore opening by substrate binding in the EutL microcompartment shell protein. *Protein Science*. 10.1002/pro.2672 (2015)

L Jiang, C Liu, **D Leibly**, M Landau, M Zhao, MP Hughes and D Eisenberg. Structure based discovery of fiber-binding compounds that reduce the cytotoxicity of amyloid beta. *eLife*, 2:e00857. (2013)

S Shen, **DJ Leibly**, RR Ogorzalek Loo, JA Loo (2012). Extensive protein multiple charging with 2 -Nitrophloroglucinol for MALDI Time-of-Flight mass spectrometry. American Society of Mass Spectrometry Annual Meeting, Vancouver BC, Canada, May 2012.

DJ Leibly, TN Nguyen, LK Kao, SN Hewitt, LK Barrett and WC Van Voorhis. Stabilizing additives to aid in the solubilization of recombinant proteins during cell lysis. *PloS One* 7(12): e52482. (2012)

KK Ojo, C Pfander, NR Mueller, C Burstroem, ET Larson, CM Bryan, AM Fox, MC Reid, SM Johnson, RC Murphy, M Kennedy, H Mann, **DJ Leibly**, SN Hewitt, CL Verlinde, S Kappe, EA Merritt, DJ Maly, O Billker and WC Van Voorhis. Transmission of malaria to mosquitoes blocked by bumped kinase inhibitors. *J Clin Invest*. 122(6): 2301 (2012)

DJ Leibly, J Abendroth, CM Bryan, B Sankaran, A Kelley, LK Barrett, L Stewart and WC Van Voorhis The Crystal Structure of the Thymidylate Kinase from *Erlichia chaffeensis. Acta Cryst F*. F67 (Pt. 9): 1090-4. (2011)

DJ Leibly, PA Newling, J Abendroth, W Guo, A Kelley, LJ Stewart and W Van Voorhis. Crystal Structure of the Cyclin-Dependent Kinase of *Giardia lamblia. Acta Cryst F*. F67 (Pt. 9): 1084-9. (2011)

Chapter 1

Introduction

1.1 Synthetic Symmetrization as a Crystallization Tool

X-ray crystallography has established itself as the leading method for determining the three-dimensional structure of macromolecules. With the elucidation of a protein's structure, one gains insight into the catalytic mechanism of an enzyme, cellular structures and new drug targets. X-ray crystallography has undergone dramatic advancements in the technologies used to determine macromolecular crystal structures since the earliest experiments, which were performed over 100 years ago. Today, a scientist with a protein crystal and minimal training can solve a structure in an afternoon, a feat once unimaginable. As structure solution technologies have advanced to where a novice can successfully solve novel crystal structures, technologies to improve the outcome of the crystallization experiment lag behind. Still it is unknown why certain proteins crystallize while others do not. Protein crystallization remains the bottleneck of x-ray crystallography[1,2]. Once the trivial issues are addressed (protein purity, excision of poorly folded sequences and broad condition screening) many seemingly well-folded, stable proteins fail to crystallize.

Varied approaches to increase crystallization success rates have been attempted in previous efforts. Reductive methylation[3], surface entropy reduction mutations[4] or carrier protein fusion[5] have been used with some success to create proteins that are more favorable for crystallization. Here modifications are performed to the protein of interest (POI) creating new versions of the protein, which may more readily form a stable crystal lattice. In each of these instances, the goal is to modify the surface of the protein to increase the probability of forming crystal contacts and a stable crystal lattice. Although this approach has proven to be successful for numerous proteins, it does not take advantage of symmetry, an inherent property of protein crystals. It has previously be proposed that induction of synthetic symmetry may serve as an alternate means of increasing the probability of forming stable protein crystals[6-7]. This approach is graphically summarized in figure 1.1. In short, asymmetric proteins may be made into dimers through surface point mutations to cysteines[6] or metal chelating mutations[8–10]. A series of mutations were made with both methods. It was found that different mutations lead to different space groups for the proteins with unique crystal contacts. This approach leads to numerous versions of the POI, each with a unique opportunity to sample different crystal contacts. It also induces internal symmetry that may be translated to crystallographic symmetry. These factors both will increase the likelihood of a protein successfully crystallizing. Subsequent to these early experiments, this approach was successfully applied to a protein that had previously failed to crystallize, *Thermotoga maritima* CelA[11]. This supported the use of synthetic symmetrization as a crystallization tool for difficult-to-crystallize proteins. It also revealed the difficulty to adapt this to a tool that can widely implemented.

1.1.1 Obstacles for Individual Proteins

Although the synthetic symmetrization of CelA led to its crystal structure, the difficulty in obtaining disulfide dimers via surface point mutations was highlighted. Synthetic symmetrization is most powerful when it is applied to novel proteins that have never crystallized. Key to the success of this approach is the ability to accurately select surface exposed residues to mutate. For novel proteins this is usually based on homology models that can vary dramatically in accuracy. For CelA, eight predicted surface residues where mutated to cysteines in order to form the disulfide dimers; only one worked. Only the structure of CelA revealed that the other mutants failed because the other cysteine residues were not as exposed as predicted.

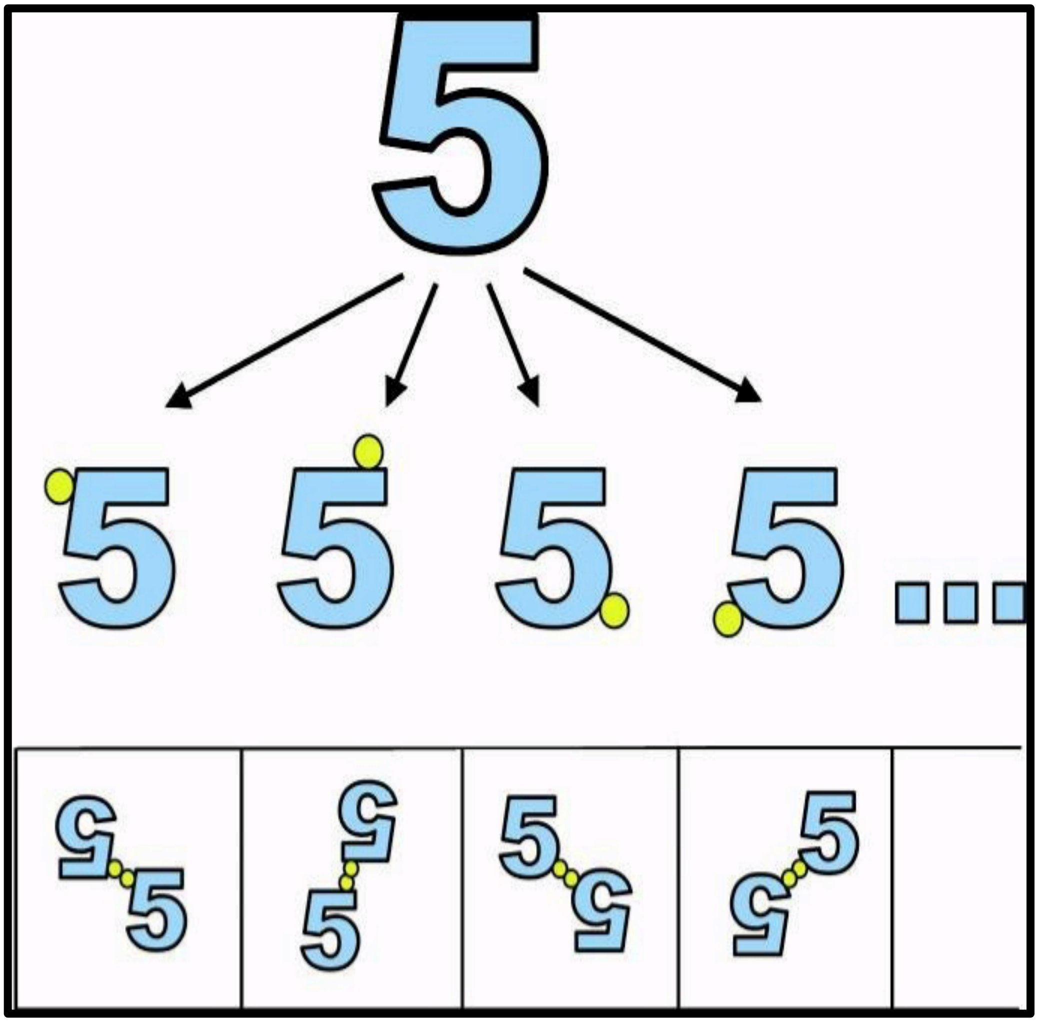


Figure 1.1 Concept of synthetic symmetrization. Adapted with permission from Banatao *et al*, 2006 (DOI 10.1073/pnas.0607674103). An asymmetric molecule such as a '5' can be made symmetric by introduction of a point of attachment (yellow circle). This can be translated to an asymmetric protein or protein complex and the introduction of a point mutation such as a cysteine or metal binding site. As the position of this mutation is changed on the surface of the molecule different oligomers are formed, each with unique opportunities to form a crystal lattice. Although the overall geometry of oligomer changes the tertiary structure of the asymmetric subunit remains unchanged.

From this it is clear that for this approach to be successful multiple mutants would have to be made, screened for oligomer formation, and subjected to crystal trials. This leads to a significant increase in the number of experiments performed for each protein that is being symmetrized. While this additional work is justifiable for a few high value proteins, it is not practical for high throughput crystallization rescue efforts. One must ensure the starting POI lacks native cysteines or remove existing cysteines through mutagenesis and obtain a reasonably accurate homology model all before introducing mutations that may disrupt protein folding. These restrictions limit practicality of this approach. Work presented in this first part of this dissertation is focused on attempts to overcome these difficulties with the use of symmetric scaffolds, assemblies that will induce their symmetry onto POIs with limited modifications needed.

1.1.2 Pre-formed Symmetric Scaffolds

Difficulties with novel proteins symmetrization may be eliminated if the oligomeric mutations are performed on a scaffold of known structure. Here, all the mutations are applied to the protein that readily purified and crystallized. A suite of numerous mutations can be formed, each with a unique geometry and potential to crystallize. The POI can then be attached to the symmetric scaffold through terminal fusions or complementation methods. This will induce the scaffold symmetry on to the POI. Now the POI will have multiple opportunities to crystallize with each different scaffold. It should be noted that the scaffold-POI complexes do not need to, and are not expected to, crystallize in the identical manner of the scaffold alone. Instead each complex will have its own unique opportunities to crystallize. Ideally this complex formation would occur *in vitro*, through some means of complementation of protein fragments. Scaffold oligomers would be pre-formed, and one version of the POI that can complement the scaffold made in advance. These pre-formed oligomers would be stored until needed. Once the POI is purified it would be mixed with the oligomers immediately prior to crystal trials. Hence, one version of your POI can be made into numerous unique versions from mutants in the scaffold suite. Through the induction of symmetry and the use of unique versions of the POI, the potential for crystallization increases.

In recent years our collaborators have created a version of green fluorescent protein (GFP) to serve as a biosensor. Known as split GFP; the 11-beta strand GFP was first split after strand 10 and subsequently strand 9 (GFP1-9), the remaining strands such as GFP10/11 (GFP hairpin) could then be attached to a target POI. Once this target POI containing the hairpin is mixed with the GFP1-9 core, complementation occurs forming the GFP chromophore.[12] This provides a means of verifying *in vitro* complex formation (figure 1.2). The ability of "native" GFP1-9 (containing no oligomeric mutations) to form diffraction quality crystals in complex with another protein has been confirmed[13]. With these promising results, efforts were undertaken to form a suite of GFP oligomers to serve as symmetric scaffolds (Chapter 2).

1.2 Higher Order Scaffolds

Simple point mutations are limited in their ability to form complex oligomers with high symmetry. It becomes difficult with the increasing number of unintended oligomers that can form. If more than one cysteine is present then multiple oligomers will form (figure 1.3). When metal chelation motifs are present, surrounding residues may contribute to ion binding, throwing off the intended oligomer as well (figure 1.4). The ability to successfully develop high-order scaffolds is desired. They involve a wider variety of crystal contacts and symmetries or new geometries for bioengineering applications. Existing oligomers may serve this purpose, yet this is limiting, as pre-existing proteins might not exist in the desired symmetries or having termini in the correct orientation for fusions.

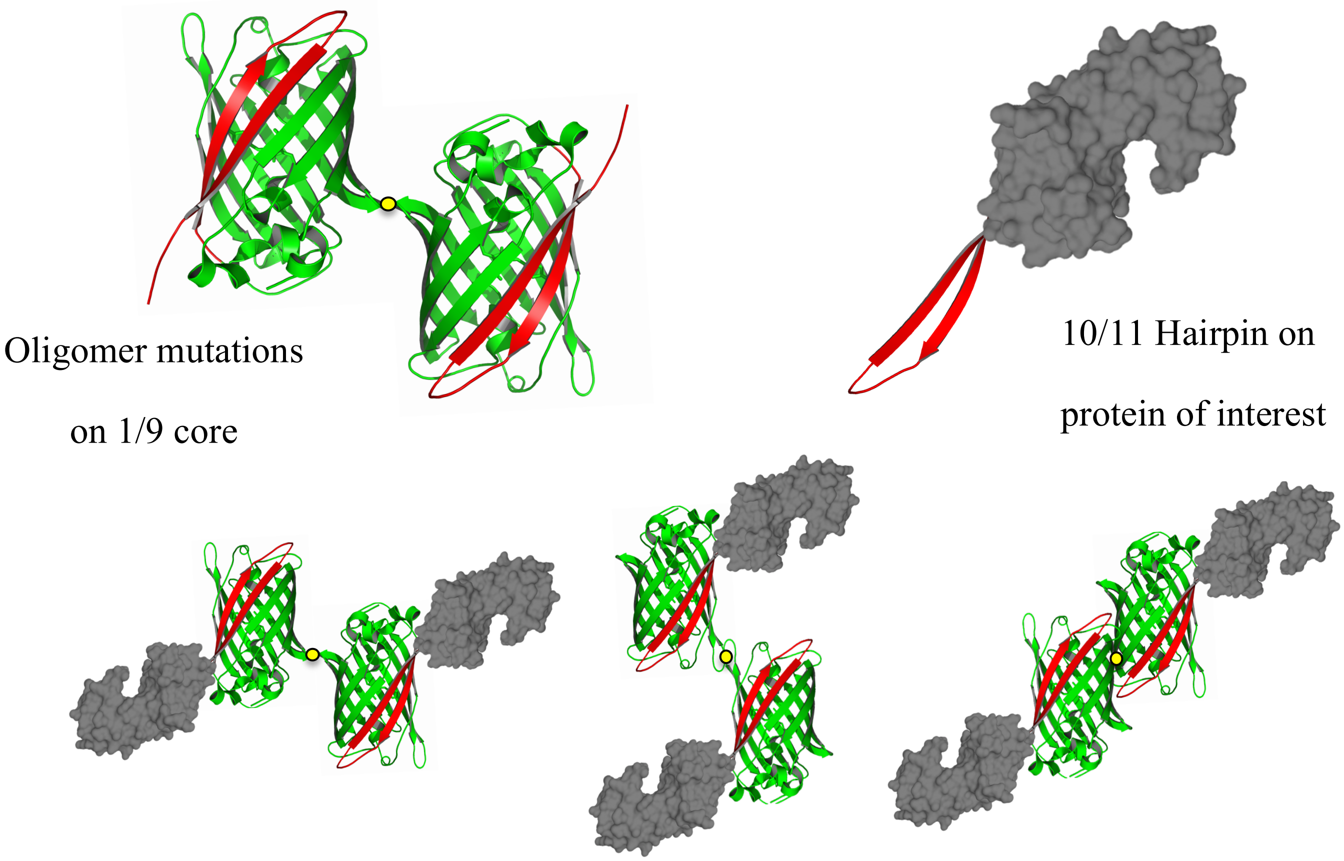

Oligomer mutations
on 1/9 core
10/11 Hairpin on
protein of interest

Figure 1.2 Scaffold Induced Synthetic Symmetrization. Instead of introducing symmetrizing mutations onto a protein of interest, the mutations can be made on a core protein, in this case GFP 1-9. The protein of interest can then be fused to the GFP 10/11 hairpin segment. Multiple GFP 1-9 cores with unique oligomer mutations can be made, once complemented the symmetry of the core will be induced onto the protein of interest.

Previous experiments with metal-mediated oligomerization of T4-lysozyme resulted in symmetric trimers.[8] However, these trimers are unpredictable. The metal-chelation efficiency is highly dependent on the conditions of the crystallization solution. Changes to pH or presence of chelating agents from crystallization screens impact how the metal ions are chelated. Thus, it is exceptionally difficult to design robust metal-mediated oligomers without the addition on non-natural chelation moieties attached to the protein molecules[14,15].

Computational programs including Rosetta[16] and Rosetta MatDes[17], have potential to overcome these challenges through designing specific self-assembling oligomers. As these computational methods improve it is conceivable that any desired symmetric protein assembly can be made. With Rosetta protein building blocks, typically native oligomers are docked in the desired symmetry. The resulting interfaces are then designed to make the assembly stable in solution. This technology has recently been used to design symmetric protein cages of tetrahedral and octahedral symmetry with one or two protein components[17,18].

Utilization of computational protein design efforts have been undertaken to design novel symmetric protein scaffolds for crystallization. Including trimeric GFP scaffolds (chapter 4) and protein cages designed by the David Baker lab at the University of Washington[17,18] (chapter 6). Rosetta Design is still in its infancy and currently suffers from high failure rates. These are a result of the hydrophobic interfaces currently designed. As our understanding of this technology evolves and with future implementation of designed hydrogen bonding patterns and electrostatic interactions, the applications of these higher order oligomers can fully be evaluated in future experiments.

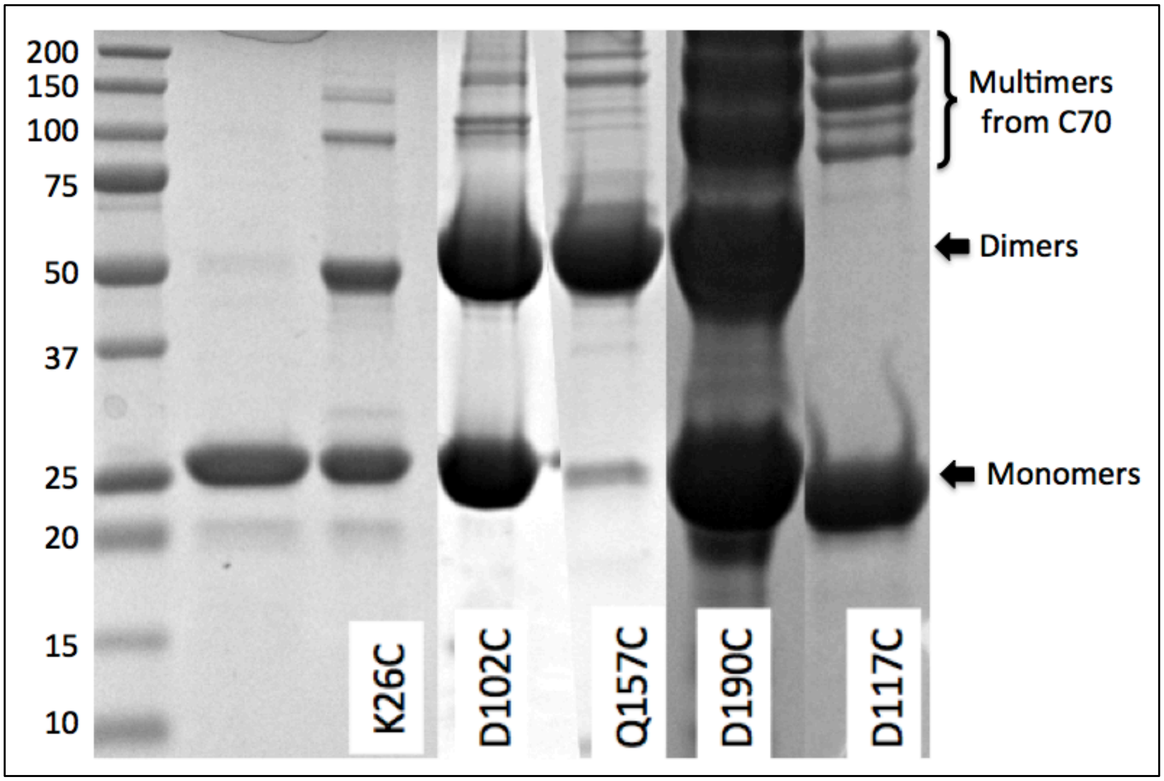
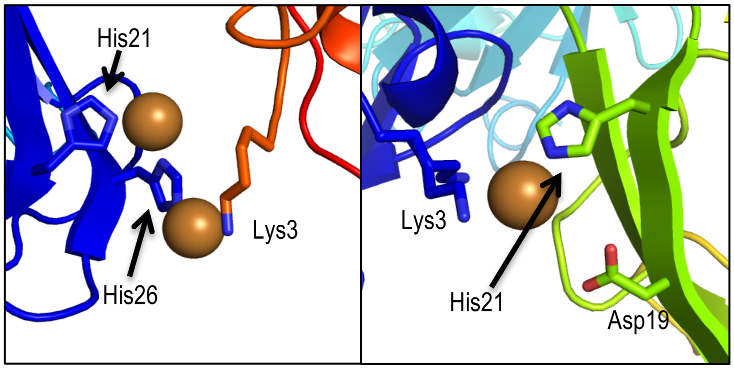
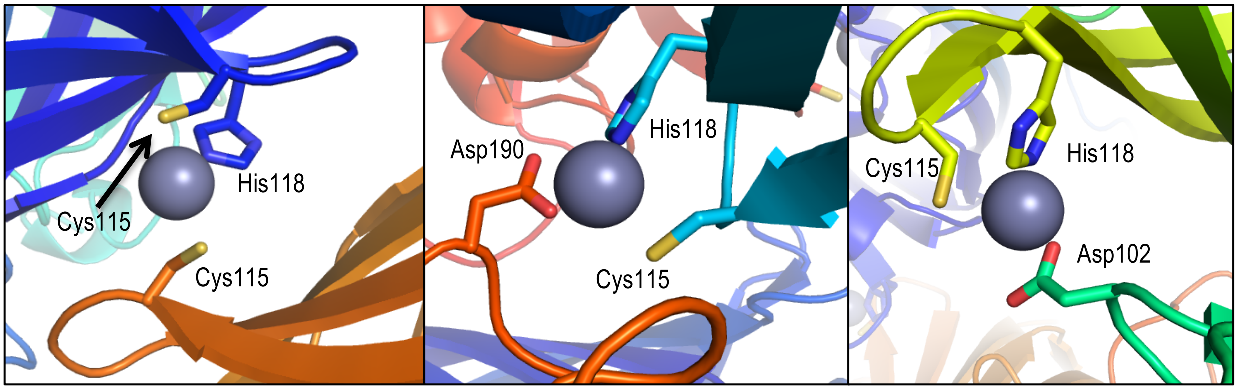

200
150
100
75
50
37
25
20
15
10
K26C
D102C
Q157C
D190C
D117C
Multimers from C70
Dimers
Monomers

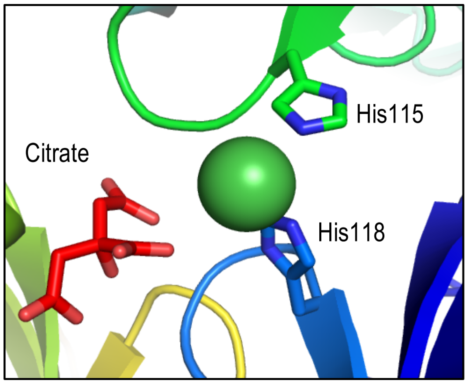
Figure 1.3 Presence of Multiple Cysteines Results in Unpredictable Oligomerization. Although it was thought that cysteine 70 was buried in the core of split GFP based upon the sfGFP structure, it was reactive and formed mulitmers. These are likely hetero-oligomers between the C70 and the introduced mutation. The unexpected reactivity of C70 indicates that it is essential that only one cysteine is present during the cysteine oxidation step. This would require the GFP dimer cores to be oxidized to dimers prior to complementation with a protein containing native cysteines.


His21
His26
Lys3
Lys3
His21
Asp19

21H/26C - Cu


Cys115
His118
Cys115
Asp190
His118
Cys115
Cys115
His118
Asp102

115C/118H - Zn


His115
Citrate
His118

115H/118H - Ni

Figure 1.4 Metal Chelation is Buffer Specific. Examples of unintended metal chelation events due to the crystallization conditions. In each case the histidine or cysteine side chains where point mutations to bind the metal ion. It was found that acidic side chains or lysine side chains contributed to one half of the metal contact depending on the buffer conditions. Small molecule chelators like citrate can also alter the metal –chelation when present in crystallization screens.

1.3 Alternate Applications of Symmetric Scaffolds

Suites of symmetric oligomers are applicable to synthetic biology and enzymatic materials as well. Instead of utilizing the symmetry to induce crystallographic symmetry and crystal contact variations, they can bring POIs into close proximity for enzymatic reactions. The split GFP scaffold discussed throughout this work has previously been shown to bring two proteins into close proximity.[19] One can envision a suite of GFP dimers can serve this purpose as well, bring up to four proteins into close proximity with two proteins fused to GFP hairpins. The mutant suite can allow alternate geometric arrangements of these molecules and facile combinations tested. When the oligomeric scaffolds are combined with oligomeric POIs, amorphous gels would be formed. If the POI is an enzyme, a purely proteinaceous enzymatic material will result. It has been shown that enzyme can have increased stability and reaction rates when encapsulated into gels[20–23]. Again, a suite of preformed oligomer and *in vitro* complementation of the components will allow rapid screening of oligomers for efficient gel formation.

1.4 Pdu Microcompartment Protein PduA

Although prokaryotes are defined, in part, as organisms with no membrane organelles, emerging research suggests purely proteinaceous assemblies serving similar functions are widespread[24,25]. These proteinaceous assemblies are referred to as a bacterial microcompartmen (MCP). MCPs are composed of an outer shell encapsulating a metabolic pathway[26] and resemble viral capsids in morphology. It is hypothesized that MCPs have evolved to contain volatile or toxic intermediates. Carboxysomes were the first and simplest MCP to be discovered; these primarily encapsulate RuBisCO and carbonic anhydrase and increase the local concentration of CO_2[27]. Later MCPs discovered to have evolved in diverse bacterial species to contain novel pathways.

One such MCP is for 1,2-propanediol-utilization (Pdu), where the shell is thought to isolate the cytotoxic propionaldehyde intermediate of the pathway (figure 1.5). As more MCPs have been discovered, there has been an increasing interest to convert them into bio-reactors[28,29]. The principal would be to exploit the shells of the MCP to encapsulate heterologous enzyme pathways in *E. coli.* Through this, pathways that are either inefficient, contain cytotoxic intermediates or contain enzymes sensitive to the cytosolic environment could potentially be active in *E. coli*. Ideas for these non-native applications have included carbon fixation or improved biofuel production, and are currently being explored.[30] Key to the success of this is the ability to engineer these MCPs to carry out new functions. This involves two distinct obstacles that need to be overcome. First, the MCPs shells need to be permeable to the substrate of the heterologous pathway. Second, the new enzymes need to be encapsulated efficiently. Shells of MCPs are composed of numerous proteins of differing function[31–33]. The Pdu MCP contains seven such proteins, six of which are hexagonal and referred as the BMC domains and one vertex pentamer known as the BMV domain (figure 1.6). The atomic structures of many of these shell proteins have been elucidated and PduA[34,35] has been extensively studied for its role in substrate transport into the MCP lumen[35,36].

Key to the transport of pathway substrates and products through the MCP shell is the presence of pores found in the center of certain shell proteins including PduA. It has been found that a residue, serine 40, is key to the selectivity of this pore. Mutations of Ser40 have different 1,2-propanediol permeability[36]. To achieve MCP based bioreactors, this transport needs to be further investigated. It is very likely that simple point mutations will not be sufficient to allow non-native substrate permeability. Efforts to change the topology of PduA to allow more dramatic changes to the pore have been undertaken (chapter 8).

1.5 Targeting Enzymes to the Microcompartment

The mechanism of luminal protein encapsulation in MCPs is another area of active investigation. Heterologous proteins encapsulated by MCPs are essential to engineered bioreactors. Evidence is emerging that in many cases the luminal cargo proteins of the MCPs are pre-assembled and the shell forms around them. This idea is emerging for the case for the carboxysome where RuBisCO assembles into a large aggregate and accessory proteins recruit the shell to the cargo[37]. A similar mechanism has been proposed in the case of the Pdu MCP[38] where smaller assemblies of cargo proteins associates prior to shell assembly. However, there is no consensus for Pdu MCP assembly mechanism.

It has been established that encapsulated Pdu enzymes have N-terminal sequences that target these enzymes to the MCP. These short sequences from PduD[39] and PduP[40] target both the native proteins to the MCP, and fusion of these sequences to non-native proteins can target them to the lumen of the Pdu MCP as well. Further, the interaction between the PduP N-terminus and PduA was established from this work.[40]

Additional luminal Pdu proteins are though to contain targeting sequences with the only common characteristic being the presence of a predicted amphipathic helix that associates with the shell[41]. These helices are thought to either associate directly with the shell proteins[41] or self-associate depending on the enzyme[38]. Currently, the only interaction that has been experimentally established has been that of PduP and PduA. Efforts to determine the x-ray crystal structure of these interactions have been attempted (Chapter 9) although this interaction has yet to be structurally determined.

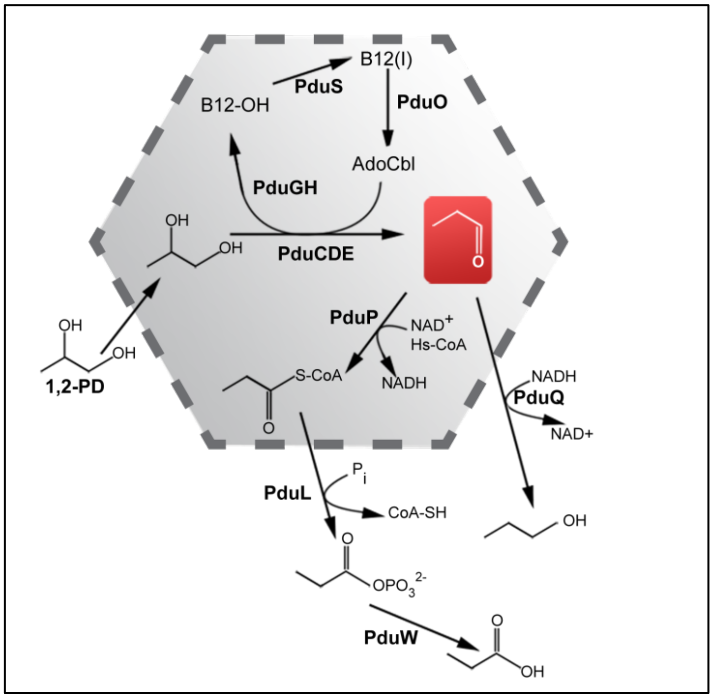

B12(I)
PduS
B12-OH
PduO
AdoCbl
PduGH
OH
OH
PduCDE
O
OH
OH
1,2-PD
PduP
NAD+
Hs-CoA
S-CoA
NADH
O
NADH
PduQ
NAD+
Pi
PduL
CoA-SH
O
OH
OPO3 2-
PduW
O
OH

Figure 1.5. The Pdu MCP. Adapted with permission from Jorda *et al* 2006 (DOI 10.1002/pro.2196). The 1,2-propandiol utilization microcompartment encapsulates the enzymes for 1,2-proapandiol catabolism to propionate. Part of this pathway involves the intermediate propionaldehyde (red box). This DNA modifying cytotoxic intermediate is sequestered within the shell and away from the bacterial cytoplasm.

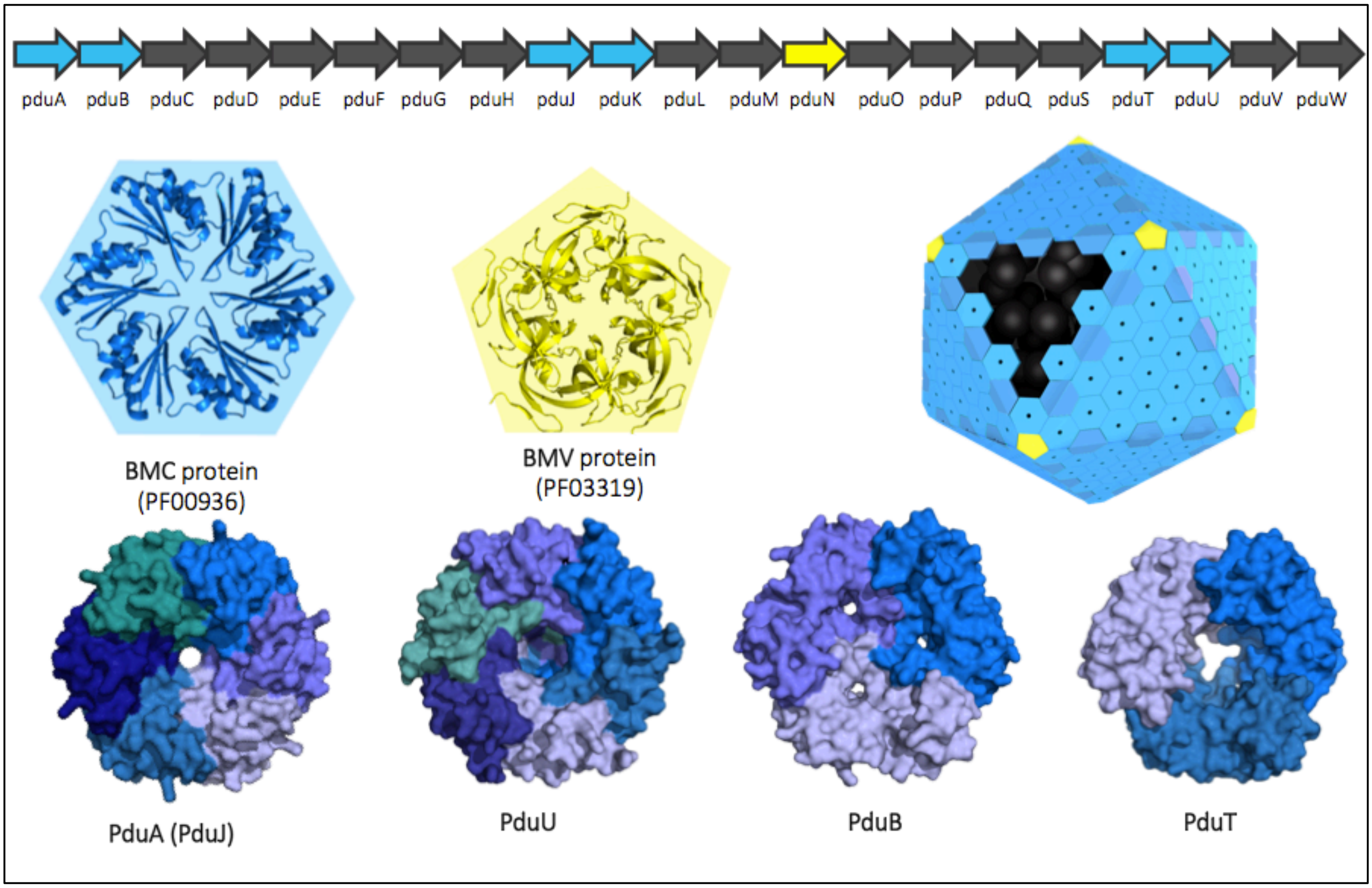

pduA
pduB
pduC
pduD
pduE
pduF
pduG
pduH
pduJ
pduK
pduL
pduM
pduN
pduO
pduP
pduQ
pduS
pduT
pduU
pduV
pduW
BMC protein
(PF00936)
BMV protein
(PF03319)
PduA (PduJ)
PduU
PduB
PduT

Figure 1.6. The Pdu MCP operon and shell proteins. The BMC proteins assemble into hexagonal tiles that are the building blocks of the facets of the microcompartment while the pentameric BMV are supposed to close the vertices. BMC are typically hexameric and perforated by a rather narrow pore that is believed to facilitate the transport of 1,2-propanediol across the shell. There are multiple paralogs of the Pdu BMC shell proteins including PduA and its highly homologous twin PduJ. These are canonical BMC shell proteins which assemble into hexamer and central pore.

1.6 Summary of the Dissertation

Research in this dissertation focuses on two areas of research. First, chapters 2-7 address the general topic of synthetic symmetrization and the applications of these symmetric scaffolds. Chapters 8 & 9 discuss research into structural changes to the Pdu MCP shell protein PduA.

Chapter 2 is a manuscript that was published in *Structure* in 2015. This manuscript reports on thirty-three novel GFP oligomer structures intended to serve as crystallizations scaffolds initially, inducing symmetry onto POIs. The suite of mutations that resulted in these diverse structures is presented followed by analysis of the structures. The background of the concept of synthetic symmetrization in general and potential applications for these mutants is discussed.

Chapter 3 delves into the efforts to validate the scaffolds described in chapter 2 as crystallization scaffolds. The difficulties in adapting the split GFP into a working scaffold are discussed. The proof-of-concept experiments that were conducted in an attempt to determine the viability of these GFP oligomers to serve as scaffolds are presented within this chapter. Although these experiments did not lead to diffraction quality crystals, the split GFP – POI complexes were optimized and purified, and micro-crystals that could be optimized in future efforts were produced.

Chapter 4 reports on efforts to computationally design GFP trimers with C3 symmetry through Rosetta protein design. It was intended that these designs would supplement the oligomers from chapter 2, and serve as a starting point to design GFP oligomers with higher symmetry. These efforts proved difficult as many of the designs failed to yield soluble protein. The resulting crystal structures from these efforts are reported.

Chapter 5 expands on the idea of scaffold-based synthetic symmetrization using T4-lyzozyme mutants fused to a coiled coil sequence to allow for *in vitro* complex formation, applications to crystallizing target proteins are discussed. Within this chapter the design of the lysozyme coiled coil constructs is described, along with the process for selecting test proteins. The difficulties of this approach are presented and future efforts for improved proof of concept experiments are proposed.

Chapter 6 is the final chapter on synthetic symmetry as a crystallization tool. Here, a computationally designed cage was fused to GFP and a novel kinesin. Optimization of both constructs and resulting crystals are discussed. The resulting crystals did not diffract to high enough resolution to determine the structures in either case. Improvements to this approach by alterations to the construct design are proposed.

Chapter 7 focuses on an alternative application of the oligomeric scaffolds previously described. Experiments involving the GFP oligomers from chapter 2 and a trimeric bromperoxidase in an attempt to create an amorphous gel are reported. The results of the initial work and resulting activity assays are reported as well as proposed modifications to improve this system. In the second part of this chapter the initial experiments to created multi-enzymes materials involving a biphenyl degradation pathway are discussed. Two failed approaches involving GFP-mediated linkages and coiled-coil fusions are explored.

Chapter 8 is focused on a different research area, the circular permutation of PduA to change the protein's topology. A manuscript describing a serendipitous cage from the initial experiments is presented. The latter part of this chapter describes in detail the efforts to rescue the initial permuted design and the resulting structure and additional modifications are presented. Additionally, efforts to strengthen the serendipitous cage are explored.

Chapter 9 describes the efforts to obtain a crystal structure of PduA in complex with the PduD targeting tail. The difficulties in obtaining this elusive structure and hypotheses why are discussed. New evidence that our original hypothesis of a direct PduA-PduD interaction may be incorrect is addressed and experiments to structurally validate the interaction between the shell proteins and luminal enzymes of the Pdu MCP are proposed.

Chapter 2

A Suite of Engineered GFP Molecules for Oligomeric Scaffolding

Reused with permission from Elsevier license 3877321093784,

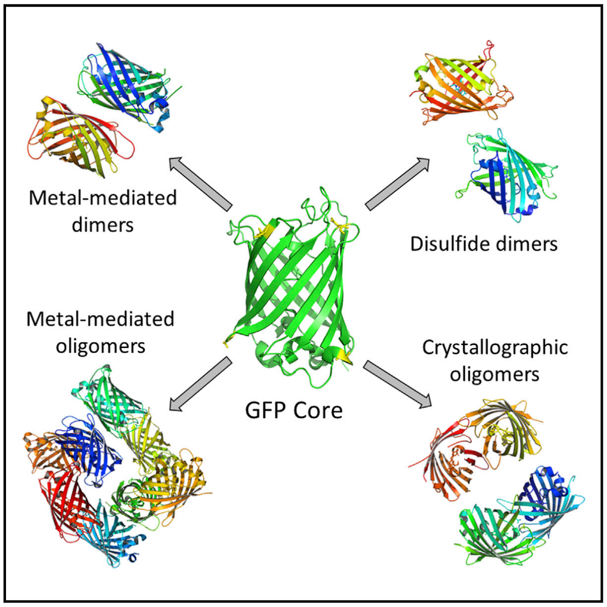
DOI http://dx.doi.org/10.1016/j.str.2015.07.008[42]

Resource

Structure

A Suite of Engineered GFP Molecules for Oligomeric Scaffolding

Graphical Abstract



Authors

David J. Leibly, Mark A. Arbing,
Inna Pashkov, ..., Geoffrey S. Waldo,
Thomas C. Terwilliger, Todd O. Yeates

Correspondence

yeates@mbi.ucla.edu

In Brief

Leibly et al. have generated a large suite of GFP oligomers with varying spatial arrangements of subunits, which are elucidated in 33 distinct crystal forms. These new GFP forms have potential applications ranging from synthetic biology to protein crystallization by connecting multiple proteins together in predefined spatial relationships.

Highlights

- A large suite of GFP mutants has been developed to allow diverse oligomerization
- 33 distinct crystal structures of designed GFP oligomers have been determined
- Target proteins can be oligomerized in diverse ways by attachment to the GFP molecules

Accession Numbers

4W69
4W6A
4W6B
4W6C
4W6D
4W6F
4W6G
4W6H
4W6I
4W6J
4W6K
4W6L
4W6M
4W6N
4W6O
4W6P
4W6R
4W6S
4W6T
4W6U
4W72
4W73
4W74
4W7X
4W75
4W76
4W77
4W7A
4W7C
4W7D
4W7E
4W7F
4W7R

Leibly et al., 2015, Structure *23*, 1754–1768
September 1, 2015 ©2015 Elsevier Ltd All rights reserved
http://dx.doi.org/10.1016/j.str.2015.07.008

Structure
Resource

A Suite of Engineered GFP Molecules for Oligomeric Scaffolding

David J. Leibly,[1,2] Mark A. Arbing,[2] Inna Pashkov,[2] Natasha DeVore,[3] Geoffrey S. Waldo,[3] Thomas C. Terwilliger,[3]
and Todd O. Yeates[1,2,*]
[1]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA
[2]UCLA-DOE Institute of Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA
[3]Bioscience Division, Los Alamos National Laboratory, MS M888, Los Alamos, NM 87545, USA
*Correspondence: yeates@mbi.ucla.edu
http://dx.doi.org/10.1016/j.str.2015.07.008

SUMMARY

Applications ranging from synthetic biology to protein crystallization could be advanced by facile systems for connecting multiple proteins together in predefined spatial relationships. One approach to this goal is to engineer many distinct assembly forms of a single carrier protein or scaffold, to which other proteins of interest can then be readily attached. In this work we chose GFP as a scaffold and engineered many alternative oligomeric forms, driven by either specific disulfide bond formation or metal ion addition. We generated a wide range of spatial arrangements of GFP subunits from 11 different oligomeric variants, and determined their X-ray structures in a total of 33 distinct crystal forms. Some of the oligomeric GFP variants show geometric polymorphism depending on conditions, while others show considerable geometric rigidity. Potential future applications of this system are discussed.

INTRODUCTION

The general idea of connecting and spatially organizing multiple proteins is an emerging theme in synthetic biology. Notable applications include the spatial organization of multiple enzymes for metabolic pathway optimization (Conrado et al., 2008; Dueber et al., 2009; Lee et al., 2012), the organization of signaling molecules (Good et al., 2011; Zeke et al., 2009), and the creation of large self-assembling protein architectures (Lai et al., 2012). Another area under exploration is the synthetic organization of protein molecules into various symmetric forms to expand the chances of being able to induce them to form well-ordered crystals (Laganowsky et al., 2011). Facile systems for enabling the specific spatial organization of arbitrary proteins of interest could therefore advance research along various lines.

Ongoing efforts toward engineering proteins for improved crystallization stem from the generally low success rate and unpredictability of macromolecular crystallization (Sundstrom et al., 2006; Stacy et al., 2011). Regardless of the varied explanation for why many proteins are difficult to crystallize, the chances for a successful outcome might be improved by promoting the formation of intermolecular contacts that are compatible with crystal symmetry. Various methods for engineering proteins to improve their likelihood of forming good crystal contacts through surface residue mutations or fusion to a carrier protein have been described and reviewed (Banatao et al., 2006; Derewenda and Vekilov, 2006; Salgado et al., 2008; Forse et al., 2011; Corsini et al., 2008; Moon et al., 2010; Zou et al., 2012) including fusion to engineered GFPs (Suzuki et al., 2010).

Synthetic symmetrization, the engineering of artificially symmetric forms of a given protein molecule, has been promoted as one method for explicitly increasing the likelihood that a protein will be able to form a crystal lattice (Banatao et al., 2006). Two potential advantages have been articulated. First, geometric arguments and analysis of observed crystallization patterns suggests that a modest advantage can be gained by building symmetry into an otherwise asymmetric protein molecule by forcing it to oligomerize. Second and perhaps more important, the ability to produce multiple distinct symmetric forms of a target protein is a major advantage for crystallization. If the protein under study is the subject of crystallization trials, then each of the oligomeric constructs (e.g. specific dimers) is in effect a distinct molecular species with new opportunities to form lattice contacts in the context of a crystal. Distinct dimeric forms of a protein, for example, can be constructed by introducing single cysteine residues at various surface-exposed residues in a protein (Banatao et al., 2006; Forse et al., 2011). In another approach, metal-binding half-sites can be designed by introducing two potential metal-ligating residues (e.g. histidines) at proximal positions on the protein surface (Laganowsky et al., 2011). These experiments have shown that proteins engineered in such ways form oligomers that are rigid enough for facile crystallization, and that many new opportunities are opened up for the crystallization of a single given protein. In many cases, the new interactions introduced into the target protein contribute to the symmetry of the crystal (Banatao et al., 2006; Chruszcz et al., 2008).

Despite the potential for synthetic symmetrization to expand the opportunities for growing protein crystals, the method as it has been applied so far is experimentally burdensome. Its advantages are offset by the need to engineer multiple variants of the target protein, whose structure may be unknown, leading to potential challenges in conferring favorable properties without disrupting its fold. In this study, we explore a route for circumventing this obstacle. The essential idea is to apply the protein engineering work (i.e. to introduce varied forms of synthetic

©2015 Elsevier Ltd All rights reserved

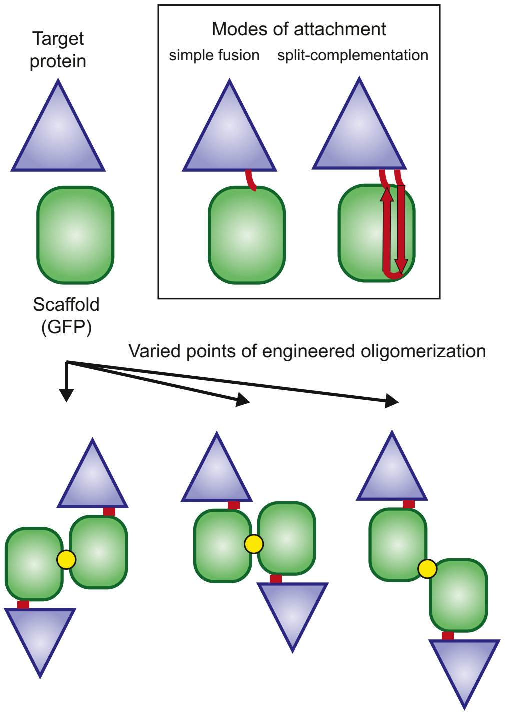


Figure 1. Concept of Scaffold-Mediated Synthetic Symmetrization
Here, GFP serves as a scaffold to induce synthetic symmetry. (Top) Multiple modes for attaching a target protein to GFP are indicated, including simple fusion and split-form complementation where the target protein is fused to a fragment of GFP, either strand 11 or strands 10–11. (Bottom) GFP (or another scaffold) is engineered in multiple ways to create varied oligomeric forms. When a target protein is connected (by fusion or complementation) to the engineered GFP molecules, varied oligomeric forms of the target protein are created.

symmetrization) to a model protein that can subsequently serve as a general carrier or scaffold for attaching otherwise arbitrary proteins. In this way a target protein can be driven into varied oligomeric forms with distinct opportunities to crystallize, without having to substantially compromise its native sequence. As long as the target protein is not much smaller than the scaffold to which it is attached, both components can be expected to participate in ordered contacts in a crystal.

Multiple strategies are possible for attaching a target protein to a scaffold protein, including by direct genetic fusion. Other possibilities are presented by a scaffold that can be produced and then reconstituted from two separate fragments. Here, we investigate the use of GFP as a scaffold for oligomerization, since GFP, particularly when accompanied by stabilizing mutations, can be expressed in split form and then functionally reconstituted from a large fragment and a small fragment (Cabantous et al., 2005, 2013; Huang and Bystroff, 2009; Nguyen et al., 2014). The key elements of the approach are illustrated in Figure 1. The use of monomeric split-GFP to complement and then crystallize another protein that is fused to the small GFP fragment has been already demonstrated in recent work (Nguyen et al., 2014). Here, the oligomerization element of the overall strategy is demonstrated by the construction and crystallographic investigation of several engineered variants of GFP. This large suite of engineered GFP proteins provides a foundation for various future developments, including those in the broad area of synthetic biology as well as in protein crystallization.

RESULTS

Rationale for GFP-Mediated Symmetrization

Engineered "split" forms of GFP have gained widespread use in the laboratory setting as biosensors (March et al., 2003) or fusion partners to probe for protein solubility (Cabantous et al., 2005, 2013). These mutants of GFP can be expressed without one or more terminal β strands of the 11 strands composing the GFP β barrel. Using circular permutants of a full-length GFP containing mutations developed for the split form of GFP (Cabantous et al., 2005), Huang and Bystroff (2009) created additional split-GFP pairs (with other tagging or "left-out" strands such as β strand 7). The partial core can then be complemented by addition of another protein that has been engineered to carry the missing GFP β strand(s), as either a terminal fusion or a loop insertion. Once complementation occurs the full β barrel is restored, and formation of the native chromophore provides a convenient readout of complex formation.

These previous developments make GFP well suited as a general carrier protein for implementing a new approach to the idea of synthetic symmetrization. The particular form of GFP used in our study can be split after strand 9, resulting in the GFP (strands 1–9) core and GFP (strands 10–11) hairpin (Cabantous et al., 2005; Nguyen et al., 2014). With this system, the hairpin formed by strands 10 and 11 can be engineered into a target protein, which will then complement GFP(1–9). In the simplest scenario, the (10–11) hairpin can be fused as an extension at either the N or C terminus of the target protein. However, the two-stranded hairpin allows for another particularly advantageous kind of construction. If the hairpin can be inserted at an internal sequence position on an exposed loop in the target protein, then the protein complex formed upon complementation will possess a two-chain crossing between the reconstituted GFP domain and the target protein structure (Figure 1). This is expected to enforce a much more rigid spatial arrangement between the two components, which could be an advantage, particularly where crystallization is the ultimate goal. In fact this has been demonstrated in one recent study, where a crystal structure revealed two copies of the molecular complex in the asymmetric unit in very nearly the same configuration, suggesting a limited range of motion when using the (10–11) hairpin insertion approach (Nguyen et al., 2014). Anticipating the ultimate advantage of the GFP(1–9) plus (10–11) hairpin approach, we focused our efforts on engineering oligomerizing variants of GFP in the strand 1–9 core region at positions remote from the (10–11) hairpin.

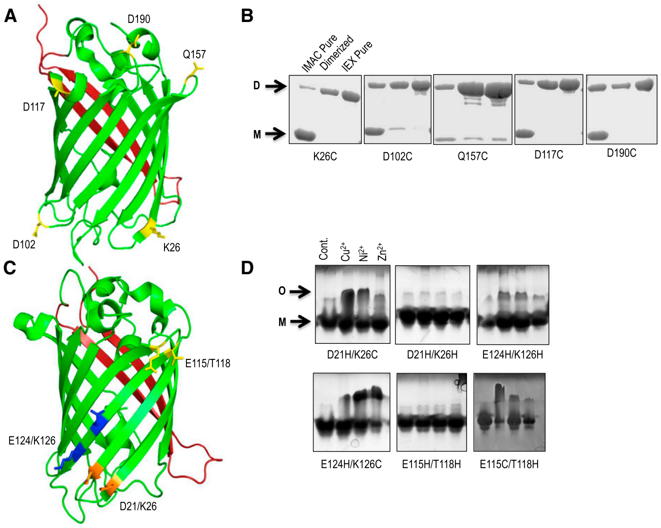


Figure 2. Locations of Point Mutations Introduced on Full-Length Split-GFP to Induce Oligomerization

(A) Locations of the individual point mutations to cysteines (yellow) on the GFP(1–9) core (green) on the opposite face of the β barrel from the GFP(10–11) hairpin (red).

(B) Each cysteine point mutant was purified in non-reducing conditions, and dimer formation was visualized on a non-reducing SDS-PAGE gel. After an initial IMAC step, GFP variants were dimerized with Cu^{2+}. The dimeric form, D, was then separated from the monomer, M, via anion exchange chromatography and used for crystallization experiments.

(C) Locations of the metal-half-site mutations on GFP (yellow, orange or blue); each site involves a pair of spatially proximal mutations (indicated). Color coding as in (A).

(D) Native PAGE screening of each metal-chelating mutation in the presence of Cu^{2+}, Ni^{2+}, and Zn^{2+}. This screen showed apparent oligomer formation for the D21H/K26C, E115C/T118H, E124H/K126H, and E124H/K126C variants, as determined by a mobility shift from the monomeric (M) band to the assumed oligomeric (O) band.

Design and Structure of Cysteine-Based GFP Dimers

In our first approach to engineering oligomerizing variants of GFP, individual cysteine residues were introduced at surface positions. Each such engineered protein was expected to produce a distinctly different dimeric structure upon oxidative disulfide formation. The utility of the disulfide-based approach to synthetic symmetrization has been demonstrated previously (Banatao et al., 2006; Forse et al., 2011). For application of the idea to GFP, we created five cysteine point mutations—K26C, D102C, D117C, Q157C, and D190C—as well as two sets of mutations to serve as either disulfide or metal-mediated oligomers (discussed subsequently): E115C/T118H and E124H/K126C. These amino acids were selected for mutation based on their polarity, surface location, and distance from strands 10–11 in order to limit interference with complementation when ultimately expressed in the split form (Figure 2). As the starting or wild-type sequence for design of the point mutations, we chose the sequence of split-GFP in its full-length form (Cabantous et al., 2013) using the superfolder GFP structure as a reference for point mutations in solvent-exposed locations (Pédelacq et al., 2006). Two native cysteines at positions C48 and C70 were first mutated to alanine to prevent subsequent interference with disulfide-based dimerization; one exception was an initial experiment and crystal structure of the K26C mutant of the superfolder form (PDB: 4W6B) in which only the cysteine at position 48 had been removed. The ultimate goal of our study is to use engineered versions of the truncated GFP(1–9) to synthetically symmetrize target proteins bearing the (10–11) hairpin, but we judged it prudent to first conduct the GFP engineering experiments against the background of the complete GFP(1–11) construct. Full-length GFP constructs bearing the single engineered cysteine residue were therefore expressed, purified, and oxidized to form homogeneous dimers (Figure 2). For all five of the cysteine sites chosen, pure dimers could be obtained in good yield with ~20–50 mg of protein obtained from 2 l of auto-induction media.

With the exception of Q175C, crystals grew readily in 1–7 days. Depending on the mutant, diffraction-quality crystals grew in as few as one condition for K126C or in more than 20 for D102C and D190C. Due to the large numbers of crystals that grew in the initial experiments, it was not feasible to screen X-ray diffraction in all crystals or to optimize all the crystal hits that were observed. We took the approach of screening crystals that appeared morphologically unique and large enough to mount for X-ray diffraction experiments. In some cases where initial crystals did not diffract despite having good morphology, minor optimization was performed, but otherwise crystals were taken directly from initial screens. Across the many crystal forms examined for the various mutants, the diffraction resolution ranged from 1.7 Å to poorer than 3.5 Å (Table 1). Rather than striving to maximize the resolution for the many crystal forms obtained, we focused on investigating the variety of crystal packing arrangements that these dimers could explore, and the degree to which they appeared to have well-ordered modes of dimerization.

In addition to the cases where we intentionally designed a disulfide bond to make GFP dimers, there were cases whereby we had anticipated the formation of a metal-binding site between GFP monomers involving a combination of an inserted histidine and cysteine pair, but obtained instead GFP dimers connected by a simple disulfide bond when the metal ion was added. These were mutant pairs D21H/K26C, E115C/T118H, and E124H/K126C. In these cases a disulfide bond was seen in the electron density map, but without evidence for metal binding at the dimer interface. These fortuitous dimers were not explored in depth to try to produce additional crystal forms, so their abilities to form alternative crystal lattices were not established.

Table 1. Summary of New GFP Crystal Forms

PDB	Mutation	Type	Space Group	Resolution (Å)	ASU[a]
4W69	Q157C	disulfide	P 43 21 2	3.98	2
4W6A	Q157C	disulfide	P 32 2 1	2.99	2
4W6B	K26C[b]	disulfide	P 21 21 21	1.90	2
4W6C	D21H/K26C[c]	disulfide	P 21 21 21	2.49	2
4W6D	K26C	disulfide	P 32 2 1	3.45	2
4W6F	D21H/K26C	disulfide	P 32 2 1	2.70	2
4W6G	D190C	disulfide	P 61	3.02	2
4W6H	D190C	disulfide	P 65	1.95	2
4W6I	D190C	disulfide	P 21 21 21	2.63	2
4W6J	D117C	disulfide	P 31 2 1	1.70	2
4W6K	D117C	disulfide	P 41 21 2	2.88	2
4W6L	D117C	disulfide	I 41 2 2	2.45	1
4W6M	D117C	disulfide	P 63	2.79	4
4W6N	D117C	disulfide	C 1 2 1	3.38	6
4W6O	D117C	disulfide	P 64 2 2	2.60	1
4W6P	D102C	disulfide	P 21 21 21	3.09	8
4W6R	D102C[c]	disulfide	P 1	3.47	16
4W6S	D124H/K126C	disulfide	P 43 21 2	3.10	2
4W6T	E115H/T118H	Cu-mediated contacts	P 43 21 2	1.60	1
4W6U	E115H/T118H	Ni-mediated contacts	P 21 21 21	2.28	4
4W72	E115C/T118H	disulfide + metal contacts	P 21 21 21	1.99	2
4W73	E115C/T118H	disulfide	P 21 21 21	2.79	2
4W74	E115C/T118H	Zn crystal contacts	P 1 21 1	2.10	8
4W7X	E115C/T118H	disulfide	P 1 21 1	2.80	4
4W75	D21H/K26C[c]	disulfide + metal contacts	P 21 21 21	3.47	2
4W76	D21H/K26C[c]	disulfide + metal contacts	P 21 21 21	2.35	2
4W77	D21H/K26C[c]	disulfide + metal contacts	P 21 21 21	3.10	2
4W7A	D21H/K26C[c]	disulfide + metal contacts	P 21 21 21	3.60	4
4W7C	D21H/K26C[c]	disulfide + metal contacts	C 1 2 1	2.50	4
4W7D	D21H/K26H	Cu crystal contacts	P 21 21 21	1.80	2
4W7E	D21H/K26H	Cu crystal contacts	P 41 21 2	2.59	1
4W7F	D124H/K126H	Cu crystal contacts	C 2 2 21	2.90	1
4W7R	D124H/K126H	Cu dimers	P 1 21 1	1.80	4

[a] Number of GFP chains in the asymmetric unit.
[b] Superfolder GFP C48A backbone mutation.
[c] Split-GFP C48A backbone mutation. All other sequences have the double mutations of C48A and C70A.

In all, we were able to characterize 20 distinctly different crystal forms of the GFP disulfide dimers and solve their structures (Table 1), with an additional six dimers containing both a disulfide bond and metal contacts. In all these structures, we modeled disulfide bonds into the electron density maps where possible, tabulating standard geometric terms and bond energies for the observed disulfide bonds (Tables 2 and S1) (Katz and Kossiakoff, 1986). In some cases where the resolution was limited this was not possible, and in at least two cases it appeared that the disulfide bond had been broken during the course of the X-ray diffraction experiment due to synchrotron radiation damage, as has been observed before (Carugo and Carugo, 2005; Weik et al., 2000).

The occurrence of multiple crystal forms for individual mutants, and the presence in several cases of multiple crystallographically independent GFP dimers in the unit cell, made it possible to analyze the range of conformations and degree of flexibility in these engineered dimers (Figure 3). An analysis of the symmetry and variations due to disulfide bond flexibility was performed for each cysteine mutation by comparing all dimers that were observed for a given point mutation (Figure 4; Table 2). In each case, we calculated the angle of rotation between the two subunits connected by the engineered disulfide bond to judge whether the synthetically generated dimers were nearly symmetric (i.e. related by a 180° rotation) (Table 2). Then, to evaluate how rigidly connected the two subunits were, we examined the degree of geometric variability between multiple instances of the same dimer as observed across different crystal forms or different asymmetric units of the same crystal form. A complete analysis is provided in Tables S2 and S3, and

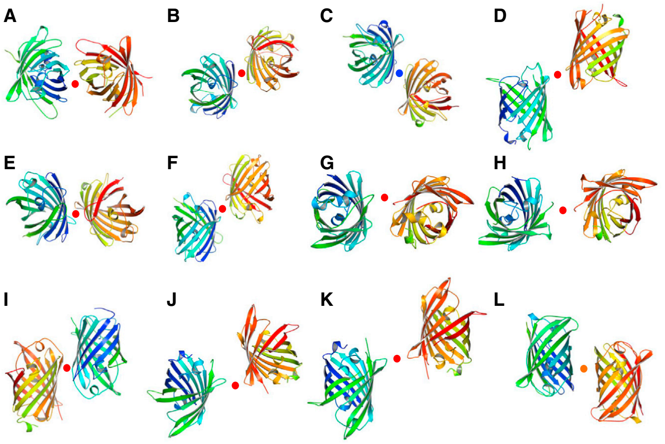


Figure 3. Observed Examples of the GFP Dimer

The internal rotation axis relating the subunits of each dimer is shown (red dot for disulfide dimers, blue for the mixed dimer (L), and orange for the metal-mediate dimer (C)). For each dimer the rotation axis corresponds to the location of the engineered disulfide bond, or metal-mediated crystal contact.

(A–L) The 12 dimers shown are from structures PDB: (A) 4W6B, (B) 4W6C, (C) 4W7C, (D) 4W6R, (E) 4W7X, (F) 4W6M, (G) 4W6G, (H) 4W6I, (I) 4W6S, (J) 4W69, (K) 4W6K, and (L) 4W7R. These dimers are representative of the complete set of 43 total dimers visualized in this work.

summarized in Figure 4 and Table 2. A description of the individual disulfide-bonded GFP structures is as follows.

K26C

Four crystal forms of K26C dimers were observed (PDB: 4W6B, 4W6C, 4W6D, and 4W6F), two in each of the space groups $P2_12_12_1$ and $P3_221$. Of these, PDB: 4W6C was the most symmetric (175.6°) while PDB: 4W6F was the least symmetric (144.3°). PDB: 4W6C, 4W6D, and 4W6F were most similar to each other with a maximum variation of 33.3°, while PDB: 4W6B varied by a rotation of up to 140.4° when overlaid on the others (Table 2; Figure 4B). Two of the structures (PDB: 4W6F and 4W6C) in which GFP dimers were obtained through a disulfide bond at position 26 arose from a D21H/K26C double mutant initially designed for metal chelation. Unexpectedly, addition of Ni^{2+} resulted in formation of a disulfide bond between residues 26C from two protein molecules during the crystallization experiments on these variants.

D102C

Two crystal forms were observed for this mutant, one in space group P1 (PDB: 4W6R) and one in $P2_12_12_1$ (PDB: 4W6P). Crystals appearing in the P1 morphology (thin plates) were obtained in numerous conditions containing PEG polymers as the precipitant. We were able to solve the structure of PDB: 4W6R to 3.47 Å; this was the highest resolution we were able to obtain from all the crystals screened of the D102C mutant. This P1 crystal form had a total of eight disulfide-bonded dimers in the crystal asymmetric unit with an average angle between the chains of 167°. The eight dimers were remarkably similar, with a maximum angular variation of only 8° (Figure 4C; Table 2). Due to this small range of variation, the CCP4 program Zanuda (Winn et al., 2011) was used to investigate and rule out the possibility that some higher crystallographic symmetry had been missed in the initial structure determination. The PDB: 4W6P structure also contained four dimers in the asymmetric unit of $P2_12_12_1$. These dimers are less symmetric

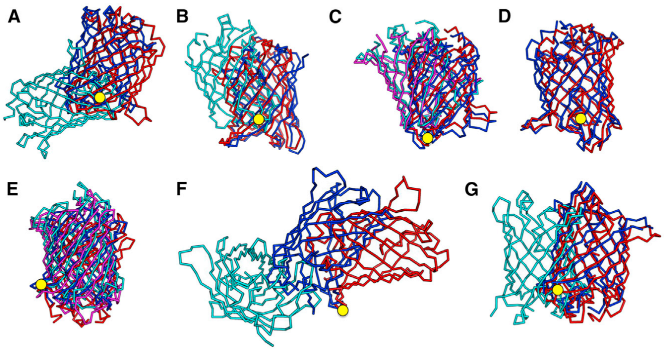


Figure 4. Chain Angle Ranges for Dimers

Depicted is the range of variation between the chain orientations for each disulfide-bonded dimer. Chain A of each dimer was first aligned to visualize the difference in the orientation of the distinct versions of chain B. Only the chain B backbone traces are depicted. Each panel illustrates the multiple conformations observed for one specific cysteine mutant. The blue and red traces represent the range of orientations the chains adopted. When a single outlier is found it is shown in cyan. When two disparate groups of conformations are present they are shown in red and blue, and cyan and magenta. When more than one dimer was observed in the asymmetric unit, instances representing the extremes in conformation were chosen. The rotation axis that relates the two molecules, and which coincides roughly with the position of the point mutation(s), is indicated by a yellow circle.

(A–G) The PDB codes for structures and dimer chains displayed are (A) K26C: red, PDB: 4W6C; blue, PDB: 4W6F; cyan, PDB: 4W6B. (B) D21H/K26C: red, PDB: 4W7A AB dimer; blue, PDB: 4W7A CD dimer; cyan, PDB: 4W75. (C) D102C: red, PDB: 4W6P CD dimer; blue, PDB: 4W6P FG dimer; cyan, PDB: 4W6R AN dimer; magenta, PDB: 4W6R KL dimer. (D) E115C: red, PDB: 4W72; blue, PDB: 4W73. (E) D117C: red, PDB: 4W6O; blue, 4W6K; cyan, PDB: 4W6N BF dimer; magenta, PDB: 4W6J. (F) Q157C: red, PDB: 4W69; blue, PDB: 4W6A A dimers; cyan, PDB: 4W6A B dimer. (G) D190C: red, 4W6H; blue, 4W6I; cyan, 4W6G.

Table 2. GFP Disulfide Dimer Characterizations

Mutant	PDB	Dimer	Disulfide $C\alpha$ Distance (Å)	Dimer Angle (°)	Grouped PDB	Chain "B" Variation Range (°)
K26C	4W6B	AB	6.4	151.66	group: 4W6C, 4W6D, 4W6F	group 4W6C–4W6F = 33.3 maximum range:
	4W6C	AB	6.2	175.55		
	4W6D	AB	6.2	158.12		
	4W6F	AB	5.6	144.29	outlier: 4W6B	4W6B–4W6D = 140.4
D21H/K26C	4W7A	AB	5.8	169.72	group: 4W7A, 4W7C, 4W76 4W77 outlier: 4W75	group: 4W7A AB–4W7A CD = 6.3 maximum range: 4W7A CD–4W75 = 32.1
	4W7A	CD	6.2	177.95		
	4W7C	AB	5.9	173.38		
	4W7C	CD	6.4	171.85		
	4W75	AB	6.2	151.90		
	4W76	AB	6.4	174.64		
	4W77	AV	6.1	173.00		
D102C	4W6P	AB	4.5	143.38	group 1: 4W6P group 2: 4W6R	group 1: 4W6P CD–4W6P FG = 8.3 group 2: 4W6R AN–4W6R KL = 7.7 maximum range: 4W6P FG–4W6R KL = 32.4
	4W6P	CD	4.6	146.21		
	4W6P	EH	4.6	143.79		
	4W6P	FG	4.6	139.64		
	4W6R	AN	5.2	165.37		
	4W6R	BI	4.7	165.15		
	4W6R	CD	4.1	170.66		
	4W6R	EJ	4.4	167.73		
	4W6R	FO	4.7	166.16		
	4W6R	GO	4.9	163.96		
	4W6R	HM	4.9	166.20		
	4W6R	KL	4.3	170.91		
E115C	4W7X	AB	6.2	166.40	group: 4W7X, 4W72, 4W73	4W72–4W73 = 12.3
	4W7X	CD	5.4	163.93		
	4W72	AB	5.9	159.85		
	4W73	AB	6.4	170.95		
D117C	4W6J	AB	5.7	154.89	group 1: 4W6K, 4W6L, 4W6M 4W6O group 2: 4W6J, 4W6N	group 1: 4W6O–4W6K = 16.4 group 2: 4W6N BF–4W6J = 10.8 maximum range: 4W6N BF–4W6M AC = 34.8
	4W6K	AB	5.7	166.82		
	4W6L	AB	5.5	180.0		
	4W6M	AC	5.6	178.44		
	4W6M	BD	6.5	178.14		
	4W6N	AD	6.1	148.41		
	4W6N	BF	6.3	146.59		
	4W6N	CE	6.4	146.87		
	4W6O	AB	5.5	179.97		
K126C	4W6S	AB	6.00	177.96	–	–
K126H	4W7R	AB	–	179.1	–	AB–CD = 1.7
	4W7R	CD		179.15		
Q157C	4W69	AB	5.5	141.18	–	4W96–4W6A B = 129
	4W6A	A	5.78	180.0		
	4W6A	B	11.7[a]	180.0		
D190C	4W6G	AB	5.8	140.95	group: 4W6G, 4W6H outlier: 4W6I	group: 4W6G–4W6H = 6.3 maximum range: 4W6H–4W6I = 41.4
	4W6H	AB	5.8	135.23		
	4W6I	AB	6.4	171.21		

[a] Potential disulfide broken during crystallization.

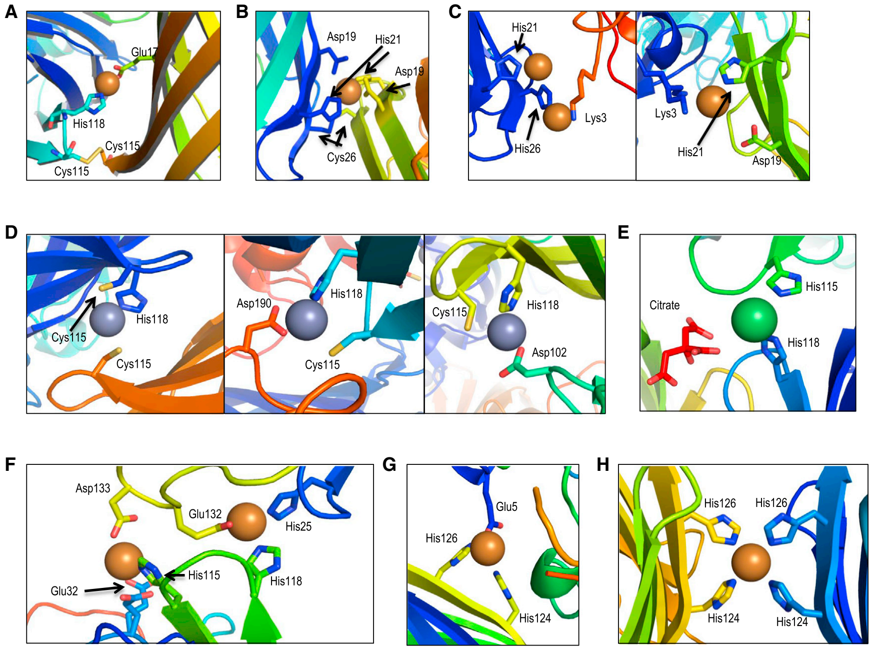


Figure 5. Observed Metal-Mediated Crystal Contacts

(A) Structure PDB: 4W72: a disulfide bond is formed in addition to the copper-binding site.

(B) Example of a mixed dimer from structure PDB: 4W76. Here the copper ion is chelated by histidine and aspartate residues from both molecules, and a disulfide bond is also formed.

(C) The two forms of metal-mediated contacts in PDB: 4W7D.

(D) The three observed zinc-mediated contacts found in PDB: 4W74. (Left) Cys115/His118 from one chain and Cys115 from another chain chelate the zinc. (Middle and right) Cys115/His118 from one chain chelate the zinc ion along with an aspartate (Asp190 or Asp102) from another chain.

(E) A nickel-mediated crystal contact in the structure of PDB: 4W6U involving histidines from the two proteins and a citrate molecule.

(F) A double copper-mediated crystal contact in the structure of PDB: 4W6T, both involving a combination of histidine and carboxylates.

(G) A copper-mediated crystal contact in the structure of PDB: 4W7F. His124 and His126 chelate the copper ion with Glu5 of the symmetry mate.

(H) Copper chelation by His124 and His126 of the symmetric dimer of PDB: 4W7R.

than those observed in the P1 form (average internal angle between subunits of ~143°). In comparison with the other dimeric forms in the same crystal asymmetric unit of this mutant, one dimer was a minor outlier, having a relative chain rotation between subunits of 5°–8°. The uniqueness of this dimer effectively rules out the possibility of any higher symmetry in the crystal.

E115C

Originally intended to serve as a metal-binding half-site, the mutated pair of residues, E115C/T118H, revealed disulfide-bonded dimer formation under crystallization conditions with the addition of metal ions. Four structures were obtained: three disulfide dimers (PDB: 4W72, 4W73, and 4W7X), and one structure with metal-mediated contacts only (PDB: 4W74, discussed subsequently). The three disulfide dimers feature an average rotation angle between subunits of 165°, with a variation up to 12° (Figure 4E; Table 2). Interestingly, PDB: 4W72 features a metal-mediated contact as well, involving the chelation of a copper ion by His118 of one chain A and Glu17 of another (Figure 5A).

D117C

This mutant resulted in six crystal forms, each in a different space group. The six dimers fall into two groups (Figure 4E; Table 2). Three of the dimeric forms observed (PDB: 4W6L, 4W6M, and 4W6O) are either perfectly symmetric with the two subunits related by crystal symmetry (PDB: 4W6L and 4W6O), or very nearly symmetric PDB: 4W6M, 179° rotation). PDB: 4W6J and 4W6M feature similarly asymmetric dimers (average internal angle of 149°), and PDB: 4W6K contains a dimer with an internal angle of 167°. It is notable that D117C dimers are rigid enough to form very well-ordered crystal lattices, diffracting up to 1.7 Å. Yet they are not locked into one conformation, and the permissible angular variation allows for multiple distinct lattices.

K126C

An intended metal-half-site pair, E124H/K126C (PDB: 4W6S) apparently underwent disulfide oxidation in the crystal drop, leading to a symmetric dimer (178°). Copper was added to the protein immediately prior to the crystallization experiment, and no copper ions were observed in the crystal structure. No further efforts were undertaken to explore the possibility of additional space groups for this dimer.

Q157C

Two structures were solved from this mutant (PDB: 4W69 and 4W6A), and only after screening and optimization of crystal conditions. This is likely a result of the point mutation being located on a somewhat flexible loop of the GFP core. The best crystals diffracted to a resolution of 4 Å (PDB: 4W69). PDB: 4W6A represents an interesting and somewhat mysterious crystal form. Two chains are in the asymmetric unit, and they contribute to two different symmetric dimers sitting on axes of crystallographic symmetry, but the expected disulfide bonds are not present. The distance between the cysteine Cα positions of the two subunits is ~11 Å. The crystals took over 6 months to grow, and we suspect that the formate in the crystallization mixture may have slowly reduced the disulfide bonds initially present (Gibson, 1969). Based on the difficulties crystallizing this mutant, we do not view it as a favorable candidate for future crystallization experiments.

D190C

As with the Q157C point mutation, D190C is located in a flexible loop that is found to be disordered in many of the GFP structures presented in this study. This mutant resulted in >20 conditions with poorly diffracting crystals. We were still able to determine the structures of three D190C mutants (PDB: 4W6G, 4W6H, and 4W6I). PDB: 4W6I was the most symmetric dimer (171°) while PDB: 4W6G and 4W6H were asymmetric at 141° and 135°, respectively (Figure 4G; Table 2).

Taking all the observed disulfide dimers together, we note that only two of these are perfectly symmetric by virtue of lying on crystallographic axes of 2-fold symmetry. Of those that did not fall on symmetry axes, another nine had internal angles between the chains >170° (11 of 36 disulfide dimers observed). The remaining majority of dimers were substantially asymmetric. This contrasts with the trend toward nearly symmetric dimers noted in earlier studies on synthetically symmetrized proteins (Banatao et al., 2006; Forse et al., 2011) that had been connected primarily through α-helical segments rather than a β-sheet conformation as in GFP.

Design and Structure of Metal-Mediated GFP Oligomers

In addition to disulfide dimerization, we explored the possibility of forming dimers or higher oligomers of GFP by designing metal-binding half-sites on its surface. This second approach follows from the work conducted by the groups of Tezcan and Kuhlman (Salgado et al., 2008, 2010; Der et al., 2012). Here, the idea is that introducing a metal half-site into the surface of a protein will lead to assembly upon addition of metal. The utility of the metal-mediated approach to synthetic symmetrization has been demonstrated before, whereby it was found that, in addition to the intended dimeric forms, varied modes of assembly can be realized upon metal addition (Laganowsky et al., 2011). Previous efforts exploring engineered metal-mediated oligomer formation have focused on mutations in α-helical proteins. In those cases, residues i and i+4 can be mutated to metal-chelating residues. The mutations are typically to His/His or His/Cys pairs in an attempt to replicate native chelation motifs. We investigated whether a variation of the approach could be applied to GFP, which consists mainly of a single β barrel. We selected residues in three distinct regions of the protein to mutate to either His/His or His/Cys pairs. These mutations were residues i and i+2 on one β strand (E124/K126) or two residues on adjacent strands (D21/K26 and E115/T118) (Figure 2C). To evaluate their ability to form oligomers in the presence of metal ions, we analyzed purified proteins in the presence of Cu^{2+}, Ni^{2+}, and Zn^{2+} salts using native gel-shift assays. Additional metal ions (e.g. Fe^{2+}, Fe^{3+}, Cd^{2+}, and Co^{3+}) were screened, but these metals either indicated no oligomer formation or had non-reproducible results by our native gel-shift assay and were not pursued for crystallization studies. We determined that mutant pairs D21H/K26C, E115C/T118H, E124H/K126C, and E124H/K126H were all able to form oligomers in the presence of each of the ions (Figure 2D). All of these mutant-metal combinations were then used for crystallization experiments to determine their ability to sample different space groups and form metal-mediated crystal contacts. Although D21H/K26H and E115H/T118H did not show shifts on the native gel assay, we proceeded with the crystallization experiments to determine whether they could still form metal-mediated contacts during the crystallization process.

From these metal-mediated variants we solved seven unique structures that were dependent on metal chelation to form. As with the disulfide and mixed disulfide-metal dimers, an ability to crystallize in a variety of conditions was observed. In a range of other cases, however, the metal ions established crystal contacts between different GFP molecules through a combination of the engineered residues and other native residues (typically Asp and Glu) on the protein surface. Only one of these structures (PDB: 4W7R) formed a symmetric dimer, whereas the other cases involved more complex spatial arrangements. In several cases, owing to low resolution and poor electron density, it was difficult to determine the exact chelation of the metal ion by the protein side chains. In some instances this likely results from exposure to synchrotron radiation, which can change the oxidation state of metal ions or damage carboxylic acid groups in the chelating aspartic acid side chains (Carugo and Carugo, 2005; Weik et al., 2000).

D21H/K26C

The designed metal half-site mutation D21H/K26C resulted in either disulfide dimers discussed previously or a mixed dimer containing the disulfide and a chelated metal ion (PDB: 4W75, 4W76, 4W77, 4W7A, and 4W7C). Adjacent to the disulfide bond, a copper ion is chelated by residues Asp19 and His21 from both participating protein chains (Figure 5B); the mutated histidine was intended for chelation whereas the aspartate was fortuitous. Some of the structures have poor electron density for the Asp19 and His21 side chains, and it appears in some instances that only one of the residues from each chain is involved in the metal chelation. Structures PDB: 4W76, 4W77, 4W7A, and 4W7C are close to being symmetric (average angle of 173.4°), with PDB: 4W75 being more asymmetric at a 152° inter-subunit rotation. The symmetric structures are very similar to each other, with a variation upon overlap of 2°–8° (Figure 4B; Table 2).

D21H/K26H

Two structures of this variant were obtained having copper-mediated crystal contacts. In PDB: 4W7E, Asp19 and His21 of one chain and Gln184 of the symmetry mate chelate the copper

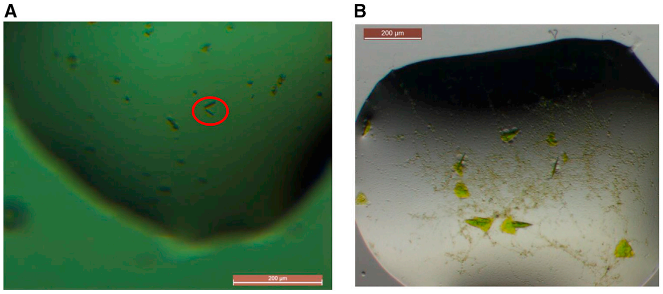


Figure 6. Crystals of Split-GFP with a Novel Crystallization Target

(A) Crystals of the STARD9-10/11-GFP1-9 (D21H/K26C) complex were obtained in a condition composed of 10% v/v 2-propanol, 0.1 M MES (pH 6.0), and 0.2 M $Ca(OAc)_2$. The protein complex was mixed in a 1:1 molar ratio with $CuSO_4$ immediately prior to the crystallization experiments. The green color of the crystals is used as an indication of the complex formation; the largest crystals observed to date (∼20 μM in the largest dimension) are highlighted by the red circle.

(B) Crystals of a designed protein with an engineered internal (10–11) hairpin in complex with GFP1-9 (D117C). The triangular plate crystals (∼50–75 μM) grew in a condition containing 0.1 M SPG buffer (pH 5.0) and 25% w/v PEG1500.

ion. This mutant crystallized in the presence of imidazole, leading to one imidazole molecule also being bound to the copper ion. Structure PDB: 4W7D features two different copper-mediated contacts (Figure 5C), and two chains are present in the asymmetric unit. Chain A makes contacts with two different protein molecules in the crystal using side chains that were engineered into this mutant. First, His21 and His26 chelate two copper ions and form a crystal contact to Lys3 of one neighboring molecule. A crystal contact to a different molecule is through Lys2 of chain A and Asp19 and His21 of the other protein, similar to the metal chelation observed in the D21H/K26C structures. The high pH (9.5) of this crystallization condition allows the lysine side chain to participate in the chelation of the copper ion.

E115C/T118H

In addition to the observed disulfide dimers of this mutant, structure PDB: 4W74 forms a complex system of metal-mediated crystal contacts between the eight protein chains in the asymmetric unit and six zinc ions via three different coordination sites (Figure 5D). The mutated Cys115/His118 half-site is found to chelate the zinc to a lone Cys115 in two cases; between chain A (Cys115/His118) and chain G (Cy115), and chain D (Cys115/His118) to chain F (Cys115). The Cys115/His118 half-site and an aspartic acid residue from a neighboring protein molecule chelate the other four zinc ions in arrangements that are generally similar to each other.

E115H/T118H

Two crystal forms of the E115H/T118H mutant with two different metal-mediated contacts were solved. PDB: 4W6U contains four chains in the asymmetric unit, yet only chains A and B feature a nickel-mediated contact. His118 of chain A and His115 of chain B are the residues responsible for metal chelation, along with a citrate molecule from the crystallization buffer (Figure 5E). A second nickel atom is chelated by residues His25 and Glu132 of chain A alone. In the structure PDB: 4W6T there is one chain in the asymmetric unit, which makes contact with other protein molecules through two copper ions (Figure 5F). His115 of the first chain and His25 and Glu132 of the symmetry mate chelate the first copper atom. His118 and Glu32 of the first chain and Asp133 of the symmetry mate chelate the second copper atom.

E124H/K126H

From the final mutant we determined two crystal structures, PDB: 4W7F and 4W7R. PDB: 4W7F contains one chain in the asymmetric unit with the copper-mediated contact formed between His124/His126 of the first chain and Glu5 of the symmetry mate (Figure 5G). The only symmetric metal-mediated dimer for which we obtained a structure is PDB: 4W7R. In this case the His124/His126 pair from chain A chelates a copper ion together with the His124/His126 pair from chain B. Two copper-mediated dimers (four subunits in total) are found in the asymmetric unit, and both dimers are nearly symmetric with chains orientated 179° apart. The two dimers are virtually identical, with only a 2° variation when aligned.

GFP Oligomers as a Crystallization Scaffold

After establishing in a previous study that a complex between the split-GFP(1–9) and a protein containing the (10–11) hairpin could form diffraction-quality crystals (Nguyen et al., 2014), we set out to crystallize a novel protein that had failed to crystallize in previous experiments. We attempted this with the motor domain of STARD9 (Torres et al., 2011), a monomeric kinesin that could serve as a target for novel anti-mitotic drug development. We co-expressed a construct of STARD9 as an N-terminal fusion to the GFP(10–11) hairpin together with the four metal-chelating GFP(1–9) mutants that consistently showed oligomerization in the native gel experiments (K26C/D21H, E124H/K126H, E124H/K126C, and E115C/T118H). Of the four experiments attempted, only K26C/D21H&E124H/K126H gave robust complementation. We were able to obtain crystals of the STARD9-10/11 and GFP1-9 (D21H/K26C) complex after approximately 3 months (Figure 6A). However, these crystals are small (∼20 μM in the largest dimension) and have not produced well-ordered diffraction to date; optimization efforts are under way.

A second computationally designed 271-amino-acid protein (to be published) containing the (10–11) hairpin as a loop insertion was co-expressed with the cysteine mutant suite of split-GFPs, all of which resulted in robust complementation. After 7 months, triangular plate crystals (∼50–75 μM) (Figure 6B) were observed containing the designed protein in complex with the GFP1-9 (D117C). As with the STARD9-10/11 constructs, optimization efforts of these crystals are under way.

DISCUSSION

The structural results presented here characterize a suite of engineered GFP molecules comprising a wide range of oligomeric

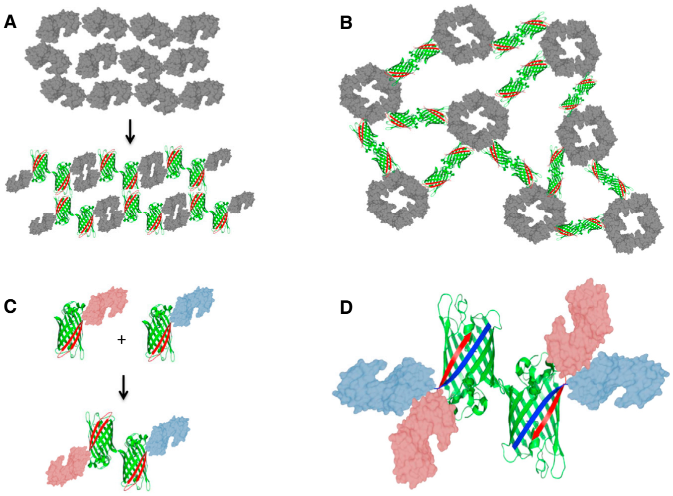


Figure 7. Alternative Applications for Engineered Oligomeric GFPs

Beyond their proposed utility as carriers for the crystallization of novel proteins, other potentially useful applications are possible.

(A) Fusion to GFP dimers could be used to change the crystal forms of existing proteins. Here a disordered crystal (top) can form a different and possibly better-ordered lattice (bottom) through fusion to one of the GFP oligomers in the available suite.

(B) Fusion to a multimeric enzyme, in this example a tetramer, could be used to create an enzymatically active amorphous gel for facile separation of enzymes and products for in vitro reaction systems.

(C) With the split form or through terminal fusions, the GFP dimers could be used to create a heterodimer for co-localization of enzymes for substrate channeling or co-crystallization experiments.

(D) Expanding on the idea from (C), two proteins can be forced into close proximity and further symmetrized, by separate genetic fusion of strand 10 to one protein and strand 11 to the other, then allowing them to complement for various applications.

forms, most of which appear highly amenable to crystallization on their own. We obtained 20 new crystal forms of seven disulfide-bonded dimers, plus 13 metal-mediated structures from five combinations of metal-chelating mutations in the presence of different metals. The 33 crystal forms are all distinct from each other (Table 1). Many of the engineered GFPs formed additional crystal forms in numerous conditions that were not pursued for structure determination. In analyzing individual GFP variants that were observed in multiple crystal forms, it was found that some of the oligomeric GFPs show strong geometric constraints between the disulfide-bonded subunits, while others display considerable geometric polymorphism. The K26C dimer was especially variable; among four instances observed for that dimer, the smallest angular deviation between any pair was 33°. The D21H/K26C and D102 mutants were the most rigid. Several instances of those dimers showed common conformations within about 8° deviation, although individual outliers were also obtained in both cases. The oligomeric GFP molecules designed here have not yet been used to successfully crystallize a target protein that was otherwise recalcitrant to crystallization. Which of the GFP constructs might ultimately be most useful in such a context is therefore presently unknown. However, it is notable that a few of the constructs formed an unusual number of distinct crystal forms readily. Among the disulfide-based dimers, the D117C construct formed the most (six) distinct crystal forms. Among the metal-mediated designs, the E115H/T118H and D21H/K26H constructs each also formed six distinct crystal forms.

The suite of oligomerizing GFP constructs designed here could be used for crystallizing target proteins by direct fusion. Alternatively, as noted above, our GFP constructs were engineered to be compatible with use in split form so that engineered variants of the GFP(1–9) construct can be reconstituted with a target protein bearing the (10–11) hairpin. In principle, this reconstitution can be performed in vivo (by co-expression) or in vitro (after separate purifications). Initial in vitro experiments using the split forms of the oligomerizing GFP constructs (not presented here) suggest that further optimization of the GFP(1–9) core may be important for stabilization in the context of the various mutations introduced into the GFP sequence. The counterbalancing advantages and disadvantages of the present system will also have to be compared with other strategies. For example, in some crystallization approaches the target protein is potentially stabilized by its fusion to an intact scaffold protein; attaching a small GFP fragment to a target protein (in the split-complementation approach) is not likely to provide such an advantage.

A principal long-term motivation for the present work is the crystallization of novel proteins, but other diverse applications in synthetic biology are likely to emerge for these oligomeric variants of GFP (Figure 7). One prospective application would be in attaching metabolically coupled enzymes together in different geometries through metal-mediated interactions or in vitro oxidized cysteines. They could be used as oligomerizing scaffolds for bringing together homo- or hetero-pairs of proteins into close proximity, in different spatial arrangements, and in ways that can be triggered by the addition of metal ions (Figures 7C and 7D). To promote formation of strictly heteromeric assemblies, future experiments would be required to design asymmetric versions of an oligomerizing carrier protein. Another avenue for future applications will be in using oligomerizing carrier proteins (GFP and others that could be developed) to drive other proteins or enzymes to form extended materials or amorphous gels (Figure 7B). While the motivating application emphasized in the present study (protein crystallization) applies primarily to target proteins that are naturally monomeric, we envision that extended materials, most likely with irregular structures, could be formed by complementing various oligomeric forms of the split-GFP(1–9) with naturally oligomeric proteins or enzymes bearing the (10–11) hairpin. In most cases this

Table 3. X-Ray Diffraction Data and Refinement Statistics

PDB:	4W69	4W6A	4W6B	4W6C	4W6D	4W6F	4W6G	4W6H	4W6I	4W6J	4W6K
Wavelength (Å)	1.0717	0.9789	0.97918	0.9789	0.9793	0.9792	0.9793	1.0717	1.0717	0.9793	1.0717
Resolution range (Å)	94.58–3.975 (4.117–3.975)	77.02–2.991 (3.098–2.991)	44.69–1.895 (1.963–1.895)	71.3–2.492 (2.581–2.492)	87.16–3.447 (3.57–3.447)	84.34–2.701 (2.798–2.701)	69.09–3.024 (3.132–3.024)	82.72–1.953 (2.023–1.953)	53.59–2.625 (2.719–2.625)	98.5–1.702 (1.763–1.702)	75.46–2.877 (2.98–2.877)
Space group	P 43 21 2	P 32 2 1	P 21 21 21	P 21 21 21	P 32 2 1	P 32 2 1	P 61	P 65	P 21 21 21	P 31 2 1	P 41 21 2
Unit cell	133.76 133.76 88.92 90 90 90	88.93 88.93 135.76 90 90 120	50.151 90.356 102.83 90 90 90	51.33 88.37 120.69 90 90 90	123.11 123.11 151.32 90 90 120	121.98 121.98 140.09 90 90 120	93.38 93.38 132.97 90 90 120	95.52 95.52 132.5 90 90 120	57.35 67.6 150.58 90 90 90	113.74,113.74 82.46 90 90,120	106.72 106.72 97.45 90 90 90
Total reflections	110,307 (10,414)	252,345 (23,254)	251,856 (13,789)	81,636 (7,060)	89,887 (8,957)	335,541 (32,238)	66,094 (6,525)	508,993 (47,571)	88,235 (6,817)	339,028 (33,029)	167,091 (13,967)
Unique reflections	7,344 (715)	12,990 (1,219)	37,298 (3,260)	19,311 (1,713)	17,869 (1,730)	33,538 (3,281)	12,859 (1,284)	49,488 (4,865)	17,334 (1,613)	67,258 (6,633)	13,200 (1,210)
Multiplicity	15.0 (14.6)	19.4 (19.1)	6.8 (4.2)	4.2 (4.1)	5.0 (5.2)	10.0 (9.8)	5.1 (5.1)	10.3 (9.8)	5.0 (4.2)	5.0 (5.0)	12.7 (11.5)
Completeness (%)	99.90 (99.31)	99.52 (95.08)	98.40 (87.52)	97.33 (88.79)	99.26 (98.69)	99.95 (99.64)	99.74 (99.46)	99.80 (98.06)	95.43 (82.54)	99.47 (98.82)	99.25 (93.51)
Mean $I/\sigma(I)$	16.9 (2.1)	20.3 (2.5)	12.3 (4.5)	7.3 (1.9)	11.91 (1.7)	5.6 (2.0)	16.7 (2.4)	16.4 (2.9)	5.9 (1.0)	16.6 (2.1)	22.1 (2.6)
Wilson B factor	162.1	72.3	20.6	60.3	112.8	64.4	95.5	25.4	54.4	25.2	91.6
R_{merge}	0.144 (1.663)	0.162 (1.385)	0.104 (0.376)	0.104 (0.660)	0.118 (1.023)	0.304 (0.578)	0.064 (0.729)	0.109 (0.819)	0.246 (1.047)	0.051 (0.702)	0.083 (1.265)
R_{meas}	0.149	0.167	0.113	0.118	0.132	0.321	0.072	0.114	0.273	0.057	0.087
$CC_{1/2}$	0.999 (0.714)	0.999 (0.767)	0.995 (0.871)	0.989 (0.881)	0.998 (0.607)	0.955 (0.871)	0.998 (0.853)	0.998 (0.811)	0.983 (0.853)	0.999 (0.782)	0.999 (0.811)
CC^*	1 (0.913)	1 (0.932)	0.999 (0.965)	0.997 (0.968)	1 (0.869)	0.988 (0.965)	1 (0.959)	1 (0.947)	0.996 (0.959)	1 (0.937)	1 (0.946)
R_{work}	0.307 (0.457)	0.191 (0.304)	0.167 (0.201)	0.248 (0.443)	0.236 (0.363)	0.204 (0.258)	0.248 (0.384)	0.166 (0.200)	0.268 (0.504)	0.189 (0.259)	0.249 (0.351)
R_{free}	0.335 (0.398)	0.240 (0.389)	0.202 (0.280)	0.276 (0.445)	0.267 (0.318)	0.238 (0.290)	0.270 (0.361)	0.190 (0.229)	0.316 (0.606)	0.212 (0.289)	0.294 (0.395)
No. of non-hydrogen atoms	3,458	3,574	3,867	3,553	3,550	3,604	3,505	3,884	3,558	3,925	3,037
Macromolecules	3,414	3,530	3,599	3,509	3,505	3,539	3,461	3,635	3,514	3,623	2,993
Ligands	44	44	47	44	45	65	44	44	44	96	44
Water	0	0	221	0	0	0	0	205	0	206	0
Protein residues	434	446	454	443	445	445	437	457	443	451	378
RMS (bonds)	0.011	0.018	0.009	0.01	0.01	0.011	0.011	0.011	0.014	0.011	0.011
RMS (angles)	1.29	1.6	1.08	1.29	1.69	1.37	1.37	1.21	1.3	1.21	0.95
Ramachandran favored (%)	97	97	98	96	98	97	97	98	97	99	97
Ramachandran outliers (%)	0	0.23	0	0.23	0.7	0.23	0.48	0	0	0	0.29
Clashscore	21.2	14.2	1.4	12.2	17.5	6.2	21.2	1.5	5.4	4.9	9.3
Average B factor	191.0	64.0	25.0	63.0	124.1	68.6	171.7	26.9	56.1	33.9	94.8
Macromolecules	191.5	64.1	24.9	63.2	124.3	68.4	171.7	26.8	56.3	33.5	35.1
Ligands	147.9	54.6	18.5	50	107.6	77.1	175.5	20.1	42.6	38.9	78.3
Solvent	–	–	27.8	–	–	–	–	30.7	–	37.1	–

(Continued on next page)

would lead to runaway oligomerization, yielding materials with potentially novel properties and uses. Other synthetic biology applications may benefit from higher-order oligomers. Based on our crystal structures, there are possible interfaces that could be mutated to achieve this purpose. As an example, a novel tetrameric form of GFP could be based on the structure of the D117C mutant PDB: 4W6M. This structure features a tetramer composed of two symmetric dimers in the asymmetric unit of the crystal. Further mutations in the region of the fortuitous interaction between dimers (residues I206, S146, and N147), either via metal-mediated interactions or by computational sequence design of a more extensive interface, could create a higher oligomeric form of GFP.

EXPERIMENTAL METHODS

Cloning

Unless otherwise stated, primers were ordered from Valuegene, enzymes were from New England Biolabs, and DNA sequencing was performed by Genewiz. The plasmid construct containing the split-GFP (Cabantous et al., 2005, 2013) was used as a template to generate a construct with a C-terminal hexahistidine tag and the C terminus: … TAAGITHHHHHH. The GFP gene was PCR amplified with Phusion DNA polymerase using the primers GFP.For and GFP.Rev, which include NdeI and HindIII restriction sites, respectively, in the primer extensions. The PCR-amplified segment was purified, digested with NdeI and HindIII, and ligated into pET24a, which had been restriction digested with the same two enzymes. Colony PCR using T7 and T7 terminator primers was performed to identify putative positive clones whose DNA sequences were subsequently confirmed by DNA sequencing. Two cysteine residues (Cys48, Cys70) were mutagenized to alanine using the primers C48A.For.New./C48A.Rev.New. and C70A/C70A_antisense to eliminate the possibility of unintended disulfide bonds. The C48A mutation was made by linear PCR amplification of the target vector with Phusion DNA polymerase, followed by DpnI digestion of the template plasmid and subsequent phosphorylation of the gel-extracted DNA with T4 polynucleotide kinase and ligation with T4 DNA ligase. The C70A mutation was made using Pfu Turbo AD polymerase (Agilent) using the Quikchange mutagenesis procedure. Additional mutations were made in the GFP construct containing the C48A/C70A mutations by the Quikchange method to generate the following GFP mutant proteins: C48A/C70A/D102C, C48A/C70A/D117C, C48A/C70A/Q157C, C48A/C70A/K26C, C48A/C70A/D190C, C48A/C70A/E124H/K126H, and C48A/C70A/E115C/T118H.

Table 3. *Continued*

4W6L	4W6M	4W6N	4W6O	4W6P	4W6R	4W6S	4W6T	4W6U	4W72	4W73	4W74
1.0717	0.9793	0.9537	0.9793	0.9793	0.9792	0.9789	0.9795	0.9795	0.9795	0.9789	0.9795
76.67–2.45 (2.538–2.45)	73.83–2.793 (2.893–2.794)	88.89–3.375 (3.496–3.375)	67.51–2.6 (2.693–2.6)	79.58–3.085 (3.195–3.085)	89.88–3.471 (3.595–3.471)	68.28–3.1 (3.211–3.1)	74.46–1.604 (1.661–1.604)	82.99–2.278 (2.36–2.278)	57.12–1.996 (2.067–1.996)	52.18–2.787 (2.887–2.787)	88.27–2.099 (2.174–2.099)
I 41 2 2	P 63	C 1 2 1	P 64 2 2	P 21 21 21	P 1	P 43 21 2	P 43 21 2	P 21 21 21	P 21 21 21	P 21 21 21	P 1 21 1
108.43 108.43 101.47 90 90 90	170.5 170.5 79.57 90 90 120	181.21 102.68 84.13 90 101.44 90	77.95 77.95 178.88 90 90 120	86.05 117.86 209.1 90 90 90	92.42 92.56 124.53 94.94 96.17 102.25	91.51 91.51 205.11 90 90 90	105.3 105.3 69.61 90 90 90	47.64 116.58 165.98 90 90 90	72.54 74.4 89.15 90 90 90	69.74 70.58 77.5 90 90 90	67.44 119.79 130.57 90 89.99 90
148,335 (14,256)	63,420 (1,852)	70,644 (6,362)	160,637 (16,471)	261,903 (23,019)	87,921 (8,257)	418,083 (42,445)	653,315 (63,113)	279,406 (25,897)	238,183 (20,955)	62,325 (4,525)	411,267 (39,095)
11,419 (1,106)	32,658 (378)	20,957 (1964)	10,518 (991)	39,143 (3,478)	46,082 (4,386)	16,549 (1,612)	51,580 (4,985)	42,930 (4,011)	33,399 (3,272)	9,864 (848)	119,482 (11,566)
13.0 (12.9)	5.7 (4.9)	3.4 (3.2)	15.3 (16.6)	6.7 (6.6)	1.9 (1.9)	25.3 (26.3)	12.7 (12.7)	6.5 (6.5)	7.1 (6.4)	6.3 (5.3)	3.4 (3.4)
99.83 (98.57)	99.15 (95.61)	97.89 (91.99)	99.94 (99.70)	97.98 (88.91)	89.44 (85.24)	99.95 (99.94)	99.78 (97.98)	99.28 (94.55)	99.74 (98.55)	98.67 (87.69)	98.81 (95.89)
23.9 (3.4)	4.8 (3.8)	6.6 (1.5)	26.0 (3.4)	10.8 (2.2)	5.7 (1.4)	18.0 (3.6)	17.8 (1.6)	14.2 (1.9)	16.6 (2.4)	10.1 (1.7)	8.1 (1.5)
61.8	71.3	84.6	68.9	70.7	101.5	76.2	26.5	44.6	41.2	73.9	32.2
0.064 (0.840)	0.766 (0.789)	0.205 (0.778)	0.072 (0.934)	0.149 (0.847)	0.095 (0.467)	0.257 (1.844)	0.072 (1.129)	0.105 (0.906)	0.060 (0.821)	0.109 (0.851)	0.104 (0.775)
0.067	0.823	0.244	0.075	0.162	0.134	0.262	0.075	0.115	0.065	0.119	0.123
0.999 (0.979)	0.683 (0.49)	0.992 (0.746)	0.999 (0.938)	0.994 (0.76)	0.989 (0.773)	0.998 (0.947)	0.999 (0.821)	0.997 (0.739)	0.999 (0.895)	0.996 (0.70)	0.996 (0.747)
1 (0.995)	0.901 (0.811)	0.998 (0.924)	1 (0.984)	0.999 (0.929)	0.997 (0.934)	1 (0.986)	1 (0.95)	0.999 (0.922)	1 (0.972)	0.999 (0.908)	0.999 (0.925)
0.252 (0.375)	0.261 (0.389)	0.316 (0.400)	0.262 (0.348)	0.232 (0.315)	0.307 (0.409)	0.223 (0.284)	0.180 (0.260)	0.210 (0.277)	0.190 (0.292)	0.221 (0.366)	0.212 (0.303)
0.278 (0.450)	0.285 (0.394)	0.363 (0.469)	0.311 (0.415)	0.279 (0.363)	0.357 (0.431)	0.276 (0.372)	0.207 (0.289)	0.250 (0.319)	0.235 (0.314)	0.297 (0.464)	0.235 (0.330)
1,635	6,752	10,419	1,662	12,960	25,002	3,538	2,074	7,317	3,817	3,519	14,583
1,613	6,652	10,331	1,637	12,828	25,002	3,442	1,846	7,083	3,570	3,469	14,200
22	100	88	22	132	0	96	69	111	45	50	182
0	0	0	3	0	0	0	159	123	202	0	201
205	842	1,306	208	1,618	3,133	436	227	890	450	437	1,793
0.01	0.011	0.009	0.013	0.009	0.008	0.012	0.019	0.012	0.011	0.013	0.013
1.23	1.2	1.49	1.85	1.42	1.76	1.55	1.7	1.36	1.24	1.42	1.51
96	99	95	98	97	96	95	99	98	98	97	98
0	0.12	0.08	0	0.13	0.44	0.47	0	0.12	0	0	0
8.1	14.0	30.0	20.8	12.6	20.0	21.6	6.3	4.5	3.2	13.7	7.2
93.1	91.1	33.4	106.4	80.1	113.2	84.2	32	47.4	46.4	67.8	39.1
93.3	91.4	99.6	106.6	80.2	113.2	84	31	47.7	46.2	67.9	39.1
79.2	72	72.8	97.2	70.6	–	91	42	36.4	42.4	55.9	36.3
–	–	–	72	–	–	–	39.7	43.8	50.5	–	37.1

(Continued on next page)

Proteins with an N-terminal tobacco etch virus (TEV) protease cleavable His6 tag were constructed by cloning the existing GFP mutants in pET24 into a modified pET28 vector with N-terminal cleavable tag to add the N-terminal sequence: MGSDKIHHHHHHHENLYFQG. In brief, the primers GFP.pMA507-star.For. and GFP.pMA507-star.Rev. were used to PCR amplify the mutated GFP DNA segments; the DNA was gel extracted and cloned into pMA507star by the Gibson ISO assembly method (Gibson et al., 2009). pMA507-star was PCR amplified with the primers PIPE.Vec.For. and PIPE.Vec.Rev. to generate compatible DNA overhangs. Primer sequences used are presented in Table S4.

Protein Expression

Plasmids containing mutant GFP genes were transformed into *Escherichia coli* BL21-DE3 expression cells (New England Biolabs). 10-ml starter cultures were grown with overnight shaking at 37°C in LB media containing appropriate antibiotics. The starter culture was used to inoculate 1 l of terrific broth medium supplemented with 20 ml 50× 5052 auto-induction sugars (Studier, 2005) and appropriate antibiotics. Cultures were grown for 4 hr at 37°C. The temperature was then reduced to 30°C, and cultures were allowed to grow for approximately 20 hr. After growth, the cultures were centrifuged at 5,000 × *g* for 30 min at 4°C. Harvested cell paste was stored at −80°C until purification.

Protein Purification

Cell paste was thawed at room temperature in a lysis buffer of 20 mM Tris (pH 8.0), 200 mM NaCl, 10 mM $MgCl_2$, 30 mM Imidazole, 400 μg/ml lysozyme, 10 μg/ml DNAse, and 1 mM AEBSF (4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride). Once the pellet was thawed, cells were lysed via sonication. Lysed cells were incubated at room temperature for 15 min prior to centrifugation to remove all insoluble material, and lysates were clarified at 25,000 × *g* for 30 min at 4°C. The soluble lysate fraction was applied to a 5 ml Ni-nitrilotriacetic acid (IMAC) column, rinsed with 10 column volumes of wash buffer consisting of 20 mM Tris (pH 8.0), 250 mM NaCl, and 30 mM imidazole. The protein was eluted from the column with wash buffer containing 250 mM imidazole. Elution fractions were pooled and then concentrated until the final volume was approximately 1 ml. For the disulfide dimers, the protein was exchanged into a buffer consisting of 20 mM Tris (pH 9.0) and 100 mM NaCl. Cysteines where then oxidized to form dimers by the addition of 10 ml of dimerization buffer (20 mM Tris [pH 9.0], 100 mM NaCl, 5 mM $CuSO_4$). This oxidation reaction was incubated at room temperature for 15 min before being quenched by the addition of 50 mM EDTA. To separate newly formed dimers from remaining monomers, the protein was dialyzed overnight at 4°C into anion exchange buffer (10 mM Tris [pH 9.5], 1 mM EDTA). The protein was applied to an anion exchange column and then eluted via a salt gradient of 0–1 M NaCl in anion exchange buffer. The major peak for each cysteine mutant was assessed for

Table 3. *Continued*

4W7X	4W75	4W76	4W77	4W7A	4W7C	4W7D	4W7E	4W7F	4W7R
0.9789	1.0717	0.9792	0.9789	0.9792	0.9795	0.9792	0.9792	0.9789	0.9789
66.77–2.8 (2.9–2.8)	69.13–3.47 (3.597–3.473)	60.5–2.345 (2.429–2.345)	60.79–3.1 (3.211–3.1)	96.28–3.603 (3.731–3.603)	96.15–2.5 (2.59–2.5)	66.57–1.799 (1.863–1.799)	67.92–2.592 (2.685–2.592)	48.76–2.9 (3.004–2.9)	92.07–1.799 (1.863–1.799)
P 1 21 1	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21	C 1 2 1	P 21 21 21	P 41 21 2	C 2 2 21	P 1 21 1
66.84 70.47 116.78 90 92.56 90	59.86 83.93 121.88 90 90 90	64.16 66.84 121 90 90 90	62.12 68.29 121.58 90 90 90	120.83 121.33 192.56 90 90 90	205.49 69.42 102.81 90 110.73 90	56.42 82.05 113.87 90 90 90	96.05 96.05 69.96 90 90 90	68.2 69.75 82.57 90 90 90	62.67 87.19 92.07 90 90.01 90
183,888 (18,230)	104,617 (8,741)	145,381 (13,147)	63,735 (6,380)	220,621 (21,172)	321,515 (31,440)	326,510 (29,374)	134,951 (13,211)	23,617 (1523)	306,630 (30,073)
26,887 (2,649)	8,254 (755)	22,306 (2,125)	9,841 (949)	33,244 (3,177)	46,757 (4,596)	49,634 (4,736)	10,580 (999)	4,558 (378)	90,790 (8,888)
6.8 (6.9)	12.7 (11.6)	6.5 (6.2)	6.5 (6.7)	6.6 (6.7)	6.9 (6.8)	6.6 (6.2)	12.8 (13.2)	5.4 (4.0)	3.4 (3.4)
99.66 (99.62)	98.78 (94.83)	99.23 (96.33)	99.87 (99.79)	99.40 (95.49)	99.06 (97.93)	99.57 (96.26)	99.68 (97.18)	99.52 (99.55)	98.83 (97.73)
8.2 (1.6)	13.6 (1.7)	10.1 (1.8)	9.9 (2.6)	13.9 (2.3)	15.6 (1.9)	8.8 (1.0)	19.8 (2.0)	8.0 (4.0)	6.4 (1.2)
57.1	125.6	47.9	69.4	112.5	63.2	25.9	58.9	82.4	23.5
0.190 (1.297)	0.150 (1.419)	0.110 (1.015)	0.151 (0.748)	0.122 (0.799)	0.073 (1.03)	0.138 (1.868)	0.113 (1.67)	0.178 (0.405)	0.122 (1.021)
0.205	0.156	0.119	0.165	0.133	0.079	0.15	0.118	0.196	0.145
0.991 (0.616)	0.999 (0.944)	0.998 (0.924)	0.996 (0.814)	0.998 (0.804)	0.999 (0.917)	0.997 (0.451)	0.999 (0.792)	0.98 (0.826)	0.994 (0.71)
0.998 (0.873)	1 (0.985)	1 (0.98)	0.999 (0.947)	0.999 (0.944)	1 (0.978)	0.999 (0.789)	1 (0.94)	0.995 (0.951)	0.998 (0.911)
0.217 (0.317)	0.301 (0.444)	0.233 (0.424)	0.217 (0.260)	0.278 (0.336)	0.226 (0.408)	0.179 (0.317)	0.207 (0.3534)	0.264 (0.404)	0.223 (0.376)
0.269 (0.386)	0.345 (0.377)	0.288 (0.459)	0.291 (0.377)	0.302 (0.337)	0.254 (0.428)	0.221 (0.342)	0.262 (0.441)	0.332 (0.439)	0.253 (0.425)
7,089	3,181	3,639	3,474	7,085	7,028	4,014	1,820	1,726	7,625
7,001	3,180	3,588	3,473	6,994	6,938	3,603	1,766	1,703	7,166
88	1	45	1	91	90	103	28	23	146
0	0	6	0	0	0	309	26	0	313
882	396	452	432	881	873	224	222	215	677
0.009	0.004	0.011	0.01	0.011	0.013	0.01	0.012	0.01	0.012
0.93	0.85	1.28	1.4	1.22	1.33	1.26	1.46	1.37	1.27
97	98	99	94	96	99	97	97	97	99
0.23	0	0	0.24	0.35	0	0	0	0	0
11.3	4.1	9.1	12.4	8.0	15.0	4.0	7.1	10.9	6.0
50.7	161.6	64.7	65.2	117.2	100.1	21.3	52.7	74.7	33.5
50.8	161.6	64.8	65.2	118.2	100.3	30.4	52.8	75	33.2
43.3	196.4	61.1	64.9	42.2	83.9	39.9	53.4	51	37.1
–	–	56.8	–	–	–	39	46.2	–	38

$CC_{1/2}$, correlation coefficient between intensities of crystallographic random half-datasets; CC*, correlation coefficient of the full dataset derived from $CC_{1/2}$.

dimer purity by non-reducing SDS-PAGE. Fractions of homogeneous dimers were pooled, buffer exchanged into GFP crystallization buffer (10 mM Tris, 100 mM NaCl), then concentrated to 20 mg/ml. Aliquots of protein were flash-frozen in liquid nitrogen and stored at −80°C for subsequent crystal trials.

Metal-mediated mutants were purified using the same method, up to the IMAC purification, where the hexahistidine tag was cleaved off with TEV protease overnight at 4°C in TEV cleavage buffer (10 mM Tris [pH 8.0], 100 mM NaCl, 5 mM DTT, 1 mM EDTA). Cleaved protein was then subject to a second IMAC step to remove the TEV protease, cleaved histidine tag, and any uncleaved protein. All unbound protein was pooled, buffer exchanged into crystallization buffer, concentrated to 40 mg/ml, flash-frozen, and stored at −80°C for future crystal trials.

Co-expression with Target Proteins

The STARD9-10/11 construct consisted of the N-terminal TEV protease cleavable His6 tag (MGSDKIHHHHHHENLYFQG) followed by the (10–11) hairpin sequence, DLPDDHYLSTQTILSKDLNEKRDHMVLLEYVTAAGITDAS, with the "DAS" serving as a linker between the hairpin and target protein as previously described (Nguyen et al., 2014). Only the first 391 amino acids (Met1–Asn391) corresponding to the putative motor domain of the protein were used in this construct.

For the prospective designed protein construct, the GFP (10–11) hairpin was inserted into a presumptive loop between Ser135 and Thr136 of the native 271-amino-acid protein. This construct features a non-cleavable C-terminal His6 tag, and as such was not used for the metal-mediated experiments.

The expression and purification methods for the co-expressed GFP(1–9) and crystallization targets with the (10–11) hairpin were essentially the same as for the GFPs alone. After size-exclusion chromatography, the fractions with approximate 1:1 molar ratio of GFP(1–9) and target protein (visualized by SDS-PAGE) were used for the crystallography experiments.

Crystallization

The GFP oligomers were crystallized using hanging-drop vapor diffusion. Initial experiments were carried out at the UCLA crystallization facility using commercial sparse matrix screens in a 96-well format. All initial screening trays were set using a Mosquito liquid handling device (TPP LabTech). Limited optimizations were performed manually in some cases using 24-well Linbro plates. Each disulfide dimer was screened initially with four commercial sparse matrix screens JCSG+ (Qiagen), SaltRx (Hampton Research), Crystal Screen I + II (Hampton Research), and Wizard I + II (EmeraldBio). Metal-mediated mutations were screened with JCSG+ and Wizard only. The final concentration of protein in all crystallization experiments was 20 mg/ml. Metal-mediated mutants were mixed with the metal ions (Ni^{2+}, Zn^{2+}, or Cu^{2+}, in three separate screens) immediately before setting crystal trays, at a final concentration of 20 mg/ml protein and 2 mM metal ion salts. Trays were set at room temperature and checked periodically over 30 days. Single crystals were mounted with CrystalCat HT Cryoloops (Hampton Research), cryoprotected as needed, flash-frozen with liquid nitrogen, and screened for diffraction. All diffracting crystals were stored for later data collection. All diffraction data were collected at 100 K at APS-NECAT beamline 24-ID-C on a DECTRIS-PILATUS 6M detector. The crystallization and cryoprotectant conditions are reported in Table S5.

Structure Determination

Datasets from individual crystals were indexed, integrated, and scaled using XDS/XSCALE (Kabsch, 2010), with the resolution limit selected to balance completeness, calculated I/σ, R_{sym}, and $CC_{1/2}$ of the highest-resolution shell with emphasis on I/σ values of >1.5 and $CC_{1/2}$ values of >0.9. Structures were solved by molecular replacement using the program Phaser (McCoy et al., 2007), with the superfolder GFP (Pédelacq et al., 2006) protein (PDB: 2B3P) as the search model. To accelerate the model building and refinement, molecular replacement solutions were initially refined with the PDB_REDO server (Joosten et al., 2011). Final iterative rounds of model building and refinement were carried out using Coot (Emsley et al., 2010) and PHENIX (Adams et al., 2010) with TLS refinement (Painter and Merritt, 2006). Structures were validated with PROCHECK (Laskowski et al., 1993), ERRAT (Colovos and Yeates, 1993), MolProbity (Davis et al., 2007), and VERIFY3D (Luthy et al., 1992). Atomic coordinates and structure factors for all 33 structures were deposited in the PDB. Figures depicting the structures were made with PyMOL (Schrödinger). Data collection and refinement statistics are given in Table 3.

Structure Comparison Procedure

To compare multiple observed instances of the same disulfide-bonded dimer, one structure was first chosen as the reference. Then one chain of a subsequent dimer was aligned to chain A of the reference dimer, and the transformation required for overlapping those two chains was applied to the second chain. Both possible assignments to chain A versus chain B were tested for each dimer, and the best match was retained for comparison. These optimal chain assignments do not necessarily correspond to chain assignments in the deposited PDB files.

ACCESSION NUMBERS

Atomic coordinates and structure factors for all 33 structures in this study were deposited, under accession numbers PDB: 4W69, 4W6A, 4W6B, 4W6C, 4W6D, 4W6F, 4W6G, 4W6H, 4W6I, 4W6J, 4W6K, 4W6L, 4W6M, 4W6N, 4W6O, 4W6P, 4W6R, 4W6S, 4W6T, 4W6U, 4W72, 4W73, 4W74, 4W7X, 4W75, 4W76, 4W77, 4W7A, 4W7C, 4W7D, 4W7E, 4W7F, and 4W7R.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five tables and can be found with this article online at http://dx.doi.org/10.1016/j.str.2015.07.008.

ACKNOWLEDGMENTS

This work was supported by NIH grant P01 GM098177 (to T.C.T.). D.J.L. was supported by Ruth L. Kirschstein National Research Service Award T32GM007185. The authors thank Michael Sawaya, Duilio Cascio, and Michael Thompson for X-ray data collection at APS beamline 24-ID-C. We thank Michael Collazo for help with the crystallization trials, and Dan McNamara for help with structure determinations. The UCLA macromolecular structure facilities are supported by the BER program of the DOE Office of Science (award DE-FC02-02ER63421). We thank David Baker and Fabio Parmeggiani for providing the designed protein as a target for fusion-based crystallization experiments. We thank the staff of the NECAT synchrotron beamline, including Jon Schuermann, Igor Kourinov, and Malcolm Capel, and for helpful discussions. X-ray data collection was supported by DOE Grant DE-FC02-02ER63421 and the NECAT beamlines of the Advanced Photon Source, which are supported by NIH Grant RR-15301(NCRR). Use of the Advanced Photon Source is supported by the DOE, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357.

Received: February 25, 2015
Revised: June 8, 2015
Accepted: July 7, 2015
Published: August 13, 2015

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Banatao, D.R., Cascio, D., Crowley, C.S., Fleissner, M.R., Tienson, H.L., and Yeates, T.O. (2006). An approach to crystallizing proteins by synthetic symmetrization. Proc. Natl. Acad. Sci. USA *103*, 16230–16235.

Cabantous, S., Terwilliger, T.C., and Waldo, G.S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat. Biotechnol. *23*, 102–107.

Cabantous, S., Nguyen, H.B., Pedelacq, J.D., Koraïchi, F., Chaudhary, A., Ganguly, K., Lockard, M.A., Favre, G., Terwilliger, T.C., and Waldo, G.S. (2013). A new protein-protein interaction sensor based on tripartite split-GFP association. Sci. Rep. *3*, 2854.

Carugo, O., and Carugo, K.D. (2005). When X-rays modify the protein structure: radiation damage at work. Trends Biochem. Sci. *30*, 213–219.

Chruszcz, M., Potrzebowski, W., Zimmerman, M.D., Grabowski, M., Zheng, H., Lasota, P., and Minor, W. (2008). Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter? Protein Sci. *17*, 623–632.

Colovos, C.Y., and Yeates, T.O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. *2*, 1511–1519.

Conrado, R.J., Varner, J.D., and DeLisa, M.P. (2008). Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. Curr. Opin. Biotechnol. *19*, 492–499.

Corsini, L., Hothorn, M., Scheffzek, K., Sattler, M., and Stier, G. (2008). Thioredoxin as a fusion tag for carrier-driven crystallization. Protein Sci. *17*, 2070–2079.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. *35*, W375–W383.

Der, B.S., Machius, M., Miley, M.J., Mills, J.L., Szyperski, T., and Kuhlman, B. (2012). Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. J. Am. Chem. Soc. *134*, 375–385.

Derewenda, Z.S., and Vekilov, P.G. (2006). Entropy and surface engineering in protein crystallization. Acta Crystallogr. D Biol. Crystallogr. *62*, 116–124.

Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A.V., Prather, K.L., and Keasling, J.D. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. Nat. Biotechnol. *27*, 753–759.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Forse, G.J., Ram, N., Banatao, D.R., Cascio, D., Sawaya, M.R., Klock, H.E., Lesley, S.A., and Yeates, T.O. (2011). Synthetic symmetrization in the crystallization and structure determination of CelA from *Thermotoga maritima*. Protein Sci. *20*, 168–178.

Gibson, H.W. (1969). Chemistry of formic acid and its simple derivatives. Chem. Rev. *69*, 673–692.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Good, M.C., Zalatan, J.G., and Lim, W.A. (2011). Scaffold proteins: hubs for controlling the flow of cellular information. Science *332*, 680–686.

Huang, Y.M., and Bystroff, C. (2009). Complementation and reconstitution of fluorescence from circularly permuted and truncated green fluorescent protein. Biochemistry *48*, 929–940.

Joosten, R.P., Joonsten, K., Cohen, S.X., Vriend, G., and Perrakis, A. (2011). Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank. Bioinformatics *27*, 3392–3398.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Katz, B.A., and Kossiakoff, A. (1986). The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. J. Biol. Chem. *261*, 15480–15485.

Laganowsky, A., Zhao, M., Soriaga, A.B., Sawaya, M.R., Cascio, D., and Yeates, T.O. (2011). An approach to crystallizing proteins by metal-mediated synthetic symmetrization. Protein Sci. *20*, 1876–1890.

Lai, Y.T., King, N.P., and Yeates, T.O. (2012). Principles for designing ordered protein assemblies. Trends Cell Biol. *22*, 653–661.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck—a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Lee, H., DeLoache, W.C., and Dueber, J.E. (2012). Spatial organization of enzymes for metabolic engineering. Metab. Eng. *14*, 242–251.

Luthy, R., Bowie, J.U., and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. Nature *356*, 83–85.

March, J.C., Rao, G., and Bentley, W.E. (2003). Biotechnological applications of green fluorescent protein. Appl. Microbiol. Biotechnol. *62*, 303–315.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Moon, A.F., Mueller, G.A., Zhong, X., and Pedersen, L.C. (2010). A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. Protein Sci. *19*, 901–913.

Nguyen, H.B., Hung, L.W., Yeates, T.O., Terwilliger, T.C., and Waldo, G.S. (2014). Split green fluorescent protein as a modular binding partner for protein crystallization. Acta Crystallogr. D Biol. Crystallogr. *69*, 2513–2523.

Painter, J.M., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr. D Biol. Crystallogr. *62*, 439–450.

Pédelacq, J.D., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat. Biotechnol. *24*, 79–88.

Salgado, E.N., Lewis, R.A., Faraone-Mennella, J., and Tezcan, F.A. (2008). Metal-mediated self-assembly of protein superstructures: influence of secondary interactions on protein oligomerization and aggregation. J. Am. Chem. Soc. *130*, 6082–6084.

Salgado, E.N., Ambroggio, X.I., Brodin, J.D., Lewis, R.A., Kuhlman, B., and Tezcan, F.A. (2010). Metal templated design of protein interfaces. Proc. Natl. Acad. Sci. USA *107*, 1827–1832.

Stacy, R., Begley, D.W., Phan, I., Staker, B.L., Van Voorhis, W.C., Varani, G., Buchko, G.W., Stewart, L.J., and Myler, P.J. (2011). Structural genomics of infectious disease drug targets: the SSGCID. Acta Crystallogr. F Struct. Biol. Cryst. Commun. *67*, 979–984.

Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Sundstrom, M., Norin, M., and Edwards, A. (2006). Structural Genomics and High Throughput Structural Biology (CRC Press).

Suzuki, N., Hiraki, M., Yamada, Y., Matsugaki, N., Igarashi, N., Kato, R., Dikic, I., Drew, D., Iwata, S., Wakatsuki, S., and Kawasaki, M. (2010). Crystallization of small proteins assisted by green fluorescent protein. Acta Crystallogr. D Biol. Crystallogr. *66*, 1059–1066.

Torres, J.Z., Summers, M.K., Peterson, D., Brauer, M.J., Lee, J., Senese, S., Gholkar, A.A., Lo, Y.C., Lei, X., Jung, K., Anderson, D.C., et al. (2011). The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly. Cell *147*, 1309–1323.

Weik, M., Ravelli, R.B., Kryger, G., McSweeney, S., Raves, M.L., Harel, M., Gros, P., Silman, I., Kroon, J., and Sussman, J.L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. Proc. Natl. Acad. Sci. USA *97*, 623–628.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.

Zeke, A., Lukac, M., Lim, W.A., and Remenyi, A. (2009). Scaffolds: interaction platforms for cellular signalling circuits. Trends Cell Biol. *19*, 364–374.

Zou, Y., Weiss, W.I., and Kobilka, B.K. (2012). N-Terminal T4 lysozyme fusion facilitates crystallization of a G protein coupled receptor. PLoS One *7*, e46039.

Structure, Volume *23*

Supplemental Information

A Suite of Engineered GFP Molecules for Oligomeric Scaffolding

David J. Leibly, Mark A. Arbing, Inna Pashkov, Natasha DeVore, Geoffrey S. Waldo, Thomas C. Terwilliger, and Todd O. Yeates

Table S1. Disulfide bond dihedral angle energy server output, Related to Table 2.

PDB	Cys1	Chi1(X1)	Chi2(X2)	Chi3(X3)	Bond Distance (Å)	Chi2'(X2')	Chi1'(X1')	Cys2	Disulfide Strain Energy (kJ/mol)
4W69	157:A:	-154.01	30.11	-178.47	2.04	-89	-77.54	157:B:	47.3
4W6A	157:A:	-177.00	73.24	38.39	2.31	73.28	-177.02	157:B:	21.6
4W6B	26:A:	-72.82	86.33	78.75	2.05	98.94	78.42	26:B:	17.1
4W6C	26:A:	75.31	78.48	106.19	2.07	75.11	67.4	26:B:	12.9
4W6D	26:A:	55.17	-177.57	143.87	2.05	93.86	-48.99	26:B:	29.1
4W6F	26:A:	-72.5	-171.2	-78.52	2.04	132.24	-75.52	26:B:	14.6
4W6G	190:A:	-70.19	-42.67	-92.93	2.17	-86.15	-59.9	190:B:	9.1
4W6H	190:A:	-31.17	-105.20	-87.99	2.05	-35.15	-86.13	190:B:	27.1
4W6I	190:A:	-64.99	142.91	100.01	2.03	151.48	-63.81	190:B:	14.7
4W6J	117:A:	-54.02	-55.14	-92.71	2.02	-109.94	-61.84	117:B:	11.3
4W6K	117:A:	-75.04	-66.27	-127.5	2.23	-18.1	-84.4	117:B:	30.9
4W6L	117:A:	-41.90	-81.68	-113.18	1.93	-81.68	-41.90	117:B:	21.2
4W6M	117:A:	-83.09	-48.39	-114.58	2.03	-73.8	-82.88	117:C:	22.5
4W6M	117:B:	-163.74	46.93	159.96	2.03	-94.13	-83.79	117:D:	41.7
4W6N	117:A:	-42.79	91.35	-150.87	2.05	-85.41	55.02	117:D:	35.9
4W6N	117:B:	-66.27	82.17	106.38	2.03	145.25	-65.11	117:F:	15.2
4W6O	117:A:	79.69	56.05	-110.65	2.03	56.05	-79.69	117:B:	16.7
4W6P	102:A:	-82.1	-7.53	-124.45	2.03	47.95	-89.95	102:B:	36.5
4W6P	102:C:	-87.56	45.31	-121.68	2.02	-5.68	-80.3	102:D:	34.0
4W6P	102:E:	-78.91	-28.57	-128	2.01	87.91	-152.25	102:H:	35.3
4W6P	102:F:	-65.00	-37.24	-62.86	2.02	5.98	-64.95	102:G:	17.5
4W6R	102:A:	-83.29	-8.38	-87.58	2.11	-124.65	70.4	102:N:	25.1
4W6R	102:B:	-69.22	16.41	-114.52	1.96	15.91	-70.74	102:I:	26.1
4W6R	102:C:	-66.76	15.51	-110.58	2.02	21.99	-69.48	102:D:	22.8
4W6R	102:E:	-66.97	26.66	-124.7	2.18	27.85	-78.86	102:J:	28.3
4W6R	102:F:	-68.49	-16.22	-164.19	1.99	93.94	-99.35	102:O:	53.0
4W6R	102:G:	-27.15	-88.45	135.51	1.99	-179.47	-69.5	102:P:	33.6
4W6R	102:H:	-55.13	-32.47	-127.85	1.98	111.91	-158.85	102:M:	32.4
4W6R	102:K:	-67.4	18.31	-113.65	2.02	17.26	-67.04	102:L:	24.1
4W6S	126:A:	-80.49	-84.83	-70.22	2.04	-79.74	-167.69	126:B:	14.9
4W7X	115:A:	-60.4	-30.1	-96.94	2.04	-132.62	161.69	115:B:	19.0
4W7X	115:C:	-72.82	-80.96	105.38	2.23	106.57	55.01	115:D:	18.1
4W72	115:A:	174.93	-130.88	-90.68	2.05	-58.13	-62.57	115:B:	10.7
4W73	115:A:	-176.42	-137.86	-86.8	2.14	-53.57	-68.97	115:B:	10.1
4W75	26:A:	-74.50	110.78	152.70	2.05	-121.01	-26.35	26:B:	53.8
4W7C	26:A:	66.77	107.3	-162.69	2.01	-102.12	-171.58	26:B:	43.1
4W7C	26:C:	78.93	71.48	80.82	1.99	92.15	74.27	26:D:	13.5
4W76	26:A:	86.6	82.03	91.34	1.92	72	75.12	26:B:	15.4
4W77	26:A:	-161.27	-84.86	-39.65	2.04	-98.47	-157.73	26:B:	36.5
4W7A	26:A:	-150.96	-114.06	-177.41	2.03	85.41	165.7	26:B:	50.8
4W7A	26:C:	174.67	-96.21	-159.9	2.03	124.31	81.24	26:D:	45.8

Values of the disulfide bond dihedral angles and calculated bond energies of each disulfide dimer as determined by the Disulfide Bond Dihedral Angle Energy Server (http://services.mbi.ucla.edu/disulfide/).

Table S2. Angular variations between dimers in pairwise comparisons, Related to Table 2.

D102C	4W6R_BI	4W6R_CD	4W6R_EJ	4W6R_FO	4W6R_GP	4W6R_HM	4W6R_KL	4W6P_AB	4W6P_CD	4W6P_EH	4W6P_FG
4W6R_AN	4	7.3	2.7	3.5	3.4	6	7.7	25.7	24	25.2	28.5
4W6R_BI		5.7	3.9	1.7	2.1	2.3	5.9	23.8	21.8	23.4	27.2
4W6R_CD			5.1	4.8	7	4.6	1.2	28.6	26.3	28.1	32
4W6R_EJ				3.2	4.6	5.2	5.3	27.4	25.5	26.9	30.5
4W6R_FO					2.2	2.7	5.3	24.9	22.9	24.4	28.1
4W6R_GP						3.9	7.5	23	21.1	22.4	26
4W6R_HM							4.8	24.1	21.8	23.7	27.7
4W6R_KL								28.8	26.5	28.4	32.4
4W6P_AB									3	2.6	6
4W6P_CD										3.7	8.3
4W6P_EH											4.7

D190C	4W6H	4W6G
4W6I	41.4	38.1
4W6H		6.3

K26C	4W6B	4W6F	4W6D
4W6C	136.1	33.3	20.7
4W6B		137.1	140.4
4W6F			19

Q157C	4W6A_A	4W6A_B
4W69	49	129
4W6A_A		95.8

D117C	4W6N_AD	4W6N_BF	4W6N_CE	4W6M_AC	4W6M_BD	4W6L	4W6K	4W6J
4W6O	32.6	34.2	34	12	11.1	7	16.4	25.7
4W6N_AD		2	1.6	33.4	31.9	33.2	20.2	9.5
4W6N_BF			1	34.8	33.3	34.7	21.6	10.8
4W6N_CE				34.7	33.4	34.6	21.7	10.3
4W6M_AC					4.1	5.3	13.7	29
4W6M_BD						4.8	11.8	27.7
4W6L							14.1	27.8
4W6K								17.6

E115C	4W7X_AB	4W7X_CD	4W72
4W73	8.7	10.3	12.3
4W7X_AB		6.7	8.6
4W7X_CD			4.6

21hc	4W7A_AB	4W7A_CD	4W76	4W7C_AB	4W7C_CD	4W75
4W77	6.9	6.1	4	6.9	5.4	30.7
4W7A_AB		8.6	5.5	5.2	2.5	24.2
4W7A_CD			3.3	6.7	6.3	32.1
4W76				4.6	3.3	29.5
4W7C_AB					4.9	27.8
4W7C_CD						26.3

The values shown are in degrees. The structures being compared are designated by their PDB code followed by the chain identifiers for the two subunits in a dimeric arrangement.

Table S3. RMS coordinate deviations between dimers arising from rotational angle variations, Related to Table 2.

D102C	4W6R_BI	4W6R_CD	4W6R_EJ	4W6R_FO	4W6R_GP	4W6R_HM	4W6R_KL	4W6P_AB	4W6P_CD	4W6P_EH	4W6P_FG
4W6R_AN	1.7	2.2	0.9	1.6	0.9	1.7	2.7	8.5	8.3	6.8	9
4W6R_BI		1.9	1.3	0.6	1.1	0.8	1.9	8.1	7.8	6.8	9
4W6R_CD			1.6	2.1	2	1.5	0.9	9.8	9.5	8.3	10.5
4W6R_EJ				1.2	0.9	1.2	2	8.7	8.4	7.2	9.3
4W6R_FO					1	1	2.1	7.9	7.6	6.6	8.8
4W6R_GP						1.2	2.4	8.1	7.9	6.5	8.7
4W6R_HM							1.5	8.5	8.2	7.1	9.3
4W6R_KL								9.7	9.4	8.4	10.7
4W6P_AB									0.6	2.7	2.8
4W6P_CD										2.7	3.1
4W6P_EH											2.5

D190C	4W6H	4W6G
4W6I	20	19
4W6H		3.9

K26C	4W6B	4W6F	4W6D
4W6C	36.4	8.2	6.2
4W6B		35.3	36.74
4W6F			7.7

Q157C	4W6A_A	4W6A_B
4W69	19.6	50.9
4W6A_A		38

D117C	4W6N_AD	4W6N_BF	4W6N_CE	4W6M_AC	4W6M_BD	4W6L	4W6K	4W6J
4W6O	10.9	12	11.7	2.6	2.7	1.6	6	10.1
4W6N_AD		1.5	1.3	11.7	9.9	11	5.8	3.8
4W6N_BF			0.4	12.8	10.9	12	6.6	3.5
4W6N_CE				12.5	10.7	11.8	6.4	3.4
4W6M_AC					2.5	1.7	6.5	11.2
4W6M_BD						1.9	4.4	9.3
4W6L							5.7	10.3
4W6K								5.3

E115C	4W7X_AB	4W7X_CD	4W72
4W73	2.5	3.4	3.8
4W7X_AB		2.7	2.7
4W7X_CD			1.5

21hc	4W7A_AB	4W7A_CD	4W76	4W7C_AB	4W7C_CD	4W75
4W77	2.6	2.6	1.8	2.5	2.2	16.3
4W7A_AB		3.7	2.1	1.3	0.8	14
4W7A_CD			1.7	3.1	3.1	17.3
4W76				1.6	1.5	15.9
4W7C_AB					1.2	14.7
4W7C_CD						14.6

The values shown are in Angstroms. The RMSD values represent the deviation of the Cα alignments used for the pairwise comparisons presented in Table S2.

Table S4. Cloning primers, Related to Methods.

Primer Name	Primer Sequence	
GFP.For.	5’-ggaattacatatgaggaaaggagaagaac-3'	
GFP.Rev.	5’-ttttttaagcttctattaatggtgatggtgatgatgtgtaatcccagcagcagttac-3'	
C48A.For.New.	5’-gccactactggaaaactacctgttcc-3'	
C48A.Rev.New.	5’-aataaatttaaggctgagttttccg-3'	
C70A	5'-tctgacctatggtgttcaagccttttcccgttatccggat-3'	
C70A_antisense	5'-atccggataacgggaaaaggcttgaacaccataggtcaga-3'	
D21H	5'-caattcttattgaattagatggtcatgttaatgggcactgctttttt-3'	C48A/K26C/D21H
D21H_antisense	5'-aaaaaagcagtgcccattaacatgaccatctaattcaataagaattg-3'	
D102C	5'-ttatgtacaggaacgcactatatatttcaaatgtgacgggacctacaag-3'	(C48A/C70A/D102C)
D102C_antisense	5'-cttgtaggtcccgtcacatttgaaatatatagtgcgttcctgtacataa-3'	
D117C	5'-tgctgaagtcaagtttgaaggttgtacccttgttaatcgtatcgag-3'	(C48A/D117C/C70A)
D117C_antisense	5'-ctcgatacgattaacaagggtacaaccttcaaacttgacttcagca-3'	
Q157C	5'-cacaaagtatacatcacggcagacaaatgcaataatggaatcaaagctaacttcaca-3'	C48A/C70A/Q157C)
Q157C_antisense	5'-tgtgaagttagctttgattccattattgcatttgtctgccgtgatgtatactttgtg-3'	
K26C	5'-gatggtgatgttaatgggcactgctttttttgtccgtggagagggt-3'	(C48A/K26C/C70A)
K26C_antisense	5'-accctctccacggacaaaaaagcagtgcccattaacatcaccatc-3'	
D190C	5'-aacaaaatactccaattggctgtggccctgtccttttaccag-3'	(C48A/D190C/C70A)
D190C_antisense	5'-ctggtaaaaggacagggccacagccaattggagtattttgtt-3'	
E124H.K126H.For.	5'-agtttgaaggtgatacccttgttaatcgtatccatttacatggtattgattttaaagaagatggaaacattc-3'	C48A/C70A/E124H/K126H
E124H.K126H.Rev.	5'-gaatgtttccatcttctttaaaatcaataccatgtaaatggatacgattaacaagggtatcaccttcaaact-3'	
E115H.T118H.For.	5'-acaagacgcgtgctgaagtcaagtttcatggtgatcaccttgttaatcgtatcg-	C48A/C70A/E115H/T118H
E115H.T118H.Rev.	5'-cgatacgattaacaaggtgatcaccatgaaacttgacttcagcacgcgtcttgt-3	
H115C.For.	5'-caagacgcgtgctgaagtcaagttttgtggtgatcacctt-3'	Used with above primers to make: C48A/C70A/E115C/T118H
H115C.Rev.	5'-aaggtgatcaccacaaaacttgacttcagcacgcgtcttg-3'	
GFP.pMA507-star.For.	5’- aaaacctgtacttccagggcatgaggaaaggagaagaacttttcac-3'	
GFP.pMA507-star.Rev.	5’-aacgagttaattaagtcgcgttatgtaatcccagcagcagttacatac-3'	
PIPE.Vec.For.	5’-cgcgacttaattaactcgtttaaacggtctccagc-3'	
PIPE.Vec.Rev.	5’-ctggaagtacaggttttcgtgatgatgatgatgatg-3'	

Sequences of the primers used for cloning the suite of GFP mutants

Table S5. Crystallization and cryo-protectant conditions, Related to Table 1.

PDB	Crystal Condition	Cryo Protectant
4W69	0.4M MgFormate, 0.1M Acetate pH 4.6, 2%w/v benzamidine	25%v/v Glycerol
4W6A	2.0M NaFormate, 0.1M NaAcetate pH 4.6	25%v/v Glycerol
4W6B	14%w/v PEG-4000, 0.2M MgCl2, 0.1M Tris pH 8.5	25%v/v Glycerol
4W6C	35%v/v MPD, 0.1M Imidazole pH 8.0, 0.2M MgCl2	--
4W6D	1.5M MgSO4, 0.5%w/v Glycerol, 0.1M MES pH 6.75	30%v/v Glycerol
4W6F	10%v/v 2-propanol, 0.1M Imidazole pH 8.0	25%v/v Glycerol
4W6G	20%v/v 1,4-Butanediol, 0.1M Acetate pH 4.5	20%v/v Glycerol
4W6H	0.1M SPG Buffer pH8.0, 25%w/v PEG-1500	--
4W6I	1.4M MgSO4, 0.1M BTP pH 7.4	25%v/v Glycerol
4W6J	35% MPD, 0.1M NaAcetate pH 4.5	--
4W6K	0.5M KSCN, 0.1M NaAcetate pH 4.6	25%v/v Glycerol
4W6L	1.5M NaNO3, 0.1M NaAcetate pH 5.0	25%v/v Glycerol
4W6M	10%w/v PEG3350, 0.1M NaAcetate pH 4.6, 0.2M NaCl	25%v/v Glycerol
4W6N	10%w/v PEG6000, 0.1M Hepes pH 6.5	20%v/v Glycerol
4W6O	20%w/v PEG6000, 0.1M Bicine pH 8.5	20%v/v Glycerol
4W6P	1.3M NaNO3, 0.1M NaAcetate pH 5.0	25%v/v Glycerol
4W6R	20%w/v PEG3350, 0.2M NaSCN	25%v/v Glycerol
4W6S	40%w/v PEG300, 0.1M Phosphate-citrate pH 4.2	20%v/v Glycerol
4W6T	0.15M Kbr, 30%w/v PEG MME 2000	25%v/v Ethylene glycol
4W6U	0.2M NaCl, 0.1M Phosphate-citrate pH 4.2, 20%w/v PEG8000	25%v/v Ethylene glycol
4W72	20%w/v PEG3000, 0.1M Acetate pH 4.5	25%v/v Glycerol
4W73	20%w/v PEG1000, 0.1M Imidazole pH8.0, 0.2M Ca(OAc)2	--
4W74	17%w/v PEG10000, 0.1M NH4(OAc), 0.1M Bis-Tris pH 5.5	25%v/v Glycerol
4W7X	1.0 M (NH4)2HPO4, acetate pH 4.5	25%v/v Glycerol
4W75	30%w/v PEG MME 2000, 0.15M KBr	20%v/v Glycerol
4W76	3M NaCl, 0.1M Bis-Tris pH 5.5	25%v/v Glycerol
4W77	50%v/v PEG200, 0.2M MgCl2, 0.1M NaCacodylate pH 6.5	--
4W7A	3M NaCl, 0.1M Bis-Tris pH 5.5	25%v/v Glycerol
4W7C	30%v/v PEG400, 0.1M Cacodylate pH 6.5, 0.2M Li2SO4	--
4W7D	20%w/v PEG8000, 0.1M CHES pH9.5	25%v/v Glycerol
4W7E	0.1M Imidazole pH 8.0, 10%w/v PEG8000	25%v/v Glycerol
4W7F	20%w/v PEG8000, 0.1M CHES pH9.5	25%v/v Glycerol
4W7R	20%w/v PEG3350, 0.2M Potassium formate	25%v/v Ethylene glycol

Chapter 3

Split GFP as a Crystallization Scaffold

3.1 Introduction - A Split GFP Mediated Crystallization Scaffold

The suites of GFP oligomers presented in chapter 2 are all based upon the backbone of split GFP, a version of GFP specifically engineered to serve as a reporter molecule[12]. When used for this original application the split GFP has proven to be an effective reporter, especially for *in vivo* applications[43–47]. The ability of split GFP to complement target proteins both *in vitro* and *in vivo* and a crystal structure of split GFP – sfCherry complex[13] made it a tempting protein core for scaffolding. Obtaining the split GFP - sfCherry complex (figure 3.1) provided evidence that a complex formed this way was rigid enough to form a stable crystal lattice. When combined with the variety of GFP oligomers that were made utilizing the split GFP sequence, (in the full length format but not the truncated strand 1-9 form), the potential to exploit these scaffolds as symmetric cores was envisioned. However, the full realization of this approach has been difficult to achieve. Formation of the preformed cores has yet to be successful, and purification of spilt GFP – POI complexes *in vivo* are hindered by low complementation efficiency and difficult purifications. Some of these experiments are detailed in this chapter.

3.2 Results - Difficulties in Dimer Formation and Purification

A key factor in the success of a protein crystallization experiment in the purity of the protein used for the experiments. Many bench crystallographers aim for a purity of their protein to be >95%, although the amount of impurities can vary depending on the crystallization target and contaminating proteins[48]. Proteins chosen for synthetic symmetrization experiments (previously mutants of T4-lysozyme (T4L) or mannose-binding protein (MBP) and now GFP) all have the advantages of being well expressed, highly soluble and yield highly pure protein.[6,8]

Oligomeric impurities typically arise during the dimerization process where the cysteines are oxidized with agents such as Cu^{2+} in alkaline conditions (pH 8.0 – 9.0). This limits disulfide oligomers to dimers.

When more than one cysteine is present, multiple oligomers are created. This was the case with the initial dimerization of split GFP (figure 3.2). The engineered split GFP backbone featured two cysteines at positions 40 and 70. A Cys40Ala mutation was used for initial construct design as Cys70 was though to be inaccessible based upon a related sfGFP structure.[49] From these preliminary dimerization experiments, it was apparent the presence of Cys70 lead to multiple oligomers formed by both Cys70 and the newly introduced cysteine. Non-reducing SDS-PAGE analysis indicated that oligomers from dimers to putative octamers were present. It is impossible to ensure that the species corresponding to the molecular weight of a dimer (~50 kDa) was homogenous for only the intended dimer. It is very likely that the purified dimer peaks from size exclusion chromatography (SEC) were a mixture of intended dimers, Cys70 homo-dimers and hetero-dimers of the two cysteines. Further, SEC was ineffective to fully resolve the monomeric and dimeric species present (Figure 3.3). Crystallization experiments were conducted, but only one structure was solved for the D102C mutant (PDB 4W6R).

3.2.1 Cysteine Null Constructs

Upon the discovery that Cys70 was not sufficiently buried inside the GFP core, but was in fact reactive, it was subsequently mutated to an alanine resulting in a new backbone (C48A/C70A). Initially a D190C mutant was tested with this new backbone and it resulted as expected in only dimers and monomers (figure 10). These results have significant implications for future efforts with this approach. First, the mutations from cysteine to alanine to create the new backbone are modifying a protein that had been painstakingly engineered to exist in a split form and exist in a semi-soluble state.

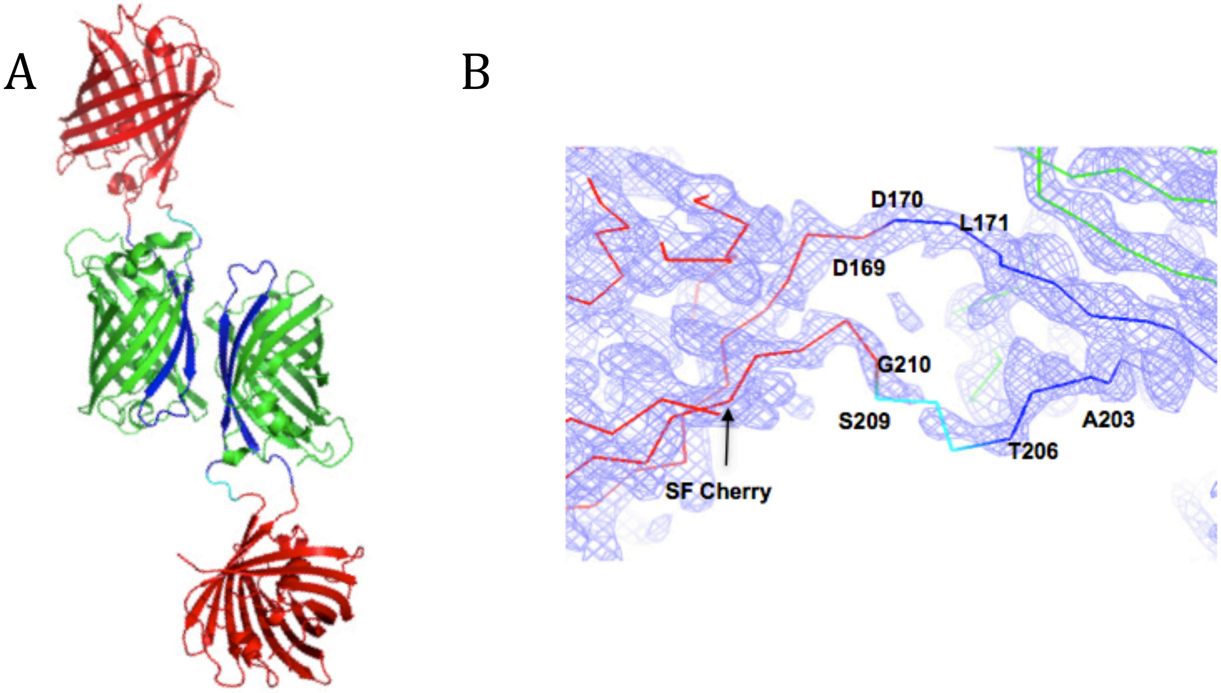


Figure 3.1. Structure of split GFP in complex with sfCherry. Adapted from Nguyen *et al.* 2013, Reproduced with permission of the International Union of Crystallography (http://journals.iucr.org/). (A) Crystal structure of split GFP 1-9 in complex with the sfCherry revealed the presence of a crystal artifact dimer interface along strands 10/11 of the GFP barrel. (B) The linkage between sfCherry and the GFP 1-9 core is a rigid linkage as seen but the backbone electron density (right), this is a result of a double pass linkage from strands 10/11 being interested in a loop of sfCherry. A terminal fusion would result in a single pass linkage.

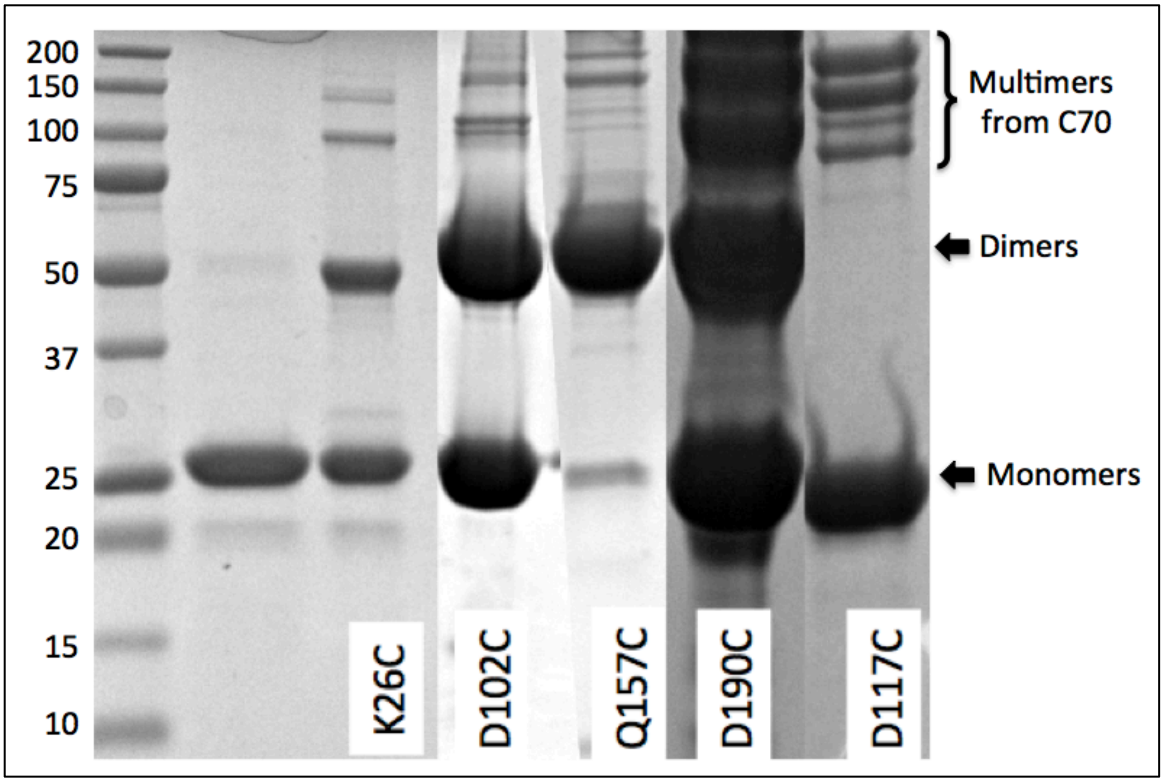

200
150
100
75
50
37
25
20
15
10
K26C
D102C
Q157C
D190C
D117C
Multimers from C70
Dimers
Monomers

Figure 3.2. Multiple Oligomers Formed When Two Cysteines Are Present. The presence of cysteine 70 in addition to the introduced cysteine mutants resulted in a predominant dimer species for all except the D117C mutants. In all cases higher order oligomers were also observed. It is likely that he dimeric species are contaminated heterodimers of the introduced mutant and Cys70 and Cys70 homodimers.

Changes to the split GFP backbone can have significant impact on solubility and complementation efficiency. When combined with the new single cysteine mutations and double metal-chelation mutations, this impact becomes more significant. Second, and more significant to the work presented in this dissertation, is the fact that it is exceedingly difficult to ensure homogenous samples for crystallization when more than one cysteine is present. Although this complication was resolved with the new GFP backbone for the oligomers in chapter 2, it presents a very real concern when combining the GFP scaffold with POIs. Difficulties with the *in vitro* purification of the mutant split GFPs resulted in a situation where the POI must be cysteine null as well. These issues are addressed in section 1.4. These disulfide issues are likely to reoccur with other proteins used as disulfide scaffolds as well. Design of symmetry scaffolds through computational design may be better suited for symmetric scaffolds and is explored in chapter 3.

3.2.2 Alternate Purification Methods

The previous disulfide symmetrization work with T4L[6] and CelA[11] used SEC to resolve the dimers from monomers after the dimerization step. This was ineffective for the GFP dimers (figure 3.3), a result of the high expression level of the GFP constructs. Successive SEC runs with low sample injection volumes may have alleviated this issue but were not practical at the time. Additionally the GFP versions that were engineered into split GFP, features a weak native dimer interface seen in the crystal structure[13,49] (figure 3.1) and is present as a minor species in solution (figure 3.5). This intrinsic dimer interface contributes to non-disulfide dimers contaminating the SEC dimer peak. While this appeared to be inconsequential for GFP disulfide dimers alone given the crystallization of D102C with monomer contamination, it would likely impede purification of homogeneous GFP-POI complexes.

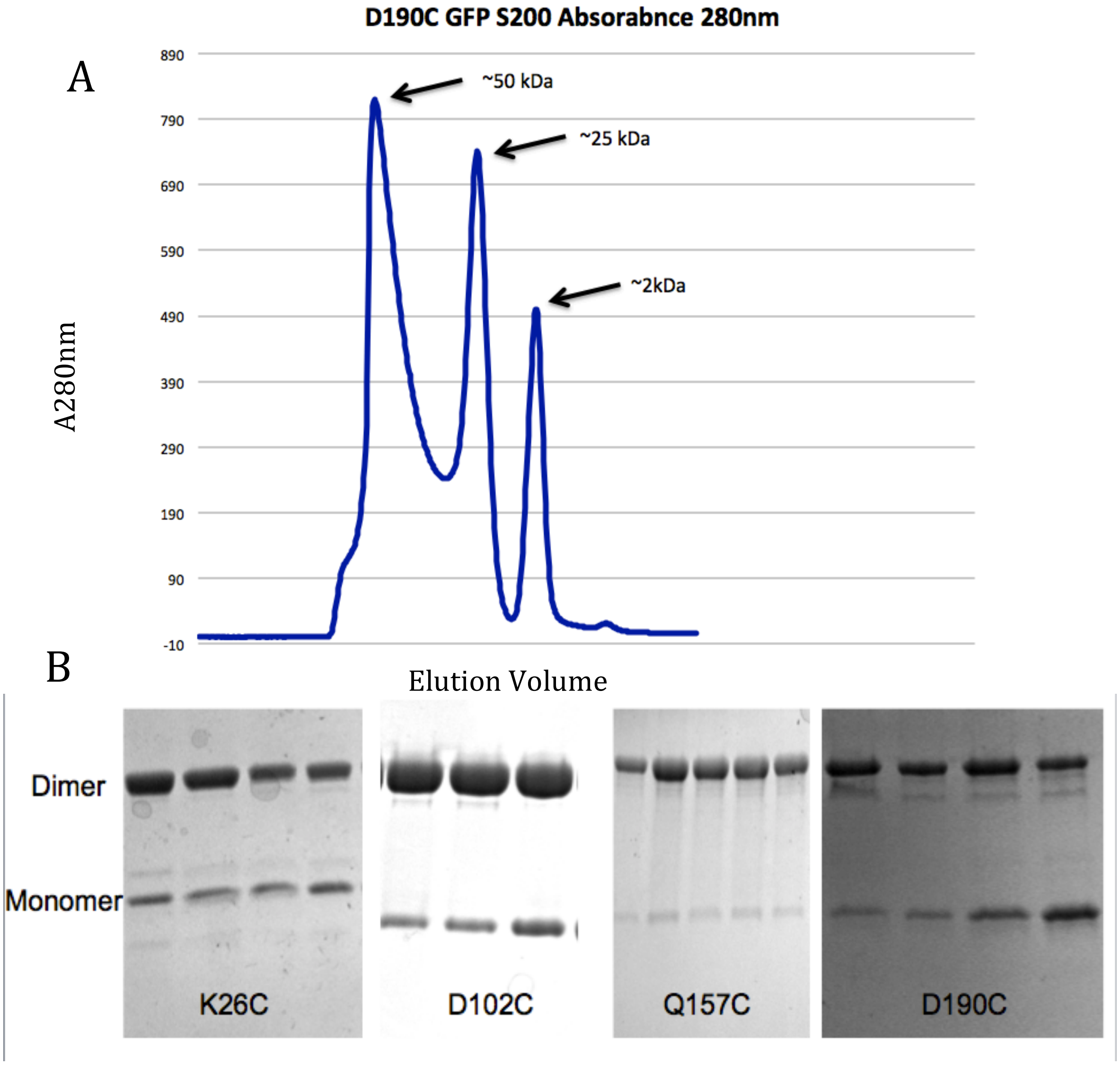


Figure 3.3. Size exclusion purification of GFP disulfide dimers. (A) Size exclusion chromatography failed to fully resolve GFP disulfide dimers from monomers. (B) The major peak that corresponded to the molecular weight of a dimer (~50 kDa) also contained GFP monomers when fractions were analyzed by non-reducing SDS-PAGE. The absorbance at 280nm trace from the D190C SEC purification represents the typical resolution between dimer and monomer peaks when run on a s200 16/60 sizing column.

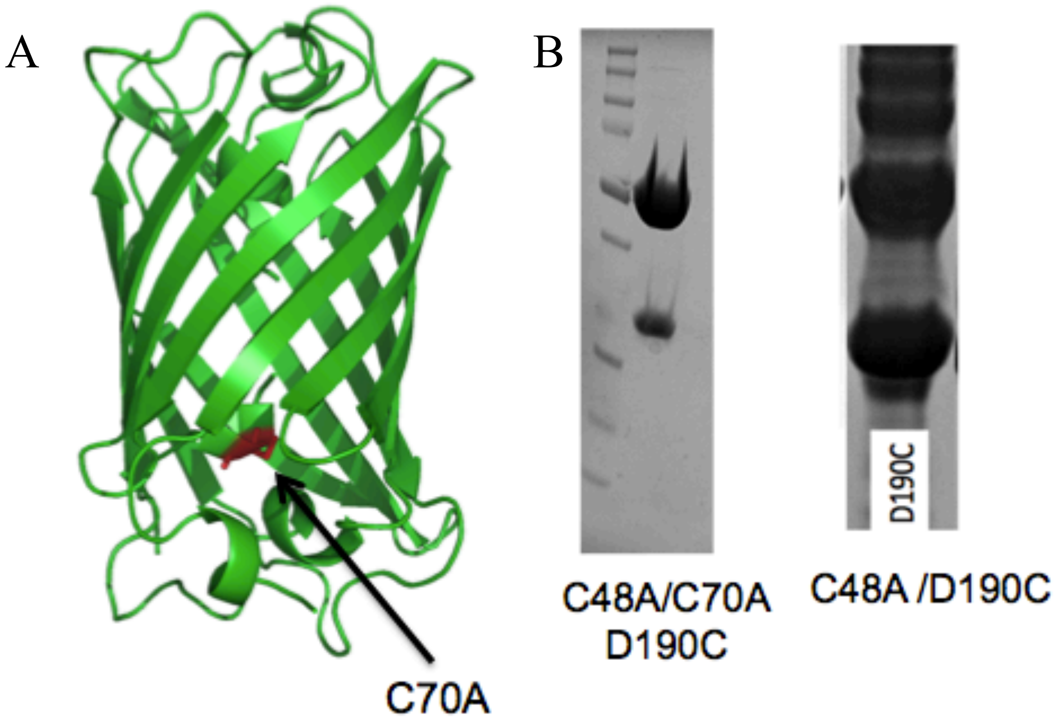


Figure 3.4. The presence of cysteine 70 leads to heterogeneous oligomers. (A) Cysteine 70 (red residue) was thought to be sufficiently buried in the core of the GFP barrel to be nonreactive. (B) Non-reducing SDS-PAGE gels indicated that while Cys70 was present, high-order oligomers were formed. Upon mutation to alanine the only bands present on non-reducing SDS-PAGE corresponded to monomers and dimers.

This effect would be exacerbated if incomplete complementation occurred *in vitro* and if the molecular weight of the different species (i.e. fully complemented and partially complemented GFP) are similar. This led to the exploration of different techniques to obtain pure dimers suitable for crystallization, which could be applied to the oligomeric GFP-POI complexes in the future. It should be noted this is only of consequence for disulfide oligomers; the metal chelating mutants could be purified with immobilized metal affinity chromatography (IMAC) since these oligomers form once the protein is in the crystallization experiment.

3.2.2.1 Thiopropyl Sepharose Purification

One potential way to separate disulfide dimers from contaminating monomers is based on the assumption that the monomer contains an unmodified reactive cysteine. Here a thiol reactive moiety is conjugated to a resin. Common commercially available resins are maleimide or thiopropyl sepharose resin (GE Healthcare). The latter was selected since it was a reversible reaction and the resin could easily be regenerated.[50] An experiment was performed where the GFP D190C was dimerized with Cu^{2+} in Tris buffer at pH 9.0 (chapter 2), the Cu^{2+} was dialyzed out overnight and the protein was run over the column. What resulted was the unfortunate occurrence of all the protein binding to the resin. The disulfide bonds were all reduced and new bonds formed to the resin resulting in only monomeric GFP eluting off of the column (figure 3.6). Coupled with the fact that these resins posses exceedingly low binding capacities (3-5mg protein per mL of resin) it is not a feasible means to purify disulfide oligomers of any protein.

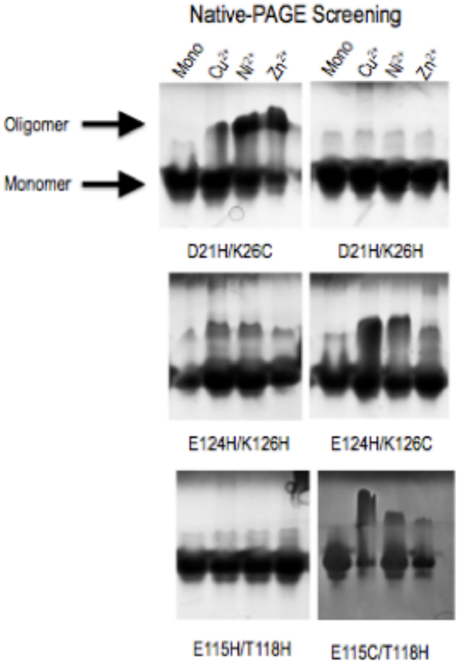

Native-PAGE Screening
Mono
Cu2+
Ni2+
Zn2+
Mono
Cu2+
Ni2+
Zn2+
Oligomer
Monomer
D21H/K26C
D21H/K26H
E124H/K126H
E124H/K126C
E115H/T118H
E115C/T118H

Figure 3.5. Native PAGE of GFP metal binding mutants indicating native oligomer presence. Native PAGE gels from assays of metal chelating mutants of GFP showed residual dimers in solution at higher concentrations (protein was at 10mg/mL when loaded onto gels). The faint high molecular weight bands in the "mono" lanes and the high molecular weight bands for the D21H/K26H mutant.

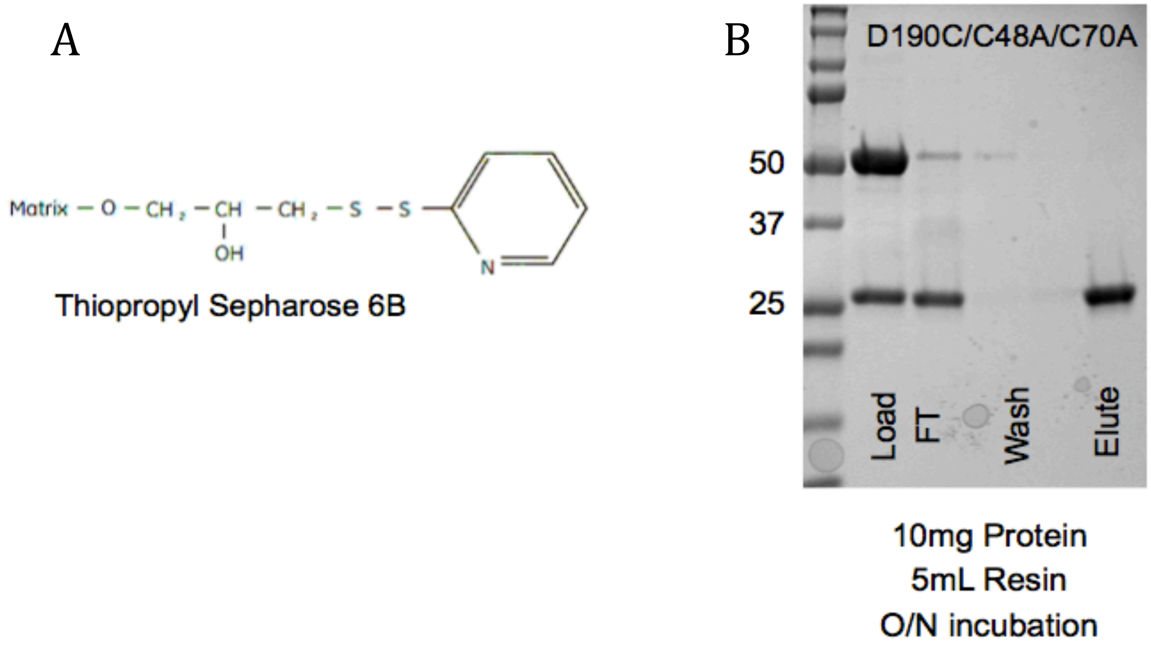

A
Matrix — O — CH₂ — CH — CH₂ — S — S
OH
N
Thiopropyl Sepharose 6B
B
D190C/C48A/C70A
50
37
25
Load
FT
Wash
Elute
10mg Protein
5mL Resin
O/N incubation

Figure 3.6. Thiopropyl sepharose purification of GFP D190C. (A) Thiopropyl sepharose was used in an attempt to remove unreacted cysteine monomers from the dimers instead of SEC. (B) When this resin was used on a mixture of D190C dimers and monomers, the monomers did not bind the column (FT lane). A minor band from unbound dimers was present in the wash and only monomers were present in the elution. When the dimers were mixed with the resin, the resin reacted with the disulfide bond, reducing the dimers to monomers that were bound to the resin.

3.2.2.2 Oxidizing Cytosolic *E. coli* Expression

An alternative approach to form homogenous dimers is to allow the recombinant *E. coli* expression system to do it. *E. coli* secretes its native disulfide containing proteins into the periplasm where the disulfide bond is formed. This has been exploited for the recombinant expression of disulfide containing proteins utilizing the PelB leader sequence to direct the recombinant protein to the periplasm.[51,52] Efforts to engineer *E. coli* expression strains to allow cytosolic disulfide bond formation have been undertaken by commercial companies. One of the newer strains known as SHuffle was developed by New England Biolabs and purported to efficiently form even complex disulfide bonds.[53] *In vivo* disulfide bond formation was attempted with the SHuffle cells. When the proteins were purified, the yield was dramatically reduced compared to expression levels in standard BL21 (DE3)[54] (figure 3.7). Furthermore, this experiment indicated that when the IMAC purification was performed in the absence of reducing agents, the Ni2+ catalyzed significant disulfide bond formation. This allowed the disulfide mutants to be expressed with high yields in standard BL21 cells, partially dimerized on-column before subsequent oxidation with Cu^{2+} prior to further purification.

3.2.2.3 Anion Exchange Chromatography

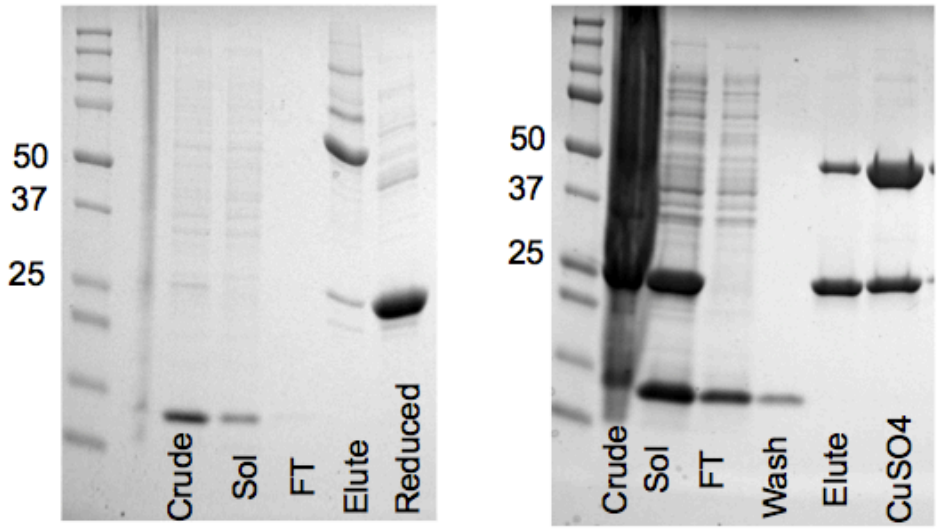
Ion exchange chromatography (IEX) relies on the charge differences between proteins to separate similar proteins by different charge states. This can efficiently resolve like proteins that differ only by the presence of one charged phosphorylation event.[55] In the case of a mixed species of GFP monomers and dimers, the charge difference is significant enough between the two species to allow them to be resolved by anion IEX. Mixed species were buffer exchanged into a start buffer composed of 100mM Tris pH 9.5 and 5% glycerol. The mixed protein was then loaded onto an anion IEX column, washed with start buffer, and then a gradient of start buffer with NaCl from 0M – 1M was used to elute separate monomer and dimer peaks (figure 3.8).

Efficient separation of the two species is crucial for crystallization and the ability of IEX to do this is paramount for future purification of disulfide scaffold – POI complexes. This technique is applicable to these complexes regardless of the POI identity since the principle of separation remains the same and buffer conditions can change based on the isoelectric point (pI) of the POI and complex. IEX separation can be combined with other techniques such as SEC to further purify difficult complexes.

3.2.3 Metal-Mediated Oligomers are Unpredictable

Although metal-mediated oligomers are easier to purity, they suffer from several issues as well. All the metal bound structures solved in chapter 2 were simply purified by a two-step process of IMAC followed by SEC. Purification of *in vivo* complexes of a metal-chelating GFP core complemented with a POI is a facile task. To accomplish this a hexa-histidine affinity tag can be applied to the GFP core only. With only the GFP tagged, only the complex with POI with be purified as non-complemented GFP cores are insoluble.[12] *In vivo* complementation would require additional purification that could be achieved with IEX, SEC or a combination of the two.

The issues with metal-mediation arise from the difficulty in predicting and designing the metal-chelation site geometry. Of all of the GFP metal-mediated oligomers designed, only one (PDB 4W7R) forms a closed dimer where the intended residues from contacts with the metal ion as a symmetric dimer. In many cases, residues adjacent to the point mutations, or from different GFP chains, take part in the metal chelation (figure 4). Depending on the crystallization conditions, these can be aspartate, glutamate or lysine side chains. This is not detrimental to crystallization experiments. In all cases the metal ion was chelated partially by the introduced point mutations, creating novel crystal contacts and new crystal forms. These metal-chelation mutants are applicable for crystallography but not scaffolding of proteins in solution.



D117C/C48A/C70A
SHuffle cells

D190C/C48A/C70A
BL21 DE3

Figure 3.7. Expression test and non-reducing IMAX purification of GFP disulfide dimers in SHuffle and BL21 cell lines. SHuffle and BL21 cell lines were used to compare expression levels of monomers and dimers. Both the D117C and D190C mutants were screened in both cell lines and gels with the most prominent overexpression are presented. For the SHuffle cell lines low protein expression resulted in poor purity of GFP (Elute & Reduced lanes) and no apparent over expressed bands in the crude or soluble fractions. The BL21 had clear overexpressed GFP monomers at 25 kDa in the soluble fraction. Upon elution from the Ni^{2+} resin in non-reducing conditions GFP dimers were formed (50 kDa band). The fraction of dimers was increased when treated with $CuSO_4$.

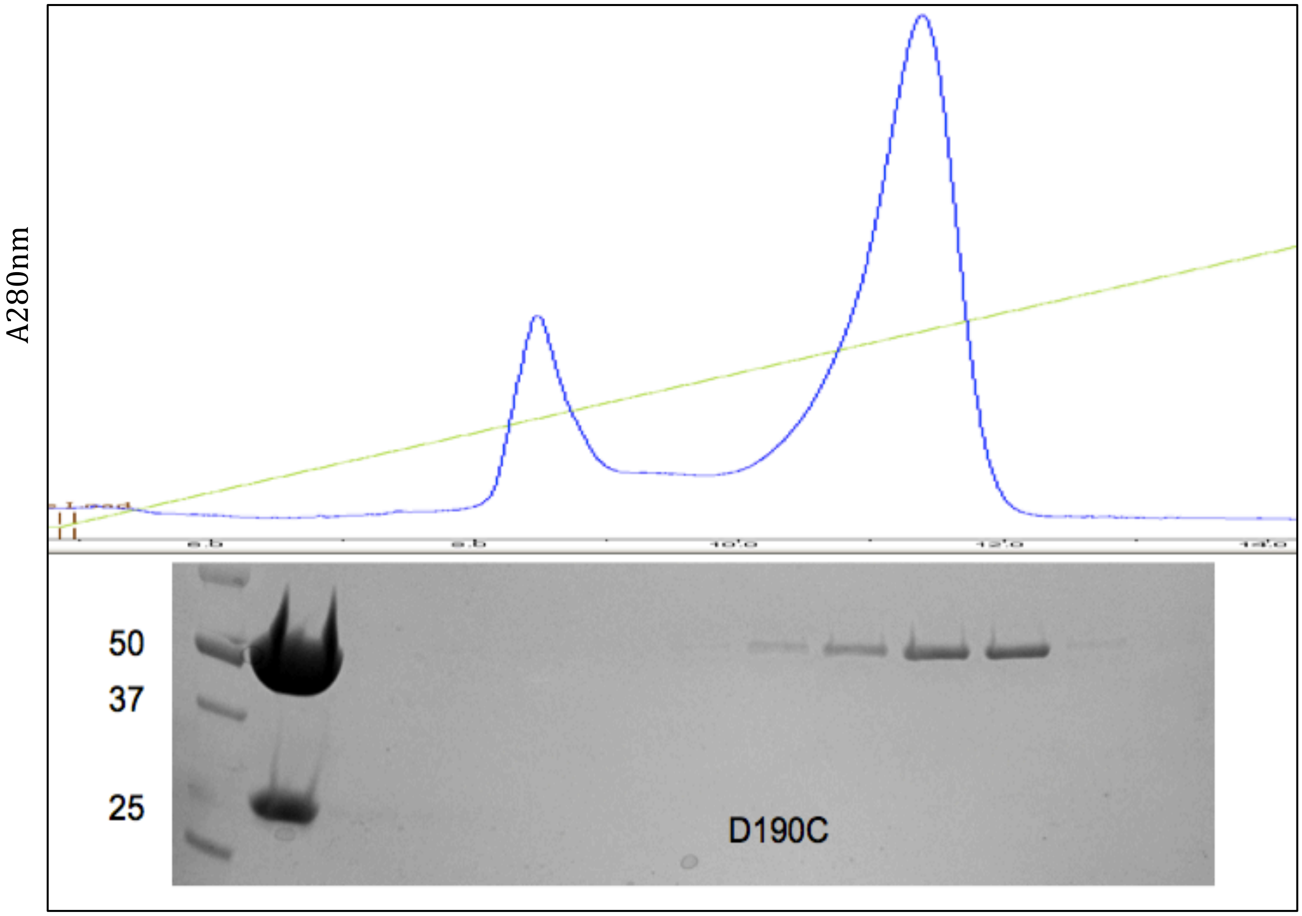


Figure 3.8. Anion exchange chromatography separation of GFP D190C monomers and dimers. GFP dimers and monomers could be resolved with anion exchange at pH 9.5. Two peaks were resolved when a gradient from 0M-1M NaCl was run. The first peak corresponded to monomers that were too dilute to be observed by SDS-PAGE. The second peak corresponded to pure dimers with no contaminating monomers. This was reproducible for all five disulfide dimers.

When complemented to POIs, the complex can participate in the unintended metal chelation events through the POI. This may be beneficial for novel crystal contacts but it can also lead to protein aggregation. When working with GFP alone, the addition of metals (Cu^{2+}, Zn^{2+}, Ni^{2+}) would occasionally produce visible aggregation that was reversible with the addition of small molecule metal chelators (ethylenediaminetetraacetic acid), dilution or through careful titration of the metal ions. This has also been observed with other proteins[8,56]. Although this is typically reversible, it created a solution that is difficult to work with as the aggregates may precipitate. If multiple cysteine residues are present, covalently bound aggregates may be formed depending on buffer conditions. This results in protein solutions that must be carefully handled to ensure reproducibility of results.

3.3 Refolding Attempts of Split GFP Mutants

All crystal structures presented in chapter 2 were from full-length GFP molecules, i.e. the split GFP sequence expressed as a single polypeptide chain instead of in two parts. In all cases the protein was well expressed and well behaved. A 2L culture of *E. coli* grown in auto-induction media[57] would overload a 5mL column of IMAC resin leading to exceedingly pure protein that could be concentrate over 300mg/mL in a standard crystallization buffer (50mM Tris pH 8.0, 100mM NaCl) at 25°C. Once the mutations were applied to the truncated split GFP 1-9 sequence, all five cysteine mutations were totally insoluble (figure 3.9). It was expected that solubility would be an issue, as the designed split GFP sequence itself required refolding for *in vitro* applications.[12] In short, 75mg inclusion bodies were denatured in 1mL TNG buffer (50mM Tris pH 7.4, 100mM NaCl, 10% glycerol) supplemented with 9M urea. This would then be rapidly diluted into 25mL of TNG buffer, filtered through a 0.22 micron filter and added to samples containing a protein fused to the GFP hairpin sequence. Complementation would then be allowed to proceed overnight. Complementation efficiency could then be monitored by fluorescence. A typical yield was ~10mg of refolded protein by this refolding method. While this yield was sufficient for bio-reporter applications it was not for *in vitro* complementation with the GFP dimer suite. The published structure of split GFP in complex with sfCherry was obtained by adding the refolded split GFP to the sfCherry cell lysate and co-purification was performed.[13] This strategy works on an individual basis, yet it eliminates the benefits of *in vitro* complementation (i.e. ease of use, disulfide dimer scaffolds in complex with cysteine containing POIs). Efforts were undertaken to optimize the refolding to obtained pre-formed dimers for *in vitro* complementation.

In discussions with the collaborators who developed split GFP (Geoffrey Waldo, personal conversations), anecdotal evidence suggested that *in vitro* complementation would only occur when strands 10/11 were fused to a soluble protein, the 10/11 hairpin was insoluble on its own. This means that all evaluation of complementation involved a fusion protein. These fusions were not always fully evaluated for complementation efficiency on their own, leading to ambiguity about the causes for complementation issues. In instances where there is low efficiency for loop insertions of the hairpins, it could be steric hindrance issues preventing complementation from hairpins that are not fully solvent exposed, or fit could reflect unexpected interference from the POI.

3.3.1 Refolding Attempts of Larger Amounts of Inclusion Bodies

First, larger amounts of inclusion bodies were refolded per 25mL refolding reaction; this would result in higher concentration of refolded protein. Up to 250mg of inclusion bodies were refolded at once. In all cases, close to 100% of protein remained soluble after subsequent filtration per SDS-PAGE gels (figure 16). However, all protein precipitated upon any concentration, regardless of the starting amount of protein.

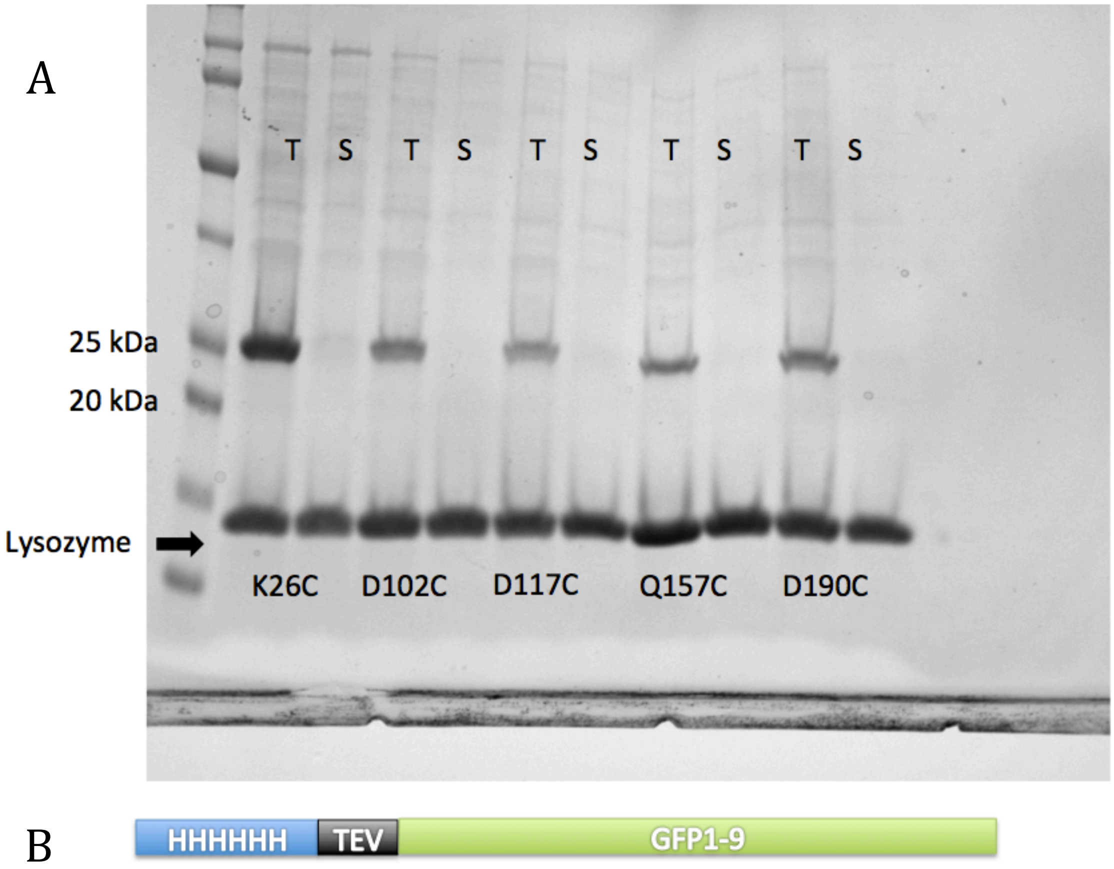

A
T S T S T S T S T S
25 kDa
20 kDa
Lysozyme
K26C
D102C
D117C
Q157C
D190C
B
HHHHHH
TEV
GFP1-9

Figure 3.9. Expression Test of Disulfide Dimers Mutations on Split GFP 1-9 backbone. (A) All five cysteine mutants were well expressed (T, total cell fraction lanes) but were completely insoluble (no 25 kDa band present in the S, soluble fraction lanes). (B) The original constructs to express GFP 1-9 with cysteine mutants contained a N-terminal hexa-histidine affinity tag and TEV cleavage site.

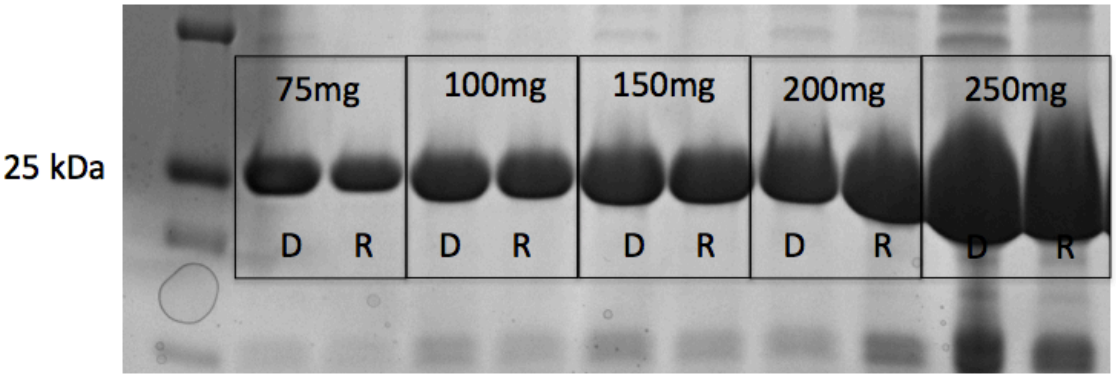


Figure 3.10. Rapid-dilution Refolding of Split GFP 1-9 K26C Inclusion Bodies. Rapid dilution refolding of was performed with varying starting amounts of inclusion bodies dissolved in 1mL of TNG buffer + 9M urea (D lanes). After refolding in to 25mL of TNG buffer and subsequent filtration to remove aggregates the majority of protein appeared soluble (R lanes). In each lane (D and R) the same amount of protein was loaded for direct comparison of protein remaining in solution after refolding. Concentration of refolded protein resulted in total precipitation of protein.

3.3.2 IMAC Purification of Refolded Protein

Even with no mutations, refolded split GFP is plagued by low refolding efficiency. Typical proteins yields from refolding are up to 80% with similar refolding techniques.[58] Additionally, the precipitation of refolded split GFP; it was hypothesized that a majority of the protein existed as micro-aggregates in solution. To separate these presumed micro-aggregates from properly refolded protein IMAC was performed. For this and the previous experiments in chapter 1.3.1, constructs were designed featuring a N-terminal hexa-histidine affinity tag, a tobacco etch virus (TEV) protease cleavage site[59] on the GFP 1-9 sequence with appropriate mutations (figure 3.9). With this the GFP cores could be efficiently separated from the aggregated protein and contaminates from the inclusion bodies. Cores could then be dimerized; the affinity tag removed with TEV protease, and the cores could then be stored for future complementation.

This was attempted with refolded split GFP K26C. After rapid dilution refolding, the protein was loaded onto a Ni^{2+} column in TNG buffer supplemented with 30mM imidazole. The majority of the protein failed to bind the resin. The column was washed with TNG + 30mM imidazole and bound protein eluted with TNG + 250mM imidazole. A buffer composed of 10mM Cu^{2+} in TNG pH 9.0 was used to oxidize the disulfide bonds. From this, pure dimers at low concentrations were obtained (figure 3.11). Again, these dimers fully precipitated upon centrifugal concentration through 30 kDa molecular weight cutoff membranes. The experiment was repeated but sfCherry containing the 10/11 hairpin sequence[13] was added after dimerization and dialysis into TNG buffer. No complementation was observed and the protein still precipitated during concentration.

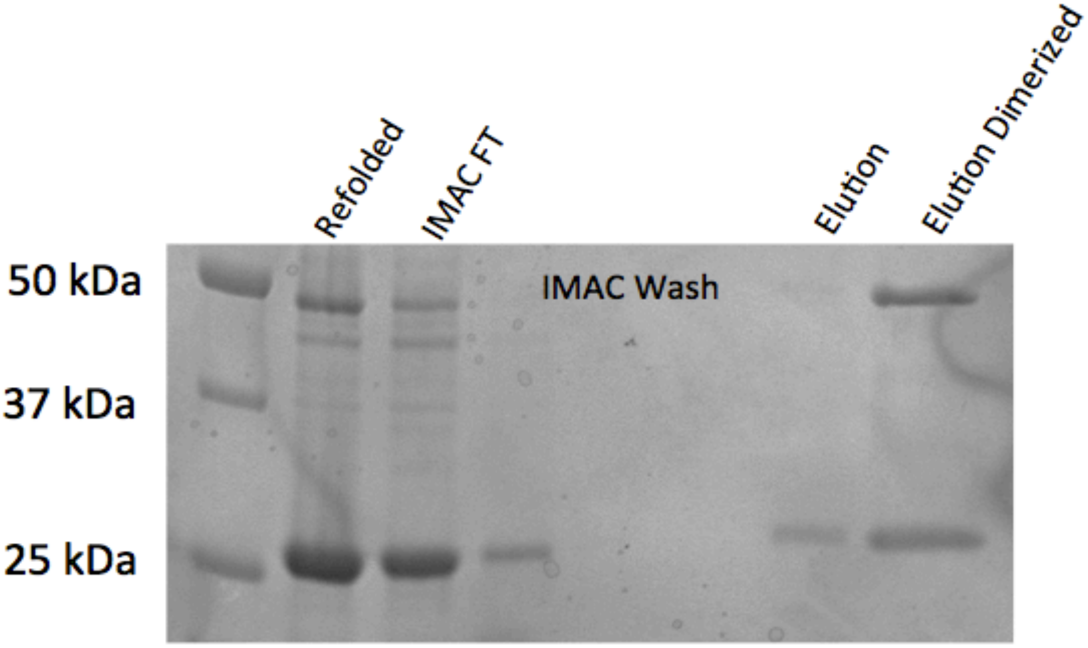
3.3.3 IMAC On-column Refolding Attempts

A final attempt to refold the histidine tagged protein involved on-column refolding of K26C. With on-column refolding, denatured protein is bound to the IMAC resin. Then a gradient from denaturing to non-denaturing conditions is used to refold the protein. This has been shown to be a highly effective technique[60]. The denatured K26C protein was loaded on the column in TNG + 9M urea and a gradient to TNG only was run. Much of the protein eluted off the column during this gradient (figure 3.12). The protein that was refolded suffered from the same solubility and complementation issues as the previously refolded protein. In discussions with the creators of split GFP, it was concluded that the N-terminal affinity tag and TEV cleavage site might be impeding proper refolding of the protein. New constructs with no affinity tags were constructed and refolding again was attempted.

3.3.4 IEX Purification of Refolded Protein

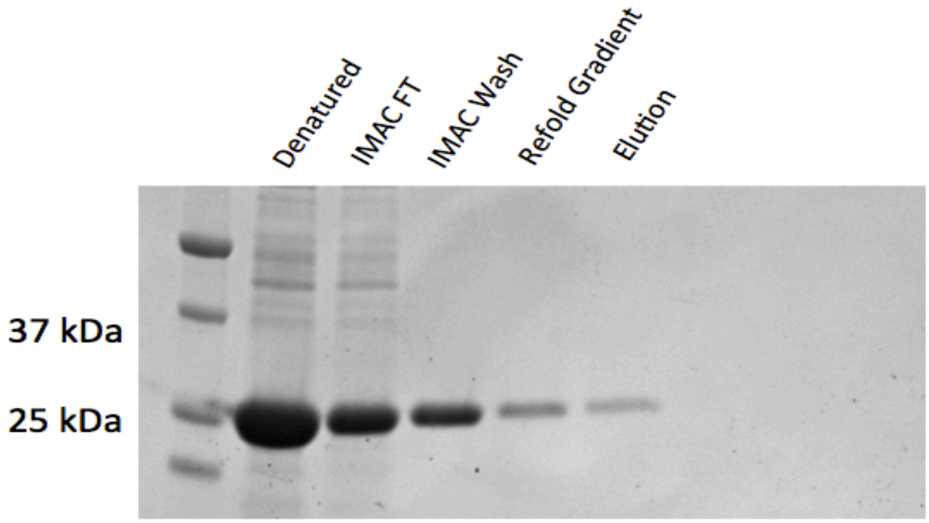
In a final attempt to create soluble dimerized GFP 1-9 cores, new constructs were cloned with no affinity tags. Because the double mutations needed for metal chelation may further destabilized the complementation efficiency when used in combination with the C48A/C70A backbone, only the disulfide dimer mutations were used. The metal chelation mutants required the cysteine null backbone as the cysteines would be reactive and have the potential to form large oligomers through hetero disulfide bond formations (figure 3.8).

Again, a purification step was needed to obtain crystallization quality purity. To achieve this, anion IEX was performed under the same conditions as the disulfide dimer purification (pH 9.5, 0-1M NaCl gradient elution).

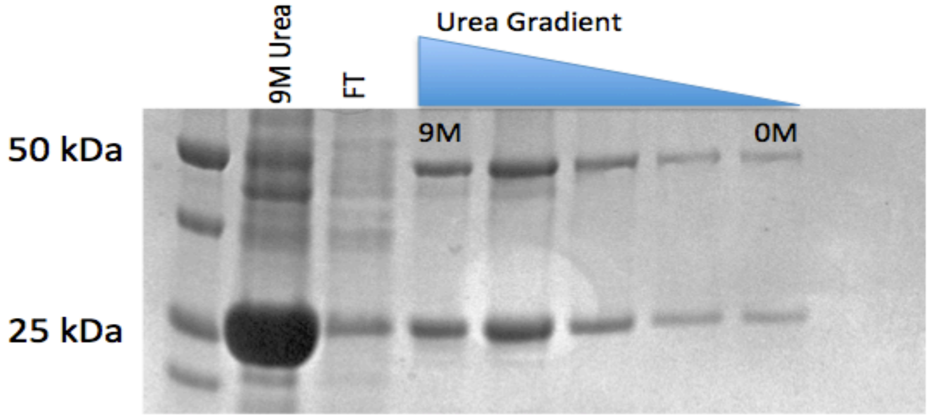

Refolded
IMAC FT
Elution
Elution Dimerized
50 kDa
IMAC Wash
37 kDa
25 kDa

Rapid Dilution Refold with IMAC

Figure 3.11. Rapid Dilution Refolding of split GFP K26C and Subsequent IMAC Purification. IMAC was attempted on refolded GFP 1-9 K26C, this was used as a way to concentrate and purify the refolded protein. The majority of the protein fialed to bind the Ni^{2+} resin (IMAC FT and wash lanes). The eluted protein was pure but predominately monomeric. The protein could be dimerized with $CuSO_4$. However, the protein failed to complement proteins containing the 10 -11 strands and could not be concentrated without precipitation.


Denatured
IMAC FT
IMAC Wash
Refold Gradient
Elution
37 kDa
25 kDa

30mM Imidazole in buffers


9M Urea
FT
Urea Gradient
9M
0M
50 kDa
25 kDa

No Imidazole in buffers

Figure 3.12. SDS-PAGE On-column refolding of Split GFP K26C. Denatured protein was bound the $Ni2^{+}$ IMAC resin in TNG + 9M urea buffer, protein was refolded by running a gradient to TMG + 0M urea. Initial runs contained 30mM imidazole in the buffers to prevent non-specific protein binding to the resin. When it was observed that the protein was eluted during the wash the experiment was repeated with no imidazole present in the binding or wash buffers. The GFP continues to leach off the column during the refolding gradient. The top gel was performed in reducing conditions while the bottom gel was non-reducing.

IEX was able to more efficiently bind refolded protein compared to the previous IMAC affinity samples (figure 3.13). Despite the new constructs and purification methods, the protein remained highly susceptible to aggregation when any concentration was attempted. This is detrimental to the originally proposed use of these GFP dimers as preformed cores that could simply be complemented *in vitro* immediately prior to crystallization. The inability to concentrate the GFP 1-9 cores prevents the formation of dimers, as it is highly impractical to perform the disulfide oxidation and subsequent buffer exchanges with large volumes of dilute protein.

3.3.5 Dynamic Light Scattering Experiments

To determine if the refolded protein was aggregated after refolding or only upon concentration, light scattering was used to measure the particle size in solution.[61] Multi-angle light scattering coupled to SEC (SEC-MALS) was attempted on freshly refolded GFP 1-9 K26C. SEC-MALS provides an accurate measure of particle size[62] but the preceding SEC column diluted the sample to the point where no protein absorbance was observed. Dynamic light scattering (DLS) was then attempted, where the distribution of particle sizes in solution is measured.[63] All five GFP 1-9 disulfide mutants were freshly refolded per ideal split GFP refolding protocols[12] and immediately analyzed by DLS. The resulting histograms of size distributions indicated that all five mutants were predominantly found in large, soluble aggregate (Figure 3.14). Due to the difficulties with refolding and *in vitro* complementation for crystallographic applications, *in vivo* complementation by coexpression was pursued.

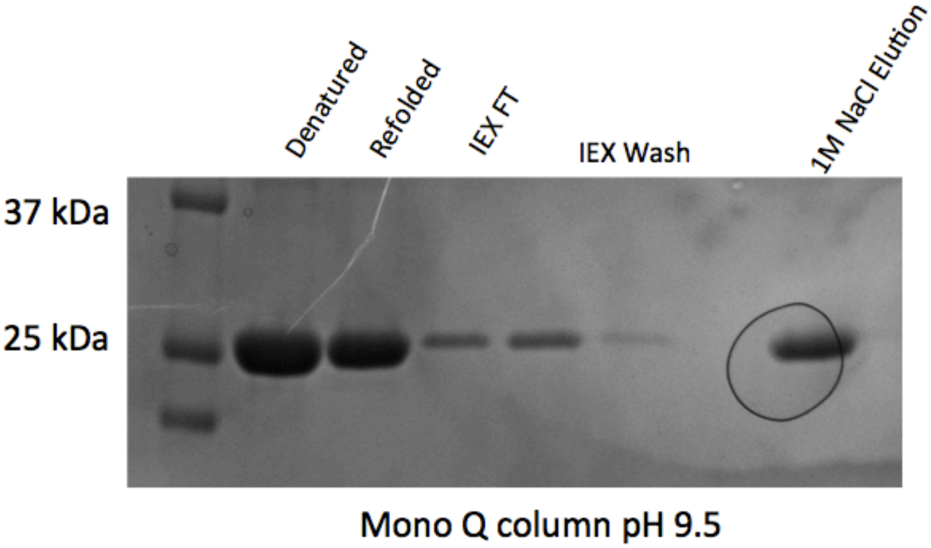

Denatured
Refolded
IEX FT
IEX Wash
1M NaCl Elution
37 kDa
25 kDa
Mono Q column pH 9.5

Figure 3.13. IEX Purification of Refolded Split GFP 1-9 K26C. Refolded protein with no hexa-histidine tag could be purified and partially concentrated with anion exchange chromatography at pH 9.5. This resulted in more efficient protein binding and higher yields then IMAC purification but the protein failed to efficiently complement control proteins *in vitro*.

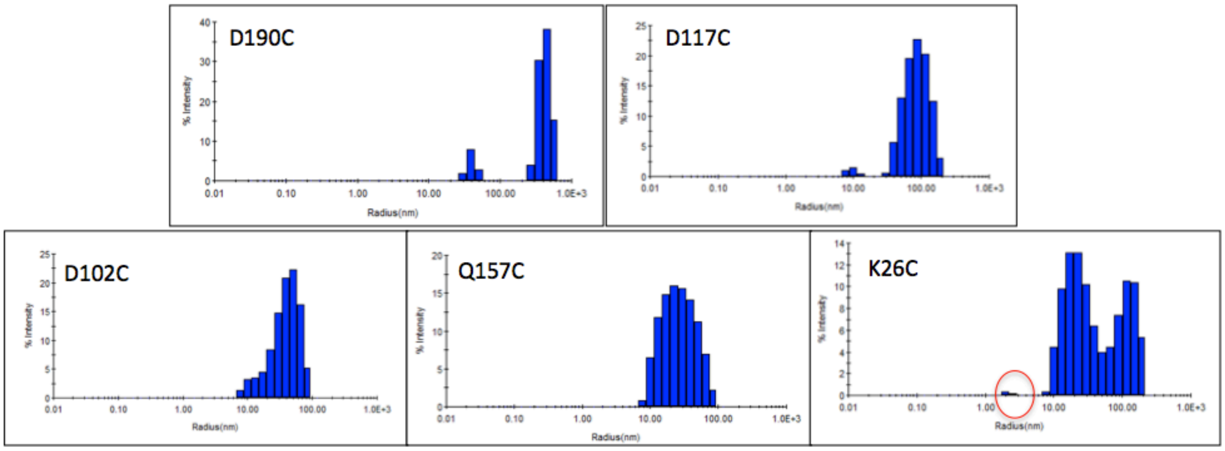


Figure 3.14. Dynamic Light Scattering Results from Freshly Refolded Split GFP 1-9 Mutants. DLS was performed on freshly refolded and filtered mutants of GFP 1-9. All five mutants existed primarily as large soluble aggregates. Only K26C had any protein with a radius consistent with monomeric GFP in solution (red circle).

3.4 Results - *Mycobacterium tuberculosis* Rv2658c *in vitro* complementation

Concurrent experiments performed by collaborators at Los Alamos National Laboratory investigated the ability for split GFP 1-9 cysteine mutants to complement *in vitro* with a control protein, *Mycobacterium tuberculosis* Rv2658c (MtbCon). This protein does not have a solved structure, but the construct used contained a C-terminal fusion to the GFP hairpin, which was well exposed. Crystals in complex with the original split GFP sequence were obtained, but these crystals were poorly ordered and failed to diffract. Optimization failed as well. The poor outcome of the crystallization experiments was likely a result of >20% of the protein is disordered according to XtalPred[64] server results (figure 3.15). Although this is a poor crystallization target, it was used to determine relative complementation efficiency compared to the original split GFP sequence.

For this complementation experiment, freshly refolded GFP mutants were mixed in a 6:1 ratio to purified MtbCon. The fluorescence was then monitored overnight. This experiment was only conducted once, but it showed all five cysteine mutants were able to complement with the MtbCon protein. The efficiency varied with D102C only complementing at ~40%, but 60-80% was observed for the other four mutants (figure 3.16). This would equate to ~4-8mg of protein per 75mg of inclusion bodies for *in vitro* complementation attempts. Crystallization screening experiments typically require milligram amounts of protein. With the inherent loss of protein with additional purifications, the yield from *in vitro* complementation of refolded protein does not possess an advantage over coexpression of the split GFP mutants and crystallization targets to obtain yields for multiple crystallization experiments.

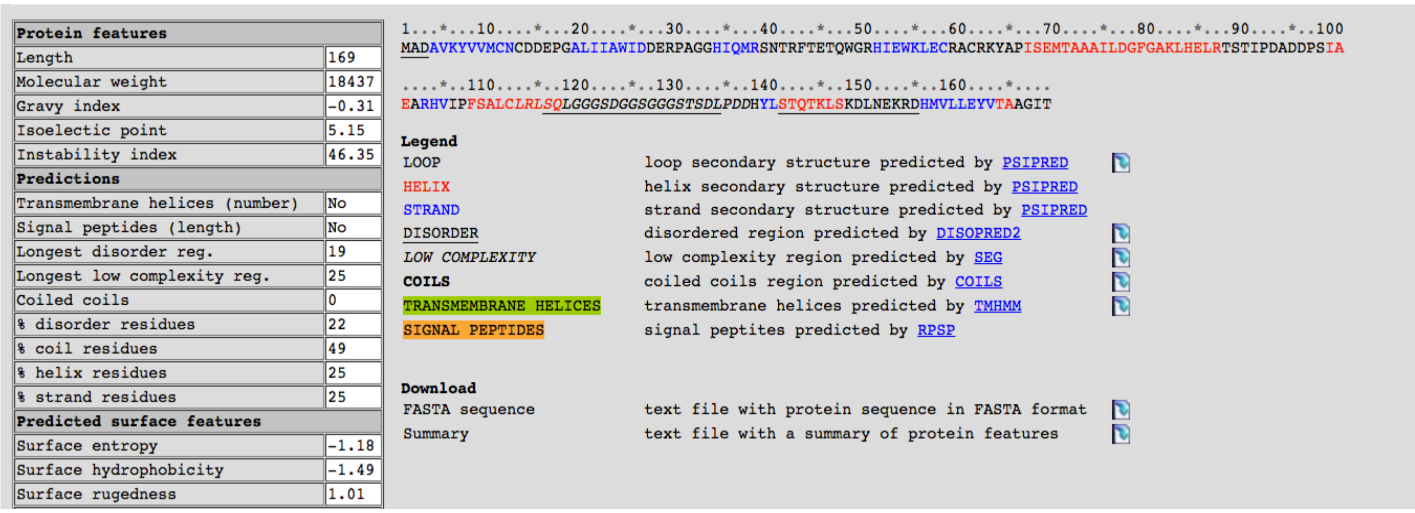
3.5 Results - Coexpression Attempts

To evaluate the ability of the split GFP mutants to act as symmetric scaffolds, non-tagged split GFP 1-9 mutants were co-expressed with novel crystallization targets containing the GFP hairpin (figure 3.17). Cell pastes that were green in color would indicate successful complementation. The complex could then be purified, dimerized and used for crystallization experiments. Through this approach, more protein could be readily obtained for more diverse crystallization experiments compared to *in vitro* complementation. However, dimerization after complementation would limit the use of disulfide dimers to POIs with no native cysteines.

3.5.1 H3L

The first protein to be used for coexpression with GFP 1-9 disulfide mutants was in collaboration David Baker from the University of Washington. The protein, H3L, was a designed alpha helical protein derived from clathrin.[65] This protein was designed to have a tetratricopeptide repeat motif, a common motif to regulate protein-protein interactions.[66] However, unlike clathrin, H3L was designed to exist as a monomer in solution and not a large protein network. Proteins with this structural motif but higher symmetry have since been published by the Baker lab.[67] H3L was selected as a coexpression candidate. It had previously formed micro-crystals during crystallization experiments.

Second, the H3L sequence contained no cysteines allowing the GFP-H3L complex to be purified then dimerized. Disulfide mutants D117C and D190C were selected initially for complex formation as they crystallized in the most diverse space groups and crystal conditions out of the five-disulfide dimers. Subsequent experiments involved the other GFP dimer mutants after a method of purification was established. Two constructs of H3L were made for these experiments, with the internal GFP hairpin linking two of the H3L constructs together. A single H3L construct with an N-terminal 10/11 hairpin was also utilized.



Protein features	
Length	169
Molecular weight	18437
Gravy index	-0.31
Isoelectic point	5.15
Instability index	46.35
Predictions	
Transmembrane helices (number)	No
Signal peptides (length)	No
Longest disorder reg.	19
Longest low complexity reg.	25
Coiled coils	0
% disorder residues	22
% coil residues	49
% helix residues	25
% strand residues	25
Predicted surface features	
Surface entropy	-1.18
Surface hydrophobicity	-1.49
Surface rugedness	1.01

```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80....*...90....*..100
MADAVKYVVMCNCDDEPGALIIAWIDDERPAGGHIQMRSNTRFTETQWGRHIEWKLECRACRKYAPISEMTAAAILDGFGAKLHELRTSTIPDADDPSIA

....*..110....*..120....*..130....*..140....*..150....*..160....*....
EARHVIPFSALCLRLSQLGGGSDGGSGGGSTSDLPDDHYLSTQTKLSKDLNEKRDHMVLLEYVTAAGIT
```

Legend

LOOP	loop secondary structure predicted by PSIPRED
HELIX	helix secondary structure predicted by PSIPRED
STRAND	strand secondary structure predicted by PSIPRED
DISORDER	disordered region predicted by DISOPRED2
LOW COMPLEXITY	low complexity region predicted by SEG
COILS	coiled coils region predicted by COILS
TRANSMEMBRANE HELICES	transmembrane helices predicted by TMHMM
SIGNAL PEPTIDES	signal peptites predicted by RPSP

Download

FASTA sequence	text file with protein sequence in FASTA format
Summary	text file with a summary of protein features

Figure 3.15. XtalPred Results for *M. tuberculosis* Rv 2658c. The sequence for the MtbCon protein used as a complementation control is predicted to be 22% disordered (underlined residues). Although this protein efficiently complemented split GFP 1-9, the percentage of disorder and low complexity make it a poor crystallization target.

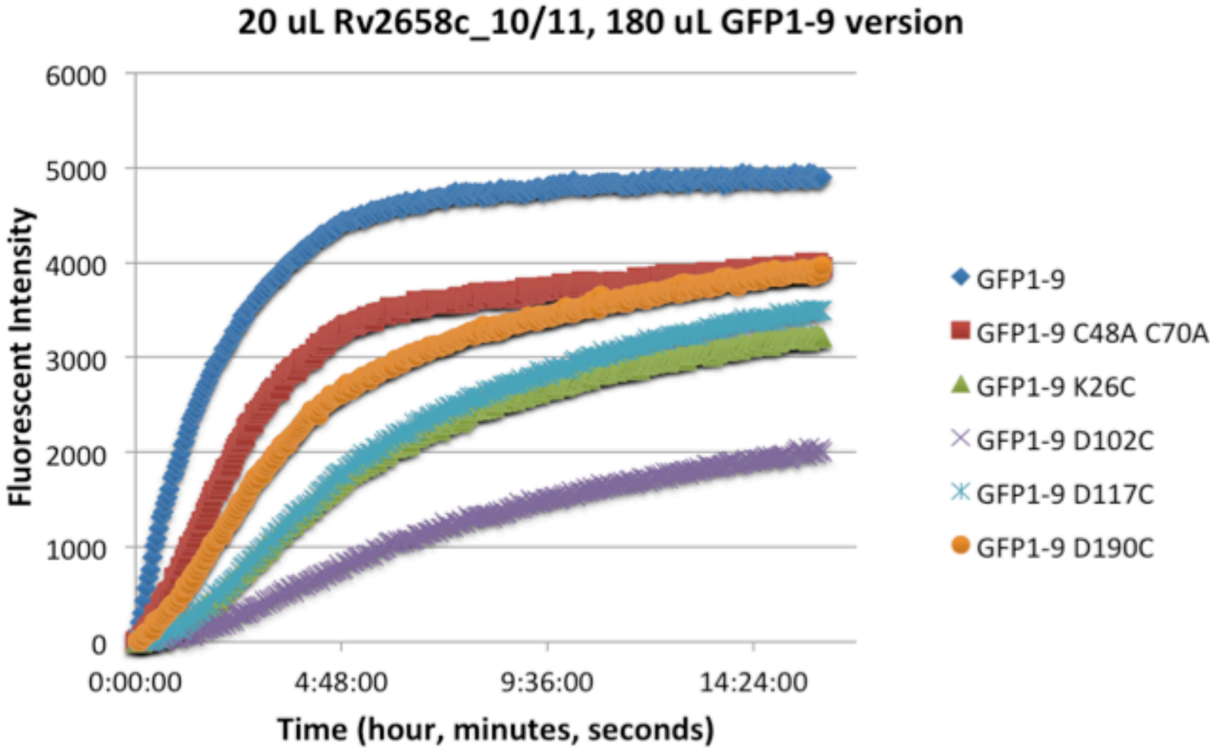


Figure 3.16. Complementation Efficiency of MtbCon with GFP 1-9 Mutants. MtbCon was able to complement all five mutants of GFP 1-9. The D102C mutant was the least efficient with only 40% of the fluorescent intensity of GFP 1-9 with no mutations. Introduction of the C48A/C70A backbone mutations reduced the efficiency to 80% of GFP 1-9 alone. This is roughly equivalent to 10% of the total refolded protein being viable for complementation (~8mg from 75mg of denatured protein).

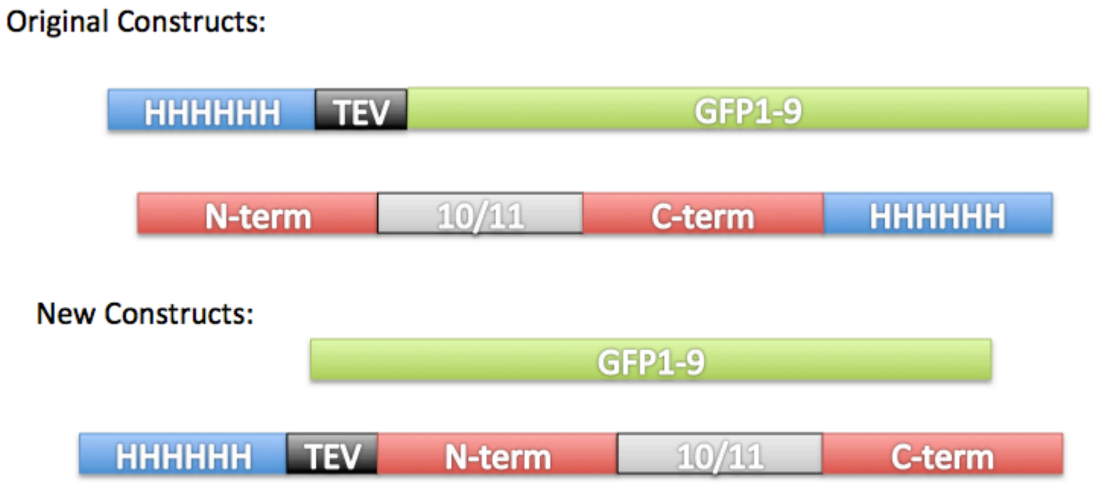

Original Constructs:
HHHHHH
TEV
GFP1-9
N-term
10/11
C-term
HHHHHH
New Constructs:
GFP1-9
HHHHHH
TEV
N-term
10/11
C-term

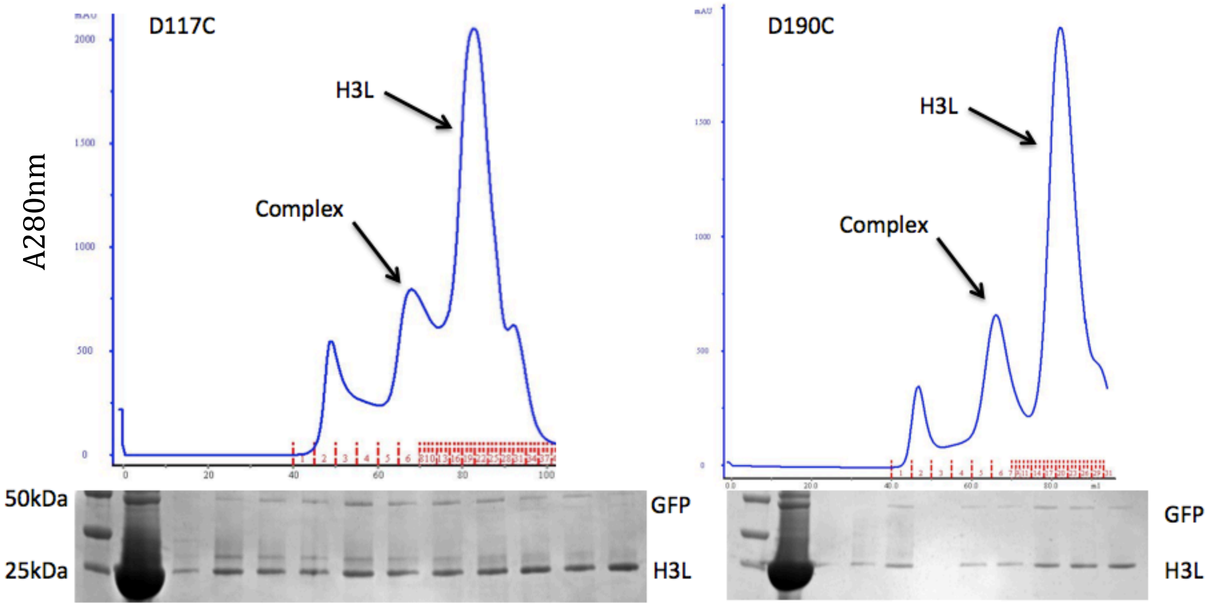
Figure 3.17. GFP 1-9 and Crystallization Target Constructs. Original constructs were intended to be expressed and purified separately then mixed *in vitro* once the GFP core was dimerized. Once it was concluded that the N-terminal hexa-histidine tag and TEV cleavage site could interfere with complementation, new constructs for coexpression and purification were made. The GFP 1-9 has no affinity tags and contains the C48A/C70A backbone and applicable oligomer mutations. The crystallization construct features the N-terminal hexa-histidine tag and TEV cleavage site for complex purification and the internal 10/11 strand sequence for complementation. Successful complementation would result in a green color of the cell pellet. To allow for coexpression the crystallization target were cloned into a pET vector and GFP 1-9 was cloned into a pACYC vector.

After coexpression of the two proteins, IMAC was used to purify the complex prior to dimer formation. Since the affinity tag was on the H3L protein, the IMAC purified protein contained the GFP-H3L complex and H3L alone. Tuning the expression levels by plasmid copy number, promoter or ribosome binding site variations may have eliminated this. However since what controls protein expression levels in *E. coli* is still not fully understood,[68] this would have to be tested for every coexpression construct. Instead it was decided to determine a purification method that could be used as a general purification scheme for all GFP-POI complexes.

SEC was performed on the dimerized GFP-H3L complex, with two resolved peaks. The first peak was green in color and contained the complex. The second major peak was only H3L (figure 3.18). A native PAGE gel was run on the pooled green fractions (figure 3.19). It was found that the complex was contaminated with monomeric H3L and that H3L was subject to proteolysis. This proteolysis was significant when the H3L was left at 4°C for one week.

The susceptibility of H3L to proteolysis was also observed when our collaborator worked with the protein. An unknown protease was purified with H3L, which could be heat inactivated at 100°C as H3L is a thermal stable protein (Fabio Parmigianni, personal communications). To eliminate the effects of proteolysis with the complex, a protease inhibitor, (4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride) was kept in all buffers until SEC and IEX was performed. Ethylenediaminetetraacetic acid (EDTA) was added to the IMAC-purified protein to chelate any metal ion for metal-dependent protease inhibition. These steps resulted in no evident proteolysis by SDS-PAGE gels for subsequent purifications.

To further purify the complex away from the contaminating H3L monomers, a tandem purification scheme of IEX followed by SEC was performed. The pI of H3L was similar to that of GFP, so anion exchange was performed in the same buffers as GFP alone.



'Green Peak' SEC Fraction SDS-PAGE

'Green Peak' SEC Fraction SDS-PAGE

Figure 3.18. SEC Purification of Dimerized GFP-H3L complexes. Three peaks observed for all GFP-H3L SEC runs. The first minor peak corresponded to the void volume of the column and contained aggregated protein. The second peak was green in color (indication of complex formation) and the third major peak was H3L alone. The non-reducing SDS-PAGE gels are presented showing the prominent band from H3L overexpression contaminating the complex after IMAC. The fractions from the complex peak showed the presence of a ~ 50 kDa band corresponding to the GFP 1-9 dimer core and a ~25 kDa band for the H3L.

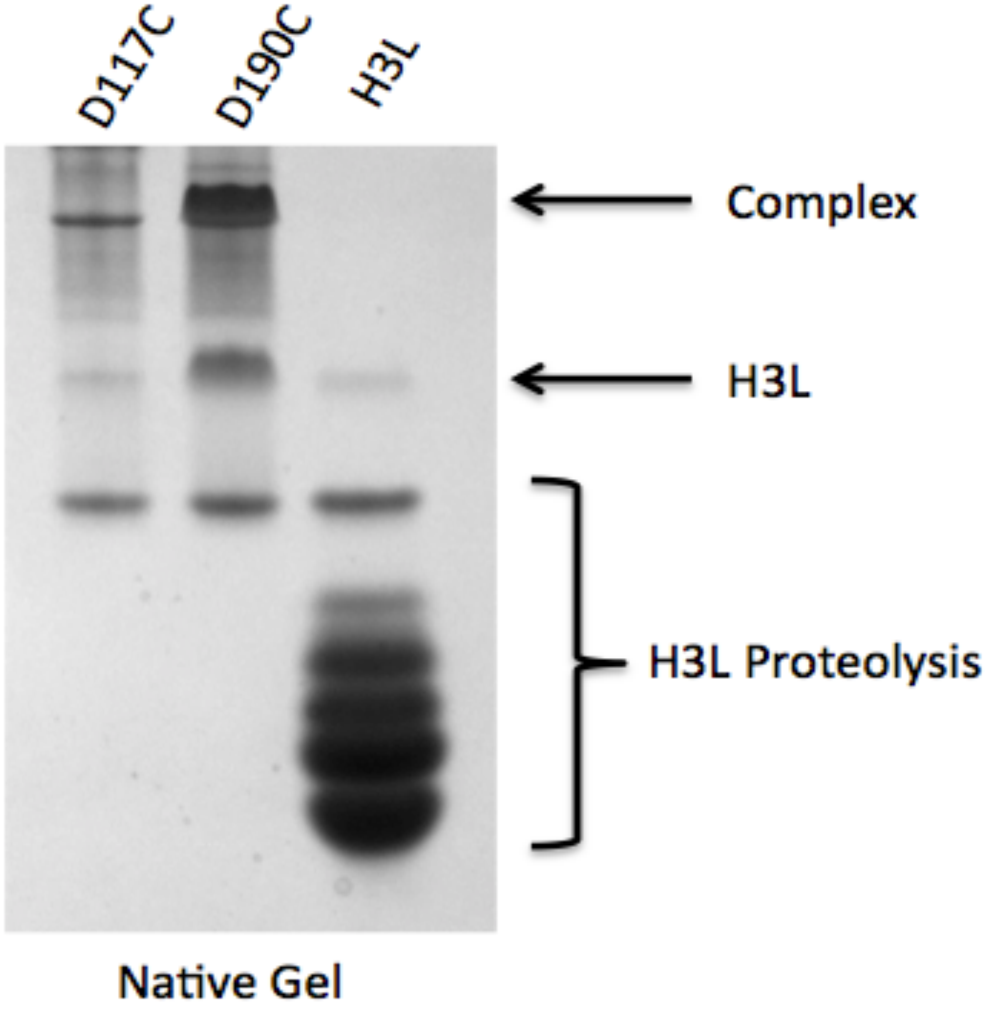

D117C
D190C
H3L
Complex
H3L
H3L Proteolysis
Native Gel

Figure 3.19. Native PAGE Gel from SEC Purified GFP-H3L Complex. Native gels analyzed the fractions pooled from the SEC runs for complex purity. For both mutants the complex was contaminated with monomeric H3L and a proteolysis production from H3L. The complexes were from thawed protein that was flash frozen immediately after purification. The H3L lanes is protein that was left at 4°C for one week and was significantly degraded by a contaminating protease.

IEX was performed first as it was more applicable to the large volume of the sample after the dimerization step. IEX alone was not sufficient to remove the contaminating H3L (figure 3.20). Once the IEX complex peak was run over the SEC column, all contaminating H3L was removed and a single band was present by native gels analysis (figure 3.21). Once this method was developed it was repeated for all five GFP cysteine mutants for both the internal and N-terminal hairpin H3L constructs. After approximately seven months, crystals were obtained in only one condition: the internal 10/11 sequence with the D117C mutant in 0.1M SPG Buffer pH 5.0, 25% w/v PEG1500. These crystals only diffracted to ~10Å and could not be reproduced or optimized (figure 3.22). Optimization experiments resulted in protein precipitation in the crystallization drops.

3.5.2 StarD9 Kinesin

The Kinesin StarD9 has been studied by the Jorge Torres lab at UCLA as a potential anti-mitotic target for novel cancer therapies.[69,70] The structure of the protein was desired as a basis for structure-guided drug design efforts. Having failed to crystallize on its own it was used as a test protein for the metal-chelating GFP mutants as it contained native cysteines and therefore incompatible with the disulfide mutants. Out of the metal-chelation mutants only D21H/K26C and E124H/K126C complemented during coexpression. Low yields due to very low expression of StarD9 limited the crystallization experiments. Small crystals grew in one condition (figure 3.23), but they failed to diffract and could not to be optimized. Alternate means to crystallize StarD9 were undertaken, as discussed in chapter 6.

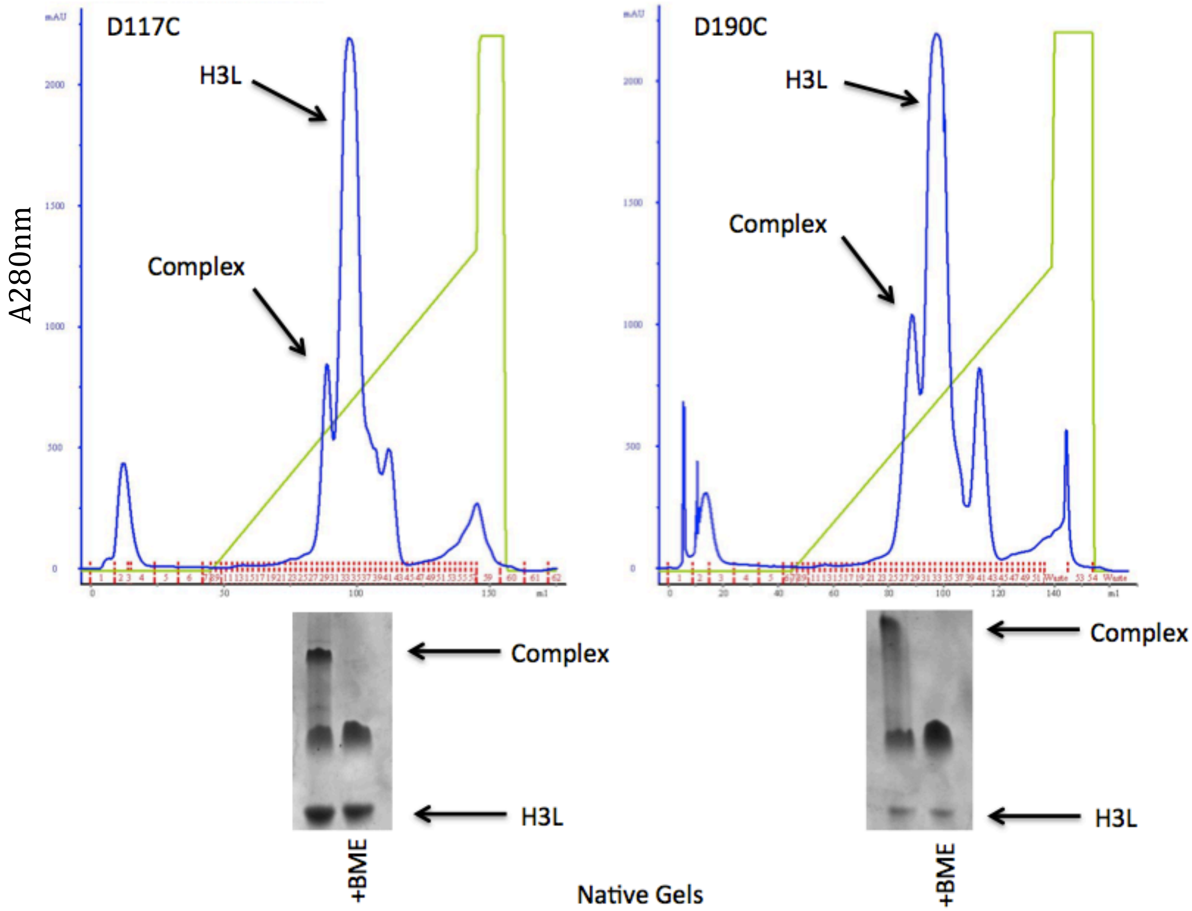


Figure 3.20. Anion Exchange Chromatograms and Native PAGE Gels of GFP 1-9 – H3L Complexes. IEX of the GFP-H3L complexes resulted in a minor peak for the complex and a major peak for H3L monomers. The complex peak contained contaminating H3L and monomeric complex.

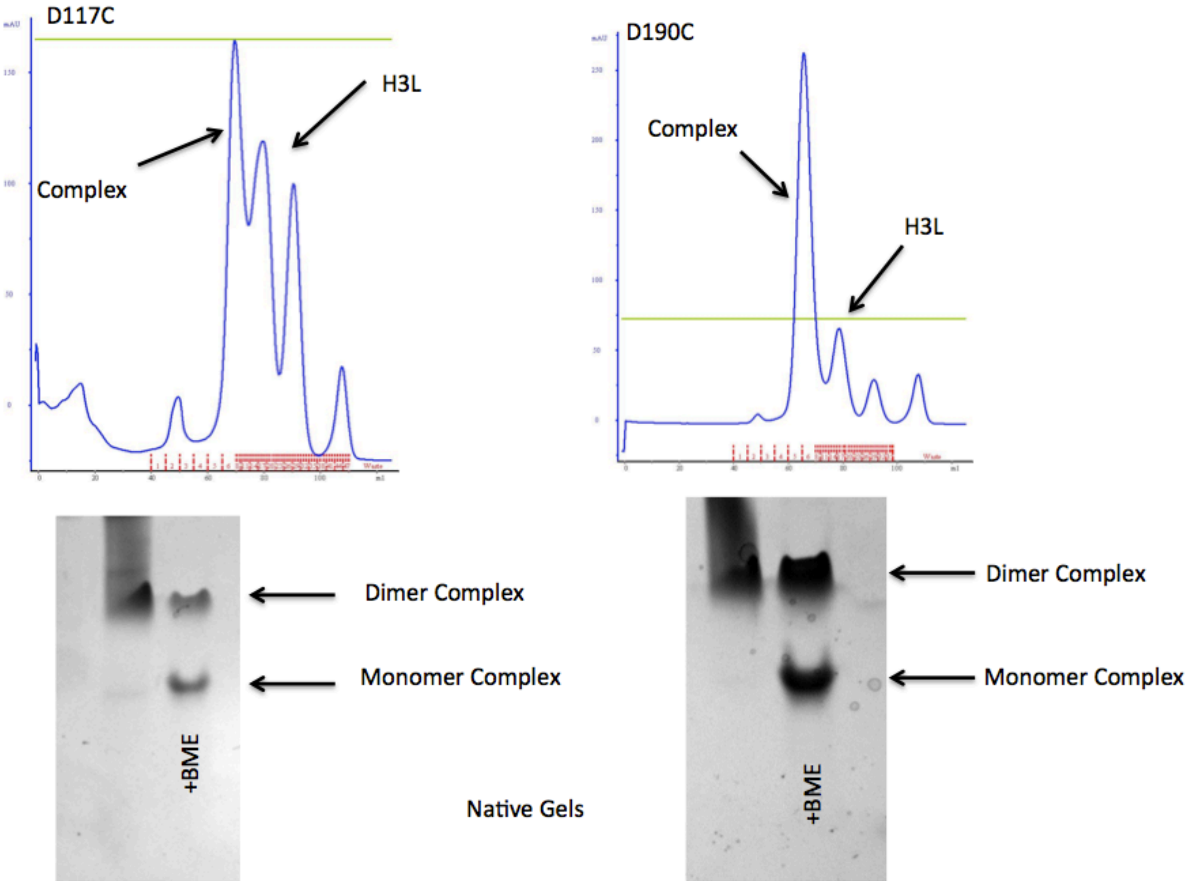
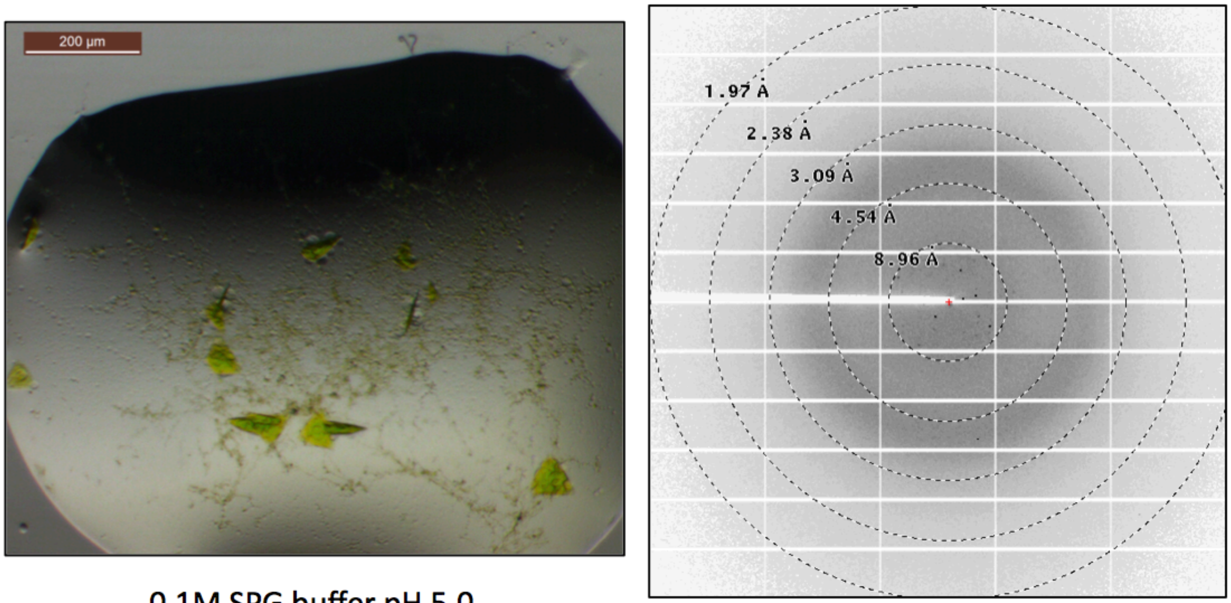
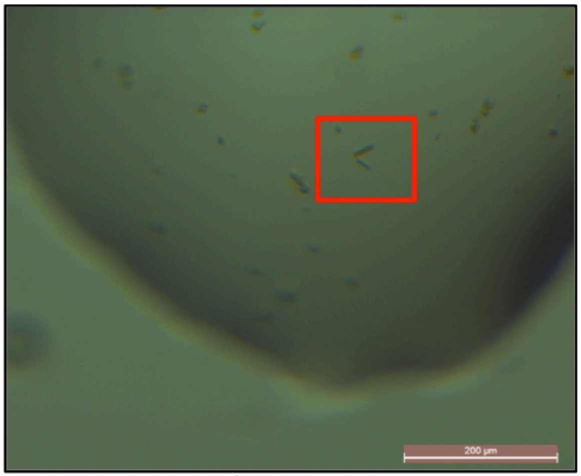


Figure 3.21. SEC After Anion Exchange of GFP 1-9 –H3L Dimer Complexes. SEC of the complex peak from IEX was able to fully resolve the GFP-H3L dimer complex from contaminating monomeric complex and monomeric H3L. Native PAGE gels showed smeared band for the SEC complex peak, smearing of the band is likely due to overloading the gel. With the addition of β-mercaptoethanol the dimer complex was partially reduced to the monomeric state.



0.1M SPG buffer pH 5.0
25% w/v PEG-1500.

Figure 3.22. Crystals of GFP D117C – H3L Dimers. After approximately seven months crystals of GFP-H3L dimers from a D117C mutation were obtained in a single condition. These this triangle plate crystals of 50-75μm diffracted weakly to ~10Å at the Advanced Photon Source synchrotron.



10% v/v 2-Propanol
0.1M MES pH 6.0
0.2M $Ca(OAc)_2$

Figure 3.23. Crystals of GFP D21H/K26C - StarD9 With Cu^{2+}. Crystals grew in approximately three months. Cu^{2+} ions were added immediately prior to crystallization experiments in order to induce oligomer formation. These small rod crystals (~2µm x20 µm) did not diffract.

3.5.3 sfCherry I206K

The structure of the complex of GFP 1-9 – sfCherry (figure 3.1) featured a crystallographic dimer along strands 10/11. This dimer had previous been observed in GFP structures and it was predicted that a mutation of I206K on GFP strand 10 would disrupt this native dimer.[71] The sfCherry – GFP 10/11 I206K complex failed to crystallize (Natasha Devore, personal communication). Due to the difficulties in obtaining crystals for this I206K mutant it was decided to pursue this as a target for crystallization with the GFP disulfide dimers. With the sfCherry I206K, both proteins (sfCherry and GFP) were known to have previously formed diffraction quality crystals, unlike MtbCon, H3L and StarD9. The ability of the suite of dimers to rescue the mutant construct would provide a proof-of-concept for the scaffold-induced symmetry as a rescue method.

Coexpression proved to be difficult for this complex; the sfCherry I206K construct failed to readily form a complex with GFP 1-9. The majority of the protein after IMAC was sfCherry alone. The position of the I206K on one of the complementing strands may have reduced the complementation efficiency resulting in the low yields. SEC was performed after IMAC without oxidation and formation of the disulfide bond. The minor peak corresponding to the GFP-sfCherry complex showed prevalent cleavage of sfCherry into ~20 kDa and 11 kDa bands (figure 3.24). It was brought to our attention the sfCherry undergoes light-induced cleavage into the two smaller bands (Geoffrey Waldo, personal communication). Protein was pooled for all five GFP cysteine mutants in complex with sfCherry I206K and one sparse matrix crystal screen (JCSG+) was set for each mutant at the maximum concentration (~2 – 5mg/mL) that could be obtained with the minimal volume (40 µL) needed for each tray. No crystals were ever observed for these trays. Because of the low efficiency and light induced sfCherry cleavage, no further work was performed.

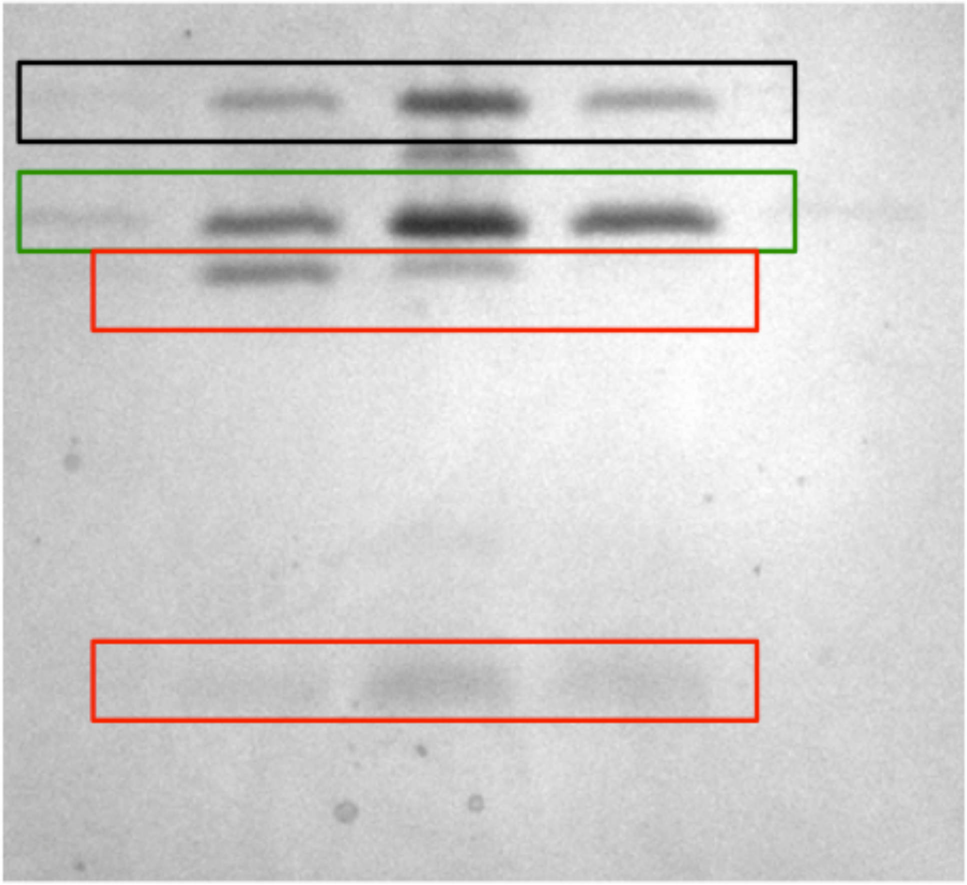


Figure 3.24. Size Exclusion Chromatography Fractions of sfCherry I206K – GFP 1-9 K26C Complex. Fractions that corresponded the complex peak indicated bands for intact sfCherry/10/11 I206K (black box), GFP 1-9 K26C (green box) and light induced cleavage products from sfCherry (red box). Presence of the cleavage products made it difficult to obtain intact complexes. For each mutant the fractions containing the least contamination were pooled for crystallization experiments.

3.6 Discussion

3.6.1 Viability of *In Vitro* GFP Complex Formation

For synthetic symmetrization to be viable for wide spread use as a rescue pathway for failed crystallization experiments it must be a facile approach. As structural biology and x-ray crystallography are increasingly being employed by those in biological sciences the importance of protein structures is ever growing. Difficulty in obtaining the structures remains a roadblock. Facile approaches to improving the outcome of a crystallization experiment can easily be performed by modifying an existing protein (e.g. lysine methylation) or by changing the crystallization conditions (additives, oil overlays).[3,72–74] With these techniques, no genetic changes need to be made the crystallization target construct. The potential for *in vitro* complex assembly using pre-formed cores of various geometries spurred the development of the split GFP oligomers. *In vitro* complementation is key for this approach to be widely used, as it would create a technique that is no more difficult than lysine methylation.

Unfortunately, the difficulty in obtaining the pre-formed GFP disulfide dimer cores prevents the *in vitro* complementation approach. With the current version of split GFP, the introduced mutations cannot serve our intended purpose because of the limitations in the robustness of split GFP. GFP 1-9, even with no mutations, is a difficult protein to work with. The largest hurdle to overcome is the low efficiency of refolding. At best, ~15% of refolded GFP 1-9 can be complemented. This is acceptable for a biosensor application where only the presence of the target protein is probed. For these experiments an excess of refolded protein can be used to overcome the low efficiency. However, this low efficiency makes it impossible to make the preformed dimer as the dimers will be have a large portion consisting of nonreactive GFP 1-9 (>85% of GFP 1-9 fails to complement or exists as soluble aggregates). SDS-PAGE gels of dimerized, refolded GFP 1-9 indicated that nearly half of the protein in solution is readily reactive for disulfide formation. Although it would be possible to purify intact *in vitro* dimer complexes (i.e. two GFP 1-9 complemented by two POIs) the low efficiency and necessary purification eliminates any benefits of the *in vitro* approach over coexpression of the two components.

Decreased efficiency for the mutated GFP 1-9 constructs to complement further complicates the *in vitro* applications. The required C48A/C70A mutations contribute to the most prominent reduction in efficiency (~20%). Directed evolution of the split GFP sequence to tolerate these may rescue the complementation efficiency. A more robust core could eliminate many of the contamination issues from non-complemented GFP 1-9. Evolution of the spilt GFP 1-9 sequence into a cysteine free backbone, similar to what was done to create split GFP initially, should be pursued.

Mutations with more dramatic reductions in complementation efficiency (D102C, metal-chelation mutants) could be eliminated from the suite and replaced with new mutants. The selected mutations (chapter 2) are only a few of the potentially viable locations, new mutations in different location on the GFP structure may be better tolerated. These should be empirically validated for complementation efficiency and only locations that are the most tolerant pursued. The efforts to determine the individual propensity for new mutants to crystallize readily and in different space groups do not need to be pursued. All work to date for synthetic symmetrization mutation suites (T4L, MBP, GFP)[6,8] have shown each mutation crystallizes in a different manner. The work in chapter 2 is indicative of the ability of GFP oligomer mutants to readily crystallize, as all mutants resulted in diffraction quality crystals.

Further, the inability to sufficiently concentrate the GFP 1-9 mutants after refolding is detrimental for *in vitro* complentation with preformed cores. Since the GFP 1-9 cores exist in a semi folded state they will remain prone to aggregation. This will likely remain an issue even if the cores are evolved to regain, or improve, the complementation efficiency to tolerate the introduced mutations. The semi-soluble state in which the refolded protein exists is extremely sensitive to centrifugal concentration. When the protein is concentrated through IEX purification and elution, the refolded protein remains soluble for some time but still remains prone to aggregation. Evolution of the GFP 1-9 will not eliminate this; the cores would still be plagued by unintended and unpredictable intermolecular interactions.

As the suite of mutants exists today, *in vitro* complementation with preformed cores is not practical. Significant modification to the mutant suite needs to be performed to fully establish complementation efficiency. Unless the aggregation issues can be addressed, which as of no is unlikely, it is not possible to have the GFP 1-9 dimers in stable frozen aliquots ready for use as needed.

3.6.2 Purification of GFP 1-9 –POI Complexes

Although *in vitro* complementation is not practical, the suite of oligomers is still useful for crystallization experiments. Use of the mutants of GFP 1-9 for the crystallization of POIs can provide an easy way to induce symmetry onto novel proteins where direct mutation of the POI to create oligomers is difficult. The efforts to establish a purification scheme of IMAC, IEX and SEC to obtain pure dimer complexes is applicable to any GFP 1-9 – POI complex created through coexpression and *in vivo* complementation. The H3L crystals are a clear example of this approach working. If time allowed for further condition screening and optimization it is likely the crystals could have been improved. The ability to obtain highly pure complexes (e.g. a single band by native PAGE) is essential for crystallization success, and the purification difficulties have been overcome.

3.6.3 *In Vivo* Complementation of Split GFP for Complex Formation

In vivo complementation of the GFP 1-9 – POI complexes will likely be required for future efforts involving the GFP dimer cores. As of now, this has been the only way that the complexes have been formed. The structure of GFP 1-9 – sfCherry was obtained from adding refolded GFP 1-9 to cellular lysate of an sfCherry expression. From a technical perspective for purification and complex formation it is no different from coexpression of the two components. Coexpression of the complex would likely be more efficient than the addition of refolded protein to cell lysate.

In vivo complementation also allows for rapid screening of both new mutants of split GFP and POI constructs containing the hairpin sequence. When complementation occurs, the green color of the cell pellet is a clear indication of successful complementation, with the intensity of the color dependent on the amount of complex formed. This provides a simple screen for variations of both parts of the complex without the need for protein purifications.

3.7 Conclusions

The difficulty in the validation of the GFP oligomer cores has primarily resulted from the solubility and refolding issues of split GFP. Alternative methods for oligomer formation and complex formation need to be evaluated. Split GFP complementation relies on the partially folded GFP 1-9; even with improvements to the backbone these issues will likely continue. GFP has proven to be accommodating to oligomer formation and crystallization when expressed as a full-length protein but has failed to perform well as an *in vitro* crystallization tool. However, this concept should not be abandoned due to the difficulties split GFP has. Instead alternate means of covalent and non-covalent protein assembly can mediate symmetry scaffolds. This becomes a more viable option, as computational protein interface design is refined.

3.8 Materials and Methods

3.8.1 Cloning of POI Constructs

The two H3L constructs were cloned into the multiple cloning sight of the pET21_NESG vector through restriction enzyme cloning. The synthesized gene contained the GFP 10/11 hairpin sequence “DLPDDHYLSTQTILSKDLNEKRDHMVLLEYVTAAGITDAS” either N-terminally or inserted between two full-length H3L sequences. The synthesized constructs were cloned into the NdeI and XhoI restriction sites of the vector.

The STARD9-10/11 construct consisted of the N-terminal TEV protease cleavable His6 tag (MGSDKIHHHHHHENLYFQG) followed by the 10/11 sequence, with the ‘‘DAS’’ serving as a linker between the hairpin and target protein as previously described[13]. Only the first 391 amino acids (Met1– Asn391) corresponding to the putative motor domain of the protein were used in this construct. The 10/11 hairpin sequence was inserted into a presumptive loop between Ser135 and Thr136 of the native 271-amino-acid protein. This was cloned via Gibson assembly[75] into a custom modified pET28 vector with N-terminal cleavable tag to add the N-terminal sequence: MGSDKIHHHHHHENLYFQG. The gene was PCR amplified with the overhangs 5'-AAAACCTGTACTTCCAGGGC added to the forward primer and AACGAGTTAATTAAGTCGCGTTA added to the reverse primer. This was then gel purified and cloned into vector that was prepared by PCR amplification.

The sfCherry I206K was obtained from collaborators and was cloned via previously described methods.[13]

MtbCon was obtained from collaborators with the C-terminal 10/11 hairpin sequence. This was then sub cloned into the same modified pET28 vector as the StarD9 construct via the same method.

3.8.2 Protein Expression

Vectors were all transformed into BL21 (DE3) expressions cell lines, with the exception of the SHuffle cell experiments where those cells were used. In all cases overnight cultures were grown with the appropriate antibiotics, 10 mL of culture was inoculated into 1L flasks of LB media with appropriate antibiotics. For GFP 1/9 constructs only auto induction sugars[57] were added and the cultures grown at 37°C for 24 hours before harvesting by centrifugation at 5000 x g for 20 minutes. Coexpression of GFP 1/9 with H3L and sfCherry 1206K was grown at 37°C until the A_{600nm} reached 0.6-0.8, then the temperature was reduced to at 30°C, expression induced with 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown overnight (~16 hours) then harvested. Coexpression cultures of GFP 1/9 and StarD9 grown at 37°C until the A_{600nm} reached 0.6-0.8, induced with 0.1mM IPTG then grown for four hours until harvesting.

3.8.3 Protein Purification

Refolding of GFP 1/9 proceeded as follows. Prior to refolding the inclusion bodies cell paste was lysed in 50mM Tris pH 7.4, 100mM NaCl, 10% Glycerol (TNG) buffer via sonication. Inclusion bodies were pelleted by centrifugation at 12,000 xg for 30 min then suspended in a buffer of 50mM Tris pH 8.0, 100mM NaCl, 1mM EDTA, 2% Triton X-100, then pelleted. This was repeated until the inclusion body pellet was white. In all experiments the inclusion bodies were denatured in TNG buffer + 9M urea at room temperature. Then refolding was performed as previously described.[12]

For binding to the Ni^{2+} resin binding of refolding protein, the refolded protein was applied to the resin, washed with TNG + 30mM imidazole and eluted with TNG + 250mM Imidazole. During on-column refolding attempts the denatured protein was loaded onto the resin in the presence of TNG + 9M urea, a gradient to TNG with no urea was run over 20 column volumes. Protein was eluted with TNG + 250mM imidazole. For on-column refolded the TNG + 9M urea and TNG were supplemented with 30mM imidazole for the initial experiments and subsequently no imidazole was added to the buffers.

For refolded protein anion exchange, refolded protein was diluted with 10x volume 100mM Tris pH 9.5, 10% glycerol (IEX buffer A); this was loaded onto the IEX column. A gradient was ran of 0 – 100% IEX Buffer B (IEX Buffer A + 1M NaCl) over 20 column volumes.

For all instances of *in vitro* dimerization the protein was concentrated with centrifugal concentrators (Amicon) until precipitation was observed (if it occurred) otherwise it was concentrated to 1mL, buffer exchanged into 50mM Tris pH 9.0, 100mM NaCl, a 10x volume excess of 50mM Tris pH 9.0, 100mM NaCl, 5mM $CuSO_4$ was used to oxidize the disulfide bond. This reaction was allowed to proceed at room temperature for 15 min and quenched with the addition of 10mM EDTA. The protein was then analyzed by SDS-PAGE or concentrated and buffer exchanged into the start buffers of the any additional purification steps.

Purification of the GFP 1/9 – sfCherry I206K constructs proceeded as previously described[13] with the exception of both proteins were coexpressed to allow for *in vivo* complementation. The complex was purified via IMAC and SEC over a s200 10/300GL column due to the low protein yields. H3L and StarD9 coexpressions were purified as described in chapter 2. Of note the StarD9 buffers were supplemented with 5mM $MgCl_2$ and 100µM ADP. For both H3L and StarD9 the purification scheme of IMAC, IEX and SEC in that order were performed. Both complexes were concentrated to 10mg/mL, flash frozen in 20-40µL aliquots (depending on yield) and stored at -80°C until used for crystallization experiments.

3.8.4 Dynamic Light Scattering

Freshly refolded GFP 1/9 mutant protein was filtered through a 0.22μm filter then immediately analyzed by DLS. DLS was performed on a Wyatt DynaPro Plate Reader II at room temperature. Default settings were used for data collection, 10 individual reading were measured per sample. Readings with acceptable auto-correlation functions per Wyatt automatic data processing were included in the presented histograms.

3.8.5 Crystallization of Complexes

Fresh or thawed protein was centrifuged at 10,000 xg for 5 min to remove any precipitate. Crystallization screening trays were set at a protein concentration of 10mg/ml or the maximum concentration obtainable for the protein samples. Commercial screens were set utilizing primarily JCSG+ (Qiagen), Wizard (Emerald BioStructures), Index (Hampton), and PACT (Qiagen). Additional screens were used when more protein was available. In instances where widespread protein precipitation was observed in the crystal screens, the protein was diluted to 5 mg/mL and new screens were set.

Chapter 4

Computationally Designed Trimeric GFPs

4.1 Introduction – Computationally Design GFP Oligomers

For oligomeric protein scaffolds for synthetic biology applications to be viable, higher order symmetry (trimers, tetramers, etc.) may be necessary. It was hoped that the metal-chelating mutants would adopt higher order oligomers than just dimers as has previously been observed.[8] Large metal contact mediated oligomers were found in some of the crystal asymmetric units, but they lacked rotational symmetry and did not form a defined symmetric unit. To achieve these desired higher order oligomers, computational design through Rosetta MatDes[16,17] was pursued.

Recently, this approach has been used to successfully design self assembling oligomers with rotational symmetry,[67] and more complex single[18] and multi[17] component cages. Through Rosetta Design, the protein subunits are docked together into the desired symmetry, and then the contacting subunit's surface is designed to form an interface. To date, it has primarily been alpha-helical domains that have been used to design these assemblies. Utilizing Rosetta Design to create GFP oligomers would the first instance where it has been attempted on beta-sheet proteins. Designed oligomers could then be used for the higher order assemblies, involving the mutants discussed in chapter 2 or through previously published rigid assembly techniques such as helical fusions[76–81] or further Rosetta Design.

Current versions of Rosetta MatDes create interfaces by side chain mutation to hydrophobic residues. In both published[17,18] and unpublished work from the David Baker lab, these hydrophobic mutations are problematic in two ways. First, the newly mutated hydrophobic patches can lead to protein misfolding and aggregation. Because of this, numerous designed interfaces must be screened for each designed assembly. GFP allows for rapid screening as green color in the cell pellet and soluble lysate indicates soluble folded protein. Second, the hydrophobic interfaces can be non-specific leading to individual subunit that can arrange in alternate geometries. In collaboration with computational postdoctoral scholar Dr. Julien Jorda, we set about designing a series of GFP mutants docked to a C3 symmetry axis.

To create the series of designs, the structure of superfolder GFP[49] (PDB 2B3P) was used as the starting sequence. This version of GFP was the precursor to spit GFP previously discussed. Superfolder GFP was engineered to be a robust, readily folding GFP variant to serve as an indication of protein expression. Being one of the most robust versions of GFP, it was thought this sequence might be more tolerant to mutations than other versions (EGFP[82], split GFP etc.). The chains were arranged around the C3 symmetry axis with the central axis of the GFP barrels perpendicular or parallel to the symmetry axis. Subunits were allowed to dock together, then rotated 5° and docked again (figure 4.1). Orientations where the subunits formed energetically favorable arrangements were then used for interface design. Sequences were then evaluated and those with the most favorable interface statistics were evaluated by x-ray crystallography.

4.2 Results – Perpendicular Docked C3 Trimers

Perpendicularly docked barrels were initially evaluated. In this orientation the GFP termini point away from the three-fold symmetry axis. This would allow terminal fusion to proteins with a reduced chance of steric hindrance since all three N or C –termini point in different directions. Clashes may still occur between the trimeric GFP core and the fused proteins. This could be addressed in future experiments once a stable trimeric core is established by engineering the termini through truncations or insertion of rigid linkers[78] to lock the geometry of fusions.

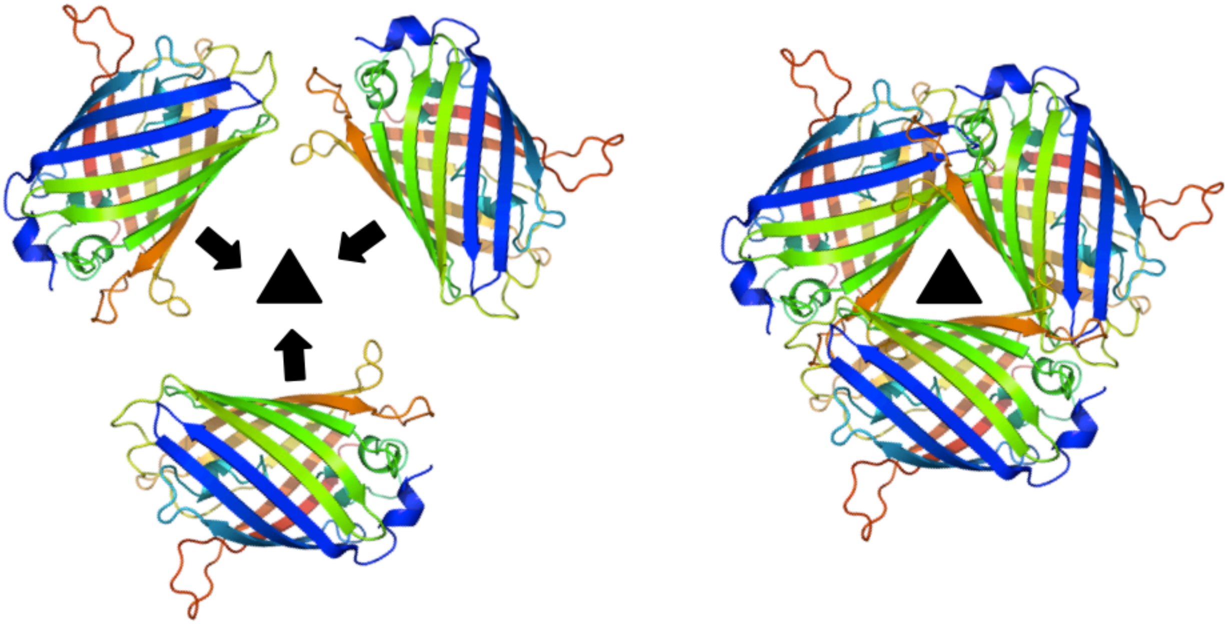


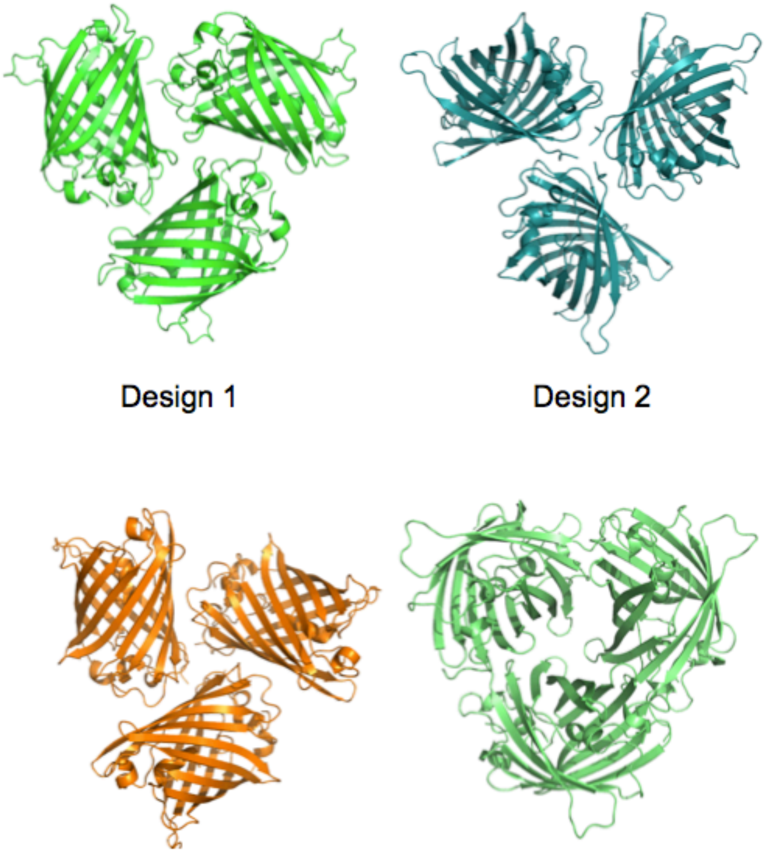
Figure 4.1. Docking of GFP Monomers to a Three-Fold Symmetry Axis. Initial design strategy focused on docking the GFP barrels perpendicular to the symmetry axis (triangle). The GFP barrels start away from the axis and are allowed to dock together towards the axis. If the barrels dock in an energetically favorable manner, interfaces are designed. The starting GFP barrels are then rotated 5° and docked again in the new orientation.

Ten docking runs were performed, and the most stable orientation from each run was then used to generate 20 sequences. Out of the ten runs, four distinct chain orientations were obtained (figure 4.2). Sequences (table 4.1) were then manually curated to five sequences per orientation. Sequences were selected based on Rosetta energy score, shape complementarity and buried surface area (table 4.2) following previously published guidelines.[17] As the designed interfaces for GFP timers are smaller then the previous examples, the corresponding energy values are smaller in magnitude. A final check was done on the sequences to ensure they contained key residues for chromophore formation and stabilization.[83–85]

4.2.1 Purification and Crystallization of C3 Designs

All twenty sequences were successfully cloned. From small screen expression tests, four designs resulted in green cell pellets (figure 4.3), while other showed indications of protein expression, and when they were purified only a small fraction of the protein was soluble. The four green pellets (indication of proper folding) were expressed and purified by IMAC and SEC. Each construct sized as a monomer by SEC elution volume (figure 4.3). In order to determine the oligomeric state of the protein, native-PAGE gels were run and compared to known GFP dimers. All four were putatively monomers on the native gels and one (C3_14) was partially aggregated.

All four proteins were used for crystallization experiments. Crystals grew rapidly for three of the four designs (C3_14 failed to crystallize). Crystals grew in diverse conditions, with diverse morphologies (figure 4.4). One crystal (C3_33 JCSG+ D12) had data collected at UCLA. All crystals that could be harvested were subsequently sent for data collection at the Advanced Photon Source. No efforts were taken to optimize crystal growth conditions after these initial crystals were obtained.



Design 1

Design 2

Design 3

Design 4

Figure 4.2. The Four Orientations of Perpendicular C3 Trimers. From the ten individual computational docking runs, each run generated one trimer design. Four general classes of trimer designs were found to be energetically favorable in from these 10 runs. The design numbered by occurrence (Design 1 most common, Design 4 the least from the 10 runs) Five sequences from each design when then selected for experimental characterization.

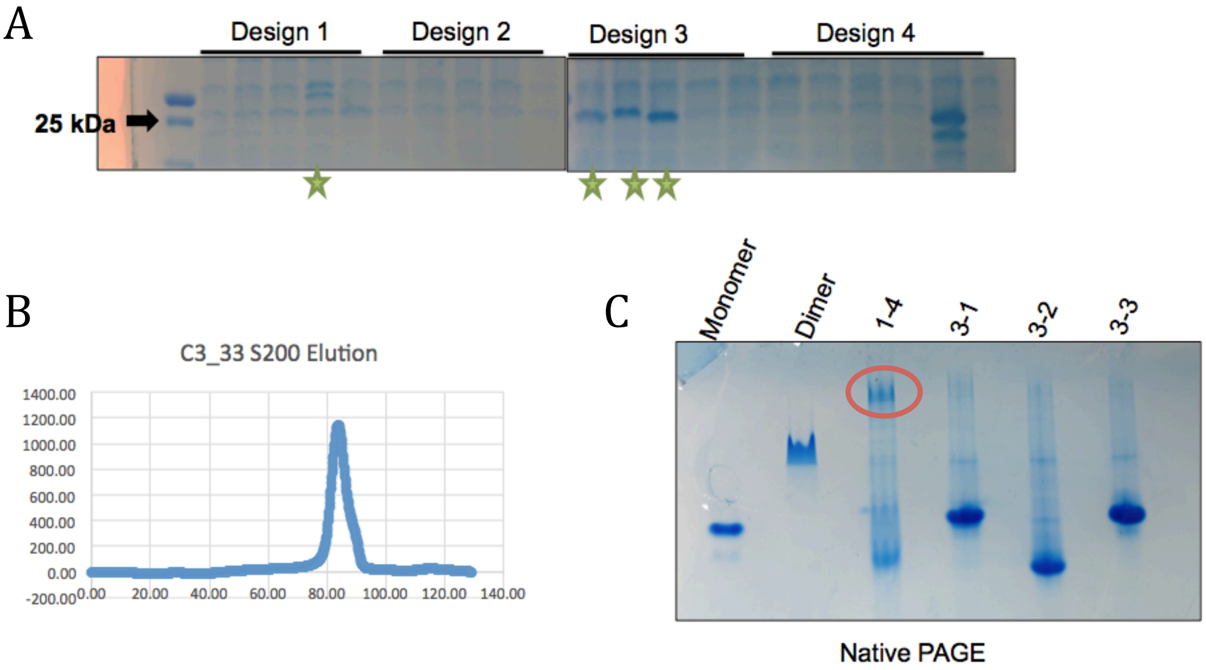
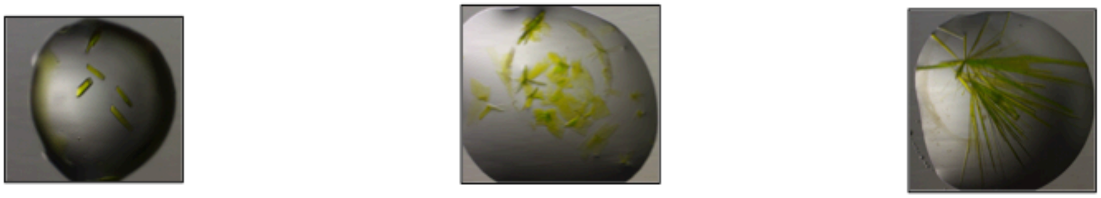


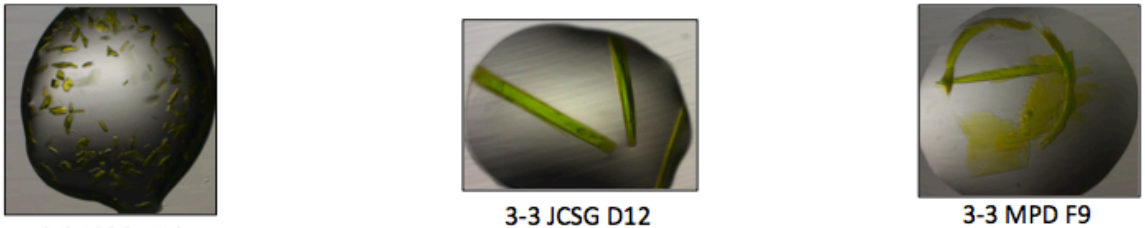
Figure 4.3. Expression Tests and Purification of Designs. (A) SDS-PAGE gels of the soluble from expression of the four designs. The 25 kDa bands correspond to the molecular weight of the GFP barrels. Designs that resulted in soluble, properly folded protein. The final two lanes were point mutations that emerged during cloning and did not result in properly folded protein. (B) All four SEC elution profiles from the purifications featured a single peak that eluted at the volume consistent with monomers. (C) Native-PAGE gel of the four SEC purified proteins compared to known monomers and dimers of GFP. Sequence variations caused the proteins to run at different rates, but all are putatively monomeric. Design C3_14 (1-4) was partially aggregated, as the protein did not enter the gel (red circle).



3-1 JCSG A9
20% PEG 3350
0.2M (NH4)Cl

3-1 MPD F9
65% MPD
0.1M MES pH 6.0

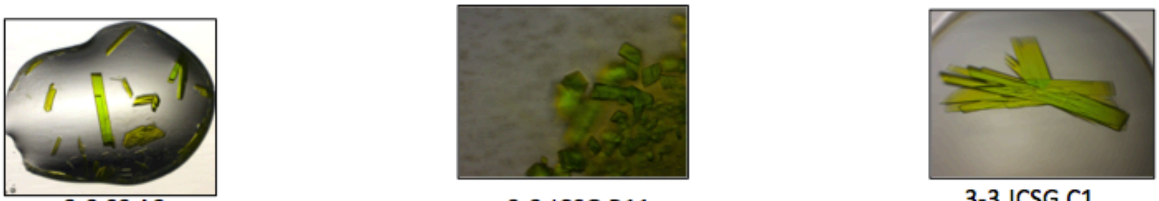
3-1 MPD F10
65% MPD
0.1M HEPES pH 7.0



3-3 JCSG H10
25% PEG 3350
0.1M Bis-Tris pH5.5
0.2M (NH3)OAc

3-3 JCSG D12
16% PEG 8000
20% Glycerol
0.04M KPO4

3-3 MPD F9
65% MPD
0.1M MES pH 6.0



3-3 SS A2
30% PEG 400
0.1M NaOAc
0.2M (NH3)OAc

3-3 JCSG B11
1.6M Na Citrate pH 6.5

3-3 JCSG C1
20% PEG 8000
0.1M PO4-Citrate pH 4.2
0.2M NaCL

Figure 4.4. Crystals of C3 Designs. The crystals observed from the C3_31 (3-1) and C3_33 (3-3) designs, the only two that crystallized. These represent all the crystal morphologies observed.

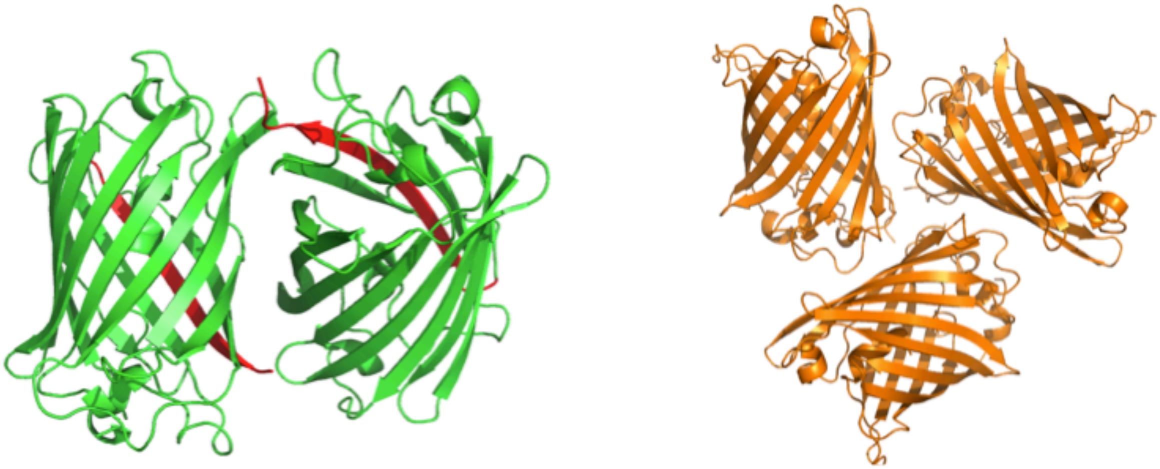
4.2.2 Structure of C3_33 JCSG+ D12 Dimer

Large crystals of the construct C3_33 grew rapidly in the 16% PEG-8000, 20% glycerol 0.04M KPO_4 (C3_33 JCSG+ D12). Crystals were rapidly cooled in liquid nitrogen and a complete data set was collected. Data was indexed to 3.0Å in space group C2. The Matthews coefficient[86] predicted[87] two GFP chains in the asymmetric unit, the C3_33 design monomer was then used to determine phases through molecular replacement program Phaser.[88] The structure was found to have formed a dimer instead of the intended trimer (figure 4.5). To confirm that this was the true solution the structure was partially refined with Refmac.[89] When the R_{work} and R_{free} dropped below 30% it was determined to be the correct placement of the chains in the electron density. This was an asymmetric dimer that was formed by the designed hydrophobic patches associating in an unpredicted manner. Interactions involving the end of the GFP to form the trimer were not present. The crystallographic dimer was significantly different from previously observed GFP crystal artifact dimers (figure 4.6).

All other crystals (figure 4.4) that diffracted had data sets collected and processed. A total of six other conditions had crystals that diffracted between 1Å – 3Å. Five of the six were in space group C2 (with unit cell dimension variations), and the sixth was in C222. Structures were determined with Phaser and refined with the PDB_Redo server.[90] All structures refined to R-factors less then 30%. In each instance the same dimer as the C3_33 JCSG+ D12 structure was found. No additional experiments were performed with these designs.

4.3 Results – Parallel Docked C3 Trimers

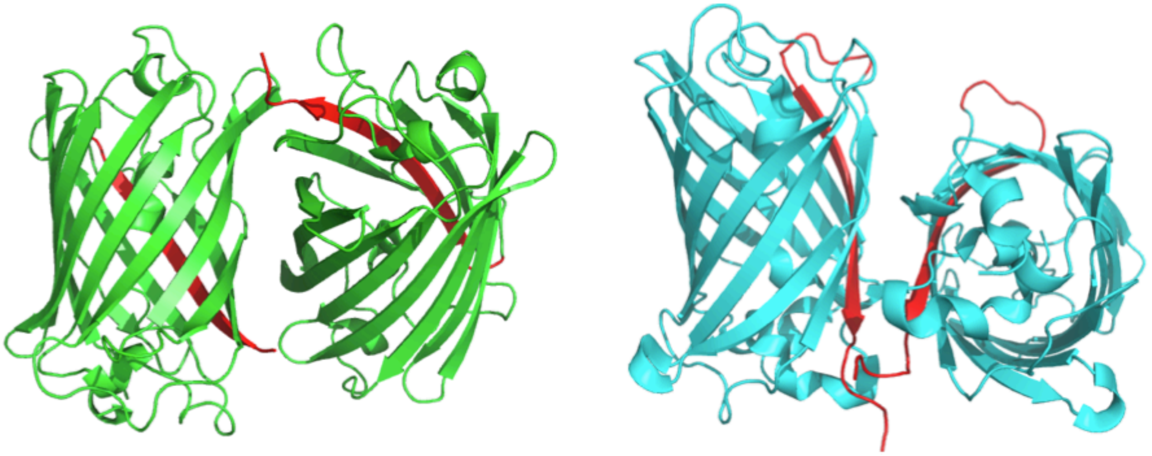
The GFP barrels were docked parallel to the symmetry axis (figure 4.7) and then used to make new mutants. The same approach for design was used, but now 10,000 sequences were designed and 20 automatically selected from the most favorable Rosetta statistics (table 4.3).



Structure

Design

Figure 4.5. The C3_33 JCSG D12 Dimer Compared to Design. The dimeric structures varied significantly form the intended design with none of the designed contacts preserved.



3-3 JCSG D12

sfGFP Crystal Dimer – "Native" GFP Dimer

Figure 4.6. Observed Dimer is a New GFP Crystallographic Artifact Dimer. The observed dimer is different from the traditionally observed "native" GFP crystallographic dimer. The previously observed dimer is a c2 symmetric dimer with the interface primarily along strand 11. Strand 11 has been colored in red for both dimers to highlight the differences in the structures.

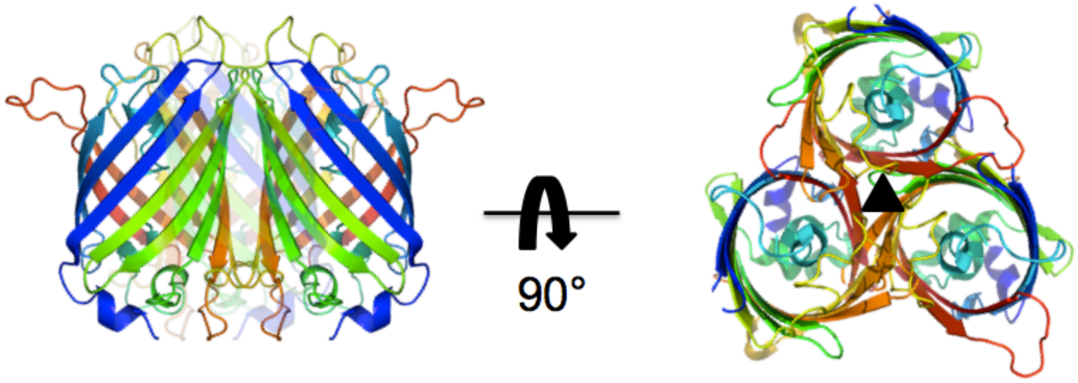


Figure 4.7. General Design Orientation of Parallel Trimers. With the failure of the perpendicular dimers, new designs were based on trimers with the GFP barrels parallel to the symmetry axis. With these designs, the interface would be larger and no rely on any interactions from the loops on the end of the barrels. However, all the termini would be on the same end of the bundled trimer. All the termini in the same location have the potential to create steric hindrance for fusion proteins.

19 of 20 were the identical barrel orientation, the final design the barrels were rotated 5° about the GFP barrel axis.

4.3.1 Purification of C3_51 Design

Only one of the 20 designs was soluble (C3_51), a large-scale expression was purified via IMAC and SEC. The main peak from sizing column had a predicted mass of ~75 kDa consistent with a trimer in solution (figure 4.8). Cubic crystals grew with in a week in 0.1M citrate pH 5.5, 2.0M $(NH_4)_2SO_4$. Crystals were cryo-protected with glycerol and diffracted 1.5Å at the Advanced Photon Source.

4.3.2 Structure of C3_33 Trimer

Data was indexed in the cubic space group P 2_1 3 to 1.5Å with a unit cell edge of 107.53Å. Phenix[91] was used to refine the structure with a final R_{work}/R_{free} of 16.9%/18.6%, complete statistics are presented in table 4.4. Once GFP chain was in the asymmetric unit, a trimer was present through crystallographic symmetry operations. However, this trimer deviated significantly from the design (figure 4.9). Each chain was rotated ~17° from the designed model. Because of this variation the structure was fully refined to accurately see the trimer interface. The mutated residues from the design form the trimeric interface but none of the computationally designed interactions were present.

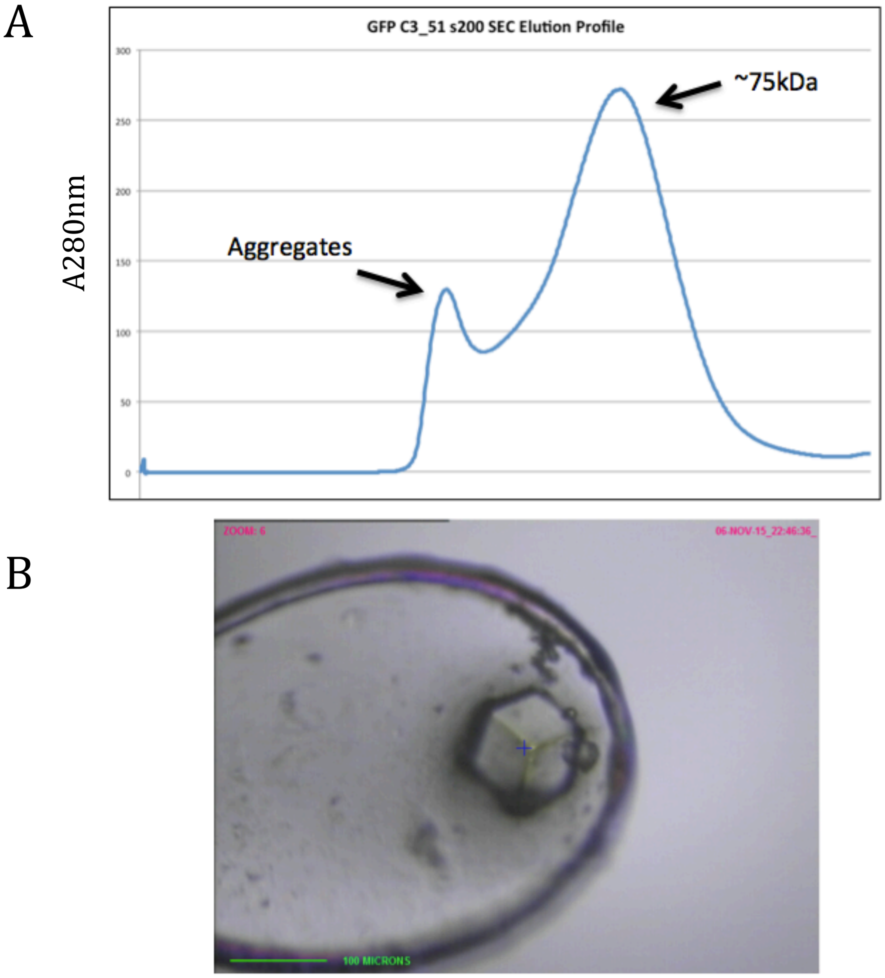
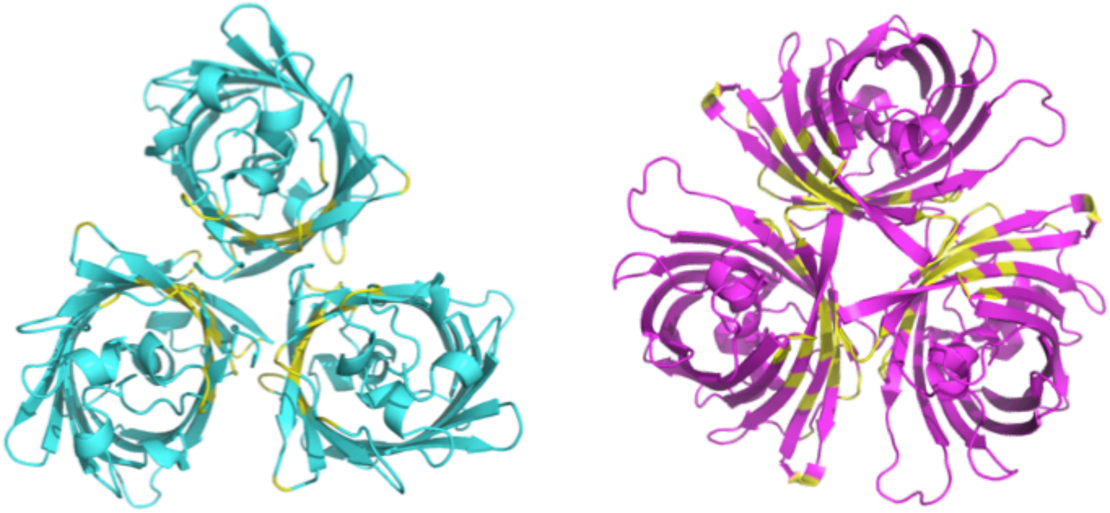


Figure 4.8. C3_52 SEC Elution Profile and Crystal. (A) The protein was run on an analytical sizing column resulting in poor resolution between the void peak and trimer peak. Only fractions of the trimer peak were used for crystallization. (B) The looped crystal used for data collection had cubic morphology. The crystal is colorless due to the lighting when the picture was taken. The crystal was faintly green in the crystallization drop due the acid conditions (pH 5.5).



Rosetta Model

Crystal Structure

Figure 4.9. Comparison of the C3_51 Design and Structure. The Rosetta Designed structure features all the barrels parallel to each other and the symmetry axis. In the crystal structure, the barrels were rotated approximately 17° from the design. The residues that were mutated for the design are highlighted in yellow. These residues formed the trimer interface for the crystal structure, but none of the specifically designed interactions were retained.

4.4 Discussion

4.4.1 Issues With GFP Computational Assembly Design

Rosetta's designed interfaces rely on the introduction of large hydrophobic patches with current approaches. These large patches of hydrophobic mutations have resulted in several issues. First, the vast majority (90% for the GFP designs) of the designed proteins failed to express, were insoluble or were misfolded. Second, the hydrophobic interactions tend to be non-specific, at least for the case of GFP. Previous unpublished work has experienced similar failures, with low solubility rates and a tendency to form incorrect geometries. Only one of forty mutants had the correct rotational symmetry, albeit the wrong interactions. It has proven exceedingly difficult to accurately design these interfaces.

The approach taken with the design of the GFP trimers diverged from traditional computational protein design. The published successes from the Baker lab have primarily come from designing new interfaces on the surface of an existing oligomer. One of the first successes was the design of a cage with tetrahedral symmetry constructed from a natural trimer.[18] Oligomeric building blocks are some what limited the orientations, in which they can assemble compared to a monomer like GFP. All previous design work has chosen the geometry of the assembly first, and then chosen the building blocks. The oligomeric building blocks have the monomers already locked into a starting geometry resulting in less freedom for individual monomers to form unintended interactions like was seen for the GFP designs. Building block subunits can then be computationally screened from structures in the protein data bank (PDB),[92] allowing the most ideal starting proteins to be used. Opposed to the approach we undertook with GFP, first selecting the building blocks and then attempting to impose the desired geometry onto it. The approach we took has the potential to allow more freedom to what is designed since it is based on subunits that are not locked into a starting geometry. Further advances in computational design strategies are needed to design interfaces with higher affinity for each other to prevent off target assemblies.

GFP may simply be ill suited for computational design efforts. Predominantly beta-sheet proteins have not been used to date for design, as alpha-helical proteins have proven more suited for Rosetta Design approaches. The GFP beta barrel is also concave along the surface making it difficult to pack tightly along the barrel, thus leaving the interfaces to be designed towards the end of the barrel. Potential geometries are further limited due to the ends of the barrel being composed of loops with a central void that cannot be designed. Successful designs would likely need to be designed with the GFP barrels parallel to the symmetry axis. Thus, the designability of GFP is limited at least with current methods.

Future GFP designs may be successful as more designed sequences are screened and as computational design improves. In lieu of waiting until new design methods are implemented into Rosetta, GFP can be well suited for an iterative design process. Instead of starting with the GFP monomers, the C3_51 trimer structure could serve as the starting point and designed back to the intended structure. This has not been an approach for computational design; instead success has come from the number of constructs screened. From this it could be possible to refine the design process and achieve higher success rates. A similar approach to this reconstructed a PduA permutation hexamer from an unintended pentamer (chapter 8).

4.4.2 Applications of New GFP Oligomers

Although the resulting structures were not as designed, these sequences are not without applications. Both the dimer and trimer are new geometries of GFP. These can therefore be used for the same applications as the split GFP oligomers in chapter 2. Since these new oligomer interfaces to no rely on *in vitro* efforts to form, they would be well suited to serve as scaffolds for proteins not applicable to a crystallization target. A POI that contains native cysteines or sensitive to metal ions could be terminally fused to the C3_33, C3_51 and an existing GFP (superfolder, split GFP etc). Each one of these has a different crystallographic oligomer, and therefore unique chances to serve as a scaffold that samples different crystal contacts. The C3_51 trimer and any future oligomers based upon the parallel, bundled arrangement will have the termini all in close proximity. This could create steric hindrance problems when proteins are fused to the oligomer GFP for scaffolding efforts. If stable oligomeric scaffolds with the desired geometry can be reliably designed, circular permutation of GFP[93,94] could be used to change the termini locations and allow for fusions. These new GFP oligomers should be screened as crystallization scaffolds as they will not have the solubility and expression issues encountered with *in vitro* complementation methods (chapter 3 & 5). Even though these scaffolds are not symmetric and thus would not benefit from the induction of symmetry, each fusion will still have unique opportunities to crystallize.

4.5 Conclusions

Failure rates for these C3 GFP trimers were similar to what has occurred with previous computational design efforts. In its current state, Rosetta Design is still very much a trial and error process where successes come from persistence and the large numbers of constructs screened. The resulting structures from these efforts with differ far too much from the intended structures to be considered designed proteins. These new oligomers could have potential applications, but do not impart the desired geometry from the designs. This limits the usefulness for synthetic biology scaffolding applications. If a desired symmetry were needed, it would be significantly easier to search the PDB for an existing protein with the desired geometry for design efforts. Specifically designed geometric scaffolds hold enormous potential but need to be refined for protein oligomer design to become a widely used technique.

4.6 Materials and Methods

4.6.1 Cloning

The amino acids sequences were codon opitmized[95] for expression in *E. coli.* Optimized genes were synthesized as gBlocks (Integrated DNA Technologies) with the terminal sequences GTTTAACTTTAAGAAGGAGATATAC and TAAAATTCGAGCTCCGTCGACAAGCTTG added to the 5' and 3' ends respectively. The genes were then cloned via Gibson assembly[75] into pET22b vector cut with NdeI and EcoRI restriction enzymes and transformed into XL-2 Blue Ultracompetent cells (Agilent Technologies). Due to an inherent mutation rate with Gibson assembly cloning, multiple colonies were checked for the correct sequence. Plasmids with the correct sequences were transformed into BL21 DE3 (New England Biolabs) for protein expression.

4.6.2 Protein Expression

All proteins were expressed in LB supplemented with auto-induction[57,96] sugars and ampicillin. The 1L LB flasks were inoculated with 10mL of overnight culture and allow to grow for 24 hours at 37°C before harvesting. Green colored cell pellets was used as the indication for protein expression as this correlated well with the results of small-scale expression screening.

4.6.3 Protein Purification

The same purification methods and buffers as described in chapter 2 were utilized. For these constructs only a two-step purification of IMAC and SEC was used. Constructs C3_14, C3_31, C3_32 and C3_33 were run on a s200 16/60 SEC column, C3_51 was run on a s200 10/300 GL column, this accounts for the poor resolution between the void peak and trimeric peaks. All buffers used for both IMAC and SEC were identical to previous GFP oligomer purifications[42] (chapter 2).

4.6.4 Protein Crystallization

Crystallization experiments were set with a protein concentration of 10mg/ml at room temperature. Four screens, JSCG+, MPDs (Qiagen), Wizard (Emerald BioStructures) and Structure Screen (Molecular Dimensions) were set for each construct. No efforts were undertaken to optimize constructs that failed to crystallize or to optimize the crystals for better diffraction. All crystals were cryo-protected with 20% glycerol prior to flash freezing in liquid nitrogen. Crystals were cryogenically stored until data collection.

4.6.5 X-Ray Data Processing and Structure Solution

All data was reduced with XDS/XSCALE,[97] phases were determined with Phaser[88] with the computational model serving as the search model. Structures were refined with PDB_Redo,[90] Refmac[89] or Phenix[91] and validated for structural accuracy with Phenix. Models were manually rebuilt in Coot.[98] Models were evaluated and compared to the designs with PyMol.[99]

Table 4.1 Expressed Sequences of GFP C3 Perpendicular Designs

C3_11
gfp1_C32b3p_23.4_64.5_8454
MSLGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMAEHDFFKSAMPEGYVQKRVIVFLNDGVYYTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNSHNVYIAADSAKNGIMAYFIIVHNVEDGSHQLASHAQYNTPIGDGP
DALPAHHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGILKGHHHHHH

C3_12
gfp1_C32b3p_23.4_64.5_0169
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDWMAMHDFFKSAMPEGYVQKRVIVFLNDGVYYTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNSHNVYIRADSAKNGIAAYFIIAHNVEDGSVQAAMHAQYNTPIGDGP
DALPAHHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGILQGHHHHHH

C3_13
gfp1_C32b3p_23.3_53_0133
MSQGEDLFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMAAHDFFKSAMPEGYVQERIIWFLNDGMYYTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNAHKVWIMADVANNGIMAIFVIMHNVEDGSVQYAAHLQFNTPIGDGP
AALPWVHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGILHGHHHHHH

C3_14
gfp1_C32b3p_23.4_53.5_8212
MSRGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMAAHDFFKSAMPEGYVQERIIWFLNDGEYFTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNAHKVWIMADLANNGIMAYFVIFHNVEDGSVQRAVHAQWNTPIGDGP
AALPWVHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGHHHHHH

C3_15
MSQGADLFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMAMHDFFKSAMPEGYVQERVISFENDGVYFTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNAHYVYITADAANNGIAAVFVIMHNVEDGSVQYAAHLQANTPIGDGP
SALPADHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGILHGHHHHHH

C3_21
gfp1_C32b3p_23.4_244.5_0776
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDWMAVHDFFKSAMPEGYVQKRVIVFLNDGVYYTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNSHNVYIRADSAKNGIAAYFIIAHNVEDGSVQAAIHAQYNTPIGDGP
DALPAHHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGILQGHHHHHH

C3_22
gfp1_C32b3p_23.4_244.5_5018
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDWMAIHDFFKSAMPEGYVQKRAIFFLNDGVYYTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNSHNVYIRADSAKNGIAAFFIIAHNVEDGSVQAAMHAQYNTPIGDGP

DALPAFHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGILQGHHHHHH
C3_23
gfp1_C32b3p_23.2_233_4868
MSQGEDLFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMAAHDFFKSAMPEGYVQERIIWFLNDGMYYTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNAHKVWIYADVANNGIMAIFVIMHNVEDGSVQYAAHLQFNTPIGDGP
AALPAAHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGILHGHHHHHH

C3_24
gfp1_C32b3p_22.9_233_1980
MSQGADLFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMAMFDFFKSAMPEGYVQERVISFENDGVYFTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNAHVVWIYANAANNGIMAVFVIMHNVEDGSVQYAAHLQANTPIGDGP
SALPAAHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGVLHGHHHHHH

C3_25
gfp1_C32b3p_22.8_232.5_3199
MSQGAALFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMAAFDFFKSAMPEGYVQERWISFENDGMYVTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNAHVVYIHADVANNGIMAVFVIMHNVEDGSVQYAAHWQLNTPIGDGP
SALPAFHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGVLHGHHHHHH

C3_31
gfp1_C32b3p_23.0_20_8085
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKKHDFFKSAMPEGYVQERVIIFKDDGTYKTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNAHVVIIMANSKNAGIKAAFVIVHNVEDGSHQLALHIQQNTPIGDGP
VALPATHYLLTQSKLSKDPNEKRDHMVLLEFVTANGISEGHHHHHH

C3_32
gfp1_C32b3p_23.2_19_3112
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKKHDFFKSAMPEGYVQERTIIFEDDGTYKTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNAHVVLIMANADNKGIKAAFVIMHNVEDGSVQFALHIQQNTPIGEGP
AALPAVHYLVTQSKLSKDPNEKRDHMVLLEFVTANGISEGHHHHHH

C3_33
gfp1_C32b3p_23.0_19.5_2178
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKKHDFFKSAMPEGYVQERTIIFKDDGTYKTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNAHVVVITANSKQAGIRAAFVIAHNVEDGSVQLALHIQQNTPIGDGP
VALPAVHALVTQSKLSKDPNEKRDHMVLLEFVTATGISEGHHHHHH

C3_34
gfp1_C32b3p_24.6_356.5_7304
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDWMAAHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNFNSHKVYIKADDKKNGIKAYFTIRHNVEDGSVQLADHYQQNTPIGDGP
ESLPLWHHLKTQSKLSKDPNEKRDHMVLLEFVTAAGIAQGHHHHHH

C3_35
gfp1_C32b3p_24.5_356_8183
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPDHMAAYDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDF
KEDGNILGHKLEYNFNSHKVYITANWLKNGIKAKFIIRHNVEDGSVQLAVHYQQNTPIGDGPDSL
PLFHHLLTQSKLSKDPNEKRDHMVLLEFVVAAGVAKGHHHHHH

C3_41
gfp2_C32b3p_2_23.3_287_8761
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERIIVFAYDGVYFTRAEVKFEGDTLVNRIELKGV
NFADNGAILGHRLEYNFNSHNVYIKADLQKNGIKANFKIRHNVENGKVQLADHYQQNTPIGDGP
VWLPAHHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGIEEGHHHHHH

C3_42
gfp2_C32b3p_2_19.1_320_2430
MSKGEELFTGVVPILVELVGNVNGQKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQVRQIIFKDDGVYSTVAEVKFEGDTLVNRIELHGS
LFKEDGNILGHKLEYNFNSHVVFIVANASANGIAAFFKIRHNVEDGSVQLAVHWQWNVPIGDGP
VLLPAAHALVTQSKLSKDPNEKRDHMVLLEFVTATGISEGHHHHHH

C3_43
gfp2_C32b3p_2_19.2_319_7876
MSKGEELFTGVVPILVELVGNVNGQKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQKRQIIFKDDGVYSTVAEVKFEGDTLVNRIELHGS
NFKEDGNILGHKLEYNFNSHVVFITANAANNGIAAYFKIRHNVEDGSVQLAVHMQWNVPIGDGP
VLLPAAHALITQSKLSKDPNEKRDHMVLLEFVTATGISEGHHHHHH

C3_44
gfp2_C32b3p_2_19.1_140.5_3829
MSKGEELFTGVVPILVELVGNVNGQKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQVRQIIFKDDGVYSTVAEVKFEGDTLVNRIVLHGA
LFKEDGNILGHKLEYNFNSHVVFIVANTSANGIAAFFKIRHNVEDGSVQLAVHWQWNVPIGDGP
VLLPAAHALVTQSKLSKDPNEKRDHMVLLEFVTATGISEGHHHHHH

C3_45
run_10_gfp2_C32b3p_2_19.1_140.5_2876.pdb
MSKGEELFTGVVPILVELVGNVNGQKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPDHMKRHDFFKSAMPEGYVQVRQIIFKDDGVYSTVAEVKFEGDTLVNRIVLHGALF
KEDGNILGHKLEYNFNSHVVFITANTSANGIAAFFKIRHNVEDGSVQLAVHWQWNVPIGDGPVLL
PADHALVTQSKLSKDPNEKRDHMVLLEFVTATGISEGHHHHHH

Table 4.2 Rosetta Values Used to Select Table 4.1 Sequences

Design	Rosetta Energy	Shape Comp.	Buried Surface Å2
gfp1_C32b3p_23.4_64.5_8454.pdb.gz	-194.794	0.535	445.536
gfp1_C32b3p_23.4_64.5_0169.pdb.gz	-193.866	0.531	542.472
gfp1_C32b3p_23.3_53_0133.pdb.gz	-180.437	0.446	615.972
gfp1_C32b3p_23.4_53.5_8212.pdb.gz	-178.967	0.529	642.650
gfp1_C32b3p_22.8_53_9910.pdb.gz	-191.094	0.465	568.555
gfp1_C32b3p_23.4_244.5_0776.pdb.gz	-194.425	0.525	455.443
gfp1_C32b3p_23.4_244.5_5018.pdb.gz	-189.451	0.521	530.756
gfp1_C32b3p_23.2_233_4868.pdb.gz	-177.738	0.429	597.561
gfp1_C32b3p_22.9_233_1980.pdb.gz	-174.801	0.459	680.698
gfp1_C32b3p_22.8_232.5_3199.pdb.gz	-174.555	0.318	637.022
gfp1_C32b3p_23.0_20_8085.pdb.gz	-166.645	0.395	522.984
gfp1_C32b3p_23.2_19_3112.pdb.gz	-159.926	0.232	537.663
gfp1_C32b3p_23.0_19.5_2178.pdb.gz	-159.143	0.354	433.022
gfp1_C32b3p_24.6_356.5_7304.pdb.gz	-136.249	0.574	391.398
gfp1_C32b3p_24.5_356_8183.pdb.gz	-132.528	0.543	468.820
gfp2_C32b3p_2_23.3_287_8761.pdb.gz	-159.372	0.615	370.840
gfp2_C32b3p_2_19.1_320_2430.pdb.gz	-114.555	0.613	505.449
gfp2_C32b3p_2_19.2_319_7876.pdb.gz	-113.708	0.467	463.836
gfp2_C32b3p_2_19.2_319_1512.pdb.gz	-113.699	0.468	463.596
gfp2_C32b3p_2_19.1_140.5_3829.pdb.gz	-113.899	0.602	499.388

Table 4.3 Sequences of GFP C3 Parallel Designs

C3_51
GFP_C3_rot2b3p_z_115_-16.8_357.5_9106.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPKWMKRHDFFKSAMPEGYVQERTIIFANDGTYKTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLAYTFNAHFVWIKADEQKNGIRATFVIFHAVKNGSVQLALHIQQNTPIGDGP
VRLPATHFLETYSVLSKDPNEKRDHMVLLEFVEAAGIAEGHHHHHH

C3_52
GFP_C3_rot2b3p_z_120_-16.9_2.5_3597.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPEHMKRHDFFKSAMPEGYVQERVIIFANDGTYKTRAEVKFEGDTLVNRIELKGI
DFLEDGNILGHYLAYTFMAHAVWIAAFKEKNGIVAYFIIFHAVLNGSVQAAVHIQQNTPIGDGP
VRLPAFHYLWTYSVLSKDPNEKRDHMVLLEFVVANGISAGHHHHHH

C3_53
GFP_C3_rot2b3p_z_120_-16.9_3.5_2620.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQEREIVFADDGTYKTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHYLAYNFNAHAVWITANWEKNGIAAYFVIFHAVLNGSVQAAVHEQQNTPIGDGP
VYLPAFHFLWTRSVLSKDPNEKRDHMVLLEFVIAAGISHGHHHHHH

C3_54
GFP_C3_rot2b3p_z_120_-16.9_3.5_9650.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERVIIFAADGTYKTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHYLLYTFNAHIVWIVAFWEKNGIRAYFIIFHAVANGSVQAAVHIQQNTPIGDGP
VYLPAFHALFTMSVLSKDPNEKRDHMVLLEFVIAFGISHGHHHHHH

C3_55
GFP_C3_rot2b3p_z_120_-17.2_0_6009.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPEWMKRHDFFKSAMPEGYVQERRIIFADDGTYKTRAEVKFEGDTLVNRIELKGI
DFYEDGNILGHKLRYTFLAHVVYIFQDKQKNGIRAFFIIFHAVKNGSVQLAFHLQQNTPIGDGP
VLLPSFHFLFTESVLSKDPNEKRDHMVLLEFVEAAGISEGHHHHHH

C3_56
GFP_C3_rot2b3p_z_120_-17.2_1.5_0048.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERVIIFANDGTYKTRAEVKFEGDTLVNRIELKGI
DFLEDGNILGHKLAYTFLAHIVWIWAFTEKNGIMAWFIIFHAVKNGSVQLAIHIQQNTPIGDGP
VRLPAFHFLFTVSVLSKDPNEKRDHMVLLEFVIAAGISEGHHHHHH

C3_57
GFP_C3_rot2b3p_z_120_-17.2_2_4097.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDSMKRHDFFKSAMPEGYVQERVIVFANDGTYKTRAEVKFEGDTLVNRIELKGI
DFLEDGNILGHKLAYTFMAHVVFISAFDEKNGIRAWFIIFHAVKNGSVQLAYHWQQNTPIGDGP
VRLPAWHYLWTLSVLSKDPNEKRDHMVLLEFVEADGISTGHHHHHH

C3_58
GFP_C3_rot2b3p_z_120_-17.2_2.5_0429.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPEHMKRHDFFKSAMPEGYVQERVIVFAADGTYKTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLAYTFLAHVVFISAFKEKNGIRAWFIIFHAVKNGSVQLAIHWQQNTPIGDGP
VRLPAWHYLWTLSVLSKDPNEKRDHMVLLEFVEADGISHGHHHHHH

C3_59
GFP_C3_rot2b3p_z_120_-17.3_0.5_8776.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPEWMKRHDFFKSAMPEGYVQERVIIFADDGTYKTRAEVKFEGDTLVNRIELKGI
DFYEDGNILGHKLRYTFMAHIVWIFQFKQKNGIRAWFIIFHAVKNGSVQLAFHEQQNTPIGDGP
VKLPAFHFLYTKSVLSKDPNEKRDHMVLLEFVEAAGISEGHHHHHH

C3_510
GFP_C3_rot2b3p_z_120_-17.3_2.5_9643.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPEHMKRHDFFKSAMPEGYVQERTIRFAADGTYKTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLLYTFLAHVVFIHANKEKNGIMAWFIIYHAVKNGSVQMAIHIQQNTPIGDGP
VKLPKWHYLVTESVLSKDPNEKRDHMVLLEFVTAAGIAHGHHHHHH

C3_511
GFP_C3_rot2b3p_z_125_-16.8_0.5_6127.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPEHMKRHDFFKSAMPEGYVQERVIIFAADGVYKTRAEVKFEGDTLVNRIELKGI
KFLENGNILGHYLLYTFLAHVVWIAAFKEKNGIVAFFIIVHSVKNGSAQLAFHRQQNTPIGDGP
VKLPDHHFLVTESVLSKDPNEKRDHMVLLEFVVANGISAGHHHHHH

C3_512
GFP_C3_rot2b3p_z_125_-16.8_0.5_7186.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERRIIFAADGTYKTRAEVKFEGDTLVNRIELKGI
DFLENGNILGHKLLYTFLAHVVWIAAFLEKNGIVAFFIIVHSVKNGSAQLAFHRQQNTPIGDGP
VKLPKHHFLVTESVLSKDPNEKRDHMVLLEFVVANGIAAGHHHHHH

C3_513
GFP_C3_rot2b3p_z_125_-16.8_0_6443.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDSMKRHDFFKSAMPEGYVQERVIIFAADGTYKTRAEVKFEGDTLVNRIELKGI
DFLENGNILGHYLLYTFLAHVVWIAANTEKNGIAAFFIIVHSVKNGSAQLAFHRQQNTPIGDGP
VKLPKHHFLVTESVLSKDPNEKRDHMVLLEFVVANGIATGHHHHHH

C3_514
GFP_C3_rot2b3p_z_125_-16.9_2_7739.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERVIEFAADGTYKTRAEVKFEGDTLVNRIELKGI
DFEENGNILGHKLKYTFHAHVVWIAANADKNGIVAYFIIVHAVKNGSAQLAFHVQQNTPIGDGP
VKLPAFHFLFTMSVLSKDPNEKRDHMVLLEFVTANGIAAGHHHHHH

C3_515
GFP_C3_rot2b3p_z_125_-17.0_0_0301.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERVIIFAADGTYKTRAEVKFEGDTLVNRIELKGI
DFLENGNILGHKLLYTFNAHVVFIAANWEKNGIMAYFVIVHDVKNGSAQLAFHYQQNTPIGDGP
VKLPKWHFLWTLSVLSKDPNEKRDHMVLLEFVIANGIAHGHHHHHH

C3_516
GFP_C3_rot2b3p_z_125_-17.0_1_2354.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERVIIFAADGTYKTRAEVKFEGDTLVNRIELKGI
DFLENGNILGHKLLYTFNAHVVFIAANWEKNGIEAYFMIVHDVKNGSAQLAWHIQQNTPIGDGP
VKLPAWHFLFTKSVLSKDPNEKRDHMVLLEFVRANGIAHGHHHHHH

C3_517
GFP_C3_rot2b3p_z_125_-17.0_1.5_5190.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERVIIFAADGTYKTRAEVKFEGDTLVNRIELKGI
DFLENGNILGHKLLYTFNAHVVFIAAFWEKNGIIAYFVIVHDVKNGSAQLAFHWQQNTPIGDGP
VKLPAWHFLFTMSVLSKDPNEKRDHMVLLEFVTANGIAHGHHHHHH

C3_518
GFP_C3_rot2b3p_z_125_-17.0_2.5_0449.pdb
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERVIIFAADGTYKTRAEVKFEGDTLVNRIELKGI
DFKENGNILGHKLEYNFNAHVVFIAANAEKNGIEAYFIIMHDVKNGSVQLAFHWQQNTPIGDGP
VYLPAWHYLYTMSVLSKDPNEKRDHMVLLEFVTASGIAAGHHHHHH

C3_519
GFP_C3_rot2b3p_z_20_-19.5_4.5_2560.pdb
MSKGEELFEGWVPILVELDGDVNGHKFSVRGWGAGDADNGALVLIFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPKHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGQTLVNRIELKGI
DFKEDGNILGHKLLYMFQAHKVYITADKQKNGIKANFKIRHAVKDGSVQLADHYQQNTPIGDGP
VLLPDNHYLFTFSYLAVQPANTSDHMVLFEYVAANGITKGHHHHHH

C3_520
GFP_C3_rot2b3p_z_20_-19.6_4.5_6816.pdb
MSKGEALFKGMVPILVELDGDVNGHKFSVRGWGAGDADNGMLVLIFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDWMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGRTLVNRIELKGI
DFKEDGNILGHKLLYMFQAHNVYITADKQKNGIKANFKIRHAVKDGSVQLADHYQQNTPIGDGP
VLLPDNHYLYTFSYLAFQPANTSDHMVLFEWVVADGITEGHHHHHH

Table 4.4 Data and Refinement Statistic for C3_51

Resolution range (Å)	76.03 - 1.5 (1.553 - 1.5)
Space group	P 21 3
Unit cell	107.53 107.53 107.53 90 90 90
Total reflections	2623128 (245434)
Unique reflections	66346 (6580)
Multiplicity	39.5 (37.3)
Completeness (%)	100.00 (100.00)
Mean I/sigma(I)	26.26 (2.28)
Wilson B-factor	18.29
R-merge	0.1216 (1.986)
R-meas	0.1232
CC1/2	1 (0.753)
CC*	1 (0.927)
R-work	0.1693 (0.3033)
R-free	0.1859 (0.3198)
Number of non-hydrogen atoms	2036
macromolecules	1748
ligands	37
water	251
Protein residues	220
RMS(bonds)	0.011
RMS(angles)	1.39
Ramachandran favored (%)	98
Ramachandran allowed (%)	2
Ramachandran outliers (%)	0
Clashscore	6.51
Average B-factor	24.30
macromolecules	22.30
ligands	28.20
solvent	37.60

Chapter 5

Coiled-Coil Fusions to T4-Lysozyme Oligomers

5.1 Introduction

The difficulties in utilizing the split GFP oligomers for *in vitro* complementation as a crystallizations scaffold (chapters 2 & 3) led to investigating alternate means to implement this concept. Solubility and complementation issues limited the experiments that could be conducted for proof-of-concept experiments to validate the approach with GFP. By its nature of existing as an incomplete protein, partially folded in solution, the solubility issues were likely to continue. Low complementation efficiency may be alleviated, but this would involve extensive validation efforts, and may not be possible with the various mutations contained in the suite. Each one of these mutants possibly would need to be further evolved in the laboratory. Alternate scaffold proteins and *in vitro* complex formation technologies were therefore investigated to use the same scheme outline for GFP scaffolds (figure 1.2, chapter 2).

Lysozyme, specifically phage T4- lysozyme (T4L), was selected to serve as the new symmetric scaffold. Lysozyme has been extensively studied with x-ray crystallography as native and mutant versions readily crystallize. T4L has been previously established as a crystallization aid for G-protein coupled receptors, where loop insertions or terminal fusions of T4L facilitate crystallization.[100–103] Further, a suite of oligomeric mutants of T4L have already been validated for the ability to crystallize multiple crystal forms.[6,8] Various mutants of T4L have been studied for stability.[104] The highly stable and readily crystallizable sequence from PBD deposition 3FA0[105] was used in the previous T4L symmetrization studies and therefore seemed applicable for these new experiments.

T4L features helices on both the N- and C- termini, a structural feature amendable to rigid fusion. These helices could be extended with a rigid helical linker for genetic fusion to crystallization targets, an approach that has proved successful with MBP crystallization fusions.[106] Combination of this technique with the oligomer mutants would create a series of genetic fusions that could be explored to aid with proteins that have failed to crystallize by traditional rescue efforts. However, extension of these helices with known heterodimeric coiled-coil motifs would allow a means of *in vitro* complex formation and a facile method of inducing symmetry onto a POI using the T4L scaffold. The heterodimeric coiled-coil pair is derived from DNA binding proteins[107,108] specifically from the crystal structure of the AP-1 c-Fos/c-Jun transcription factor.[109] These sequences have previously been engineered to have 10,000-fold preference for the hetero interactions over homo interactions.[110,111] This lead to two sequences, AQLEKELQALEKENAQLEWELQALEKELAQ for fusion to the C-terminus of a protein and the cognate element AQLKKKLQALKKKNAQLKWKLQALKKKLAQ for fusion to the N-terminus of a protein. These helical fusions to T4L may also be varied in position by addition or subtraction of up to four residues (one helical turn) to change the geometry of the linkage and increase the combinatorial number of scaffolds derived from this approach.

A former graduate student in the lab, Dan McNamara Ph.D, created models of the helices from the original c-Fos/c-Jun structure with ten terminal alanine residues to allow for modeling of the coiled-coil linkage between two proteins. With these models one can fuse the coiled-coil sequence onto either the N- or C- terminus of T4L to set the orientation of the helix to reduce steric clashes between the two linked proteins. When aligned to the T4L structure, multiple positions of the helical fusion were possible. If a homology model of the crystallization target is available the helical, models could be modeled onto the predicted structure to guide construct design.

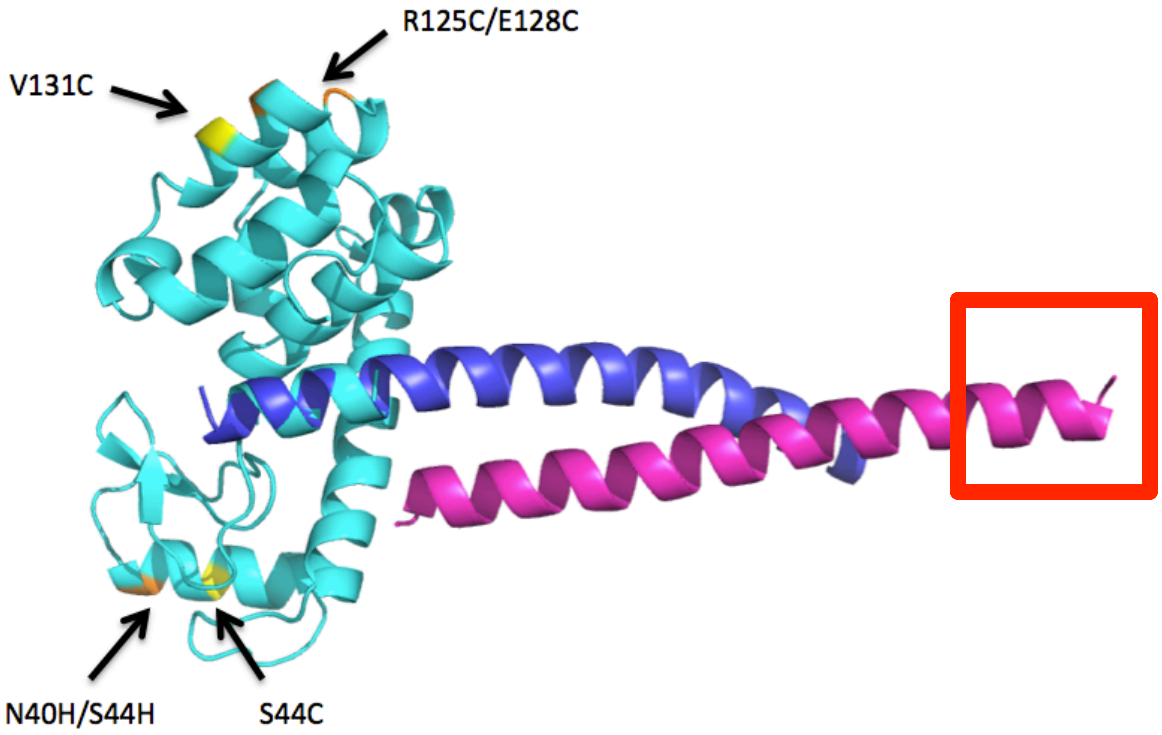


Figure 5.1. Model of the Coiled-Coil Pair Fused to T4L. The x-ray crystal structure of a highly stable version of T4L (PDB 3FA0, cyan) was aligned with the c-Fos (magenta)/c-Jun (purple) structure. The model that indicated no steric hindrance is presented. Here, the c-Jun sequence is to be fused to the N-terminus of T4L, the crystallization target would be fused to the N-terminal end of the c-Fos coil (red box). This coiled-coil model is from the native c-Fos/c-Jun sequence; the engineered heterodimeric pair for expression constructs replaces these sequences for expression constructs. Mutations are highlighted for disulfide dimers (yellow) and metal chelating sites (orange).

5.2 Results

5.2.1 Creation of T4L Coiled Coil Constructs

The model of the coiled-coil pair was aligned to both the termini of the T4L structure. Alignments were then shifted by one amino acid until a model with no steric hindrance between the T4L and coiled coil was made (figure 5.1). Previously published mutations were evaluated for their location compared to the coiled-coil position. Mutations that would put a symmetric dimer interface in a location that would not clash with the coiled-coil or a bound POI were selected. The sequences of these constructs are presented in table 5.1 and include published cysteine mutants (S44C and V131C)[6], one previously validated metal-chelation site (R125C/E128C)[8] and one novel pair of mutations for metal-chelation (N40H/S44H).

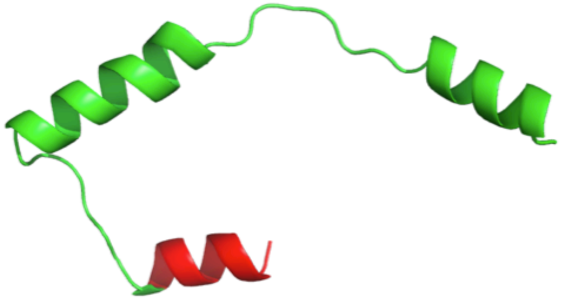
This single N-terminal helical fusion was selected, as there would be no apparent steric clashes when the complex is formed and oligomerized, although additional positions of the fusion would be possible. Linkage through the T4L N-terminus allows for more pre-existing oligomeric mutations to be utilized in this system. If the linkage was through the C-terminus, the established mutations would be on the same face of T4L as linkage resulting in severe clashes then oligomerization is attempted. To evaluate a C-terminal linkage in future endeavors new positions of oligomer formation would need to be explored.

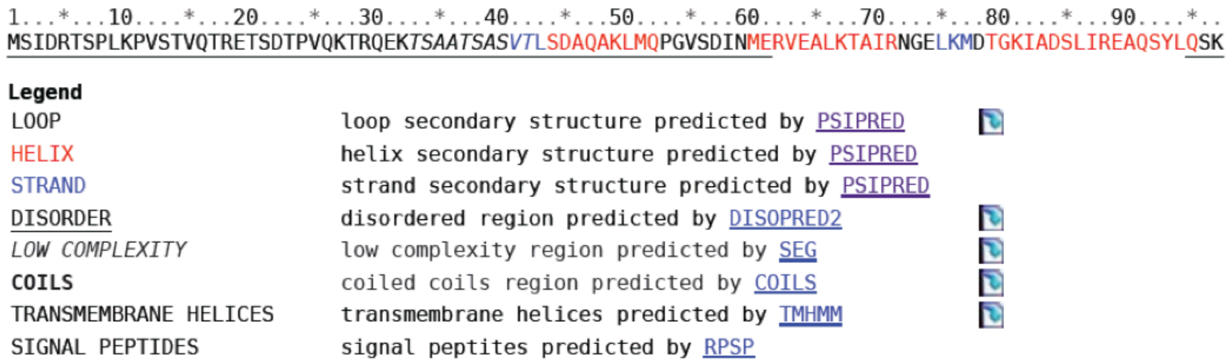
5.2.2 Selection of Crystallization Targets

Coiled-coil linkages of this type require one half of the pair to be fused to the N-terminus of one protein and the second half onto the C-terminus of the partner. Since the T4L core was the N-terminal half, the crystallization target would be required to be fused onto the C-terminus. It is thought that the coiled-coil sequence will form helical structures naturally.[110] However, fusion to an existing helix would provide a more viable system as a more predictable geometry would exist. This approach of helical fusions has been established as a means of forming rigid linkages capable of crystallization.[81,106] In order to validate this approach, novel crystallization targets were selected. As these proteins do not have known structures, there were two ways considered to ensure a C-terminal helix. First, fusion of a known helical sequence[112–114] could be used to link the target protein to the coiled-coil regardless of the secondary structure of the C-terminus of the POI. However, this leads to an ambiguity in the orientation of the helix as the termini are disordered or flexible in many protein structures.[115]

The second approach relied on homology modeling and secondary structure predictions to select targets with predicted C-terminal helices; it was decided to follow this approach. To evaluate this, first XtalPred[64] was used to evaluate the secondary structure. Sequences with a predicted C-terminal helical region where then ran through the Phyre2 structural homology server,[116] this served two purposes. Confirmations of the presences of a terminal helix and provide a homology model to allow alignment of the coiled-coil sequence to aid in construct design.

Selection of well-behaved crystallization targets was of importance due to the failures of GFP scaffolding discussed in chapter 3. To select the proteins the online target database from the Center for Structural Genomics of Infectious Disease (XSGID) (http://www.csgid.org/targets/index) was utilized. Here, experimental results of all target proteins are presented. Selection criteria included >1 year with no structure solution, have entered crystal trials (indication of high purity) , and obtainable to high concentrations (>5 mg/mL). Sequences meeting this criterion were then run through XtalPred, those with predicted C-terminal



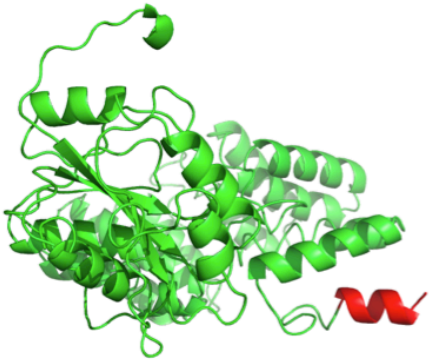


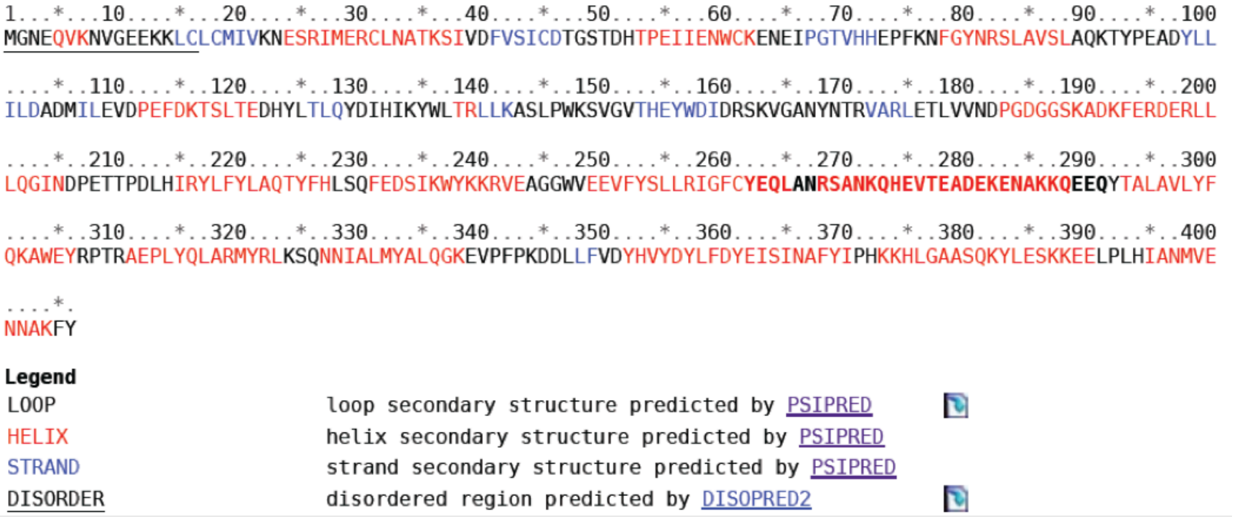
```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80....*...90....*..
MSIDRTSPLKPVSTVQTRETSDTPVQKTRQEKTSAATSASVTLSDAQAKLMQPGVSDINMERVEALKTAIRNGELKMDTGKIADSLIREAQSYLQSK
```

Legend

LOOP	loop secondary structure predicted by PSIPRED
HELIX	helix secondary structure predicted by PSIPRED
STRAND	strand secondary structure predicted by PSIPRED
DISORDER	disordered region predicted by DISOPRED2
LOW COMPLEXITY	low complexity region predicted by SEG
COILS	coiled coils region predicted by COILS
TRANSMEMBRANE HELICES	transmembrane helices predicted by TMHMM
SIGNAL PEPTIDES	signal peptites predicted by RPSP

Figure 5.2. Homology Model and Secondary Sequence Prediction of IDP01024. The homology model of IDP01024, *Salmonella enterica* anti-sigma28 factor FlgM used for the coiled-coil construct is presented; the predicted C-terminal helix where the coiled-coil is fused is highlighted in red.



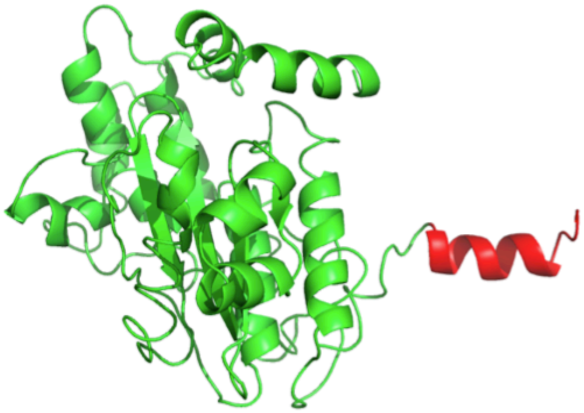


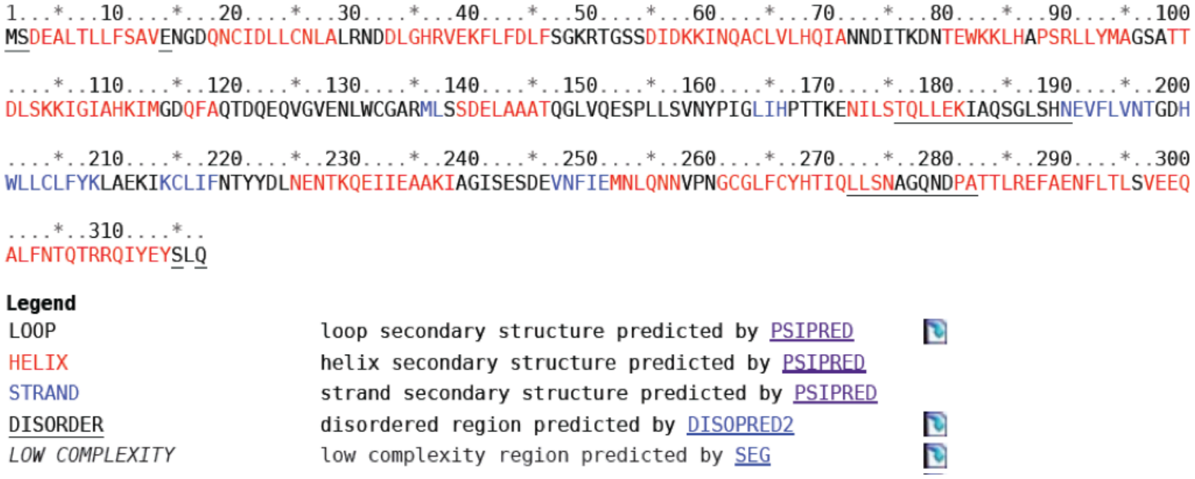
```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80....*...90....*..100
MGNEQVKNVGEEKKLCLCMIVKNESRIMERCLNATKSIVDFVSICDTGSTDHTPEIIENWCKENEIPGTVHHEPFKNFGYNRSLAVSLAQKTYPEADYLL

....*..110....*..120....*..130....*..140....*..150....*..160....*..170....*..180....*..190....*..200
ILDADMILEVDPEFDKTSLTEDHYLTLQYDIHIKYWLTRLLKASLPWKSVGVTHEYWDIDRSKVGANYNTRVARLETLVVNDPGDGGSKADKFERDERLL

....*..210....*..220....*..230....*..240....*..250....*..260....*..270....*..280....*..290....*..300
LQGINDPETTPDLHIRYLFYLAQTYFHLSQFEDSIKWYKKRVEAGGWVEEVFYSLLRIGFCYEQLANRSANKQHEVTEADEKENAKKQEEQYTALAVLYF

....*..310....*..320....*..330....*..340....*..350....*..360....*..370....*..380....*..390....*..400
QKAWEYRPTRAEPLYQLARMYRLKSQNNIALMYALQGKEVPFPKDDLLFVDYHVYDYLFDYEISINAFYIPHKKHLGAASQKYLESKKEELPLHIANMVE

....*.
NNAKFY
```

Legend

LOOP	loop secondary structure predicted by PSIPRED
HELIX	helix secondary structure predicted by PSIPRED
STRAND	strand secondary structure predicted by PSIPRED
DISORDER	disordered region predicted by DISOPRED2

Figure 5.3. Homology Model and Secondary Sequence Prediction of IDP01204. The homology model of IDP01204, *Bacillus anthracis str. Ames* glycosyl transferase, group 2 family protein used for the coiled-coil construct is presented; the predicted C-terminal helix where the coiled-coil is fused is highlighted in red. This protein was previously found to crystallize through CSGID efforts but failed to diffract.





```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80....*...90....*..100
MSDEALTLLFSAVENGDQNCIDLLCNLALRNDDLGHRVEKFLFDLFSGKRTGSSDIDKKINQACLVLHQIANNDITKDNTEWKKLHAPSRLLYMAGSATT

....*..110....*..120....*..130....*..140....*..150....*..160....*..170....*..180....*..190....*..200
DLSKKIGIAHKIMGDQFAQTDQEQVGVENLWCGARMLSSDELAAATQGLVQESPLLSVNYPIGLIHPTTKENILSTQLLEKIAQSGLSHNEVFLVNTGDH

....*..210....*..220....*..230....*..240....*..250....*..260....*..270....*..280....*..290....*..300
WLLCLFYKLAEKIKCLIFNTYYDLNENTKQEIIEAAKIAGISESDEVNFIEMNLQNNVPNGCGLFCYHTIQLLSNAGQNDPATTLREFAENFLTLSVEEQ

....*..310....*..
ALFNTQTRRQIYEYSLQ
```

Legend

LOOP	loop secondary structure predicted by PSIPRED
HELIX	helix secondary structure predicted by PSIPRED
STRAND	strand secondary structure predicted by PSIPRED
DISORDER	disordered region predicted by DISOPRED2
LOW COMPLEXITY	low complexity region predicted by SEG

Figure 5.4. Homology Model and Secondary Sequence Prediction of IDP90101. The homology model of IDP90101, *Salmonella enterica* sseL deubiquitinase used for the coiled-coil construct is presented; the predicted C-terminal helix where the coiled-coil is fused is highlighted in red.

helices were modeled with the Phyre2 server. From this, three proteins were selected (table 5.2) referred to as IDP01024, IDP01204 and IDP90101 per CSGID target identifications (figures 5.2 – 5.4). Of these, IDP01204 crystallized but failed to diffract according to CSGID records, this led this target to be the focus of the T4L scaffolding efforts. For each construct, the coiled-coil structure was aligned onto the C-terminus of the homology model. In each case the coiled coil sequence was fused after the final predicted residue of the terminal helix. Only one construct was designed for each of the target proteins.

5.2.3 Expression and Purification of constructs

Genes for the coiled-coil constructs were synthesized, cloned and subject to expression tests. All four T4L constructs were well expressed but insoluble, only one of the crystallization targets expressed and was soluble (IDP01204). The coiled-coils were simultaneously being explored for designed materials (Dan E McNamara Dissertation, UCLA). Similar results were observed. The fusion of the coiled-coil sequences on either terminus could dramatically reduce or eliminate protein solubility (Joshua Laniado, unpublished work). From these efforts it was found that solubility could be improved by buffer optimization. Neutral buffers with moderate salt and glycerol seemed to ameliorate some of the solubility issues. Therefore, efforts to purify the seemingly insoluble T4L constructs were carried out in an improved buffer of 50mM Hepes pH 7.0, 500mM NaCl, 5% Glycerol which served as the basis for purification buffers.

In these new conditions it was possible to purify the T4L constructs via IMAC. However, the protein was extremely sensitive to changes in buffer conditions. In order to oxidize the T4L cysteines to form dimers, the pH would have to increase to >8.0 and contain Cu^{2+} to allow efficient bond formation. When the protein was rapidly exchanged into a buffer of this pH, all of the protein precipitated. The experiment was repeated with an overnight dialysis from the optimized buffer to the buffer at pH 8.0. Slow buffer exchange kept the majority of the protein in solution. The oxidization reaction was allowed to occur with the addition of Cu^{2+} and quenched with EDTA. Subsequent exchange into oxidation buffer without Cu^{2+} and EDTA again caused all the protein to precipitate. Due to these difficulties and similar difficulties with the other efforts involving the coiled-coil fusions, it was decided to attempt co-purification of the T4L metal chelating constructs with IDP01204. This co-purification approach eliminated many of the solubility issues during the purification of the designed cages (unpublished work). As the only expressed and soluble crystallization target (IDP01204) contained native cysteine residues, only T4L metal chelation mutants were used for co-purification (40H/44H & 125C/128C).

These metal-chealting T4L constructs appeared to be more soluble and stable when purified without the crystallization partner, partly because the buffers did not have to be changed during purification to oxidize the disulfide bonds. Both constructs were to be used for crystallization experiments in the presence of Cu^{2+}, Zn^{2+} or no metal ions.

5.2.3.1 Co-purification of Complexes

During the cell lysis step of purification, cell paste from the T4L and IDP01204 were combined. Initially expression screening indicated the T4L expressed approximately at twice the level of IDP01204 (data not shown), therefore an excess mass of IDP01204 cell paste was used for this step. It was the intention that during this step the coiled-coil hetero-interaction would form, eliminating the solubility issues inherent to the coiled-coil fusion constructs. This was the case. The newly formed complex was stable through IMAC, TEV cleavage of Histidine affinity tag, SEC purification and concentration.

TEV cleavage of the Histidine affinity tag and subsequent removal of the cleaved tag and TEV protease through IMAC removed many of the impurities from the purifications. With SEC further impurities were removed resulting in crystallization-level purity. From the SEC runs the complex was well resolved from T4L monomers (figures 5.5 & 5.6, SDS-PAGE gels not shown). No monomeric IDP01204 was present in any fractions, likely a result of solubility issues of un-complexed protein. Surprisingly, for both preparations a significant peak of free T4L monomers was present despite the excess of IDP1204 added during lysis. Purification conditions were likely poorly optimized for IDP01204 and favorable for the T4L constructs. Co-purification with the double histidine T4L constructs resulted in approximately three-fold more protein for both the complex and free T4L monomers compared to the double cysteine mutant. Throughout all expression screening the double cysteine mutant consistently exhibited reduced expression compared to all other T4L constructs. The double cysteine mutant resulted in poorly resolved complex from larger aggregates present. When analyzed by SDS-PAGE, the void peak (presumed aggregates) and a broad peak between the void and complex peak contained bands corresponding to T4L and IDP01204. Incomplete reduction of disulfides formed during the Ni^{2+} IMAC purification subsequent to TEV cleavage of the histidine affinity tag is one possible source of these aggregates, although the presence of 1mM DTT in the SEC should have prevented this. This construct may be prone to aggregation due to this appearance of complex aggregation and reduced protein expression and solubility.

.

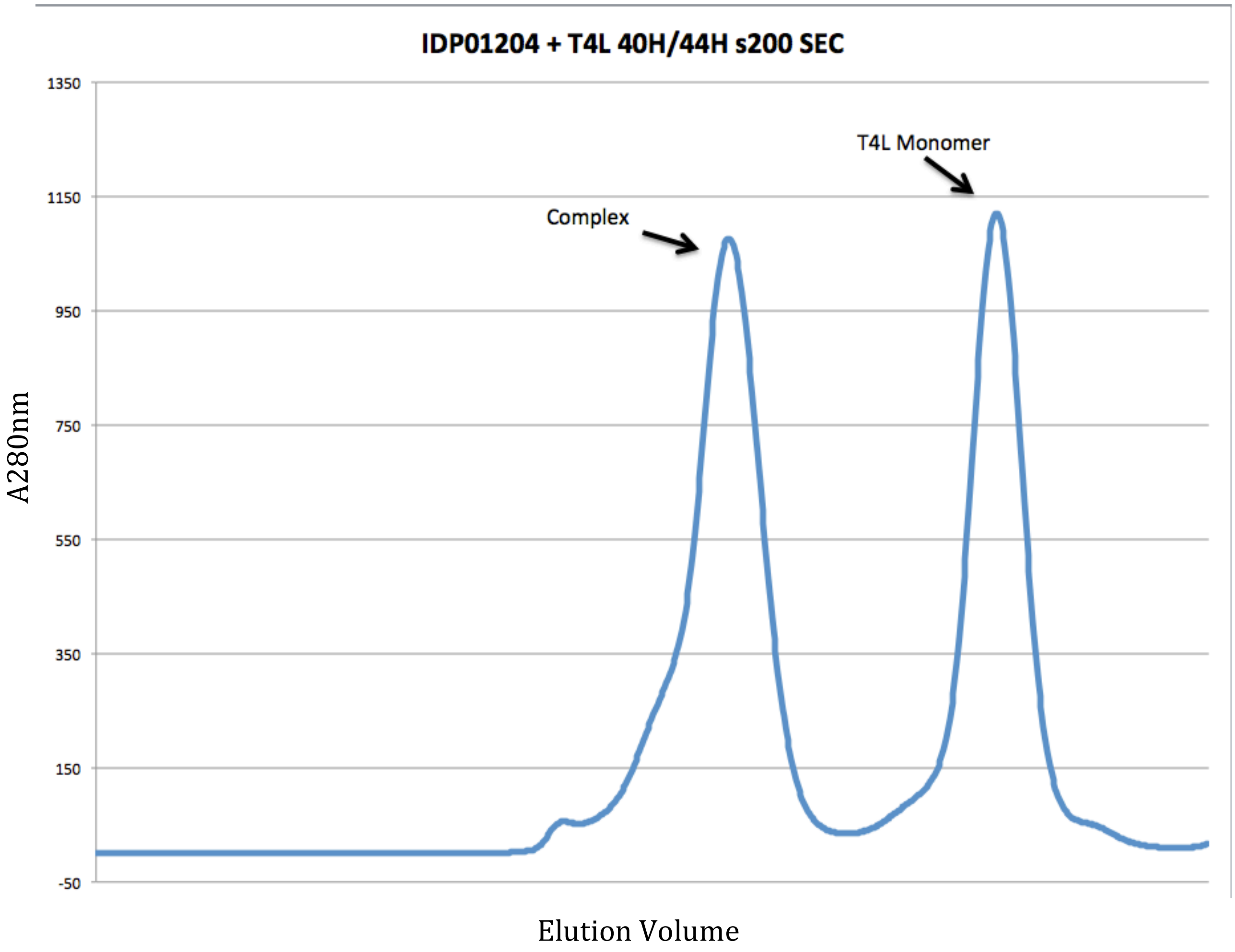


Figure 5.5. Size Exclusion Elution of IDP01204 Co-purified with T4L 40H/44H. SEC was performed on the co-purified proteins after TEV cleavage to remove the Histidine affinity tag. The two prominent peaks correspond to the T4L-IDP01204 complex and un-complexed T4L monomers. Fractions from the complex peak with equimolar amounts of the two proteins were pooled and used fro crystallization experiments.

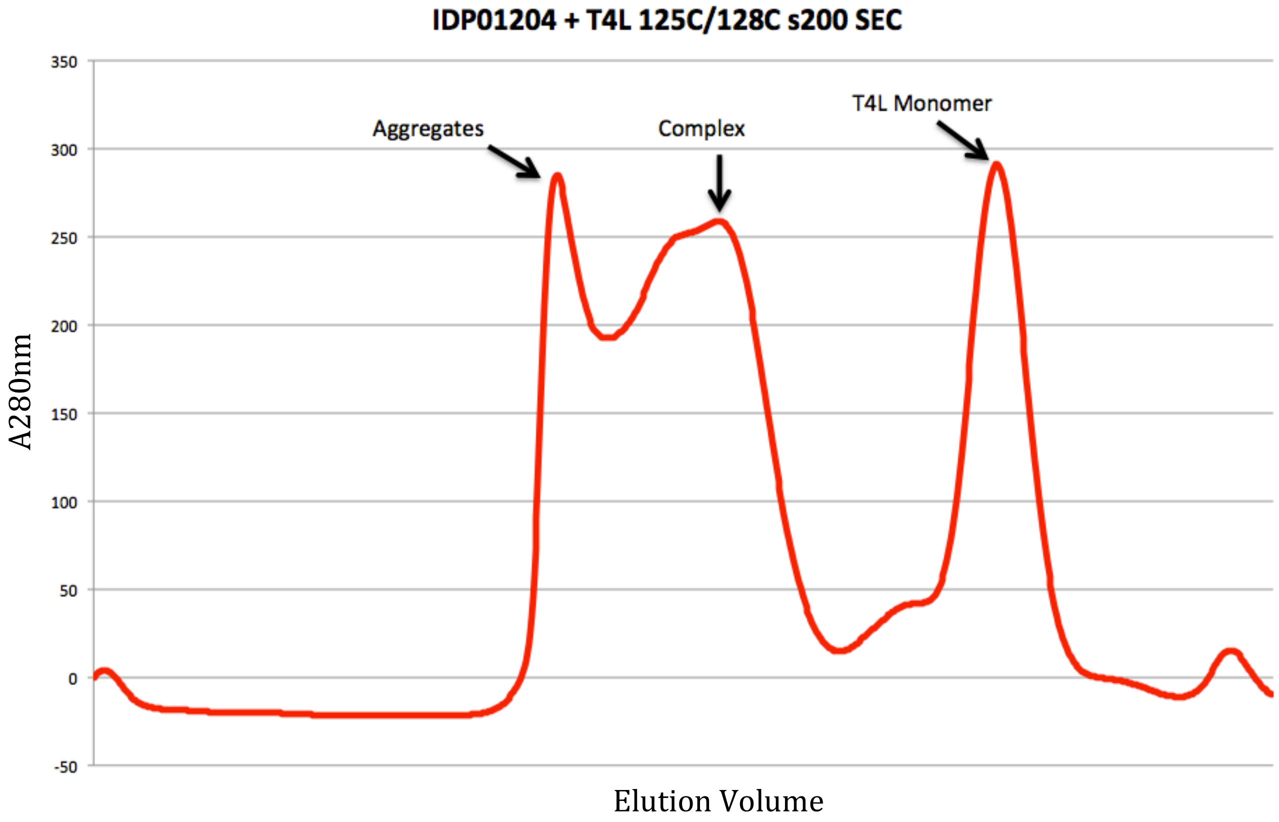


Figure 5.6. Size Exclusion Elution of IDP01204 Co-purified with T4L 125C/128C. SEC was performed on the co-purified proteins after TEV cleavage to remove the histidine affinity tag. The three prominent peaks correspond to aggregated protein, T4L-IDP01204 complex peak and un-complexed T4L monomers. Fractions from the complex peak with equimolar amounts of the two proteins were pooled and used fro crystallization experiments. Due to the significant aggregation of this construct, care was taken to choose fractions of the complex furthest from the void.

5.2.4 Crystallization Experiments

Fractions from the SEC purification that indicated an equimolar amount of both proteins by SDS-PAGE were pooled, concentrated and stored at -80°C until crystallization experiments could be set. Freshly thawed protein was originally used at a concentration of 10mg/mL for crystallization screening. At this concentration the protein widely precipitated throughout various sparse matrix and target screens for both constructs. Protein concentration was reduced to 5mg/mL and screens reset, with no crystals grown as of publication of this dissertation.

5.3 Discussion

5.3.1 Coiled-Coil Motifs Significantly Reduce Protein Solubility

Fusion of the coiled-coil sequences has significantly reduced the solubility and stability of the T4L and crystallization constructs. Optimization of the buffers has allowed for the successful purification of T4L but this was not possible with all three crystallization constructs. Thus, it was not possible to fully validate this approach as a means of *in vitro* complex formation for scaffolding as only one target was evaluated. Solubility and expression issues have been consistent with the results from other fusions to the coiled-coils; certain proteins appear to be more tolerant to the fusions (unpublished work). Fortuitously, T4L is tolerant to the fusions in optimized buffer conditions, and remains soluble when the buffer conditions are slowly exchanged. This is likely, in part, due to the stability of the starting version of T4L and the wide range of pH and salt conditions in which it is stable.[6,8,105] However, the unpredictable effects of coiled-coil fusions on a target protein is an impediment to full validation and implementation of this approach.

Two separate roadblocks need to be overcome for full implementation. First, the fusion constructs have altered expression results when compared the wild type proteins. Coiled-coil fusions result in either reduced or abolished expression or fully insoluble expression of the proteins. Second, the coiled-coils result in proteins that are soluble in a narrow range of buffer conditions, and rapid changes, even into similar conditions, results in precipitation of the proteins. Precipitation during buffer changes is predominantly hindering the formation of T4L coiled-coil disulfide dimers for *in vitro* complex formation.

5.3.2 Stability Through Co-Purifications

As with the split GFP – POI complexes formed through co-expression (chapter 3), the solubility issues were alleviated. Co-purification alone was successful in preventing the protein precipitation issues during purification, and resulted in pure protein in sufficient amounts for crystallization experiments. Complex formation in this manner however is not practical for widespread use since individual purifications for each complex must be performed, eliminating the benefit of pre-formed oligomers. Furthermore, this approach prevents the use of the disulfide dimer scaffolds in complexes where the target protein contains native cysteines. This was the case with IDP01204, preventing screening of all T4L constructs to validate this approach.

5.3.3 Rescue Strategies to Validated Coiled-Coil Mediated Oligomers

5.3.3.1 Coiled-Coil Optimization

To date, all efforts we have undertaken for coiled-coil mediated oligomers and materials have been based upon one coiled-coil hetero pair. This approach has so far proven futile to create the intended designed assemblies, primarily due to the solubility issues that have arisen in these efforts. Viability of this approach cannot be determined with so few soluble proteins to evaluate. As of now this simply cannot be used as a prevalent crystallization rescue approach. *In vitro* complex for scaffolding or symmetry induction for crystallization remains a goal and would significantly reduce the efforts needed for screening multiple geometries. Well-behaved, soluble complexes that were obtained through co-purification are an indication that this approach may be viable with simple changes.

It is unknown why these specific coiled-coils are reducing the solubility of proteins that they are fused to. It is possible that there are homodimers leading to insolubility for oligomer fusions or leading to interactions between the helices that were unforeseen. They may interfere with the folding of the fusion proteins or they may interact directly with the fusion proteins. Each case would result in the aggregation and precipitation that has been observed. Alternative coiled-coil sequences could eliminate many of the issues that have arisen from the coiled-coil fusions. Multiple unique heterodimeric coiled-coils with alternative sequences have been reported.[111,117–121] These pairs should be evaluated for their properties when fused to target proteins. If they do not inhibit protein solubility and retain their preferential hetero-dimeric binding, they would be ideal candidates for future efforts. It remains a possibility that the expression and solubility issues for two of the three crystallization targets is inherent to the proteins chosen and not due to presence of the coiled-coils. These proteins, although purportedly soluble, have previously failed in crystallization efforts and may have folding or stability issues.

5.3.3.2 Disulfide Bond Optimization

Disulfide oxidation with Cu^{2+} in alkaline conditions, as with GFP disulfide dimers, is not viable with the current constructs. The drastic changes in buffer conditions results in significant loss of protein. Alternative means of disulfide formation may be pursued to circumvent these buffer changes. The first approach would be to express the T4L disulfide mutants in *E. coli* cell lines engineered for efficient disulfide bond formation. This approach was attempted with GFP dimers (chapter 3), but was not pursued as *in vitro* dimer formation was optimized. Use of the SHuffle cell lines (chapter 3) or alternative oxidizing cell lines[122] may be applicable.

In the event that these cell lines reduce the expression levels of T4L as they did with GFP, small molecule oxidizing agents[123] may serve to form the disulfide bonds in favorable buffer conditions for solubility of the T4L constructs. This small molecule approach was initially used for the original T4L synthetic disulfide dimer work[6] but was not further pursued as Cu^{2+} catalyzed oxidation was more efficient. Alternatively, bifucntional sulfhydryl cross-linking reagents can create the dimers instead of disulfide bond formation.[124,125] Short, rigid zero-length cross linkers would be well suited for crystallographic applications due to their rigidity.[126,127] Cross-linked oligomers are an intriguing approach to explore in future efforts, even where efficient disulfide bond formation is possible, as multivalent crosslinking reagents are available. Commercially available cross-linkers are available in bi-, tri- and tetra- functional forms. With these, a single cysteine point mutation can be adapted for di-, tri- and tetrameric scaffolding applications.

5.3.4 New Synthetic Symmetry Scaffolds and Fusion Approaches

So far the two scaffolds that have been attempted have significant issues with protein solubility. These issues arise from the constraint that the scaffolds are capable for *in vitro* complex formation. It cannot be determined as of yet if the symmetry scaffolding approach is even practical for crystallization rescue. It would be prudent to focus a more conservative approach for this technology. The approach of scaffolding could be attempted with a series of mutants on a scaffolding protein that is directly fused to the POI. Although this prevents the *in vitro* complex formation that is desired, it eliminates the solubility and contaminations. A logical protein to serve as this scaffold would be MBP. There are already a suite of oligomer mutations[8] and a proven method of creating crystallization fusions through a rigid helical linker.[106] If this proves successful in aiding in the crystallization of novel difficult-to-crystallize proteins (such as the ones that have be attempted in this dissertation), then the efforts to optimize *in vitro* complex formation could be explored.

5.4 Conclusions

Coiled-coil fusions hold potential for creating protein complexes from monomeric building blocks, or even large assemblies from oligomeric starting proteins. However, the solubility issues of such fusions have slowed validation of this approach. Optimization efforts need to be undertaken, specifically alternate heterodimeric pairs in order to create a system applicable to widespread use for scaffolding as intended by this work. The ability to create soluble and stable scaffolds hampered similar efforts with split GFP. Since these issues have been eliminated by buffer optimization for the T4L constructs (without disulfide formation), T4L coiled-coil fusions holds promise for further development as a scaffold.

5.5 Materials and Methods

5.5.1 Cloning

The amino acid sequence from PDB 3FA0 was used to create all T4L constructs, amino acid sequences for the crystallization target constructs were obtained from the NCBI protein database.[128] Amino acid sequences were codon optimized[95] and synthesized as gBlocks (Integrated DNA Technologies) with the terminal sequences GTTTAACTTTAAGAAGGAGATATAC and TAAAATTCGAGCTCCGTCGACAAGCTTG added to the 5’ and 3’ ends respectively. The genes were then cloned via Gibson assembly[75] into pET22b vector cut with NdeI and EcoRI restriction enzymes and transformed into XL-2 Blue Ultracompetent cells (Agilent Technologies). Plasmids with the correct sequences were transformed into BL21 DE3 (New England Biolabs) for protein expression. The amino acid sequences used for all constructs are presented in tables 5.1 and 5.2.

5.5.2 Protein Expression

All proteins were expressed in LB media supplemented with ampicillin and carbenicillin. The 1L LB flasks were inoculated with 10mL of overnight culture and allow to grow at 37°C until OD A_{600nm} reached 0.6-0.8, the temperature was reduced to 15°C, protein expression was induced with 1mM IPTG and allowed to grow for ~16 hours before harvesting.

5.6.3 Purification

Optimized purification was performed in buffer based upon 50mM Hepes pH 7.0, 500mM NaCl, 5% Glycerol. IMAC and SEC was performed as previously published,[129] with TEV cleavage performed overnight dialysis in a buffer by dialysis of mixed IMAC purified T4L or complex and TEV protease in a 1:50 ratio in a buffer of 10mM Tris pH 8.0, 150mM NaCl, 1mM EDTA, 10% glycerol at 4°C, the TEV protease was removed but a second IMAC purification step prior to SEC.[129] Purifications in these buffers resulted in no precipitation for the T4L constructs. Disulfide bond formation was attempted per the methods presented in chapter 2; all attempts this was unsuccessful due to protein precipitation.

5.5.4 Crystallization Experiments

Initial crystallization experiments were set with a protein concentration of 10mg/mL at room temperature. Eight screens, JSCG+, MPDs, PACT (Qiagen), Wizard (Emerald BioStructures), ProPlex, Structure Screen (Molecular Dimensions), Peg, SaltRx (Hampton Research) were set for each complex. Widespread precipitation was observed for both constructs. Protein concentration was reduced to 5mg/mL and trays reset, widespread precipitation (>50% of conditions) was observed for the double cysteine mutant, while only ~25% of double histidine precipitated. After six months no crystals were observed and no further screening has been conducted.

Table 5.1. Sequences of T4-Lysozyme Coiled-Coil Oligomer Constructs

T4L Backbone from PDB deposition 3FA0:
MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITKDEAEK
LFNQDVDAAVRGILRNAKLKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDEAA
VNLAKSRWYNQTPNRAKRVITTFRTGTWDAYK

T4L with N-terminal coiled coil:
AQLKKKLQALKKKNAQLKWKLQALKKKLAQIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSP
SLNAAKSELDKAIGRNTNGVITKDEAEKLFNQDVDAAVRGILRNAKLKPVYDSLDAVRRAALINM
VFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYNQTPNRAKRVITTFRTGTWDAYK

T4L V131C Expression Construct:
MGSDKIHHHHHHENLYFQGAQLKKKLQALKKKNAQLKWKLQALKKKLAQIFEMLRIDEGLRLKIY
KDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITKDEAEKLFNQDVDAAVRGILRNAK
LKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDEAACNLAKSRWYNQTPNRAKR
VITTFRTGTWDAYK

T4L S44C Expression Construct:
MGSDKIHHHHHHENLYFQGAQLKKKLQALKKKNAQLKWKLQALKKKLAQIFEMLRIDEGLRLKIY
KDTEGYYTIGIGHLLTKSPSLNAAKCELDKAIGRNTNGVITKDEAEKLFNQDVDAAVRGILRNAK
LKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYNQTPNRAKR
VITTFRTGTWDAYK

T4L R125C/E128C Expression Construct:
MGSDKIHHHHHHENLYFQGAQLKKKLQALKKKNAQLKWKLQALKKKLAQIFEMLRIDEGLRLKIY
KDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITKDEAEKLFNQDVDAAVRGILRNAK
LKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRMLQQKCWDCAAVNLAKSRWYNQTPNRAKR
VITTFRTGTWDAYK

T4L N40H/S44H Expression Construct:
MGSDKIHHHHHHENLYFQGAQLKKKLQALKKKNAQLKWKLQALKKKLAQIFEMLRIDEGLRLKIY
KDTEGYYTIGIGHLLTKSPSLHAAKHELDKAIGRNTNGVITKDEAEKLFNQDVDAAVRGILRNAK
LKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYNQTPNRAKR
VITTFRTGTWDAYK

Table 5.2. Crystallization Targets Derived From CSGID Database

1) IDP01024 - *Salmonella enterica* anti-sigma28 factor FlgM

NCBI Sequence:

MSIDRTSPLKPVSTVQTRETSDTPVQKTRQEKTSAATSASVTLSDAQAKLMQPGVSDINMERVEA
LKTAIRNGELKMDTGKIADSLIREAQSYLQSK

Expressed Construct:

MGSDKIHHHHHHENLYFQGSIDRTSPLKPVSTVQTRETSDTPVQKTRQEKTSAATSASVTLSDAQ
AKLMQPGVSDINMERVEALKTAIRNGELKMDTGKIADSLAQLEKELQALEKENAQLEWELQALEK
ELAQ

2) IDP01204 - *Bacillus anthracis str. Ames* glycosyl transferase, group 2 family protein

NCBI Sequence:

MGNEQVKNVGEEKKLCLCMIVKNESRIMERCLNATKSIVDFVSICDTGSTDHTPEIIENWCKENE
IPGTVHHEPFKNFGYNRSLAVSLAQKTYPEADYLLILDADMILEVDPEFDKTSLTEDHYLTLQYD
IHIKYWLTRLLKASLPWKSVGVTHEYWDIDRSKVGANYNTRVARLETLVVNDPGDGGSKADKFER
DERLLLQGINDPETTPDLHIRYLFYLAQTYFHLSQFEDSIKWYKKRVEAGGWVEEVFYSLLRIGF
CYEQLANRSANKQHEVTEADEKENAKKQEEQYTALAVLYFQKAWEYRPTRAEPLYQLARMYRLKS
QNNIALMYALQGKEVPFPKDDLLFVDYHVYDYLFDYEISINAFYIPHKKHLGAASQKYLESKKEE
LPLHIANMVENNAKFY

Expressed Construct:

MGSDKIHHHHHHENLYFQGGNEQVKNVGEEKKLCLCMIVKNESRIMERCLNATKSIVDFVSICDT
GSTDHTPEIIENWCKENEIPGTVHHEPFKNFGYNRSLAVSLAQKTYPEADYLLILDADMILEVDP
EFDKTSLTEDHYLTLQYDIHIKYWLTRLLKASLPWKSVGVTHEYWDIDRSKVGANYNTRVARLET
LVVNDPGDGGSKADKFERDERLLLQGINDPETTPDLHIRYLFYLAQTYFHLSQFEDSIKWYKKRV
EAGGWVEEVFYSLLRIGFCYEQLANRSANKQHEVTEADEKENAKKQEEQYTALAVLYFQKAWEYR
PTRAEPLYQLARMYRLKSQNNIALMYALQGKEVPFPKDDLLFVDYHVYDYLFDYEISINAFYIPH
KKHLGAASQKYLESKKEELPLHIANMVENNAKFAQLEKELQALEKENAQLEWELQALEKELAQ

3) IDP90101 - *Salmonella enterica* sseL deubiquitinase

NCBI Sequence:

MSDEALTLLFSAVENGDQNCIDLLCNLALRNDDLGHRVEKFLFDLFSGKRTGSSDIDKKINQACL
VLHQIANNDITKDNTEWKKLHAPSRLLYMAGSATTDLSKKIGIAHKIMGDQFAQTDQEQVGVENL
WCGARMLSSDELAAATQGLVQESPLLSVNYPIGLIHPTTKENILSTQLLEKIAQSGLSHNEVFLV
NTGDHWLLCLFYKLAEKIKCLIFNTYYDLNENTKQEIIEAAKIAGISESDEVNFIEMNLQNNVPN
GCGLFCYHTIQLLSNAGQNDPATTLREFAENFLTLSVEEQALFNTQTRRQIYEYSLQ

Expressed Construct:

MGSDKIHHHHHHENLYFQGSDEALTLLFSAVENGDQNCIDLLCNLALRNDDLGHRVEKFLFDLFS
GKRTGSSDIDKKINQACLVLHQIANNDITKDNTEWKKLHAPSRLLYMAGSATTDLSKKIGIAHKI
MGDQFAQTDQEQVGVENLWCGARMLSSDELAAATQGLVQESPLLSVNYPIGLIHPTTKENILSTQ
LLEKIAQSGLSHNEVFLVNTGDHWLLCLFYKLAEKIKCLIFNTYYDLNENTKQEIIEAAKIAGIS
ESDEVNFIEMNLQNNVPNGCGLFCYHTIQLLSNAGQNDPATTLREFAENFLTLSVEEQALFNTQT
RRQIYEAQLEKELQALEKENAQLEWELQALEKELAQ

Chapter 6

Computationally Designed Cages as Scaffolds

6.1 Introduction

Designed protein cages are large self-assembling structures. These assemblies mimic similar assemblies found in nature such as virus capsids, clathrin and ferritin.[130–132] In recent years, methods to design these cages with specific geometries have been developed involving oligomer fusions[77–81] and computationally designed interfaces.[17,18] It has been proposed that these cages may serve as cargo delivery vehicles for targeted delivery of therapeutics[133] and recently this concept has shown indications of improving anti-cancer drug delivery.[134] We instead, looked into the ability of these cages to serve as scaffolds crystallization targets as an alternative use for these structures.

This is an extension of the work discussed in chapters 3 and 5. Here, now instead of induction of simple C2 symmetry, higher-order tetrahedral symmetry is used for scaffolding. It has been predicted[7] and confirmed[135] that internal protein symmetry tends to be replicated as crystallographic symmetry when a symmetric protein crystallizes. In the case of highly symmetric and rigid protein assemblies, the asymmetric unit of the crystal tends to only be a subunit of the assembly. The cage is then reconstructed through crystallographic symmetry operations. One example of this is presented in chapter 8, where a cage with icosahedral symmetry was found to have one of the 12 faces in the asymmetric unit. The full cage could then be constructed through the symmetry operations of the $P4_232$ space group in which the protein crystallized. Anecdotally, crystallization of these highly symmetric cages and designed assemblies also has higher success rates than asymmetric or low symmetry proteins.

As the library of highly symmetric cages from natural,[136] engineered[137] and fully designed[17,18,77] continues to increase, we looked into adapting them for crystallization scaffolding applications. Here, a cage of known structure and solvent exposed termini would be selected. Crystallization targets would them be genetically fused to the exposed termini, decorating the cage with the target protein. Provided the cage was of sufficient size to prevent steric clashes between the newly fused proteins, the cage symmetry of the cage would be retained. Cage fusions could crystallize provided the fusion protein is capable of making at least one crystal contact.

In order to test this, we went about choosing a robust cage to serve as the scaffold. It was decided to utilize a computationally designed tetrahedral cage published by King *et al.* (PDB 4EGG).[18] This cage (T3-10) is made from a single subunit that natively forms a trimer. To form the cage, four of the trimers were docked together with tetrahedral symmetry and the interfaces mutated to spontaneously assemble (figure 6.1). T3-10 is an ideal candidate cage for this purpose. The C-terminus is solvent exposed allowing facile genetic fusions to target proteins. The C-termini are sufficient separated to reduce the chances of steric clashes between fusion proteins. Finally, T3-10 has a small internal volume, resulting in a rigid cage. Cages that are highly porous have high solvent contents and are somewhat flexible, resulting in low resolution x-ray diffraction and lack of symmetry in the crystal.[17,138]

Validation of this approach was conducted by choosing two proteins to serve as crystallization targets. Novel crystallization targets proved difficult to validate previous synthetic symmetry scaffolds (chapter 3 and 5). Thus, it was decided to use two target proteins for these experiments, one of which has previously crystallized. Superfolder GFP[49] (sfGFP) would serve as the first target protein, it had previously crystallized and is a compact structure we thought would prevent steric clashes. StarD9[69,70] would serve as a novel crystallization target, that had previously indicated it was capable of crystallizing (chapter 3) although no diffraction was obtained.

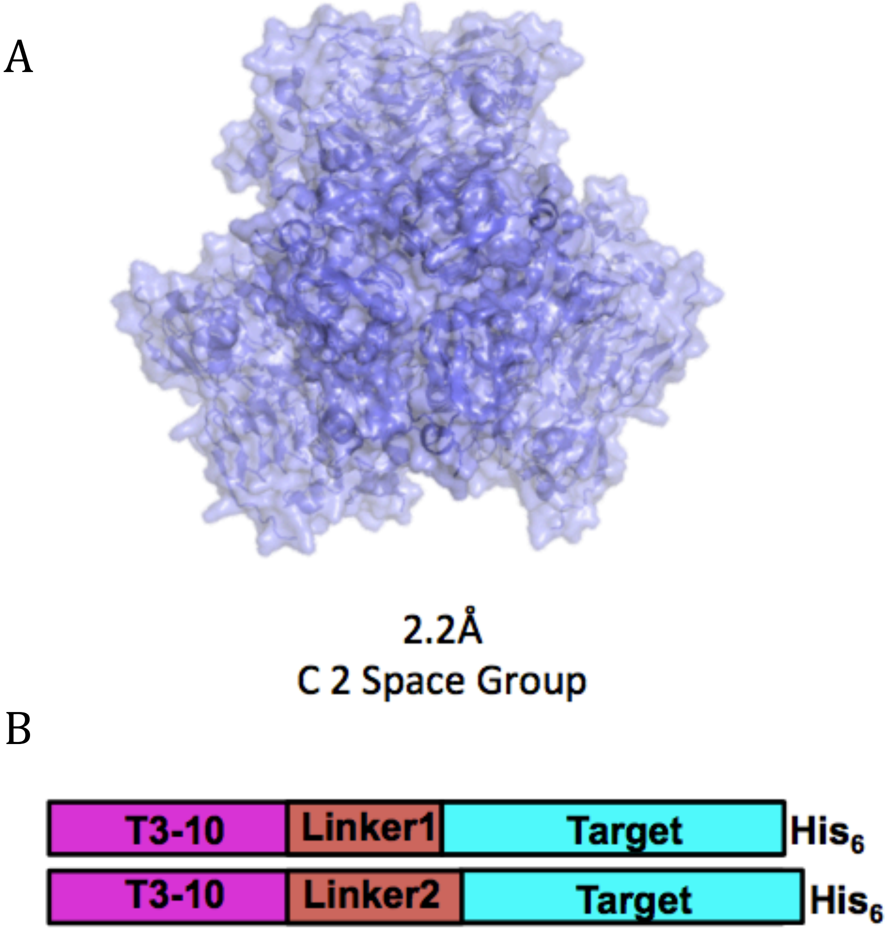

A
2.2Å
C 2 Space Group
B
T3-10
Linker1
Target
His6
T3-10
Linker2
Target
His6

Figure 6.1. Structure of T3-10 and Fusion Construct Designs. (A) The crystal structure of the tetrahedral cage T3-10 (PDB 4EGG). T3-10 originally crystallized in space group C2 with the tetrahedral 2-fold symmetry axis replicated with the 2-fold crystallographic symmetry axis. (B) General design scheme for T3-10 fusions, linker sequences could be varied between the T3-10 and target genes. A C-terminal hexa-histidine tag would be used for purification.

6.2 Results

6.2.1 T3-10 sfGFP Fusions

Initial experiments were conducted on a T3-10 sfGFP fusion construct (T3G) with the full sequences from both proteins with a flexible linker (GSGTGSG) between the two genes. In the original structure of T3-10 the final 12 amino acids were disordered and not visible in the electron density. This is a clear indication of a highly flexible region of the protein, leading a very flexible linkage between the two proteins. Although unfavorable for crystallization, we pursued this construct initially to determine the ability of the cage to form with a C-terminal fusion.

T3G was well expressed and purified to high purity by IMAC and a Superose6 SEC column with a matrix intended for the separation of large protein assemblies in the MDa range (figure 6.2). To determine if the cage was fully assembled in solution, SEC-MALS was used to determine the molecular mass of the assembly.[62] T3G has a predicted mass of 577 kDa, and the mass obtained by SEC-MALS was 565 ± 4 kDa (figure 6.2) at a protein concentration of ~1mg/mL. From this it was determined that the cage was fully assembled in solution. Crystallization experiments were set at concentrations of 1, 5 and 10 mg/mL. No crystals were obtained, but widespread precipitation was observed. Dynamic light scattering was performed to determine if there was a concentration dependent aggregation of the assembly. DLS indicated that the molecular radius of the protein in solution increased as the concentration increased (figure 6.2). This concentration dependent aggregate formation was the likely reason for the precipitation in the crystallization experiments, especially at the higher protein concentrations.

Superfolder GFP was known to form a crystallographic dimer (PDB 2B3P). If this dimer interface were present in solution at higher concentrations, it would be one explanation of the concentration dependent aggregation of T3G as amorphous assemblies.

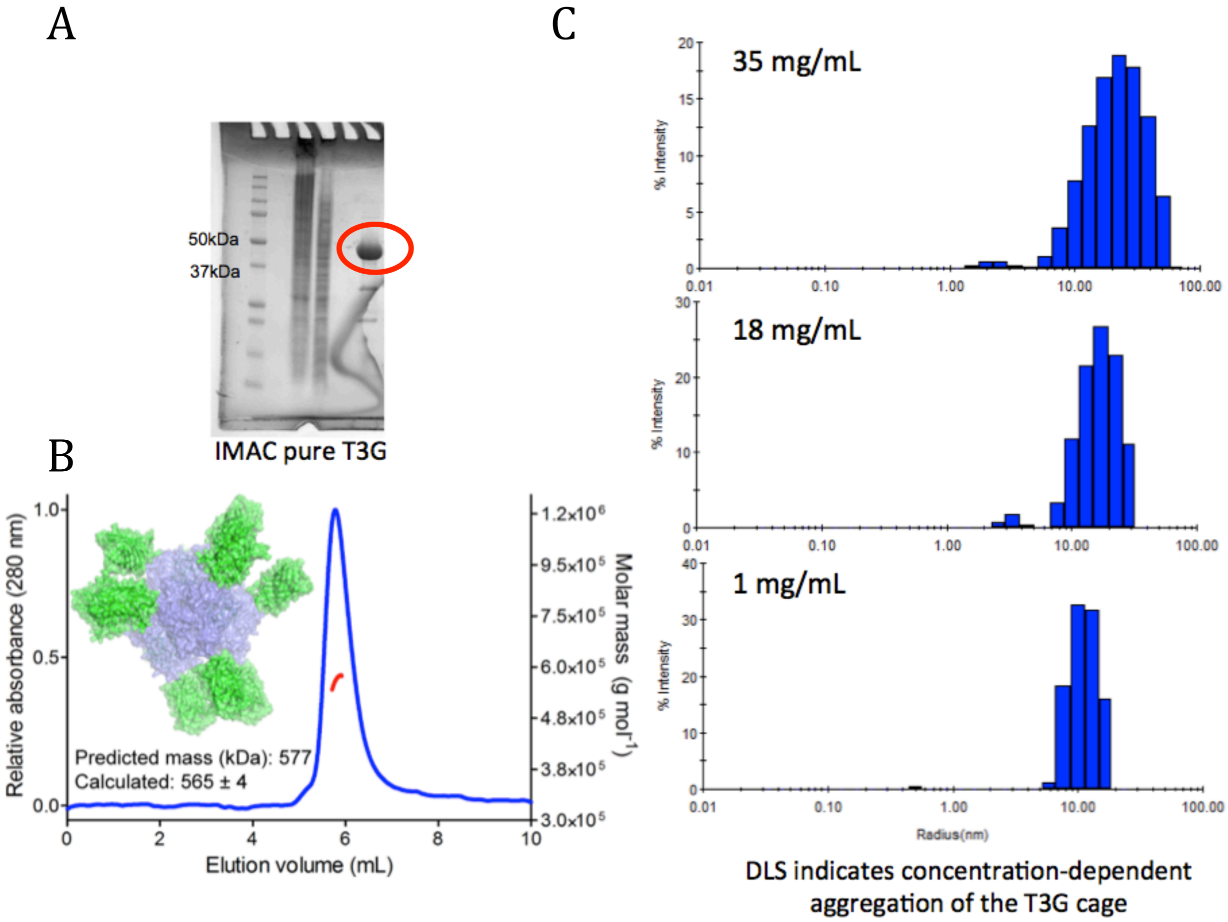
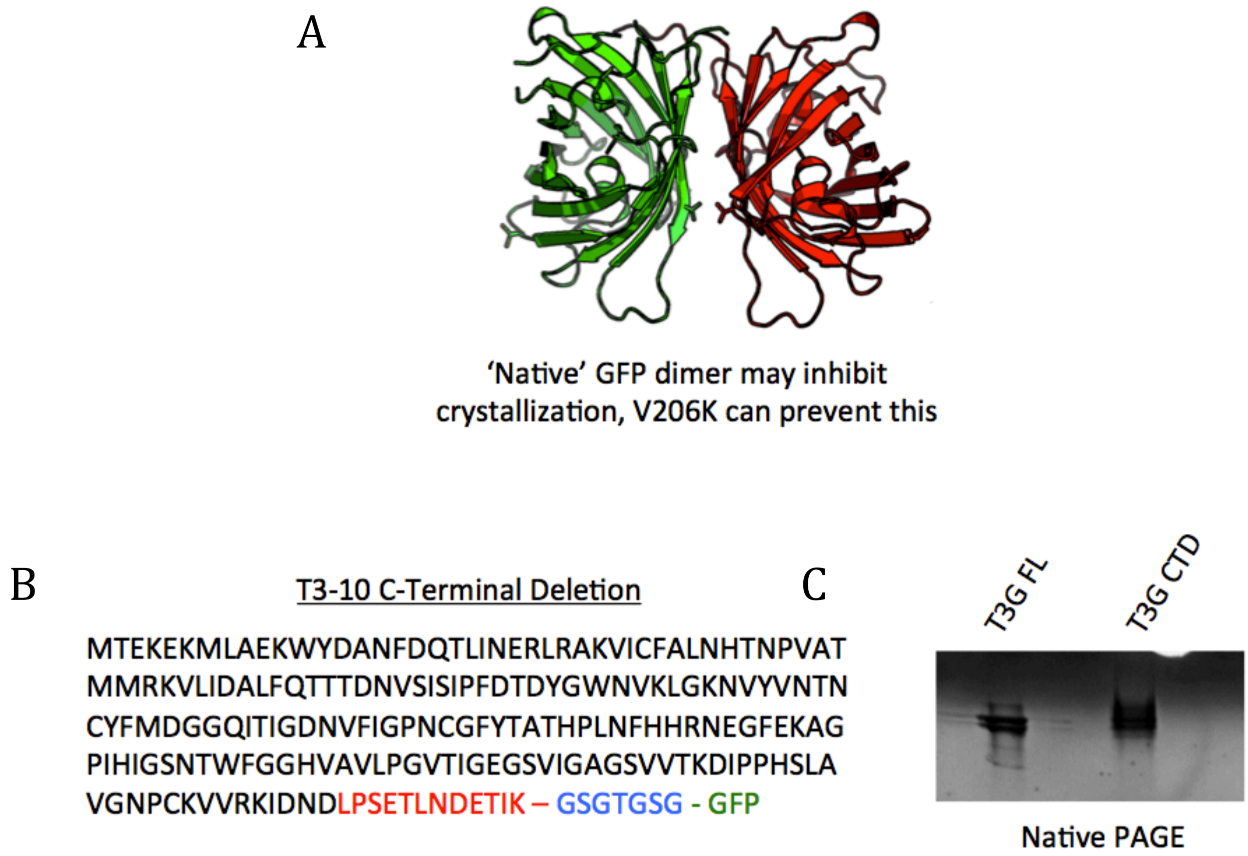


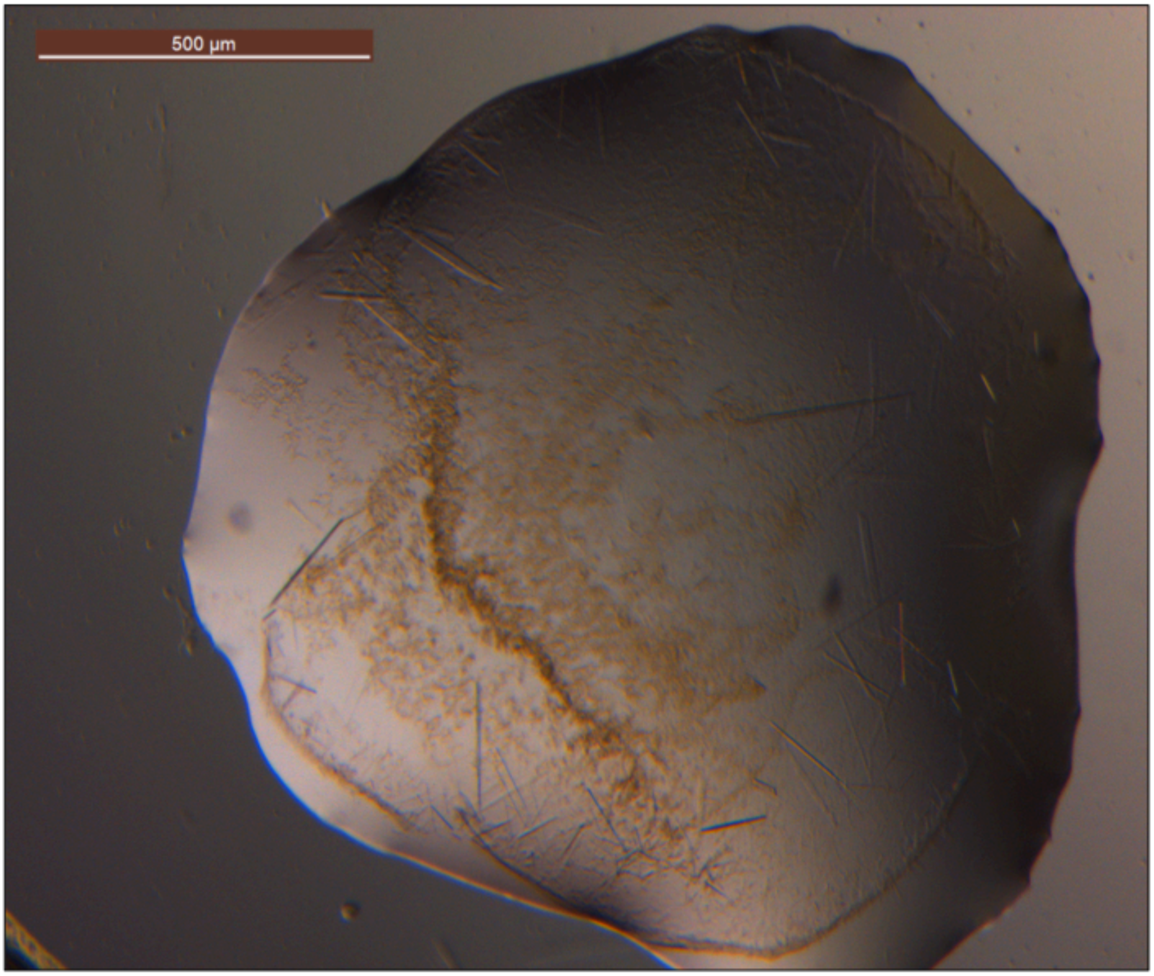
Figure 6.2. Experimental Characterization of T3G Fusion in Solution. (A) SDS-PAGE gel of IMAC purified T3G resulted in highly pure protein (red circle) subsequent SEC purification further purified the protein and gave initial indications the tetrahedral cage was still formed in solution. (B) SEC-MALS was used to validate the size of the cage in solution. The cage was predicted to have a mass of 577 kDa; the single peak from SEC had a calculated mass of 565±4 kDa indicating a fully formed cage in solution. A proposed model of the T3G cage is also presented. (C) DLS was used to determine the changes in particle size in a concentration dependent manner. As the protein concentration increased the radius of the particles also increased, an indication of aggregation.


A
'Native' GFP dimer may inhibit crystallization, V206K can prevent this
B
C
T3G FL
T3G CTD
Native PAGE

T3-10 C-Terminal Deletion

MTEKEKMLAEKWYDANFDQTLINERLRAKVICFALNHTNPVAT
MMRKVLIDALFQTTTDNVSISIPFDTDYGWNVKLGKNVYVNTN
CYFMDGGQITIGDNVFIGPNCGFYTATHPLNFHHRNEGFEKAG
PIHIGSNTWFGGHVAVLPGVTIGEGSVIGAGSVVTKDIPPHSLA
VGNPCKVVRKIDNDLPSETLNDETIK – GSGTGSG - GFP

Figure 6.3. Optimization of the T3G Crystallization Construct, T3G-CTD. (A) The crystallographic dimer of sfGFP (PDB 2B3P). This dimer is mediated by a specific interaction between Val206 from each chain. The mutation, V206K is intended to prevent this dimer from forming in new T3G constructs. (B) The structure of T3-10 (PDB 4EGG) found no electron density for the C-terminal residues highlighted in red. Lack of density indicates disorder, resulting in a highly flexible linkage between the two proteins. These disordered residues were eliminated from new constructs resulting in T3-10 C-terminal deletion (CTD). (C) Native PAGE gel of both T3G Full Length (FL) and T3G CTD. T3G CTD ran as a single band while T3G FL had multiple bands present.



JCSG+ B5
0.1M Na Cacodylate pH 6.5
40% MPD, 5% Peg 8000

Figure 6.4. Crystals of T3G CTD. Crystals of T3G CTD rapidly grew as small rods in one condition from the screen JCSG+ (0.1M Na Cacodylate pH 6.5, 40% MPD, 5% Peg 8000). Crystals diffracted to approximately 18-20Å at the Advanced Photon Source and could not be optimized.

6.2.1.1 Dynamic Light Scattering of T3-10 sfGFP

To determine if there was an aggregation of the T3G cage in a concentration dependent manner, DLS was performed on samples of the cage at three protein concentrations (1, 18 and 35 mg/mL) (figure 6.2). Only the sample at 1mg/mL resulted in a monodisperse sample with a radius consistent with the cage (~10nm). As the protein concentrations increased the sample became polydisperse with an increase in the calculated radii of the particles indicating protein aggregation.

6.2.1. T3G CTD V206K Constructs

It was hypothesized that two factors were impeding the crystallization of the T3G construct and causing in the amorphous aggregation found in many of the crystallization drops. First, the aggregation of T3G (figure 6.2) was believed to the play a predominate role in the amorphous aggregation in the crystallization drops. In order to prevent this, the mutation V206K was introduced. Residue 206 of GFP is typically a small hydrophobic residue that is key to the native dimer seen at high protein concentrations and in crystal structures. It has been established that in GFP versions where residue 206 is an alanine, mutation to lysine abolishes the dimer interface, resulting in purely monomeric protein.[71] The corresponding residue in sfGFP is valine 206; the crystal structure was examined to confirm that this residue mediated the dimer interface (figure 6.3).

The second factor that was believed to interfere with crystallization was a highly flexible linker between T3-10 and sfGFP. The designed flexible linker 'GSGTGSG' was introduced to allow some flexibility to linkage. However, the C-terminus of T3-10 is discorded as it lacks of density in the crystal structure. This motivated the design of a new T3-10 construct, known as T3-10 CTD (C-terminal deletion) to lack the residues 'LPSETLNDETIK.' The new version of T3G incorporating both these changes (T3G CTD V206K) was purified. T3G CTD V206K ran as one distinct band on native PAGE gels (figure 6.3) whereas T3G had several other bands in low abundance.

Crystallization experiments were performed on the new construct. Small rod crystals rapidly grew in one condition (figure 6.4). Diffraction experiments were performed on the crystals at the Advanced Photon Source, none of the crystals diffracted to better than 18Å. Extensive optimization was undertaken but the crystals could not be reliably reproduced or optimized.

6.2.2 T3-10 CTD StarD9 Fusions

Concurrent with the T3G CTD V206K experiments, a second cage fusion was made with T3-10 CTD StarD9 kinesin. Due to previous difficulties in recombinant protein expression of StarD9, both IPTG and auto-induction expression was tested (figure 6.5). This auto-induction procedure provided significantly more protein and was used for all future expression. Purification was performed with IMAC and SEC. Because of low expression levels the IMAC purified protein had extensive *E. coli* protein contamination and two sequential SEC purifications were required for crystallization quality protein that formed a single band by Native-PAGE gels (figure 6.6).

StarD9 cage fusions were extensively screened in crystallization experiments. Earlier unpublished work suggested that the addition of ATP and Mg^{2+} aided in the stability if the protein compared to the apo form; both cofactors were included throughout the purification and crystallization efforts. Two distinct crystal morphologies were observes (figure 6.7). The first crystals grew in polyacrylic acid, and formed extremely thin clustered plate crystals. This crystal form could not be optimized, and did not diffract with synchrotron x-rays. It was not further pursued. Rod like crystals, similar to the T3G CTD V206K crystals were the second morphology. These crystals grew in several conditions containing various concentrations of K/Na tartrate and Peg 3350. Approximately 1 in 50 of the rod crystals diffracted, none better than to 15Å. From these results extensive optimization efforts of the rod crystals were undertaken.

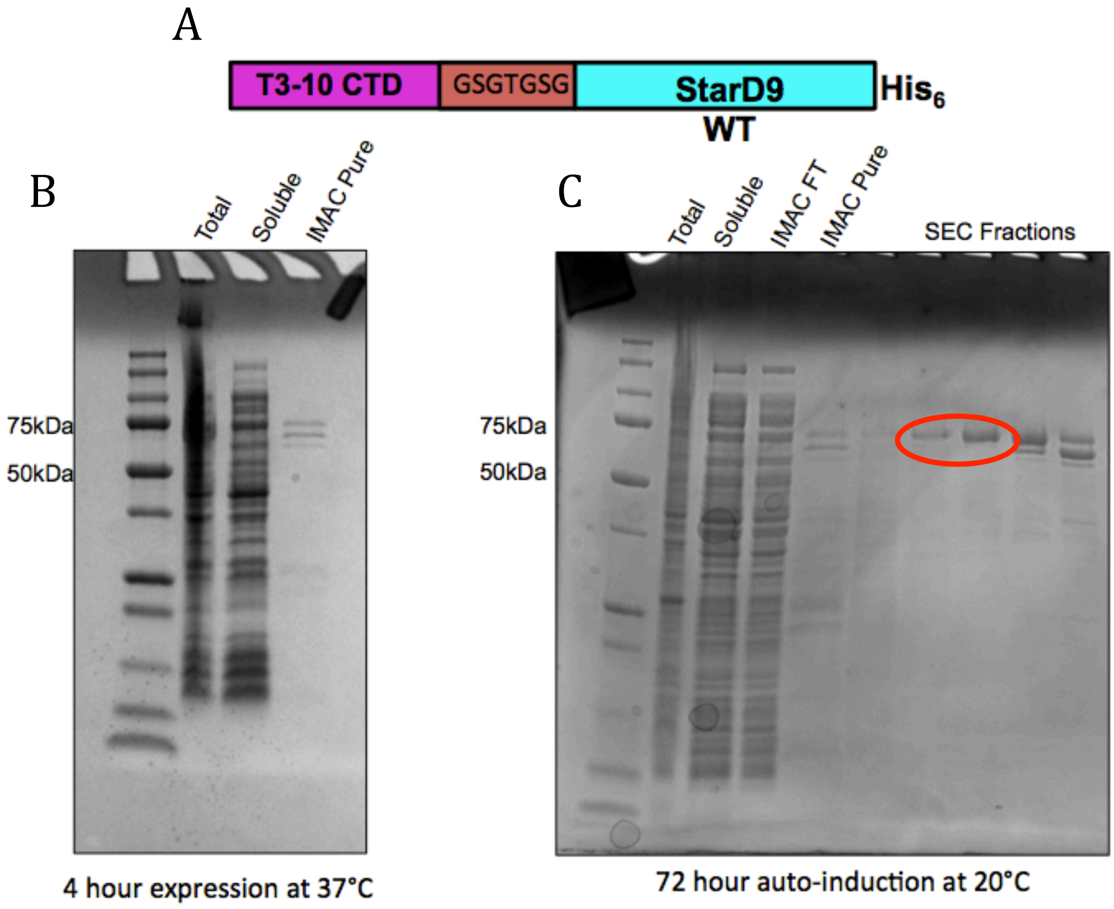
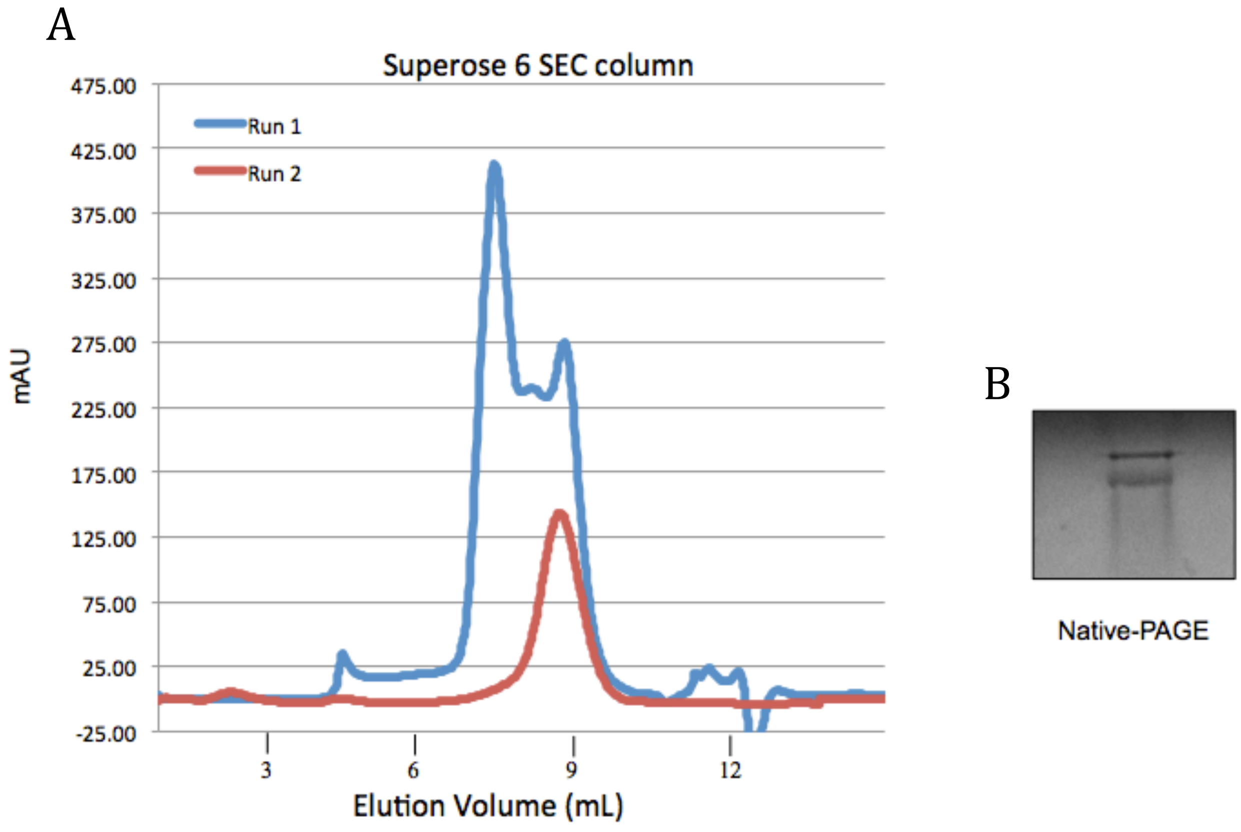

A
T3-10 CTD
GSGTGSG
StarD9
His6
WT
B
Total
Soluble
IMAC Pure
75kDa
50kDa
4 hour expression at 37°C
C
Total
Soluble
IMAC FT
IMAC Pure
SEC Fractions
75kDa
50kDa
72 hour auto-induction at 20°C

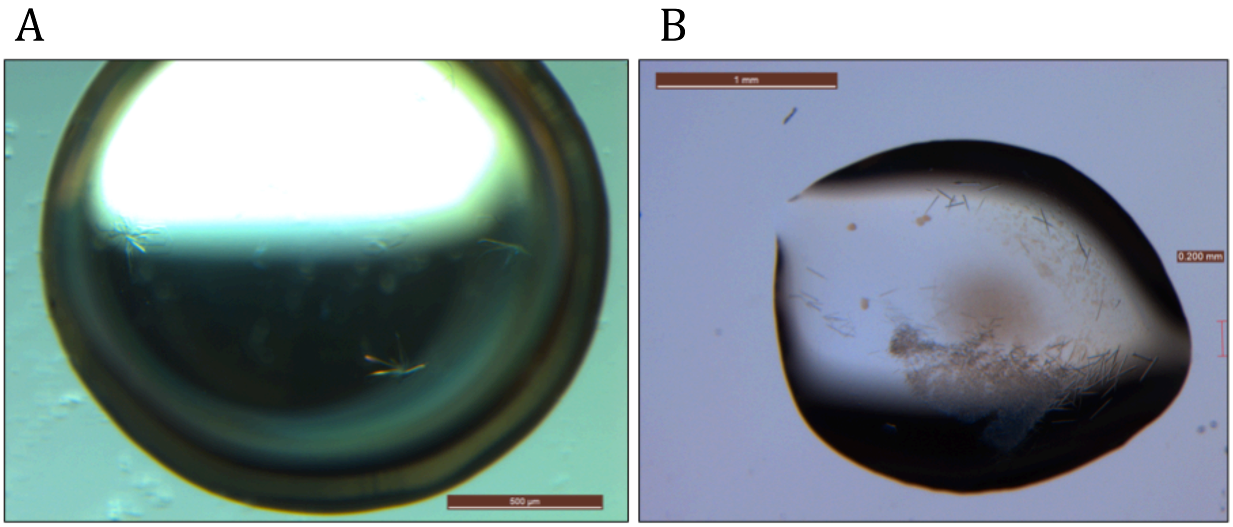
Figure 6.5. Expression Tests of T3-10 CTD StarD9. (A) T3-10 CTD was genetically fused to StarD9 with a GSGTGSG linker and a C-terminal histidine affinity tag from IMAC purification. (B) IPTG induction for 4 hours at 37°C resulted in a poorly expressed band after IMAC purification with a 70kDa protein contaminating the pure protein that could not be further purified due to protein loss during SEC. (C) Auto-induction for 72 hours at 20°C produced protein that could be purified by SEC with some fractions contaminated by a smaller molecular weight protein. The most homogenous fractions (red oval) were pooled for crystallization experiments.



Monomer : 63.6 kDa (20.5kDa T3-10, 43.1kDa StarD9)

Cage: 765 kDa

Figure 6.6. Figure 6.6. Successive SEC Purifications of T3-10 CTD StarD9. (A) SEC immediately after IMAC resulted in significant protein contamination of the T3-10-StarD9 cage (blue trace). A second SEC purification of the peak containing the cage fusion resulted in a single SEC peak of pure protein (red trace). (B) Native-PAGE of the fraction pooled from the second SEC purification. The protein ran as one distinct band in the resolving layer of the gel. The upper band is protein that remained in the stacking layer of the gel; this had also been observed with the T3G constructs on native gels.



JCSG G2
0.02M MgCl2, 0.1M Hepes pH 7.3 - 7.5,
20 -24% Polyacrylic acid 5100

Pact E9
0.15M K/Na Tartrate, 18 – 22% Peg 3350
0.1M Hepes pH 7.2 – 7.6

Figure 6.7. Crystal of T3-10 CTD Full-Length StarD9 Fusions. (A) Thin plate crystals grew in conditions based on screening condition JCSG+ G2 (0.02M $MgCl_2$, 0.1M Hepes pH 7.3-7.5, 20-24% polyacrylic acid 5100). Diffraction quality crystals were not obtained in any of the conditions and the crystals could not be improved through optimization. (B) Rod like crystals grew in conditions based upon the hits from the Pact screen in condition E9 (0.15M K/Na tartrate, 18-22% Peg 3350, 0.1M hepes pH 7.2-7.6). These crystals weakly diffracted at low resolution and were extensively optimized.

6.2.2.1 T3-10 StarD9 Crystal Optimization

Extensive optimization was undertaken for the small rod crystals originally grown in condition Pact E9. First, purifications were repeated with ADP and the non-hydrolysable ATP analog AMP-PNP. Protein purified with ADP crystallized in the same manner while the protein with AMP-PNP formed trigonal crystals and rod crystals in conditions derived from the Pact E9 original condition (figure 6.8). The triganol crystal form could not be further optimized to form larger crystals.

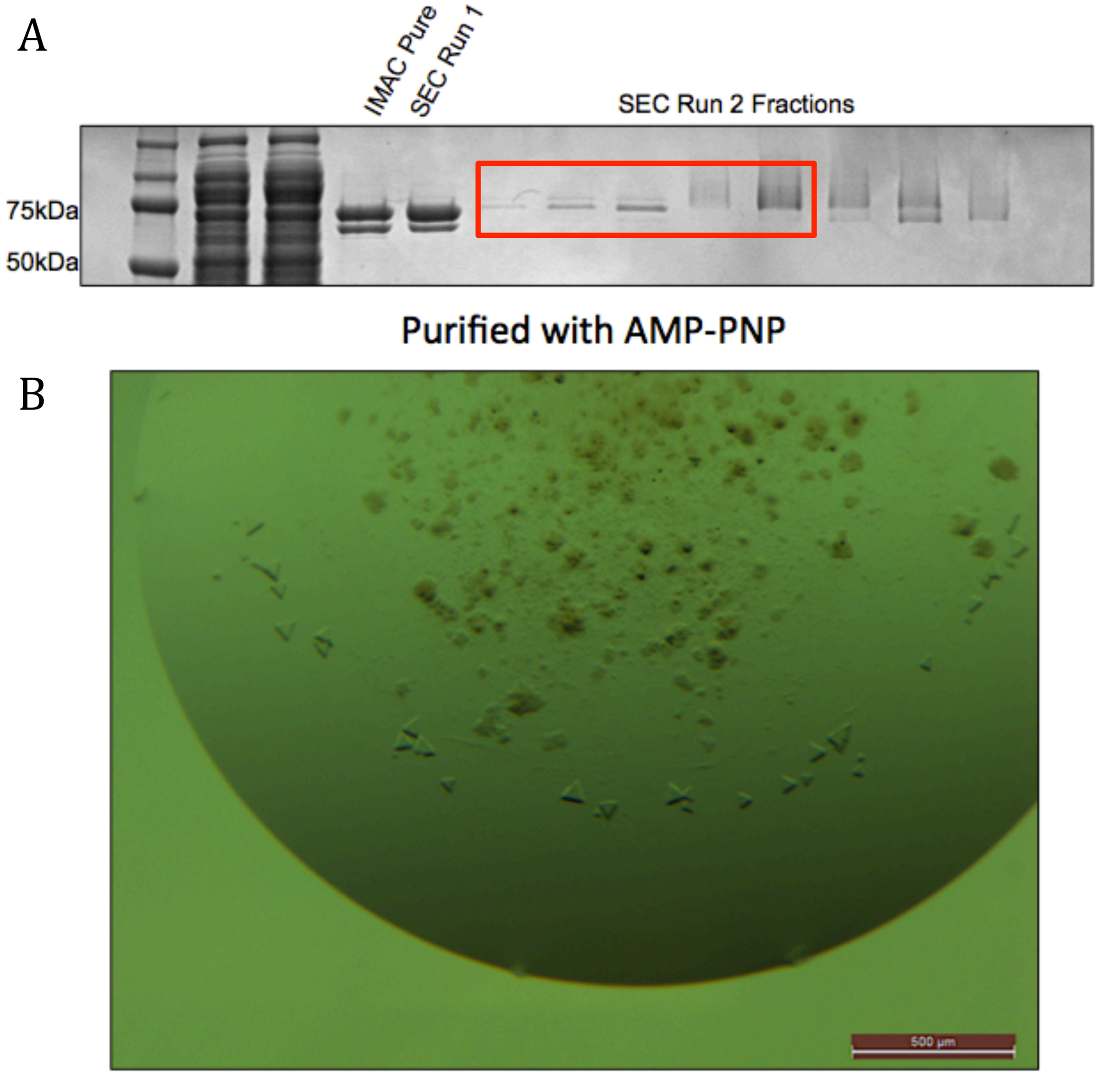
In attempts to increase the size of the rod crystals, optimization involved additive screening, variation in drop size and the ratio of protein to well solution. Larger crystals were only obtained by slowing the rate of vapor diffusion through the used of an oil overlay of the crystallization well solution[139] for proteins with both ATP or AMP-PNP present (figure 6.9).

6.2.2.2 T3-10 CTD StarD9 Crystals Forms

In total three distinct crystal forms were found in the Pact E9 derived conditions (figure 6.10). However, none diffracted to sufficient resolution to allow for structure solution through molecular replacement. The rod crystals were indexed in space group P6 2 2, unit cell of a=b= 198.1Å c=252.1Å, α=β=90° γ=120° and diffracted to 10-15Å. The trigonal crystals from the AMP-PNP purification were indexed in R3 2, unit cell of a=b= 148.5Å c=229.74Å, α=β=90° γ=120° and diffracted to 10Å. The final crystal form was a cluster of cubic crystals. These indexed in C2 with a unit cell of a=177.73Å b= 155.08Å c=177.78Å, α= γ =90° γ β =120° and diffracted to 7.3Å, however a complete data set could not be obtained because of crystal decay and radiation damage.

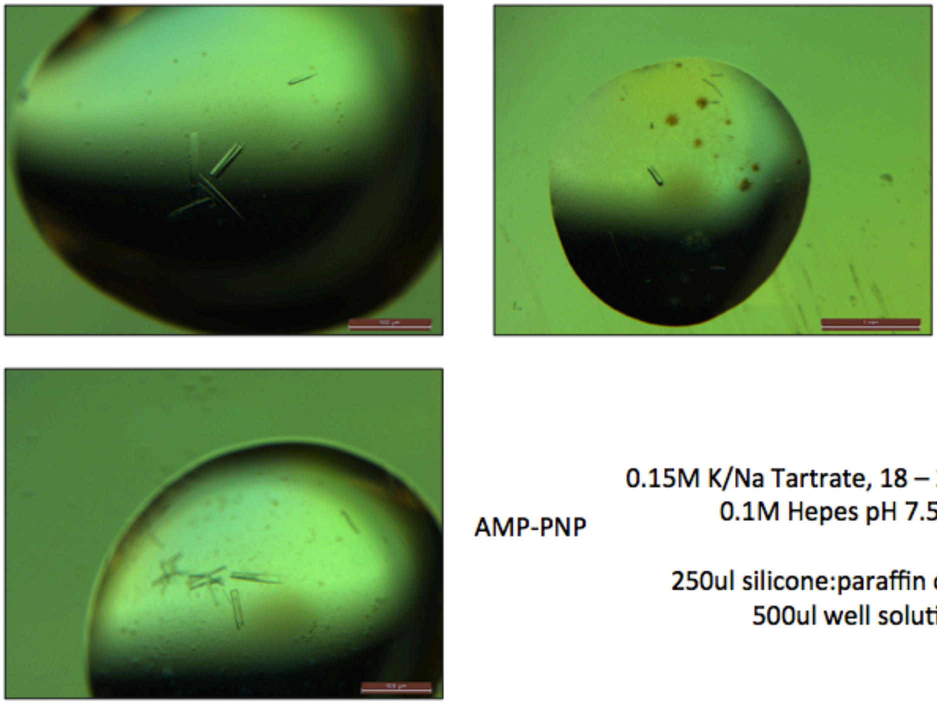
6.2.2.3 T3-10 StarD9 Loop Deletion Constructs

After insufficient resolution for structure solution from all the screened crystals, a new construct was created by removing a disordered loop from the protein (figure 6.11). This loop (Val297-Gly329) is absent from the closest StarD9 homolog Kif1A[140]. This crystallized in the same P6 2 2 crystal form as the full-length construct. These new constructs consistently diffracted to <7Å (figure 6.12).


A
IMAC Pure
SEC Run 1
SEC Run 2 Fractions
75kDa
50kDa
Purified with AMP-PNP
B
500 µm

0.2M K/Na Tartrate, 20% Peg 3350, 0.1M Hepes pH 6.5

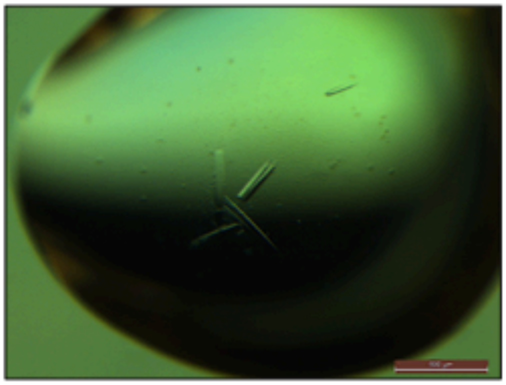
Figure 6.8. T3-10 CTD StarD9 Cage Fusion Purified with AMP-PNP. (A) In an attempt the improve crystal diffraction the fusion cage was purified in the presence of AMP-PNP instead of ATP. SEC fractions highlighted by the red box were pooled for crystallization experiments. (B) In addition to the previously observed rod crystals, new trigonal crystals that only grew with AMP-PNP were observed.



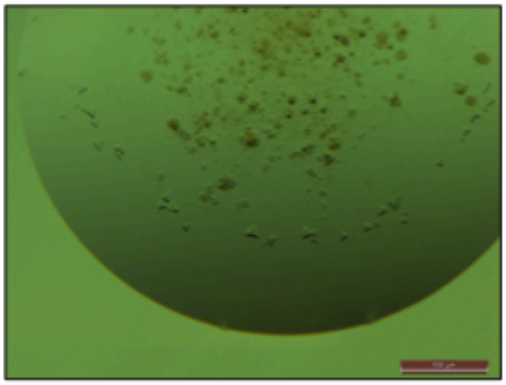
0.15M K/Na Tartrate, 18 – 22% Peg 3350
0.1M Hepes pH 7.5 – 7.6

250ul silicone:paraffin oil overlay
500ul well solution

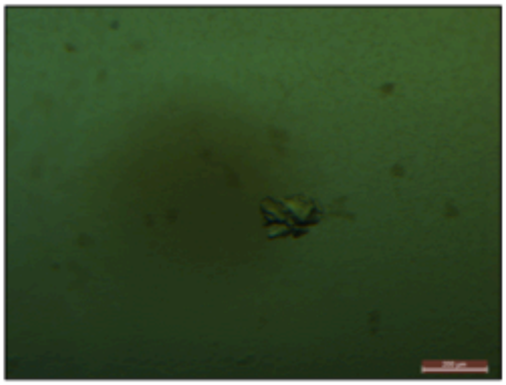
Figure 6.9. Oil Overlay Optimized Crystals of T3-10 CTD StarD9. The rod crystal form from Pact E9 conditions was only grew to larger sizes when vapor diffusion was slowed. To slow diffusion, an overly of 1:1 silicone:paraffin oil was applied to the top of the well solution.



P 6 2 2
Unit Cell: 198.1Å 198.1Å 252.1Å (or 125Å)
90.000° 90.000° 120.000°
ASU = 4 (2 for c = 125Å)
6.6Å Resolution

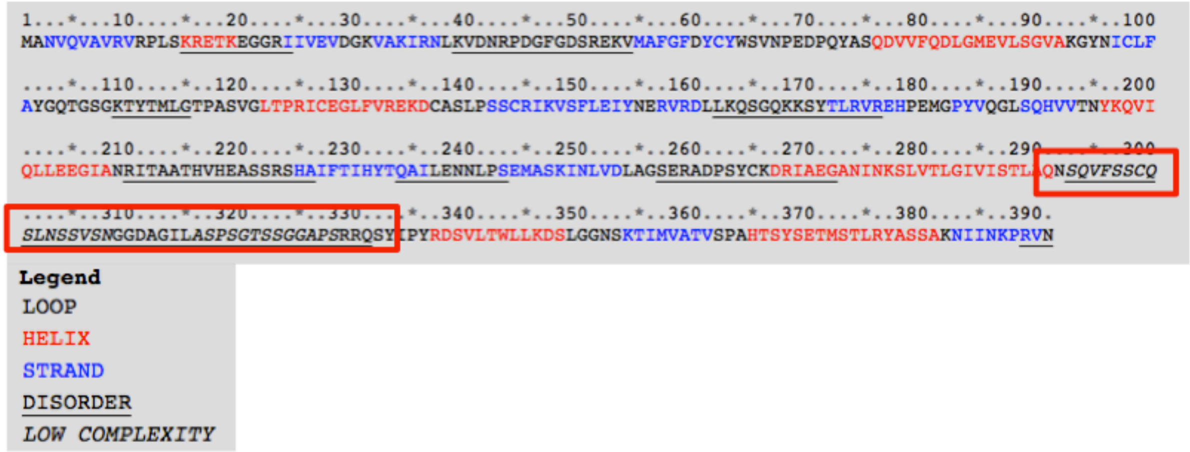


R 3 2
Unit Cell: 148.50Å 148.50Å 229.74Å
90.000° 90.000° 120.000°
ASU = 3 or 4
10.0Å Resolution



C 2
Unit Cell: 177.73Å 155.08Å 117.78Å
90.000° 100.335° 90.000°
ASU = 5 or 6
7.3 Resolution

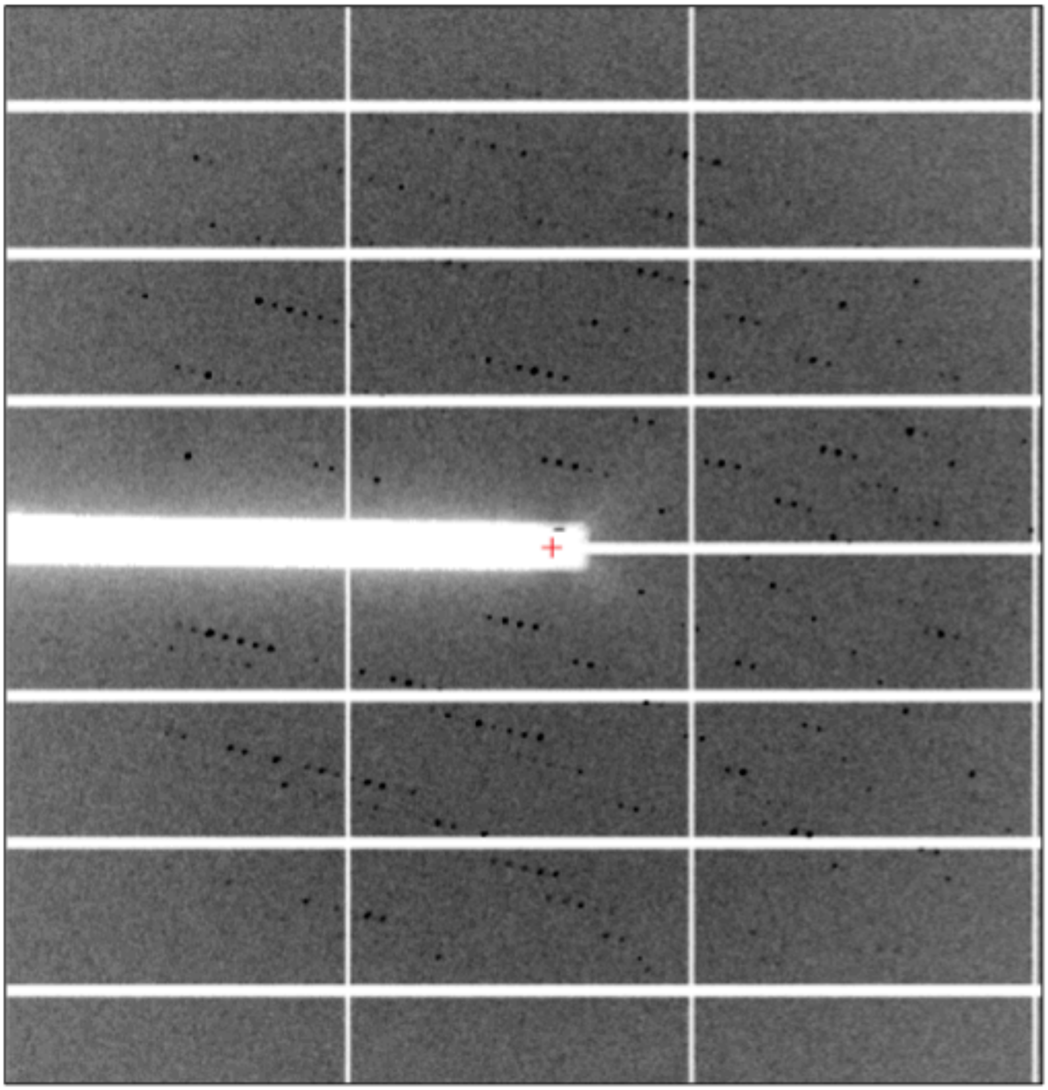
Figure 6.10. The Three Distinct Crystal Forms of T3-10 CTD StarD9. Three unique crystal forms were found, all grew in conditions based on Pact E9 with only variations to K/Na tartrate, Peg 3350 or pH changes. None of the crystals resulted in data sets that could be solved by molecular replacement.



```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80....*...90....*..100
MANVQVAVRVRPLSKRETKEGGRIIVEVDGKVAKIRNLKVDNRPDGFGDSREKVMAFGFDYCYWSVNPEDPQYASQDVVFQDLGMEVLSGVAKGYNICLF

....*..110....*..120....*..130....*..140....*..150....*..160....*..170....*..180....*..190....*..200
AYGQTGSGKTYTMLGTPASVGLTPRICEGLFVREKDCASLPSSCRIKVSFLEIYNERVRDLLKQSGQKKSYTLRVREHPEMGPYVQGLSQHVVTNYKQVI

....*..210....*..220....*..230....*..240....*..250....*..260....*..270....*..280....*..290....*..300
QLLEEGIANRITAATHVHEASSRSHAIFTIHYTQAILENNLPSEMASKINLVDLAGSERADPSYCKDRIAEGANINKSLVTLGIVISTLAQNSQVFSSCQ

....*..310....*..320....*..330....*..340....*..350....*..360....*..370....*..380....*..390.
SLNSSVSNGGDAGILASPSGTSSGGAPSRRQSYIPYRDSVLTWLLKDSLGGNSKTIMVATVSPAHTSYSETMSTLRYASSAKNIINKPRVN
```

Legend
LOOP
HELIX
STRAND
DISORDER
LOW COMPLEXITY

Figure 6.11. XtalPred Results for the Kinesin Domain of StarD9. The kinesin domain of StarD9 features a large disordered loop (red box). This loop is absent from other kinesin domains including the closest homolog of StarD9, Kif1A. This loop was removed and replaced with a 'GSG' linker. This construct crystallized in the same space group as the full-length protein but diffracted to a higher resolution.



T3-10 CTD StarD9 PactE9, collected Oct. 16, 2014, APS 14-ID-C
P 6 2 2
Unit Cell: 198.1Å 198.1Å 252.1Å 90.000° 90.000° 120.000°
6.6Å Resolution

Figure 6.12. Diffraction Image from T3-10 CTD StarD9 Loop Deletion Construct. Diffraction from the optimized construct with optimized crystal growth conditions diffracted with strong, well-defined reflections to 6.6Å. Full data sets were obtained from several isomorphous crystals to similar resolutions.

6.2.2.5 Molecular Replacement Efforts

Extensive efforts were undertaken to solve the structure of T3-10 CTD StarD9 with the loop deletion utilizing the 6.6Å data set through molecular replacement. Search models used include the structures T3-10, Kif1A and a homology model of StarD9 from Phyre2.[116] Search models in varied oligomeric states of fusion models from either Phyre2 predictions or manually made in PyMol[99] were also used without success. Additional search models with truncated loops and alanine only backbones were generated in CHAINSAW[141] were attempted. Molecular replacement was attempted with Phaser[88] and automated molecular replacement programs MrBump and MRage.[142,143] At best only partial structures of poorly packed T3-10 chains were obtained. None of these partial solutions packed without significant steric clashes and refinements failed.

6.3 Discussion

6.3.1 Viability of Cages as Scaffolds

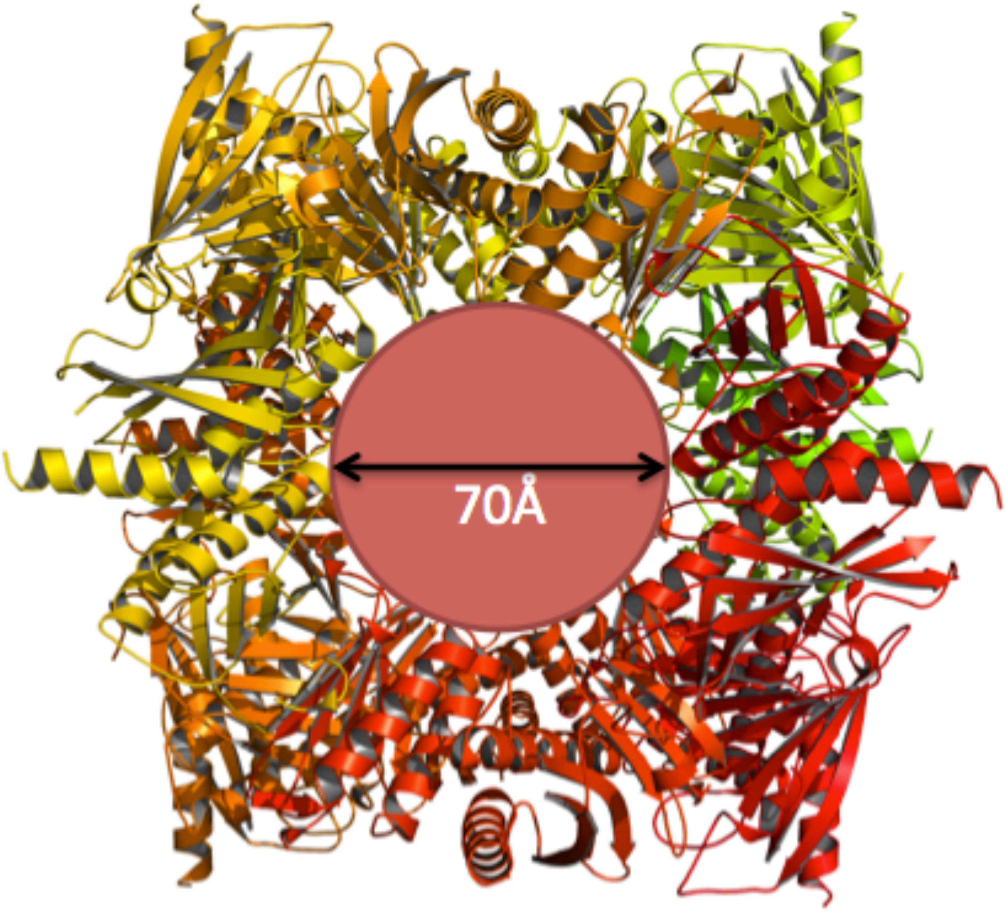
Key to the viability of protein cages, (whether natural, engineered or designed) to serve as crystallization scaffolds is for the cages to be robust enough to still form when an addition protein is genetically fused to it. With two examples we have found this is possible. Characterization of the T3G cage indicated that the cage is fully formed with close to the expected molecular mass in solution. Although the StarD9 fusion was not as extensively characterized, the size exclusion elution profile and native gels indicated the presence of a large cage in solution.

It is also essential that the crystallization targets fused to the surface of the cage are capable of making sufficient crystal contacts to form a stable crystal lattice. While this is a property unique to each crystallization target, the results for both fusion proteins presented here are promising. In both instances, crystals were obtained, and in the case of StarD9 fusions stable enough to diffract to low resolution were obtained. The crystals obtained for both fusions crystals did not crystallize in expected morphologies. Since each edge of the protein is identical, it is expected that the tetrahedral cages would crystallize in morphologies resembling cubes, since crystal growth might be expected to occur equaling in along all three axes.

The rod crystals have significant growth in only one direction, suggesting the cage fusions are not as symmetric as we predicted them to be. Flexibility in the linkers that would be one explanation for this resulting in the fusions proteins packing in unintended orientation on the surface of the T3-10 cage. While the molecular mass of the T3G cage was validated in solution, indicating assembly of twelve subunits in solution, no efforts to validate the shape of the cages was undertaken. Unexpected orientations of the fusion proteins would be one reason why molecular replacement has failed to reveal the low-resolution structure of the T3-10 CTD StarD9 cage. With the lack of a crystal structure it has remained impossible to fully validate our approach. Until the resolution of the x-ray diffraction improves, experimental phasing to solve the structure has a low chance of success. The logical next step will be to validate the shape of the cage to determine if the tetrahedral symmetry remains intact or distorted. Future experiments should be contemplated to further characterize the cage with negative-stain electron microscopy[144–147] or small angle x-ray scattering.[148,149] From these approaches it would be possible to determine if the cages are forming in the manner we have intended.

6.3.1.1 Internal Cage Fusions

Fusion of crystallization targets to the external surface of a protein cage requires the target protein to form the crystals contacts. Since this approach is intended to be used as a means of rescuing proteins that have failed to crystallize with existing methods, it remains a possibility that the proteins can simply not make these contacts. One way to overcome this is to utilize larger cages with internal cavities where the crystallization target can be placed. Steric clashes would be more significant with this approach, but recently two component cages have been developed.[17]



T33-28

Figure 6.13. Designed Cages Have Internal Cavities Amenable to Internal Fusions. The structure of the two-component cage T33-28 features an internal cavity with a diameter of 70Å, sufficient to accommodate a small crystallization target. Since the cage is composed of two separate protein components, not every subunit requires a fusion to the crystallization target reducing the possibility of steric clashes. Since the crystallization target is internal to the cage the crystal contacts on the exterior of the cage remain unchanged allowing the fusion version of the cage to crystallize in the same manner as the apo protein.

With internal fusions of the crystallization target proteins, the crystal contacts would be made from the cage itself. Since the cages that would be used for this have their x-ray structures solved, it has already been established that these cages can form the necessary crystal contacts. Unless the fusions perturb the cage assembly, these contacts would remain unchanged and the fusion cages should crystallize in the same manner as the apo cages. Internal encapsulation of proteins is protective in nature as well. It has been found that proteins internalized in cages are resistant to proteolytic degradation.[150] This can be beneficial for recombinant expression of proteins sensitive to endogenous *E. coli* proteases. One such protein is StarD9. Low expression levels due to proteolysis hindered expression of monomeric StarD9 for crystallization studies. This was partially mitigated prevented by fusion to T3-10, yet smaller molecular weight bands were observed in some instances, which may correspond to degraded protein.

Although internal fusions would retain the existing crystal contacts, which are advantageous to the external fusions, they still entail unique challenges. First, the internal space is limited in the cages. Because of this the size of potential fusion proteins is limited, there is a physical limit to the size of the proteins that can be contained in the cage. With single component cages, every subunit of the cage requires a fusion protein adding to the potential of steric hindrance. This would be reduced in two component cages where the fusion can be made to one of the subunits. The second limitation to internal fusion is the crystallization of intrinsically disordered proteins or proteins with disordered regions. These proteins fail to crystallize because they have no definite structure and cannot make crystal contacts. If these proteins are externally fused to cages they will still fail to form crystal contacts. However, if these proteins are internally fused, the cage can still form crystals but it is likely that no electron density would be observed for the protein. Furthermore, if the linkage between the crystal target and cage is too flexible, the internal proteins would be free to move and not be in one fixed position. This would again obscure the electron density of the crystallization target.

6.3.2 Linker Optimization

One aspect of these cage fusion constructs that has not yet been optimized in the linker sequences between the cage and POI. In each construct a flexible ‘GSGTGSG’ sequence was used to link the two proteins. The flexibility of this sequence was to prevent possible steric hindrance of the external fusions allowing each protein to adopt different orientations, and allowing the cage to fully form and pack POIs around the cage. Although this may aid in fusion cage assembly, it is detrimental to crystallization due to the flexibility. Several studies have investigated protein linkers and have found rigid linkers suitable to efficiently fuse two proteins or domains together.[112,113,151] Several linkers should be screened, the T3G construct would be well suited for these variations as it is well expressed and has been more extensively characterized. From these linker variations a stable cage core could be established to serve as a universal scaffold for crystallization.

6.5 Conclusions

Preliminary experiments have indicated that our protein cage still form with externally fused target proteins. In both instances, the resulting cages only crystallized with extensive optimization and only diffracted to low resolution. To fully realize this approach as a reliable and consistent crystallization method, further elucidation of a functional cage scaffold is required.

6.6 Materials and Methods

6.6.1 Cloning

The T3G construct was commercially synthesized and cloned into the expression vector pET-22b via the NdeI and XhoI restriction sites with restriction enzyme cloning. To created the T3G CTD V206K constructs, first site-directed mutagenesis[152] was used to create the mutation. Each gene was then amplified with PCR, excluding the C-terminal disordered residues from T3-10. T3G CTD V106K was assembled via Gibson assembly[75] into pET22b vector cut with NdeI and EcoRI restriction enzymes and transformed into XL-2 Blue Ultracompetent cells (Agilent Technologies). Plasmids with the correct sequences were transformed into BL21 DE3 (New England Biolabs) for protein expression. The terminal sequences GTTTAACTTTAAGAAGGAGATATAC and TAAAATTCGAGCTCCGTCGACAAGCTTG were used to clone the genes into the vector. T3-10 StarD9 constructs were synthesized as gBlocks (Integrated DNA Technologies) and cloned via Gibson assembly utilizing the same method as the T3G cloning.

6.6.2 Protein Expression

All constructs were expressed in LB supplemented with auto-induction[57,96] sugars and ampicillin. The 1L LB flasks were inoculated with 10mL of overnight culture and allow to grow for 72 hours at 20°C before harvesting.

6.6.3 Protein Purification

Both T3G constructs were purified with IMAC over Ni^{2+} and SEC on a 10/300 GL Superose6 column (GE Healthcare). IMAC was performed as previously described for GFP oligomers[42] (chapter 2). SEC was performed in a buffer of 25mM Tris pH 8.0, 150mM NaCl, 1mM DTT, the protein was then concentrated to 10mg/mL, flash frozen in liquid nitrogen and stored at -80°C until crystallization experiments were set.

StarD9 fusions were also purified through the same IMAC and SEC scheme. Optimized buffers were used through out the purification, IMAC used a wash buffer of 50mM Tris pH 7.4, 400mM NaCl, 5mM $MgCl_2$, 2mM beta-mercaptoethanol, 10% w/v glycerol, 30mM imidazole. The elution buffer was the same composition with the addition of 250mM imidazole. SEC was performed in the buffer: 25mM Tris pH 7.4, 100mM NaCl, 5mM $MgCl_2$, 1mM DTT, 5% w/v glycerol. 100uM ATP, ADP or AMP-PNP was added throughout the purification depending on the purification, each cofactor was used in at least on purification and crystallization experiment. Protein was then concentrated to 1-10mg/mL, flash frozen in liquid nitrogen and stored at -80°C until crystallization experiments were set.

6.6.4 Crystallization

Broad screening was used for all constructs, eight screens, JSCG+, MPDs, PACT (Qiagen), Wizard (Emerald BioStructures), ProPlex, Structure Screen (Molecular Dimensions), Peg, SaltRx (Hampton Research) were set for each fusion. T3G constructs were used at proteins concentrations of 1-2.5mg/mL with crystals obtained at 2,5mg/mL in 0.1M Na Cacodylate pH 6.5, 40% MPD, 5% Peg 8000 only and could not be optimized. StarD9 experiments used a protein concentration of 1mg/mL. Diffraction quality crystals grew in multiple conditions based on the screening condition Pact E9 (0.15-0.2M K/Na tartrate, 18-22% Peg 3350, 0.1M hepes pH 7.2-7.6) in a protein to well solution ratio of 2:1. To obtain larger crystals the readily diffracted, the T3-10 CTD StarD9 Loop Deletion constructs included a 250uL 1:1 silicone:paraffin oil overlay applied to the 500uL well solution.

6.6.5 X-Ray Data Collection and Processing

Diffraction of StarD9 was only obtained with intense synchrotron radiation exposure (100% transmission at Advanced Photon Source beamline 24-ID-C). All data was indexed with XDS/XSCALE,[97] attempts to determine phases were performed with Phaser,[88] MrBump and MRage[142,143] Solution from these programs had refinement attempted with PDB_Redo,[90] Refmac[89] or Phenix[91] and models were manually evaluated in Coot.[98] In no instances was a real solution or partial solution determined as no solution refined in a reasonable manner. Partial solutions had significant steric clashes or physically impossible.

CHAPTER 7

Oligomeric Enzyme Based Materials

7.1 Introduction – Sol-Gels for Protein Encapsulation

It has been established that encapsulation of immobilized enzymes in various material such as Sol-Gels,[20–22] hydrogels[23,153,154] or alternative inorganic scaffolds, results in favorable conditions for large-scale biocatalysts when compared to the same proteins in solution.[155–158] When compared to the identical proteins in solution, encapsulated proteins tend to have increased activity due to substrate channeling,[159] longer enzyme life[21,22,38] and phase separation between the catalytic material and products allowing facile remove of said products.

These observations have spurred research into large proteinaceous self-assembling materials. These would recapitulate properties of encapsulated enzymes without the need for an added matrix to support the proteins. Recent studies have reported such self-assembling two-dimensional materials from protein interface design.[160–162] Here, self-assembling interface must be of a specific geometry for the material to form, and thus is limiting to the subunits that can be used. We sought to create materials using flexible linkages, to form three-dimensional amorphous gels instead of materials with fixed geometries (figure 7.1). To form these gels, oligomeric enzymes would be selected then linked through either GFP dimers (chapter 2) or the heterodimeric coiled-coils previously discussed (chapter 5).[111,121]

7.2 Results – Candidates for Enzymatic Gels

The first candidate protein that was selected was the trimeric bromoperoxidase A2 (Bro) from *Streptomyces aureofaciens* (figure 7.1). This protein was selected as the structure is known (PDB 1BRO)[163] and had previously been expressed as a fusion protein in the Yeates' group.[80] Bromperoxidases have published activity assays[163,164] and some indications that gel encapsulation improves enzyme activity rates.[155,156] This version of Bro lacked native cysteines and had a loop (Asp133-Pro136) amendable to insertion of the split GFP complementation strand 10/11 hairpin (chapter 2 and 3). To create the Bro gels, split GFP dimers presented in chapter 2 would be used with *in vivo* complementation utilized to form the complexes. *In vitro* oxidation of the disulfide bonds would form the amorphous materials. This simple one enzyme system would be used to validate the hypothesis that proteinaceous gels increase enzymatic activity when compared to the proteins in solution when the disulfide bonds are reduced.

In order to test a more complex enzymatic pathway, the Polychlorinated Biphenyl (PCB) Degradation Pathway of BphABCD would be tested. Bioremediation of PCB contaminated soil has been extensively studied.[165–169] Through these approaches, PCB degrading organisms are added to the contaminated soil to degrade the toxic PCBs into non-toxic molecules. We sought to create materials of the PCB degrading enzymes only. These materials could be made and stored long term and then applied to contaminated materials, without the need to store and transport live microorganisms to contaminated soil. Coiled-coils can be used to created the Bph materials since the presences of essential cysteine residues for iron-sulfur clusters makes cysteine cross-linking problematic BphA.[170]

7.2.1 Bromperoxidase – GFP Dimer Materials

Bromperoxidase containing an insertion of the 10/11 hairpin (Bro-10/11) sequence was coexpressed with the five split GFP disulfide mutants (K26C, D102C, D117C, Q157C and D190C) (chapter 2). Out of the five, only three, D117C, Q157C and D190C, was efficiently complemented by the Bro-10/11 construct. Only the Bro-10/11 gene contained a histidine to allow for IMAC purification of the complex. IMAC followed by SEC was performed. The SEC elution profile consisted of a broad peaks, an indication of multiple states of the complex (i.e. complexed to one, two or three GFP beta barrels). The green fractions were pooled and concentrated for oligomerization. Disulfide bonds were oxidized with Cu^{2+} in alkaline conditions. No visible phase separation was visible due to low protein concentration. Oligomer formation was confirmed with non-reducing SDS-PAGE and native gels analysis (figure 7.2).

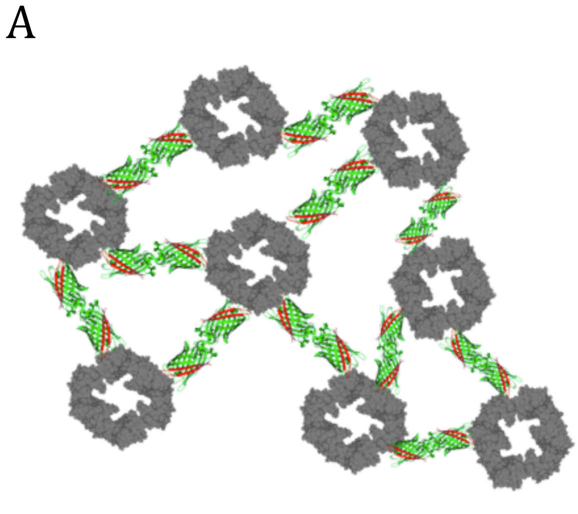


Figure 7.1. Amorphous Proteins Gels Formed Through Oligomer Linkages. (A) Connections of an oligomeric enzyme (grey) through a flexible protein linker (green) would create an amorphous gel-like material. In this example the linkage would be through GFP disulfide dimers. Flexibility comes from both a flexible sequence connecting the two proteins and the rotation of the disulfide bond allowing multiple conformations. (B) The crystal structure of the trimeric bromperoxidase 2A (PDB 1BRO). This protein lacks native cysteines and contains a permissive loop for insertion of GFP strands 10 & 11, making this an ideal candidate for material formation via split GFP dimers.

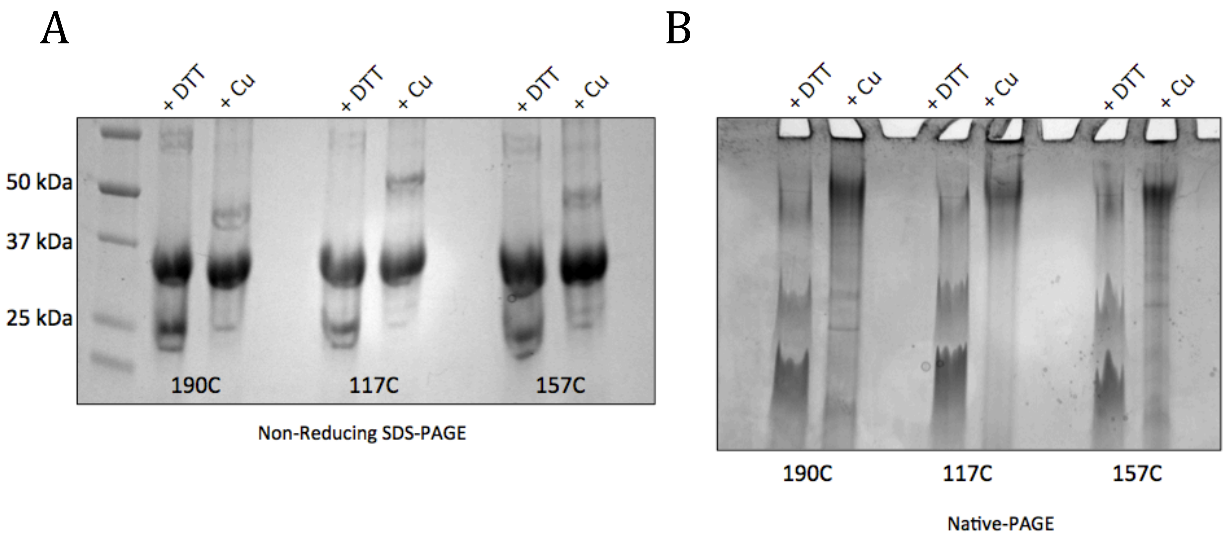

A
+DTT
+Cu
+DTT
+Cu
+DTT
+Cu
50 kDa
37 kDa
25 kDa
190C
117C
157C
Non-Reducing SDS-PAGE
B
+DTT
+Cu
+DTT
+Cu
+DTT
+Cu
190C
117C
157C
Native-PAGE

Figure 7.2. Formation of Large Bromperoxidase – GFP Oligomers. (A) Non-reducing SDS-PAGE gels of purified bromperoxidase-GFP complexes that have been reduced and oxidized. When oxidized the ~25 kDa monomeric GFP band is converted to the higher molecular weight (~45 – 50 kDa) band indicating dimer formation. The major band in each lane at 30 kDa corresponds to the bromperoxidase. For each complex there was incomplete complementation as seen by the drastic difference in band intensity between the two proteins. (B) Native-PAGE gel of the bromperoxidase-GFP complexes that have been reduced and oxidized. The three bands in the reduced (+DTT) lanes correspond to bromperoxidase trimers with differing numbers of complexed GFP. When the disulfide bonds are oxidized, large complexes are formed that fail to enter the resolving layer of the pre-cast gel.

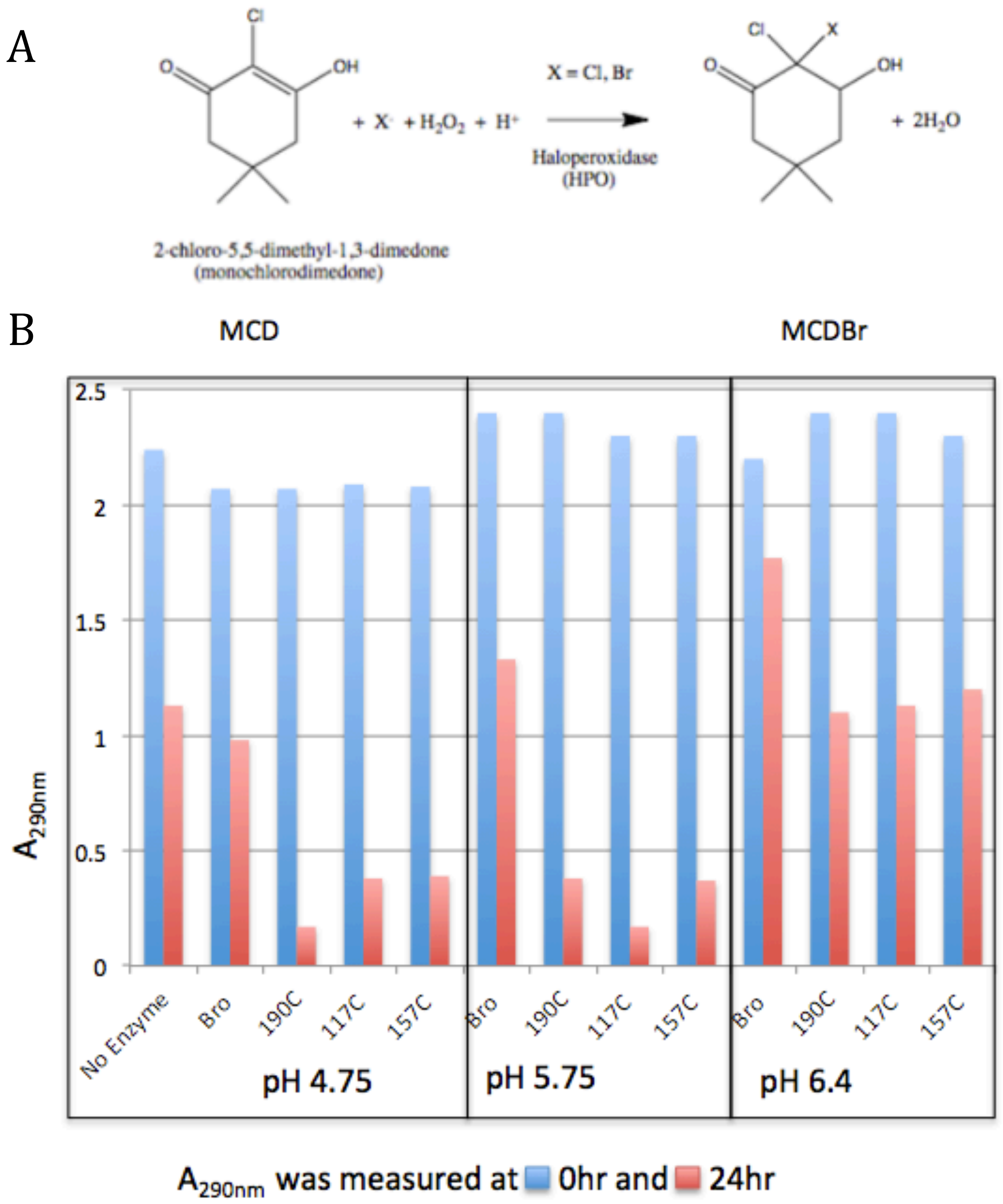


Figure 7.3. Monochlorodimedone Bromperoxidase Assay and pH Optimization. (A) Bromperoxidase activity can be assayed by monitoring the conversion of monochlorodimedone to the brominated form with a corresponding loss of absorbance at 290nm. (B) Screening of pH variations indicated that the bromperoxidase-GFP fusions were most active at pH 4.75 or 5.75. At pH 4.75 protein precipitation was observed leading to pH 5.75 to be used for all future assays.

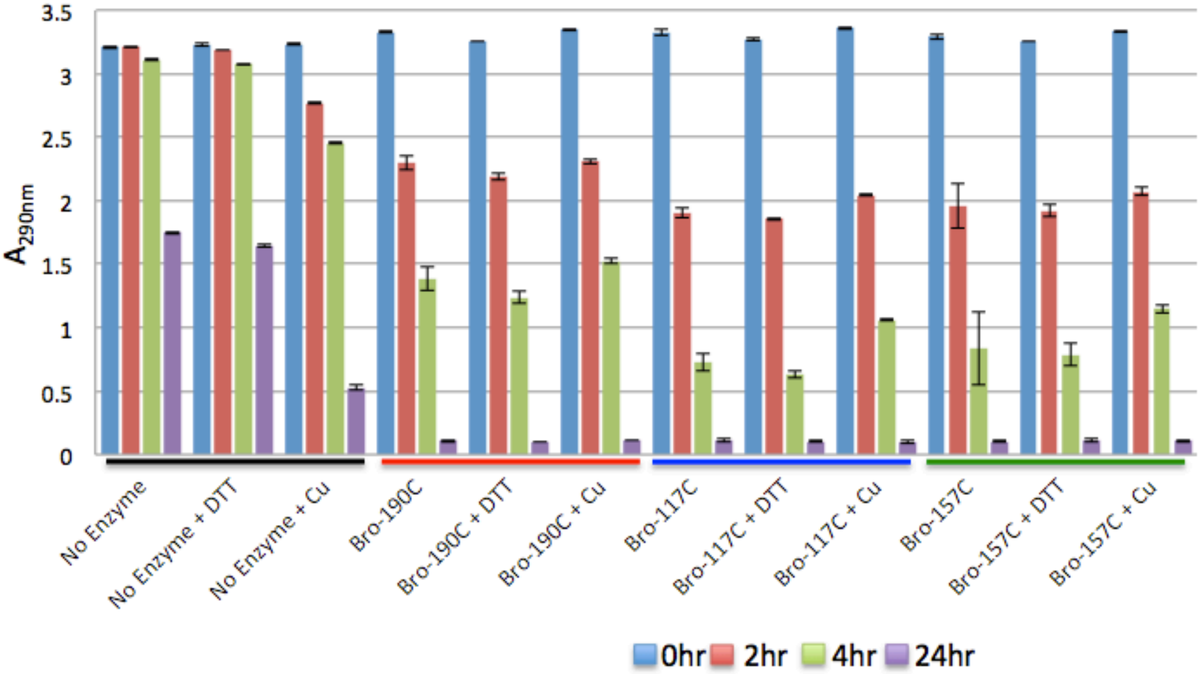


Figure 7.4. Bromperoxidase-GFP Activity Assays. Assays for each complex was performed at pH 5.75 in triplicate and compared to a control with no enzyme present as monochlorodimedone can be brominated when bromine ion are present. Each assay was performed with protein from SEC purified protein was reduced during the purification. Additional DTT was added to fully reduce any disulfides that may have inadvertently formed; Cu^{2+} was added to oxidize the disulfides. For each complex the oxidized versions had reduced activity after four hours. After a full 24 hours every protein sample fully brominated the monochlorodimedone. Error bars represent the standard deviation of the three experiments.

7.2.2 Bromperoxidase Activity Assays

Bromperoxidases are a class of enzymes with a validated activity assay when the substrate monochlorodimedone (MCD) is brominated (MCDBr). Activity can be monitored by loss of absorbance at 290nm (figure 7.3).[171–175] Each bromperoxidase has a pH preference for optimal activity. The Bro-GFP complexes were screened at three pHs, 4.75, 5.75 and 6.4. Similar activity was observed for both pH 4.75 and 5.75 which was increased over pH 6.4 (figure 7.3). When protein was added to the reaction at pH 4.75, minor protein precipitation was observed. Thus all future assays were performed at pH 5.75. Activity for the Bro-GFP complexes in oxidized and reduced forms for all three GFP dimers was assayed (figure 74.). In all instances the oxidized form of the protein showed reduced activity after four hours. All proteins assays showed complete bromination of MCD after 24 hours. Further efforts were not undertaken to optimize this assay.

7.3 Polychlorinated Biphenyl Degradation Pathway BphABCD

Crystal structures for all four enzymes from the BphABCD PCB degradation pathway have been solved[170,176–178] (figure 7.5). To form these materials the coiled-coils discussed in detail in chapter 5 were fused to the either the N- or C-terminus of one of the Bph proteins (table 7.1). In addition to the crystal structures, activity assays have been published for all four enzymes in the pathway.[170,177–180] Unfortunately, only the substrate for the enzymes BphA and BphC are commercially available, necessitating pairs of proteins to be used and two enzymes to be assayed together to fully validate the pathways. The required pairs would be BphA&B and BphC&D, for each pair one protein would have the N-terminal and the other the C-terminal coiled-coil helix genetically fused. Crystal structures were evaluated to choose the most favorable termini to prevent clashes (figures 7.6 and 7.7). It would also be possible to assay the entire pathway at once if all four proteins were mixed *in vitro* and the substrate for BphA added.

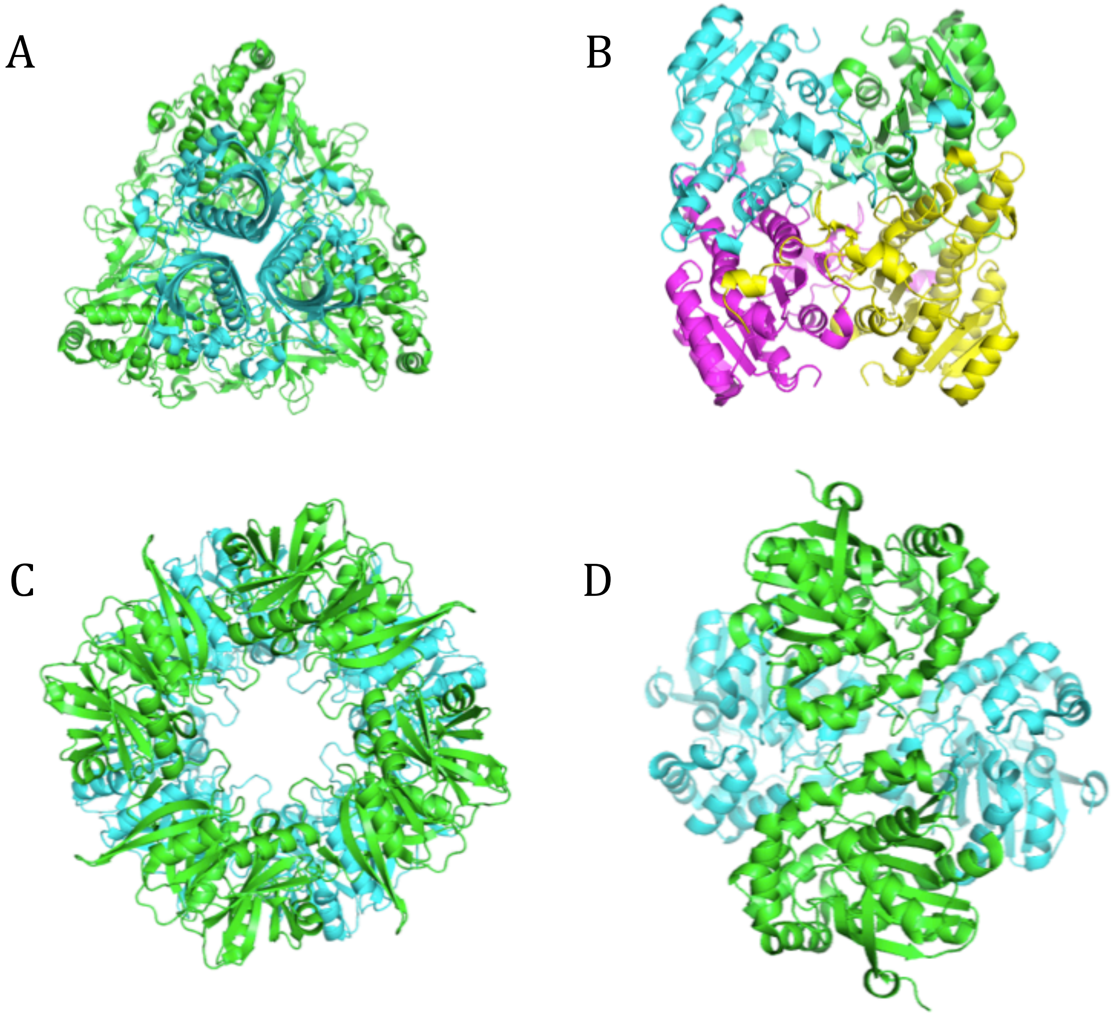


Figure 7.5. Crystal Structures of the Four Bph Enzymes. (A) Structure of BphA from *Burkholderia* (PDB 2XSH) a hexamer with C3 symmetry. (B) Structure of BphB from *Pseudomonas* (PDB IBDB) a tetramer with D2 symmetry. (C) Structure of BphC from *Burkholderia* (PDB 1LGT) an octamers with D4 symmetry. (D) Structure of BphD from *Burkholderia* (PDB 2OG1) a tetramer with D2 symmetry.

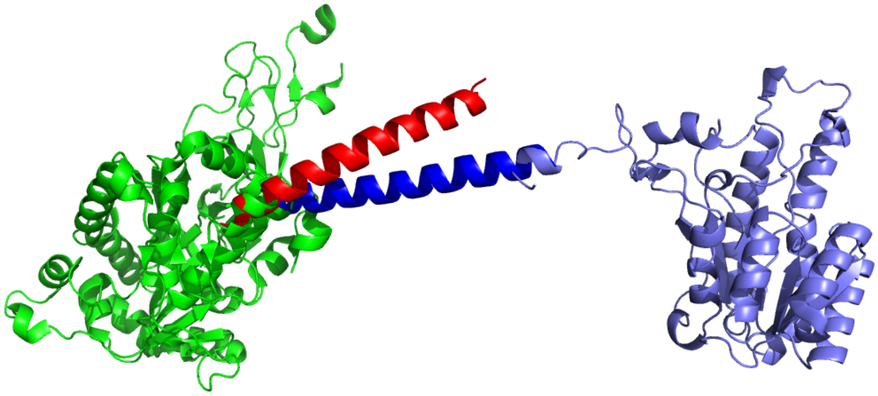


Figure 7.6. Model of the BphA-BphB Coiled-Coil Pair. Model of the coiled-coil linkage between monomers of BphA (green) and BphB (purple). The N-terminal coiled-coil helix is indicated in red and the corresponding C-terminal helix is modeled in blue.

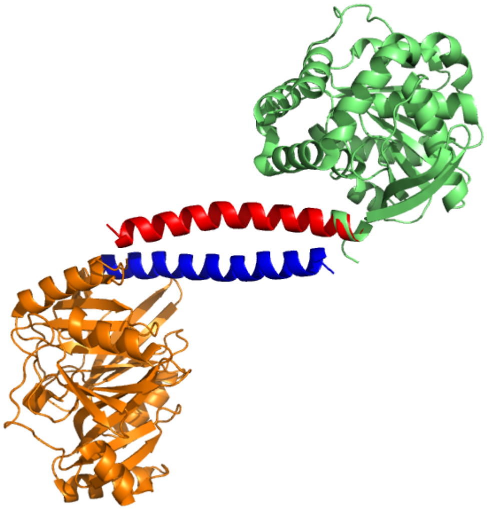


Figure 7.7. Model of the BphC-BphD Coiled-Coil Pair. Model of the coiled-coil linkage between monomers of BphC (orange) and BphD (green). The N-terminal coiled-coil helix is indicated in red and the corresponding C-terminal helix is modeled in blue.

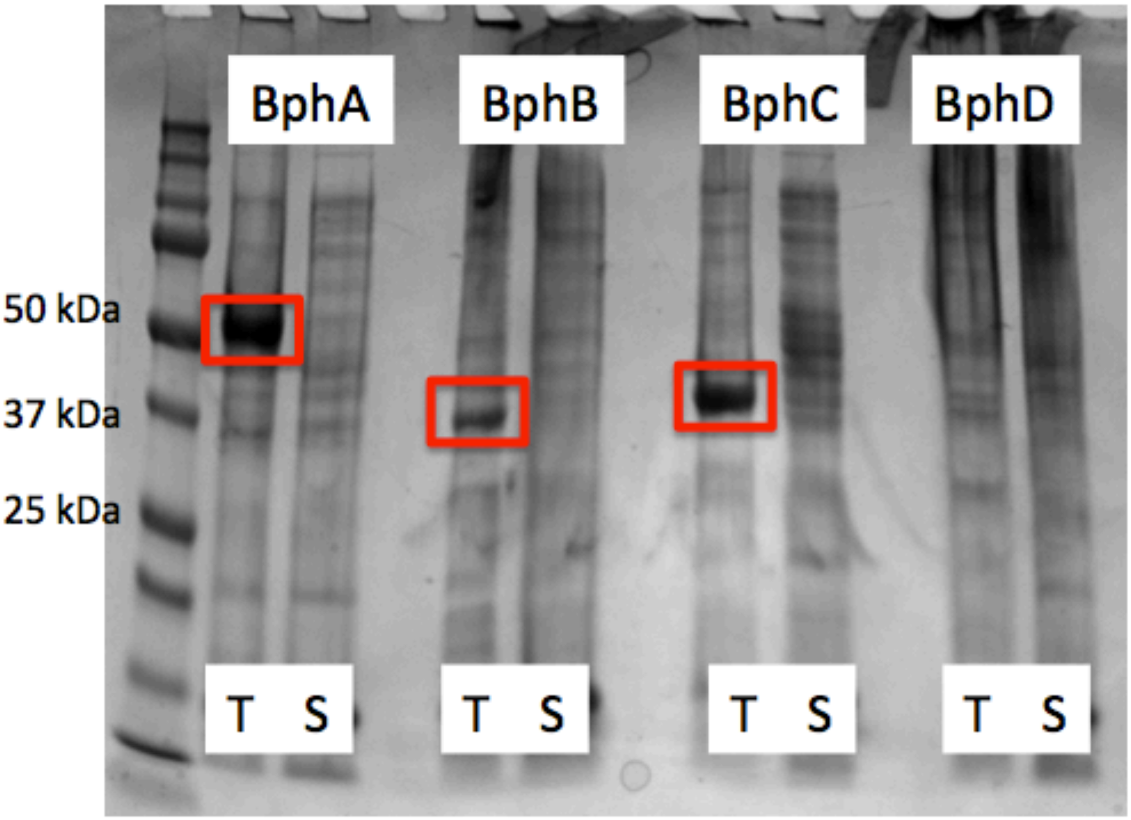

BphA
BphB
BphC
BphD
50 kDa
37 kDa
25 kDa
T S
T S
T S
T S

Figure 7.8. SDS-PAGE Gel from Bph Coiled-Coil Fusion Expression Testing. Coiled-coil fusions of BphA, BphB and BphC were overexpressed in high levels (red box) in the total cell lysate (T lanes) but totally absent from the soluble fraction (S lanes). Lack of an overexpression band for BphD suggests that the construct failed to express.

7.3.1 Bph Enzyme Constructs

To design the Bph constructs, crystal structures were examined and one terminus was chosen for each enzyme for the coiled-coil helical fusion (table 7.1). Residues that were flexible strands in the structures or missing from the structures were removed if they were on the terminus of the fusion. BhpA contains iron sulfur clusters and therefore it is not possible to form linkages to BphB via GFP dimers, as was the case with the bromperoxidase. BphC and BphD contained cysteines that were not essential for enzyme activity; these cysteines were removed from the genes to allow for future experiments with GFP disulfide dimer mediated materials. After the genes were designed, the sequences were codon optimized for expression in *E. coli* and synthesized.

7.3.2 Bph Expression Testing

All four Bph constructs were expressed at low temperature (15°C) and lysed in conditions optimized for the coiled-coil constructs (chapter 5). BphA, BphB and BphC were all well expressed but totally insoluble, BphD overexpression could not be differentiate from endogenous *E. coli* proteins bands (figure 7.8). Due to the difficulty in optimizing the coiled-coil constructs for solubility (unpublished work) no further efforts were undertaken for the coiled coil constructs.

7.4 Discussion

7.4.1 Structure Hypothesis for Lack of Bromperoxidase Activity

The reduction in enzyme activity was surprising, and the opposite of what was intended. One possible explanation is due to structural changes in the bromperoxidase when the large oligomers are formed. The location of the 10/11 hairpin insertion is immediately after a helix that adjacent to the substrate channel (figure 7.9). Movement of this helix could partially block substrate access to the active site, in turn slowing diffusion of substrate into the active site. This would be prevented by changing the location of the hairpin strands to the C-terminus of the bromperoxidase.

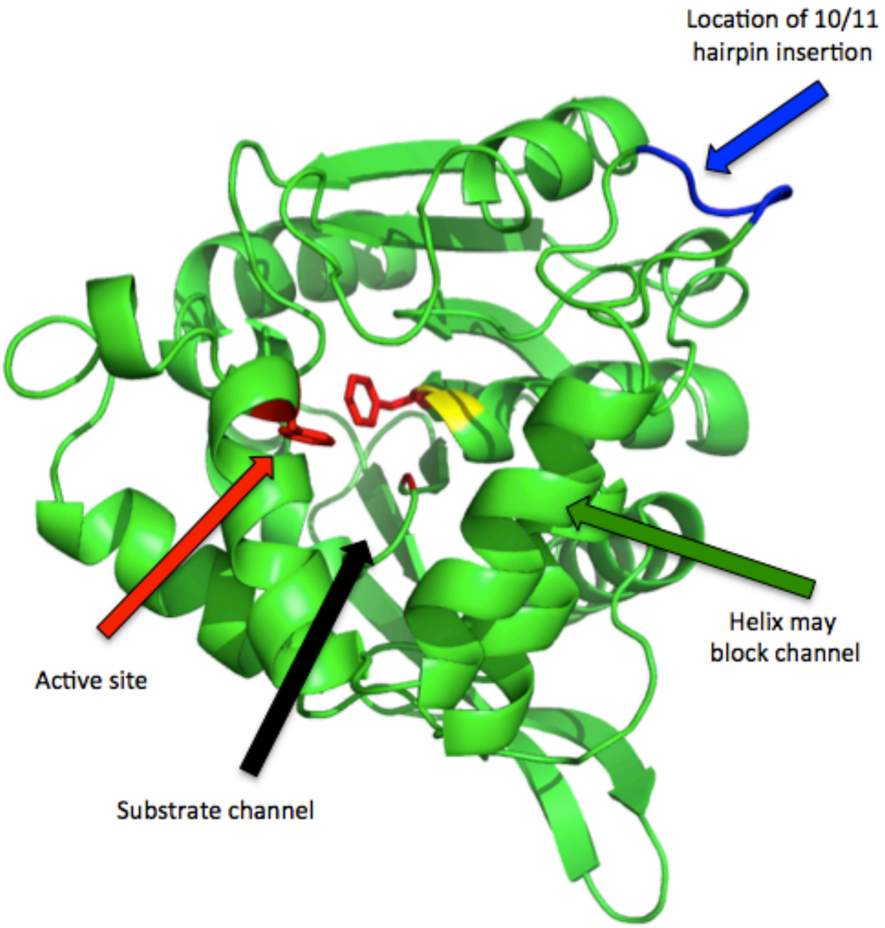


Figure 7.9. Structure of Bromoperoxidase Active Site and Substrate Channel. The crystal structure of one subunit from the bromperoxidase trimer used in the protein gel experiments (PDB 1BRO). The active site is highlighted in red; the substrate channel is indicated with the black arrow. The loop where the 10/11 GFP strands were inserted to mediate the linkage to split GFP is highlighted in blue. This loop is connected to a helix (green arrow) perturbation of this helix could partially occlude the substrate channel, reducing the apparent activity of the enzyme. Insertion of the GFP strands into the bromperoxidase sequence may have moved the helix but the effect was most apparent when the GFP cysteines were oxidized to form dimers, changing the oligomeric state of the complex in solution and potentially perturbing the helix.

7.4.2 Optimization of Bromperoxidase Materials With Multivalent Thiol Cross-linkers

An alternative means of linking the Bro-GFP complexes besides disulfide bonds would be through cysteine specific cross-linkers. This approach has been shown to work for the formation of encapsulating gels via maleimide cross-linking reagents.[154] Variations in the valency and cross-linker length would modify the physical properties of the gel. These could be rapidly screened with existing protein complexes to determine a gel composition that is most efficient as a bioactive gel. The decreased activity of the oxidized samples could originate from slowed diffusion of the substrates through the gel-like material and not from structure changes, leading to occlusion of the substrate channel. Material formation though longer cross-linking agents would create a more porous material allowing more efficient diffusion of the substrate into the material and products out.

Decreased activity in the oxidized, gel-like forms of enzymes is not necessarily detrimental to this approach. Increased enzyme stability, longevity and ease of product separation afforded by the gel formation can overcome reduced activity rates. These factors need to be evaluated in future experiments before concluded if this entirely protein-based gel formation is viable for *in vitro* enzymatic catalysis.

7.4.3 Modifications of Bph Enzymes for Soluble Expression

While the bromperoxidase-GFP gels would serve as a proof-of-concept, the practicality of the system is limited. However, bioremediation remains an import area of research for the removal of PCB contamination. Thus, further efforts need to be undertaken to modify the Bph enzymatic gel formation. Insoluble protein expression and protein precipitation have remained prevalent roadblocks with the fusion of the coiled-coil sequences. Fusion of alternate coiled-coils[117–120] would be a simple rescue approach. These sequences have not been extensively tested as protein fusion and may result in similar solubility issues, given the nature of these helices.

The essential cysteine residues of BphA prevent facile fusion to disulfide dimer mutants suites of GFP[42] or T4L.[6] Metal-mediated oligomer formation is tempting but is highly dependent of buffer conditions and protein identity. Direct fusion of Bph enzyme subunits to each other (i.e. BphA to BphB) through helix fusions[81] is one rescue approach. However, this would results in *in vivo* material formation. Being that these materials are not in a specific closed geometry (like a cage) very large complexes would be formed that would necessitate significant optimization of the purification methods.

In vitro material formation remains the desired approach to make these protein gels as it eliminate the need to isolate and purify the large pre-formed gels from *E. coli.* Site specific cross-linking has been extensively evaluated.[181–185] Here, non-native amino acids specific for the cross-linking reagents can be incorporated during recombinant translation, or specific amino acid sequences can be genetically introduced to the Bph genes. The X-ray structured of the enzymes can guide the location of these sites. This approach is applicable for the formation of gels of oligomeric enzymes, but also could allow the *in vitro* attachment of monomeric proteins to pre-formed scaffolds.

The bulk of the effort in creating these proteinaceous materials has been focused on reliable means of creating the material out of new proteins instead of exploiting existing materials. Instead of creating materials out of each enzymatic pathway that is being studied, pre-formed scaffolds should be exploited. These scaffolds could be two-dimensional sheets or porous three-dimensional gels. Once, stable and reliable materials are formed any enzyme could then be attached to it, eliminating the need for oligomeric proteins to be used. Enzymes could be attached though the site specific cross-linking or the use of split protein complementation. With split-protein complementation one portion of the system would be attached to the scaffold and the complement to the enzyme. Many of these systems have been developed including split-GFP[12] and the more robust Spy-Tag/Spy-Catcher[186] and split venus[187] systems.

7.5 Conclusions

The initial success of the Bro-GFP large oligomer formation was offset by unexpectedly low activity of the enzyme. Although, the decrease in activity was not was surprising it is a clear indication that the entirely proteinaceous, amorphous material made does in fact have an affect on incorporated enzymes. Before a definite conclusion can be made for the viability of this approach, the stability and ease of product extraction must be characterized. However, these initial results suggest that this is indeed a means of making bioactive materials. This potential will only be improved as universal scaffolds are for designed to create these materials.

676 Materials and Methods

7.6.1 Cloning

All genes were synthesized as gBlocks and cloned via Gibson assembly per the methods described in chapter 5. The bromperoxidase gene from PDB deposition 1BRO had the loop Asp133-Pro136 replaced with the split-GFP 10/11-strand sequence to allow complementation to GFP cores. The GFP cores used are identical to those discussed in chapter 3. Sequences of the modified Bph enzymes for coiled-coil fusions are presented in table 7.1. In all cases the genes were codon optimized[95] for expression in *E. coli.* Appropriate plasmids were transformed into BL21 DE3 expression cells.

7.6.2 Protein Expression

All constructs were expressed in LB supplemented ampicillin for the Bph contructs, or ampicillin and chloramphenicol for the Bro-GFP coexpressions. The 1L LB flasks were inoculated with 10mL of overnight culture and allow to grow for at 37°C until the A_{600nm} optical density reached 0.6-0.8, the temperature was then reduced to 15°C and protein expression was induced with 1mM IPTG and grown for 16 hours.

7.6.3 Protein Purification

Bro-GFP expression that resulted in a green pellet (indication of expression and complementation) we used for purifications. Protein was purified per previously published methods[79] with the exception of 1mM DTT used as the reducing agent in the SEC purifications. The protein eluted in a broad peak indicating several forms present, therefore all fractions that were green were pooled for activity assays, protein was concentrated to >10mg/mL and used immediately for activity assays. Bph coiled-coil proteins were lysed for expression testing in a buffer composed of 50mM Hepes pH 7.0, 500mL NaCl, 1mM $MgCl_2$, 10% w/v glycerol, 0.5 w/v triton X-100, with additions of lyophilized lysozyme and DNase1. Cells were then lysed via sonication, insoluble material was removed by centrifugation at 5000 xg for 15 minutes. Samples from the total and soluble cell lysates were analyzed via SDS-PAGE.

7.6.4 Bromperoxidase Material Formation

Aliquots of SEC purified Bro-GFP complexes were used for material formation. Protein was buffer exchanged into the oxidation buffer 100mM Tris pH 9.0, 100mM NaCl, 5mM CuSO, and incubated at room temperature for 30 minutes. The reaction was quenched by addition of 50mM EDTA, the protein was then buffer exchanged into 20mM Tris pH 8.0, 100mM NaCl.

7.6.5 Bromperoxidase Activity Assays

Conditions for the bromperoxidase assay were based upon the Sigma-Aldrich assay library conditions[188] (http://www.sigmaaldrich.com/life-science/metabolomics/enzyme-explorer/learning-center/assay-library.html) with the exception of 50mM acetate pH 5.75 serving as the assay buffer and performed at room temperature, these conditions were consistent with many of the published methods. Absorbance was measured at 0, 2, 4 and 24 hours.

Table 7.1. Bph Enzyme Coiled-Coil Expression Constructs.

BphA (PDB 2XSH)
Full-length Protein
MSSAIKEVQGAPVKWVTNWTPEAIRGLVDQEKGLLDPRIYADQSLYELELERVFGRSWLLLGHE
SHVPETGDFLATYMGEDPVVMVRQKDKSIKVFLNQCRHRGMRICRSDAGNAKAFTCSYHGWAYD
IAGKLVNVPFEKEAFCDKKEGDCGFDKAEWGPLQARVATYKGLVFANWDVQAPDLETYLGDARP
YMDVMLDRTPAGTVAIGGMQKWVIPCNWKFAAEQFCSDMYHAGTTTHLSGILAGIPPEMDLSQA
QIPTKGNQFRAAWGGHGSGWYVDEPGSLLAVMGPKVTQYWTEGPAAELAEQRLGHTGMPVRRMV
GQHMTIFPTCSFLPAMNNIRIWHPRGPNEIEVWAFTLVDADAPAEIKEEYRRHNIRNFSAGGVF
EQDDGENWVEIQKGLRGYKAKSQPLNAQMGLGRSQTGHPDFPGNVGYVYAEEAARGMYHHWMRM
MSEPSWATLKP

Coiled-coil Helical Fusion Construct
MGSDKIHHHHHHENLYFQGAQLKKKLQALKKKNAQLKWKLQALKKKLAQPEAIRGLVDQEKGLLD
PRIYADQSLYELELERVFGRSWLLLGHESHVPETGDFLATYMGEDPVVMVRQKDKSIKVFLNQCR
HRGMRICRSDAGNAKAFTCSYHGWAYDIAGKLVNVPFEKEAFCDKKEGDCGFDKAEWGPLQARVA
TYKGLVFANWDVQAPDLETYLGDARPYMDVMLDRTPAGTVAIGGMQKWVIPCNWKFAAEQFCSDM
YHAGTTTHLSGILAGIPPEMDLSQAQIPTKGNQFRAAWGGHGSGWYVDEPGSLLAVMGPKVTQYW
TEGPAAELAEQRLGHTGMPVRRMVGQHMTIFPTCSFLPAMNNIRIWHPRGPNEIEVWAFTLVDAD
APAEIKEEYRRHNIRNFSAGGVFEQDDGENWVEIQKGLRGYKAKSQPLNAQMGLGRSQTGHPDFP
GNVGYVYAEEAARGMYHH

BphB (PDB 1BDB)
Full-length Protein
MKLKGEAVLITGGASGLGRALVDRFVAEGAKVAVLDKSAERLAELETDHGDNVLGIVGDVRSLE
DQKQAASRCVARFGKIDTLIPNAGIWDYSTALVDLPEESLDAAFDEVFHINVKGYIHAVKACLP
ALVASRGNVIFTISNAGFYPNGGGPLYTAAKHAIVGLVRELAFELAPYVRVNGVGSGGINSDLR
GPSSLGMGSKAISTVPLADMLKSVLPIGRMPEVEEYTGAYVFFATRGDAAPATGALLNYDGGLG
VRGFFSGAGGNDLLEQLNIHP

Coiled-coil Helical Fusion Construct
MGSDKIHHHHHHENLYFQGMKLKGEAVLITGGASGLGRALVDRFVAEGAKVAVLDKSAERLAELE
TDHGDNVLGIVGDVRSLEDQKQAASRCVARFGKIDTLIPNAGIWDYSTALVDLPEESLDAAFDEV
FHINVKGYIHAVKACLPALVASRGNVIFTISNAGFYPNGGGPLYTAAKHAIVGLVRELAFELAPY
VRVNGVGSGGINSDLRGPSSLGMGSKAISTVPLADMLKSVLPIGRMPEVEEYTGAYVFFATRGDA
APATGALLNYDGGLGVRGFFSGAGGNDLLEQQLEKELQALEKENAQLEWELQALEKELAQ

BphC (PDB 1LGT)
Full-length Protein
SIRSLGYMGFAVSDVAAWRSFLTQKLGLMEAGTTDNGDLFRIDSRAWRIAVQQGEVDDLAFAGY
EVADAAGLAQMADKLKQAGIAVTTGDASLARRRGVTGLITFADPFGLPLEIYYGASEVFEKPFL
PGAAVSGFLTGEQGLGHFVRSVPDSDKALAFYTDVLGFQLSDVIDMKMGPDVTVPAYFLHSNER
HHTLAIAAFPLPKRIHHFMLEVASLDDVGFAFDRVDADGLITSTLGRHTNDHMVSFYASTPSGV
EVEYGWSARTVDRSWVVVRHDSPSMWGHKSVRDKAAARNKA

Coiled-coil Helical Fusion Construct
MGSDKIHHHHHHENLYFQGSIRSLGYMGFAVSDVAAWRSFLTQKLGLMEAGTTDNGDLFRIDSR
AWRIAVQQGEVDDLAFAGYEVADAAGLAQMADKLKQAGIAVTTGDASLARRRGVTGLITFADPF
GLPLEIYYGASEVFEKPFLPGAAVSGFLTGEQGLGHFVRSVPDSDKALAFYTDVLGFQLSDVID
MKMGPDVTVPAYFLHSNERHHTLAIAAFPLPKRIHHFMLEVASLDDVGFAFDRVDADGLITSTL
GRHTNDHMVSFYASTPSGVEVEYGWSARTVDRSWVVVRHDSPSMWGHKSVRDKAAARNKAQLEK
ELQALEKENAQLEWELQALEKELAQ

BphD (PDB 2OG1)
Full-length Protein
MTALTESSTSKFVKINEKGFSDFNIHYNEAGNGETVIMLHGGGPGAGGWSNYYRNVGPFVDAGY
RVILKDSPGFNKSDAVVMDEQRGLVNARAVKGLMDALDIDRAHLVGNSMGGATALNFALEYPDR
IGKLILMGPGGLGPSMFAPMPMEGIKLLFKLYAEPSYETLKQMLQVFLYDQSLITEELLQGRWE
AIQRQPEHLKNFLISAQKAPLSTWDVTARLGEIKAKTFITWGRDDRFVPLDHGLKLLWNIDDAR
LHVFSKSGHWAQWEHADEFNRLVIDFLRHA

Coiled-coil Helical Fusion Construct
MGSDKIHHHHHHENLYFQGAQLKKKLQALKKKNAQLKWKLQALKKKLAQESSTSKFVKINEKGF
SDFNIHYNEAGNGETVIMLHGGGPGAGGWSNYYRNVGPFVDAGYRVILKDSPGFNKSDAVVMDE
QRGLVNARAVKGLMDALDIDRAHLVGNSMGGATALNFALEYPDRIGKLILMGPGGLGPSMFAPM
PMEGIKLLFKLYAEPSYETLKQMLQVFLYDQSLITEELLQGRWEAIQRQPEHLKNFLISAQKAP
LSTWDVTARLGEIKAKTFITWGRDDRFVPLDHGLKLLWNIDDARLHVFSKSGHWAQWEHADEFN
RLVIDFLRHA

Chapter 8

Circular Permutations of PduA

8.1 Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein

Reused with permission from Royal Society of Chemistry (no license required)

DOI 10.1039/C6CC00851H[184]

Published on 09 March 2016. Downloaded by University of California - Los Angeles on 07/04/2016 19:52:40.

ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2016, **52**, 5041

Received 27th January 2016,
Accepted 9th March 2016

DOI: 10.1039/c6cc00851h

www.rsc.org/chemcomm

Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein†

J. Jorda,[a] D. J. Leibly,[ab] M. C. Thompson[b] and T. O. Yeates*[ab]

We report the crystal structure of a novel 60-subunit dodecahedral cage that results from self-assembly of a re-engineered version of a natural protein (PduA) from the Pdu microcompartment shell. Biophysical data illustrate the dependence of assembly on solution conditions, opening up new applications in microcompartment studies and nanotechnology.

Nature has evolved exquisite symmetrical structures across a range of length scales. Examples of such structures based on self-assembling proteins include virus capsids, clathrin, ferritin and bacterial microcompartments (MCPs).[1–6] These assemblies encapsulate nucleic acids, endocytic cargo, iron and metabolic enzymes, respectively. Two common features of such structures are (1) their assembly from many copies of one or a small number of distinct protein subunit types, and (2) highly symmetric arrangements of the subunits, typically having cubic or icosahedral (or related dodecahedral) forms.

Inspired by nature, an emerging emphasis within the field of bionanotechnology is the design and production of novel three dimensional protein assemblies that might serve as molecular containers. Hollow, nanoscale structures have attracted considerable interest because they have the potential to be engineered for the targeted biological delivery of cargo, including drug molecules and imaging reagents such as dyes and nanoparticles.[7,8] Molecular cages or shells built from protein subunits are privileged platforms for bionanotechnology applications because their properties can be modulated easily by changes to their amino acid sequences, and they can be produced using recombinant overexpression technologies. Efforts aiming at engineering proteins to self-assemble into complex polyhedral cages have led to a series of recent successes.[9–15] Strategies aimed at engineering proteins to form geometrically regular architectures have focused on targets obeying the symmetries of the Platonic solids: tetrahedral, cubic/octahedral and icosahedral/dodecahedral.[16] A series of designed cages based on tetrahedral symmetry (12 copies of one or two distinct subunit types) and cubic/octahedral symmetry (24 copies of one or two distinct subunit types) have been validated in detail with crystallographic studies.[10–12,17,18] This leaves icosahedral/dodecahedral architectures with 60 equivalent subunits – the highest possible cubic point symmetry in three-dimensions – as the ultimate target for designing novel protein cages.

Here we report a dodecahedral cage that self-assembles from 60 copies of a redesigned protein building block from the bacterial microcompartment shell (or BMC) family. The protein PduA is a major component of the shell of the propanediol utilization (Pdu) microcompartment in *Salmonella typhimurium*. As part of its natural function, PduA forms a 6-fold symmetric cyclic hexamer.[19] This hexamer exhibits a shape and chemical self-complementarity at its perimeter that promotes further side-by-side assembly of hexamers to form a tightly packed molecular layer about 2 nm thick, which comprises (along with other related proteins) the outer shell of the bacterial microcompartment.[20,21] In total, roughly five to 15 thousand protein subunits make up the entire shell structure, whose diameter can range from about 100 to 200 nm. We were motivated to radically redesign the PduA protein based on the observation that certain unusual members of the BMC protein family have undergone cyclic permutations during evolution, giving three-dimensional structures that are built from the same arrangement of secondary structure elements, but in a different linear order in the protein sequence.[22] An interesting aspect of the permuted type of BMC protein is that one such protein (EutS) was revealed earlier to be unusually flexible, assembling into a bent rather than a flat hexamer.[23] We therefore began with the PduA protein sequence and converted its topology to match that of the most closely related BMC protein known to have a permuted topology, PduU.[22]

The circular permutation of a protein is a topological reorganization of its sequence whereby the initial termini are linked (sometimes requiring a short intervening polypeptide), while new termini are created by a disconnection elsewhere in the sequence.

---

[a] *UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA. E-mail: yeates@mbi.ucla.edu*

[b] *Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6cc00851h

Published on 09 March 2016. Downloaded by University of California - Los Angeles on 07/04/2016 19:52:40.

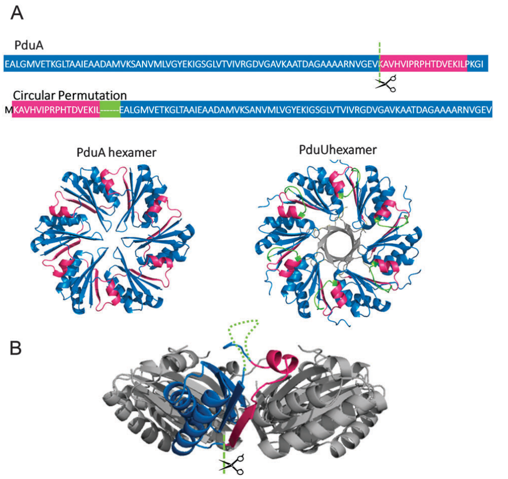

A
PduA
EALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIGSGLVTVIVRGDVGAVKAATDAGAAAARNVGEVKAVHVIPRPHTDVEKILPKGI
Circular Permutation
MKAVHVIPRPHTDVEKIL------EALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIGSGLVTVIVRGDVGAVKAATDAGAAAARNVGEV
PduA hexamer
PduUhexamer
B

Fig. 1 A redesign of the PduA sequence by circular permutation. The C-terminal segment of the BMC domain (pink) was appended to the N-terminus (blue) with a linking loop extracted from the corresponding PduU sequence, depicted in green (A). A rotation by 90° of the PduA hexameric crystal structure, illustrating in green dots where the new linker was introduced to join the termini of native PduA, and with a scissor symbol where the new termini were created (B).

Our designed circular permutation of PduA borrowed from the PduU topology, but kept the native PduA sequence wherever possible. The PduA termini were connected with the same linker sequence as seen in PduU, while new chain termini were introduced in the location where they occur in PduU (Fig. 1). The feasibility of creating a circularly-permuted PduA construct was first evaluated computationally. We created a permuted version of the protein coordinates by removing a C-terminal segment of PduA, and appending it to the N-terminus, grafting a peptide linker taken from PduU. The structure of this chimeric protein was subsequently refined with the program Modeller.[24] Out of 100 independent computer runs, the model with the best energy (DOPE) score[25] was defined as the starting model (named P1). Three additional variations on the design were considered and evaluated computationally. Design variations P2 and P3 contained amino acid sequence changes suggested by the Rosetta Matdes program.[11] A final design variation, P4, featured a short linker sequence, GGSGGS, chosen for high flexibility. Full protein sequences are given in ESI,† Table S1. To assess the quality of these models, the Rosetta energy scores of the designed hexamers were calculated after relaxing the strict symmetry constraints in a custom protocol[11] followed by a geometry validation step by the ramalyze and rotalyze routines in the program PHENIX.[26] Final Rosetta scores for each model were −1225, −1636, −1658 and −1850 Rosetta Energy Units for P1, P2, P3 and P4, respectively. Unexpectedly, the P4 design with the empirically chosen glycine and serine linker was predicted to be the most stable design.

The P1–P4 protein constructs were created *via* gene synthesis and the amplicons were cloned into the pET-22b expression vector *via* Gibson assembly.[27] Protein expression was carried out in *Escherichia coli* BL21 cells, and recombinant proteins were purified using metal affinity chromatography, facilitated by inclusion of a hexahistidine tag in the protein sequence, which was subsequently removed by treatment with TEV protease. A final gel filtration step resulted in pure protein samples (ESI,† Fig. S1). A major peak with an estimated molecular weight of 45 kDa was collected and concentrated to ∼10 mg ml^{-1} in 50 mM tris pH 9, 50 mM NaCl for each protein.

Crystallization trials were conducted by hanging drop vapor diffusion on the four variations on the designed protein. In each instance initial screening was performed at a protein concentration of 5 mg ml^{-1} with up to five commercially available sparse matrix screens. Design variation P4 was the only case that gave crystals readily. Conditions for crystal growth were subsequently optimized, and high-quality crystals were obtained by diluting the protein to 2.5 mg ml^{-1} and crystallizing by vapor diffusion against a well solution of 0.1 M tris pH 8.5, 1.8 M ammonium sulphate, and 1.25% w/v PEG-10 000. Crystals grew in 1 week, after which they were soaked in 25% 1,2-propanediol as a cryoprotectant, and flash frozen in liquid nitrogen. X-ray diffraction data were collected at the Advanced Photon Source (NE-CAT beamline 24-ID-C). X-ray diffraction extended to a resolution of 2.5 Å, and the data were reduced in space group $P4_232$ using the program XDS/XSCALE.[28] The program PHASER[29] was used to obtain phases by molecular replacement with a PduA monomer (PDB 4PPD[20]) serving as the search model. The molecular replacement solution revealed five monomers in the asymmetric unit, and a high solvent content of 56%. Structural refinement was performed by iterative rounds of model adjustment and refinement using Coot[30] and PHENIX[26] respectively. Structure validation was performed with PHENIX. The refined atomic coordinates and structure factors were deposited in the PDB under accession code 5HPN.

Surprisingly, the crystal structure of the redesigned, circularly-permuted PduA revealed a dodecahedral cage made of 60 copies of the protein subunit (Fig. 2A). The asymmetric unit of the crystal contains a single pentamer. The crystallographic symmetry operators then produce a protein assembly with icosahedral point group symmetry. Each pentameric unit constitutes one of the 12 faces of a dodecahedron (Fig. 2A). This polyhedral assembly is 13 nm in diameter and encloses an inner space with a diameter of about 7 nm and a volume of approximately 180 nm^3. This 60 subunit cage is the first reported structure of a novel synthetic protein complex with icosahedral symmetry, though a number of new icosahedral protein architectures created by design have been obtained in recent work (Jacob Bale, Neil King, and David Baker, unpublished data).

The formation of this dodecahedral protein cage results from two structural changes introduced by the circular permutation. First, there is a decisive alteration in the primary oligomerization state of the protein. A switch from the typical cyclic 6-fold hexameric arrangement observed for native BMC-family proteins to a 5-fold pentameric arrangement is critical to the architecture observed; icosahedral point group symmetry requires pentagonal units with 5-fold symmetry. Interestingly, this switch occurs in

This journal is © The Royal Society of Chemistry 2016

Published on 09 March 2016. Downloaded by University of California - Los Angeles on 07/04/2016 19:52:40.

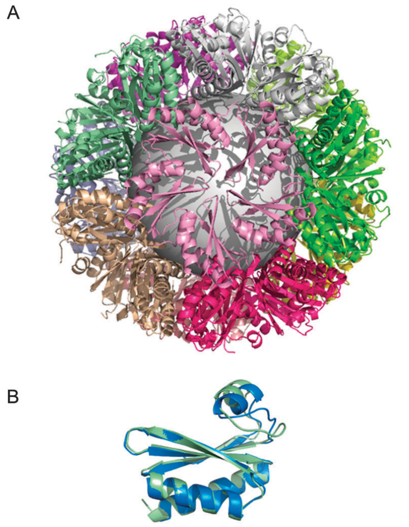


Fig. 2 Crystal structure of the circularly permuted PduA protein (design variant P4). (A) The unit cell of the crystal contains a cage of 12 self-assembling pentamers in the approximate shape of a dodecahedron with icosahedral symmetry, similar to those found in simple viral capsids. (B) A detailed look at the structure of a single protein subunit shows that the computational design (pale green) is in close agreement with the crystal structure (blue), but with notable differences in a loop region and the position of a short alpha helical segment.

the absence of mutations in the protein–protein interface that is responsible for the cyclic oligomerization; mutations near the subunit interface were avoided during the design stage. With the pentagonal shape forbidding a flat tessellation, but with the altered oligomeric unit still evidently exhibiting a tendency to self-associate further, the result is a novel dodecahedrally shaped structure with icosahedral symmetry.

Despite the dramatic architectural alterations evident in the dodecahedral cage, many of the features believed to be characteristic of bacterial microcompartment shells are recapitulated. The self-assembled pentamers form a tight, almost seamless interface. Likewise, the pore at the center of the pentamer is still present, although its diameter is reduced from roughly 6 Å in wild-type PduA to about 3.5 Å as a result of forming a smaller cyclic oligomer. Finally, a structural alignment of the monomer observed in the crystal structure with the computationally designed model reveals an overall difference of only 1.1 Å (rmsd). The agreement in the core of the protein domain is even closer; the glycine and serine loop region and a short alpha-helical segment account for most of the deviation (Fig. 2B).

The unexpected observation of the dodecahedral cage in the crystal led us to investigate whether or not that assembled form of the protein was well-populated in solution, a requisite property if the cage is to find utility in various applications in solution. The potentially reversible solution-dependence for such an assembly is also a useful property for cargo delivery and other nanotechnology applications.

In order to analyze the solution behavior of the designed protein, we carried out dynamic light scattering (DLS) experiments at varying pH and salt concentrations to assess a dependence of cage formation on these two factors (Fig. 3). All samples were prepared from the same protein stock solution used for crystallization after exchanging buffers into a 10 mM CHES pH 9, 50 mM NaCl solution. Protein samples were subsequently diluted into DLS buffers to a final concentration of 3.5 mg ml^{-1}. By combining incremental pH values from 6 to 9 and three NaCl concentrations between 50 mM and 500 mM, we obtained 12 different buffer conditions for testing. Results from the DLS experiments (Fig. 3) indicate that self-assembly of pentamers into a dodecahedral species occurs at high yield in solution under specific conditions. The cage dominates (about 95% by mass) around pH 8 and 50 mM NaCl. Conversion appears to occur between the pentameric and dodecahedral states under other conditions, suggesting reversibility of the oligomerization. This is consistent with our earlier observation that the protein elutes at low concentration from a gel filtration column as a pentamer, but assembles into the dodecahedron under specific conditions evaluated by DLS and by crystallization.

Although this study involved a deliberate and dramatic redesign of a protein molecule, the discovery of its highly unusual assembly state was serendipitous. Despite the unexpected route by which this novel protein architecture was obtained, we expect that

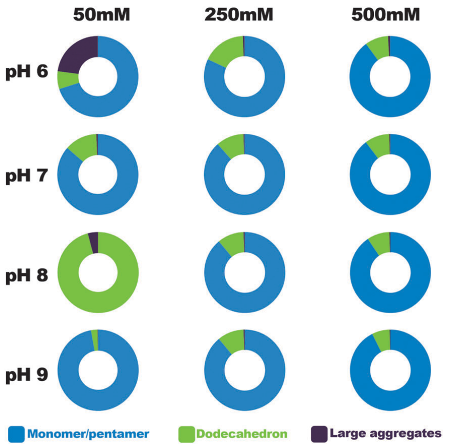


Fig. 3 Distribution of oligomeric forms of the designed protein according to dynamic light scattering experiments at varying pH and NaCl concentrations. Twelve different combinations of pH and salt concentrations are shown. For each condition, the mass percentages of the different oligomeric states are reported. An oligomeric state was defined based on the estimated mass of a detected particle. Mass peaks within the 5–30 kDa range were taken to be either monomers or pentamers, whereas mass peaks between 500 kDa and 700 kDa along with reported radii consistent with the dodecahedral cage structure were taken to be the dodecahedral cage, whose calculated mass and radius are 528 kDa and ~7 nm. Peaks of 40 MDa and above were considered as aggregates.

Published on 09 March 2016. Downloaded by University of California - Los Angeles on 07/04/2016 19:52:40.

it could have applications in the field of bionanotechnology as platform for encapsulation and targeting of various cargos. We note that a particularly distinct property of the cage obtained is the overall tightness of the protein packing and the narrowness of the pores that run between the interior and exterior regions, which would be relevant for delivery applications. This very tight packing evidently derives from the natural shape properties of bacterial microcompartment shell proteins, which in their naturally assembled states form flat, tightly packed layers.

Future experiments aimed at modifying and advancing the utility of this protein cage include structure-based redesigns of the pentamer–pentamer interface to enhance the stability of the dodecahedron. Assessing the stability of these designs over a wider array of buffer conditions would provide finer control of the assembly process. Mutations to the interior and exterior surfaces of the cage could be explored for the purposes of encapsulation and targeting, respectively. Additionally, this protein assembly offers insight into the evolution of symmetric proteins in general, demonstrating that new symmetries can arise unexpectedly. In the present case, the changes in assembly state resulted from sequence permutation rather than from the more familiar scenario of mutations to interfacial regions of the protein subunit.

This work was supported by NSF grant CHE-1332907 and NIH grant AI081146. D. J. L. was supported by Ruth L. Kirschstein National Research Service Award T32GM007185. The authors thank Michael R. Sawaya, Duilio Cascio, and Dan McNamara for X-ray data collection at APS beamline 24-ID-C, and Joshuah Laniado for helpful suggestions. X-ray core facilities at UCLA are supported by DOE Grant DE-FC03-02ER63421. The NECAT beamlines of the Advanced Photon Source are supported by NIH Grant RR-15301 (NCRR). Use of the Advanced Photon Source is supported by the DOE, Office of Basic Energy Sciences, under Contract DE-AC02 06CH11357.

Notes and references

1 J. E. Johnson and J. A. Speir, *J. Mol. Biol.*, 1997, **269**(5), 665–675.
2 D. S. Goodsell and A. J. Olsen, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 105–153.
3 E. D. Levy and S. Teichmann, *Prog. Mol. Biol. Transl. Sci.*, 2013, **117**, 25–51.
4 C. Chowdhury, S. Sinha, S. Chun, T. O. Yeates and T. A. Bobik, *Microbiol. Mol. Biol. Rev.*, 2014, **78**(30), 438–468.
5 T. O. Yeates, C. S. Crowley and S. Tanaka, *Annu. Rev. Biophys.*, 2010, **39**, 185–205.
6 T. O. Yeates, C. A. Kerfeld, S. Heinhorst, G. C. Cannon and J. M. Shively, *Nat. Rev. Microbiol.*, 2008, **6**(9), 681–691.
7 S. Howorka, *J. Mater. Chem.*, 2007, **17**, 2049–2053.
8 S. J. Tsai and T. O. Yeates, *Prog. Mol. Biol. Transl. Sci.*, 2011, **103**, 1–20.
9 Y. T. Lai, D. Cascio and T. O. Yeates, *Science*, 2012, **336**(6085), 1129.
10 N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida, I. Andre, T. Gonen, T. O. Yeates and D. Baker, *Science*, 2012, **336**(6085), 1171–1174.
11 N. P. King, J. B. Bale, W. Sheffler, D. E. McNamara, S. Gonen, T. Gonen, T. O. Yeates and D. Baker, *Nature*, 2014, **510**, 103–108.
12 Y. T. Lai, E. Reading, G. L. Hura, K. L. Tsai, A. Lagonawsky, F. J. Asturias, J. A. Tainer, C. V. Robinson and T. O. Yeates, *Nat. Chem.*, 2014, **6**, 1065–1071.
13 H. Gradisar, S. Bozic, T. Doles, D. Vengust, I. Hafener-Bratkovic, A. Merteji, B. Webb, A. Sali, S. Klavzar and R. Jerala, *Nat. Chem. Biol.*, 2013, **9**(6), 362–366.
14 D. J. Huard, K. M. Kane and F. A. Tezcan, *Nat. Chem. Biol.*, 2013, **9**(3), 169–176.
15 B. Worsdorfer, K. J. Woycechowsky and D. Hilvert, *Science*, 2011, **331**(6017), 589–592.
16 J. E. Padilla, C. Colovos and T. O. Yeates, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**(5), 2217–2221.
17 Y. T. Lai, L. Jiang, W. Chen and T. O. Yeates, *Protein Eng., Des. Sel.*, 2015, **28**(11), 491–499.
18 Y. T. Lai, E. Reading, G. L. Hura, K. L. Tsai, A. Lagonwsky, F. J. Asturias, J. A. Tainer, C. V. Robinson and T. O. Yeates, *Nat. Chem.*, 2014, **6**(12), 1065–1071.
19 C. S. Crowley, D. Cascio, M. R. Sawaya, J. S. Kopstein, T. A. Bobik and T. O. Yeates, *J. Biol. Chem.*, 2010, **285**(48), 37838–37846.
20 S. Sinha, S. Cheng, Y. W. Sung, D. E. McNamara, M. R. Sawaya, T. O. Yeates and T. A. Bobik, *J. Mol. Biol.*, 2014, **426**(12), 2328–2345.
21 A. Pang, S. Frank, I. Brown, M. J. Warren and R. W. Pickersgill, *J. Biol. Chem.*, 2014, **289**(32), 22377–22384.
22 C. S. Crowley, M. R. Sawaya, T. A. Bobik and T. O. Yeates, *Structure*, 2008, **16**(9), 1324–1332.
23 S. Tanaka, M. R. Sawaya and T. O. Yeates, *Science*, 2010, **327**(5861), 81–84.
24 B. Webb and A. Sali, *Curr. Protoc. Bioinf.*, 2014, **47**, 6.6.1–5.6.32.
25 A. Colubri, A. K. Jha, M. Y. Shen, A. Sali, R. A. Berry, T. R. Sosnick and K. F. Freed, *J. Mol. Biol.*, 2006, **363**(4), 835–857.
26 P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral and R. W. Grosse-Kunstleve, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 213–221.
27 D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchinson and H. O. Smith, *Nat. Methods*, 2009, **6**, 343–345.
28 W. Kabsch, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 125–132.
29 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658–674.
30 P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 486–501.

This journal is © The Royal Society of Chemistry 2016

Electronic Supplementary Material (ESI) for ChemComm.
This journal is © The Royal Society of Chemistry 2016

Supplementary Information for:

Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein

J. Jorda, D.J. Leibly, M. C. Thompson, and T. O. Yeates

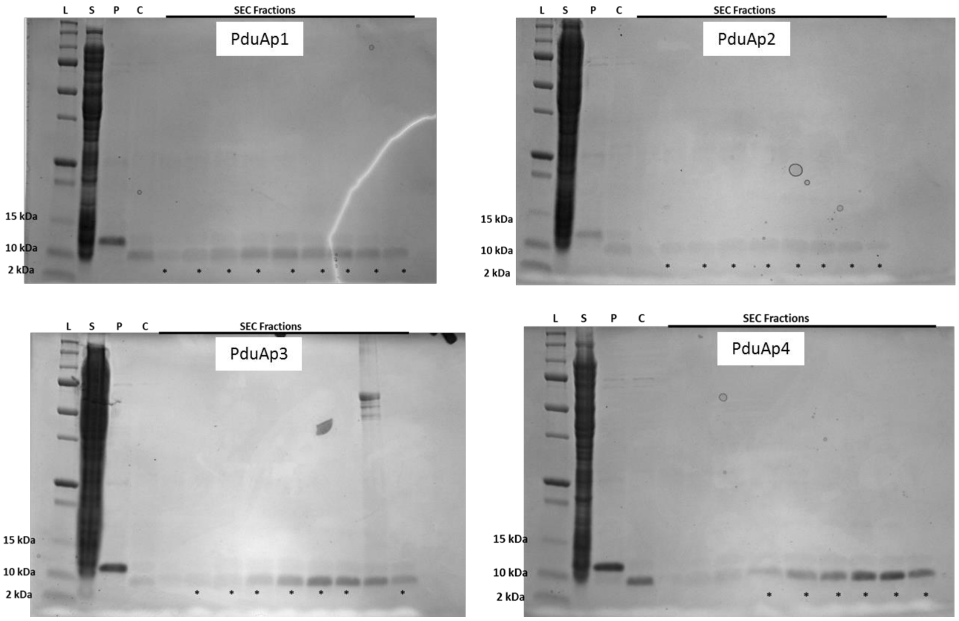


Figure S1. SDS-PAGE gels of the four PduAp purifications. Lanes are marked to indicate the protein standard ladder (L), soluble fractions (S), IMAC purified protein (P), TEV protease cleaved protein (C) and size exclusion chromatography fractions. Fractions pooled for crystallization trials are marked (*).

Table S1. Protein sequences of the four PduAp variants.

```
>P1
MKAVHVIPRPHTDVEKILPKGIDAVSAEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIGSGLVTVI
VRGDVGAVKAATDAGAAAARNVG
>P2
MKAVHVIPRPHTDVEKILGLQDAVEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIGSGLVTVIVRG
DVGAVKAATDAGAAAARNVG
>P3
MKAVHVIPRPHTDVEKILPLTGPPEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIGSGLVTVIVRG
DVGAVKAATDAGAAAARNVG
>P4
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIGSGLVTVIVRG
DVGAVKAATDAGAAAARNVG
```

8.2 Introduction – Design of a More Stable Permuted PduA Cage

The serendipitous cage that was from the permuted PduA sequence, PduAp4 spurred an effort to further strengthen this assembly. The signification geometry change from the tessellated hexagons of native PduA[35] to the closed cage of the permuted PduA (PduAp) pentagons was the result of a minor structural change (figure 8.1). While the PduAp cage is scientifically fascinating, it is impractical for proposed protein cage applications such as cargo delivery, due to the instability of the cage, as it dissociates in low protein concentrations (figure 8.2) and only forms in limited buffer conditions (figure 8.2).

8.3 Results- Attempts to Optimize the Permuted PduA Cage

Computationally redesigned versions of the PduAp mutant (PduAp4) that formed the cage were designed and expressed in order to form more stable cages. It was intended that redesigned caged would form stable cages in a wider variety of pH and salt conditions, and at lower protein concentrations, and be that the cage would more resistant to dissociation. Dr. Julien Jorda took the lead in the computational redesign of PduAp.

8.3.1 IEX Purification of Untagged PduAp

PduAp4 forms the icosahedral cage with both termini on the luminal face, the original construct was purified fortuitously as the cage dissociated into pentamers during the purification allowing the N-terminal histidine tag to be exposed. If the designed cages were stabilized sufficiently, the cage would exist throughout the purification, preventing accessible affinity tags fused to either termini. To accommodate this, ion exchange purification method was utilized. The soluble cell lysate was passed over a cation exchange column to remove non-specific proteins. The flow-though of this column was immediately applied to an anion exchange column. PduAp mutants could then be eluted from this column in high purity, similar to that of Ni^{2+} IMAC purifications.

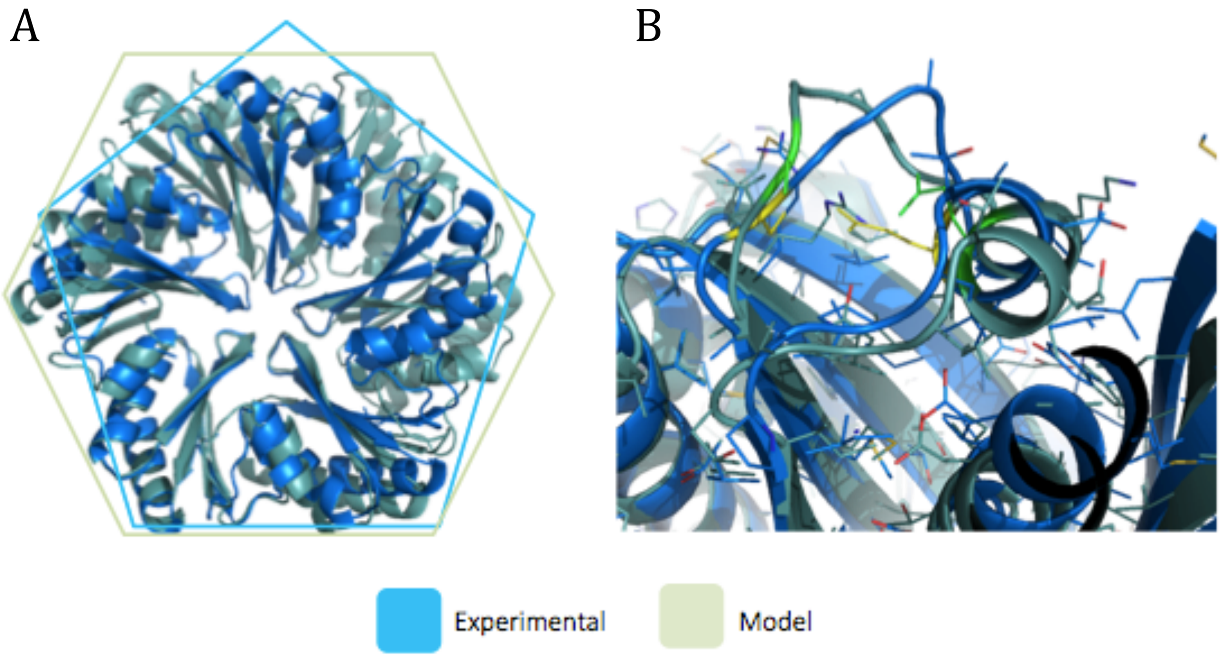


Figure 8.1. Slight Deviation of The PduA Terminal Helix Drastically Alters Geometry. (A) PduAp4 forms a pentamer instead of the intended hexamer. (B) A slight deviation in the designed loop to allow the circular permutation resulted in a shift (blue) of the PduA wild type terminal helix (green). This shift results in a more compact monomer that can pack as a pentamer.

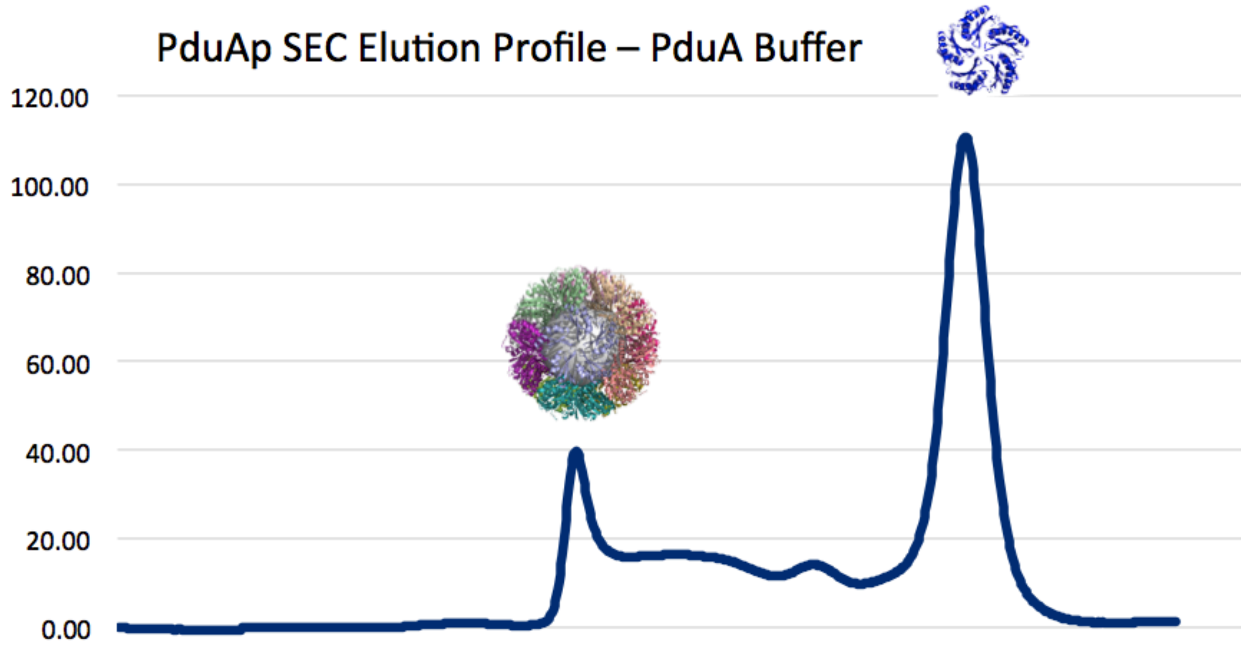


Figure 8.2. Size Exclusion Chromatography Elution Profile of PduAp4. The SEC elution profile of IMAC purified PduAp4 resulted in to peaks. The first peak corresponded to the void volume of the s200 superdex column. It was first thought this peak contained only protein aggregates. After the cage structure was solved it was determined that this peak contains the icosahedral cage. The main peak corresponded to the pentamer subunits of the cage. Dilution of the protein during SEC run caused disruption of the cage.

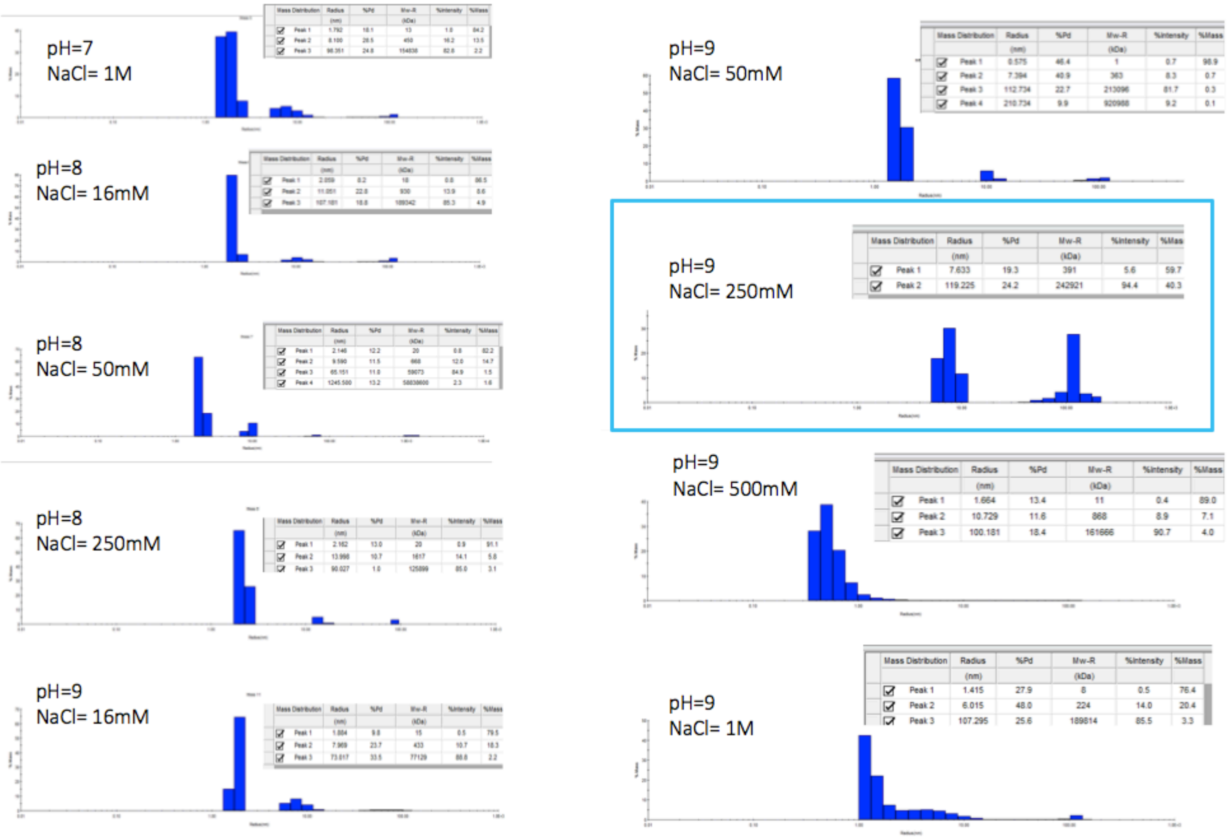


Figure 8.3. Dynamic Light Scattering Data of PduAp4 Buffer Screening. DLS was utilized to rapidly screen for conditions that readily formed the cage. In this experiment, it was determined that the cage was most abundant at pH 9.0 and 250mM NaCl (particles with a radius of ~10nm). The cage was also present as a minor species in a buffer of pH 8.0 and 50mM NaCl. It was later determined that the presence of aggregates in samples obscured the true distribution of particles in solution. Subsequent experiments concluded the cage was the predominant species at pH 8.0 and 50mM NaCl (chapter 8.1). All other buffer conditions resulted in dissociation of the cage into smaller subunits (pentamers or monomers).

8.3.2 Rosetta Designed Constructs

Initial attempts to optimize the cage used Rosetta Design[16,17] to redesign the pentamer-pentamer interface. Three separate rounds of design were performed and the top sequences from each round were selected for experimental characterization. In total, 12 Rosetta Designed cages based off of the PduAp cage structure were screened for protein solubility. None of these proteins had soluble protein expression. Although these low success rates are consistent with the results of the other Rosetta Design[17,18] studies (chapter 4), the effects were likely exacerbated by the large number of mutations introduced on PduAp structure (up to 1/3 of the residues).

8.3.3 BeAtMuSiC Hotspot Mutants

After the failure of the Rosetta redesigned constructs to express as soluble protein, a more conservative mutation approach was undertaken. It was reasoned that the expression problems with the Rosetta Designs was a direct result of the large percentage of the protein that was mutated. In order to determine key location of the pentamer-pentamer interface that could be mutated to increase stability, the program BeAtMuSiC[189] was used. From this analysis, mutations to two key interface residues, Asp43 and Lys76, were found to improve the energy scores the most. Mutation pairs were analyzed, and pairs that provided the most favorable interface and global energies (figure 8.4) were then cloned and expressed. The proteins were purified via ion exchange (figure 8.5). IEX purified protein was then further purified with SEC to determine if the cage was formed. In all cases cthe mutant proteins ran as monomers (figure 8.6). The proteins with the highest yield were analyzed by DLS to determine the size of concentrated particles in solution and to screen buffers for cage formation. In all instances the proteins remained monomeric.

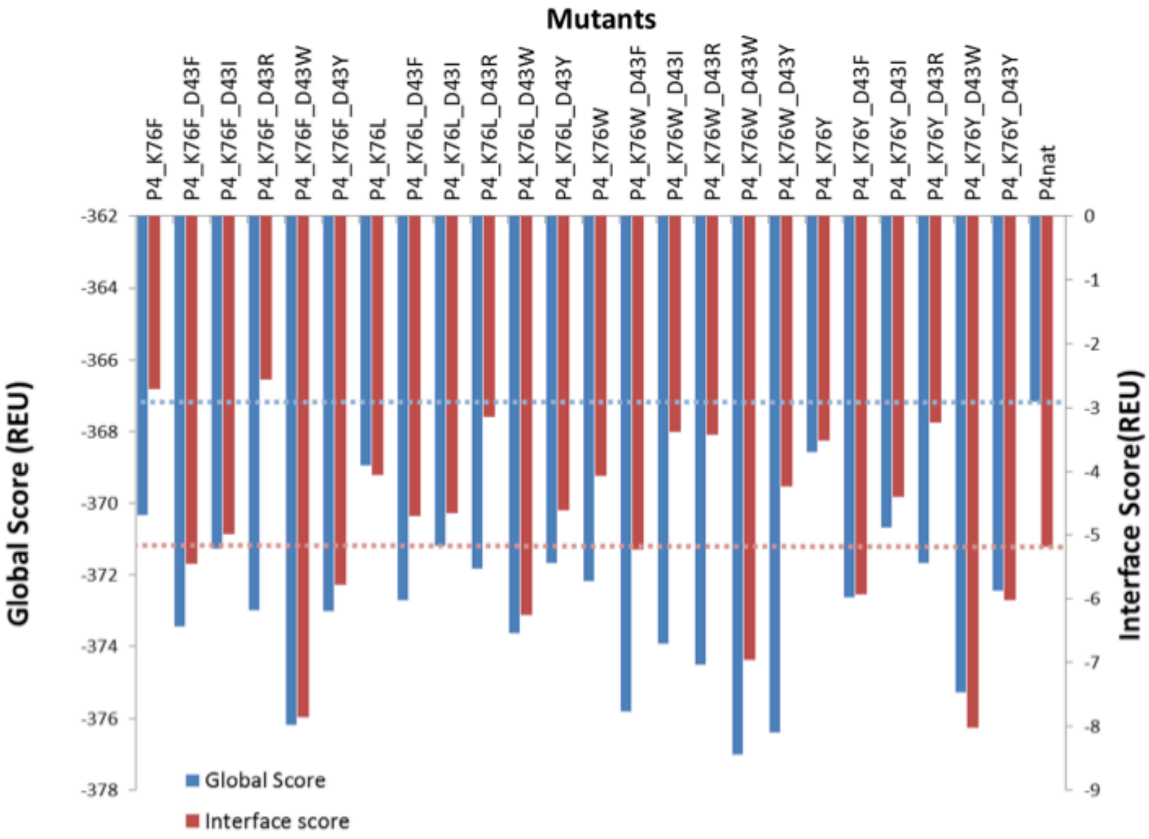


Figure 8.4. Energy Scores From Asp43/Lys76 PduAp4 Mutants. The global (blue) and interface (red) energy scores of each mutant were compared to the scores from the PduAp4 cage (dashed lines). Mutants that had lower energies then the PduAp4 sequence were thought to form a more stable assembly.

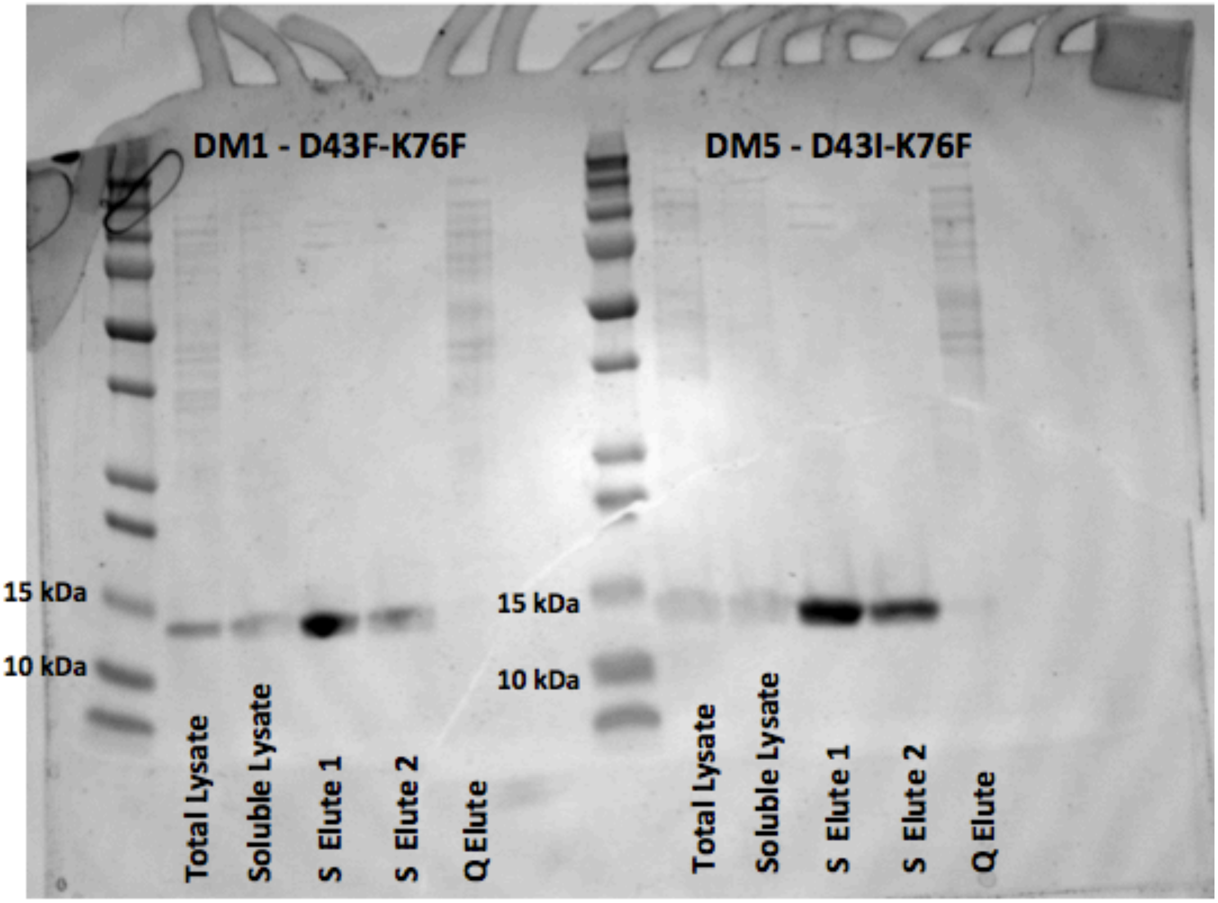

DM1 - D43F-K76F
DM5 - D43I-K76F
15 kDa
10 kDa
15 kDa
10 kDa
Total Lysate
Soluble Lysate
S Elute 1
S Elute 2
Q Elute
Total Lysate
Soluble Lysate
S Elute 1
S Elute 2
Q Elute

Figure 8.5. Ion Exchange Chromatography Purification of Two PduAp4 Mutants. IEX was employed for the purification of the PduAp constructs with no affinity tag. Cells were lysed in a buffer at pH 9.0 with only 50mM NaCl. After the cells were lysed and clarified the soluble fraction was diluted 2x in a buffer lacking NaCl. This was then applied to a Q Sepharose cation exchange to remove contaminating *E. coli* proteins (Q Elute). Flow through from this column was then immediately applied to S Sepharose anion-column and washed to remove unbound protein. PduAp mutants were eluted off with NaCl (S Elute 1, 0.5M NaCl; S Elute 2 1.0M NaCl).

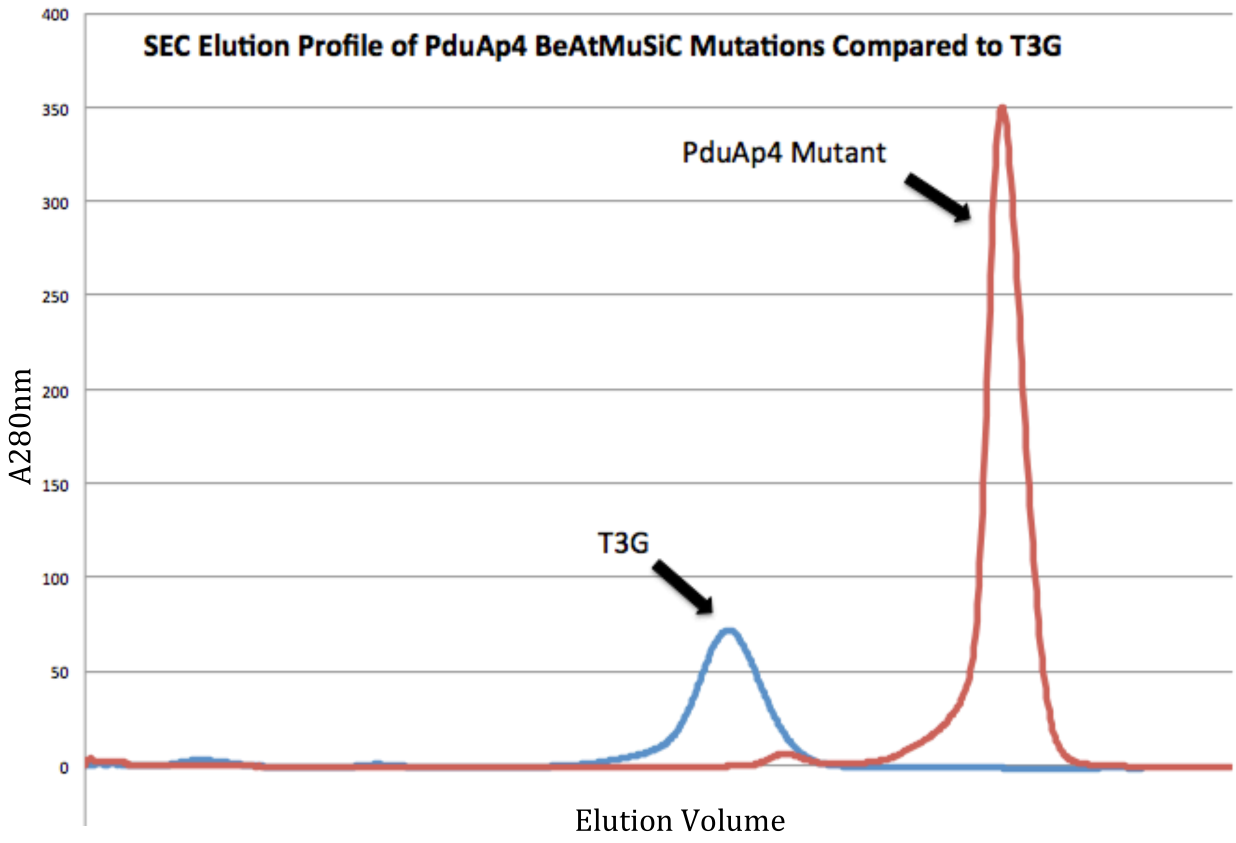


Figure 8.6. PduAp4 Mutants Are Purified as Monomers. IEX purified PduAp4 mutants were further purified on a Superos6 SEC column intended for the purification of large complexes. All PduAp4 mutants (red trace) eluted as a significantly smaller protein then the known cage T3G (blue trace). Subsequent analysis of the mutants on an s200 SEC column indicated the protein was monomeric, eluting at the volume expected of a 10 kDa protein.

8.3.3 Cysteine Mutants of PduAp

A series of cysteine mutants of PduAp4 were created (table 8.2). Cysteine mutants were placed at the pentamer-pentamer interface at locations where polar residues were in close proximity (Cα distance of <5Å). After purification, the cysteines could be oxidized to form disulfides to strengthen the cage. As with the Asp43/Lys76 mutation, all cysteine mutants were soluble but sized as monomers that could not be reassembled into cages by changing the buffer conditions.

8.4 Results - Reconstructing the PduA Hexamer

Circular permutation of PduA was performed in order to make its topology resemble that of its naturally permuted homolog PduU.[190] After the permutation was performed, the N-terminus of the PduA would be moved from the luminal surface to the external surface of the microcompartment. Once this was achieved fusions could be added to the N-terminus of PduAp, such as the occluding beta barrel native to PduU. This would create a version of PduA with a fully occluded pore, which has remained elusive,[36] allowing study of the roles of the other shell proteins with pores. Pore occlusion and external fusions to the shell would allow further engineering of the shell for non-native functions.

Once the structure of the pentamer of PduAp4 was solved, it was used to redesign sequences new permuted PduA hexamers. Rosetta was used redesign PduAp4 and enforce six-fold symmetry. From this, four new permutations were designed (table 8.3). One of these, PduAp7, had the X-ray structure solved and refined (table 8.4). PduAp7 formed the intended hexameric structure with a backbone RMSD of only 0.9Å (figure 8.7) (unpublished). This permuted PduA can serve as the basis for future modifications, including N-terminal fusion of the PduU beta barrel.

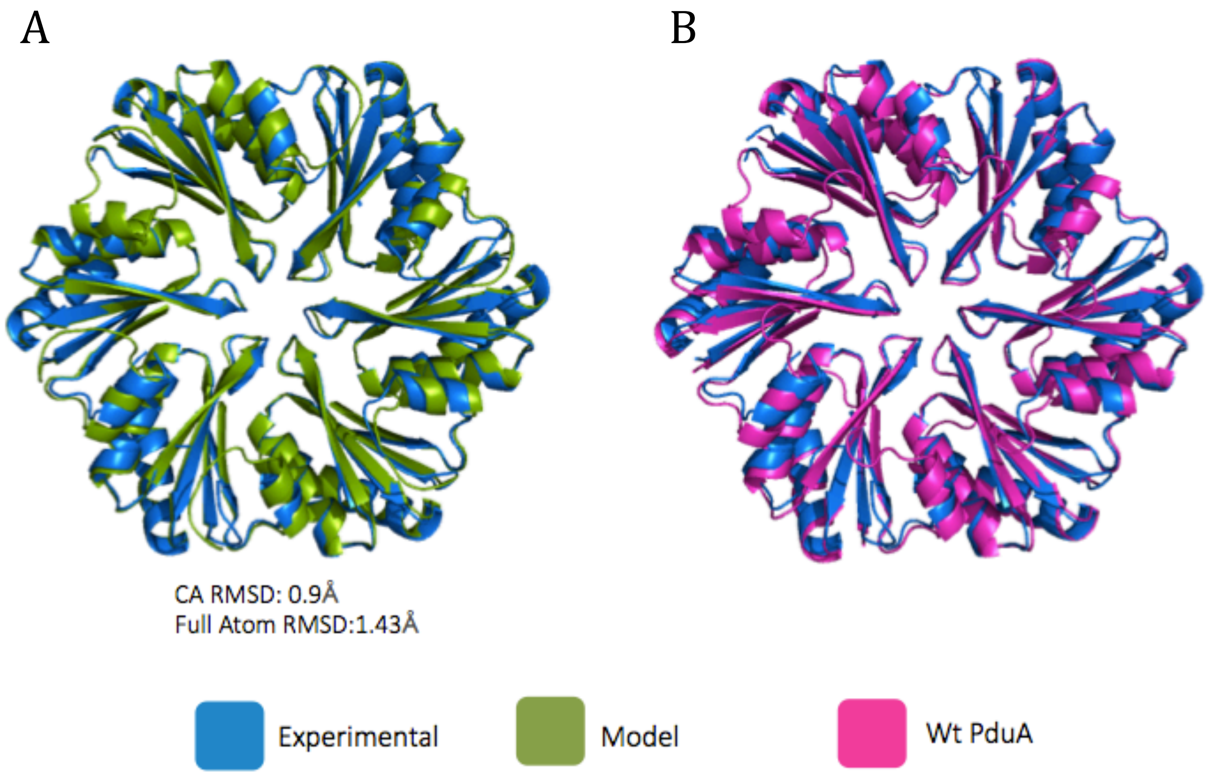

A
B
CA RMSD: 0.9Å
Full Atom RMSD:1.43Å
Experimental
Model
Wt PduA

Figure 8.7. Crystal Structure of PduAp7. (A) The experimental crystal structure of PduAp7 (blue) formed the designed structure (green), with a backbone RMSD of only 0.9Å. (B) The PduAp7 structure is nearly identical to wild type PduA (magenta) when two structures are aligned. This suggests PduAp7 can be used to replace PduA in the Pdu shell. Once this is achieved the pore may be fully occluded to study the role of the other pores in the shell or allow external fusion of proteins to the external surface of the shell via the external N-terminus of PduAp7.

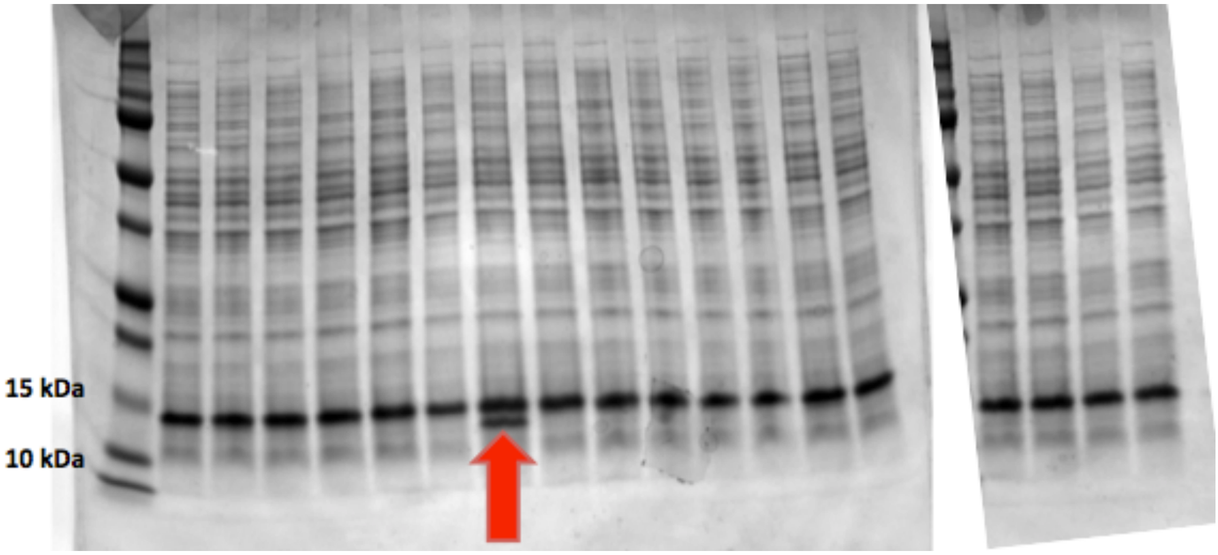


Figure 8.8. Soluble Expression Testing of PduAp7 – PduU Barrel Fusions. Soluble fractions from the first 18 of 19 PduAp7 – PduU Barrel Fusions, construct 19 was screened at a later date and was insoluble. Only one design, P7U-7 (red arrow) was soluble. The 14.5 kDa band present in all lanes is chicken egg white lysozyme added during cell lysis.

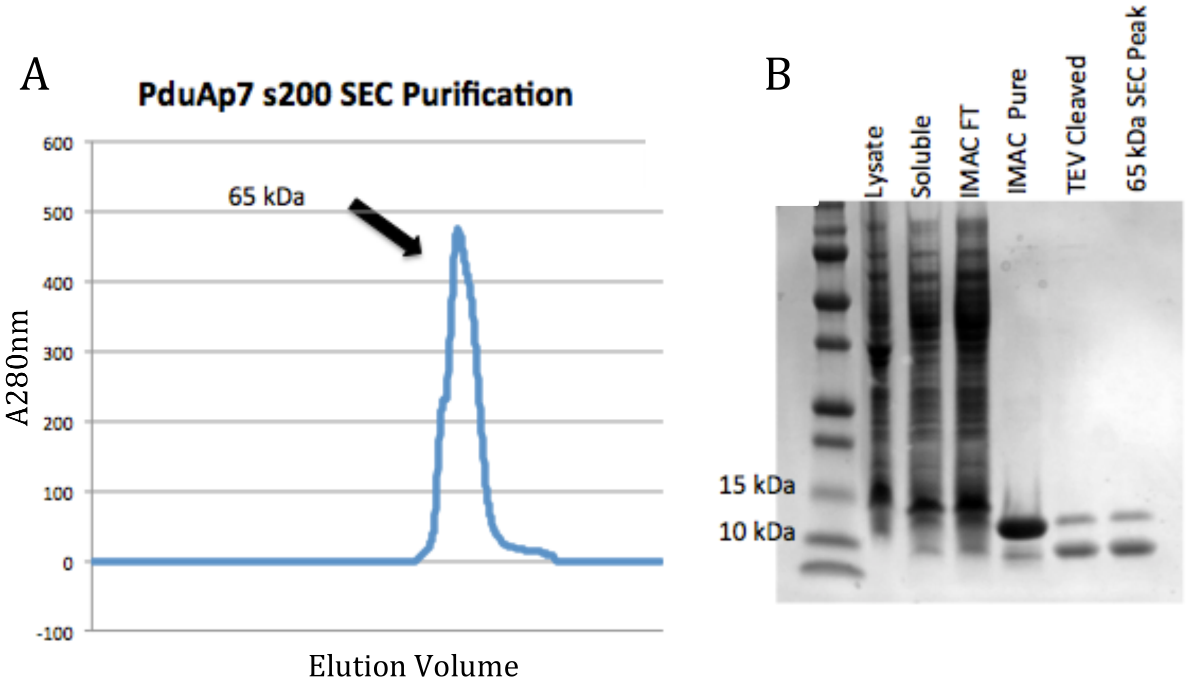


Figure 8.9. Purification of P7U-7. (A) S200 SEC elution of IMAC purified and TEV cleaved P7U-7. The major peak of the purification corresponded to a predicted molecular weight of 65 kDa. The leading fractions of the peak were pooled for crystallization experiments. (B) SDS-PAGE gel of the P7U-7 purification. A double band was apparent in the IMAC purified protein, the lower band increased and the upper band decreased upon TEV cleavage. The IMAC column used for this purification was previously used for another attempted purification and it is likely the TEV protease from that purification was not fully eluted from the resin prior to this experiment. Incomplete TEV cleave and removal of the affinity tag suggests a buried tag or TEV site for a fraction of the hexamer subunits.

8.5 Redesigned PduA-PduU Permutations

Once the hexameric PduAp7 was confirmed, efforts were undertaken to attach the PduU beta barrel to the N-terminus of PduAp7. Rosetta was used to design and refine 19 designs of the PduAp7-PduU fusion (table 8.5). Of these only one design expressed as soluble protein (figure 8.8). This construct was purified (figure 8.9) and is currently undergoing crystallization experiment optimization after small crystals were observed in one condition (figure 8.10).

8.6 Discussion

8.6.1 Experimental Characterization of Computational Designed Permuted PduA Constructs

The Rosetta redesigned PduAp4 constructs were complexly insoluble. Although solubility and expression issues plague computational interface design, the effects have been exacerbated by the size of the PduAp monomers. Shell proteins such as PduA are typically small proteins of approximately 90 amino acid residues. By forming a cyclic hexamer, interfaces are present on two distinct faces of the subunits. Rosetta specifically redesigns these interfaces to form structures with designed symmetry. For a protein like PduA, this may lead to up to 1/3 of the residues being altered. At this point the expression and solubility of the proteins are altered by changes in the folding and aggregation propensity of the subunits.

When the more conservative double mutations were characterized, in all instances they were found to be purely monomers after SEC purification. Since the double mutations involved the same two residues in each case, and given the mutation of charged residues to primarily hydrophobic resides, a common problem may be affecting these constructs. If the mutated residues are essential for the native folding of the protein, these mutations may have drastically slowed or abolished proper folding of the monomers. This does not explain the results for the cysteine point mutations. Here, mutations of polar residues to cysteines in various locations of the proteins resulted in monomers from SEC as well. This suggests an alternate cause of the monomer formation.

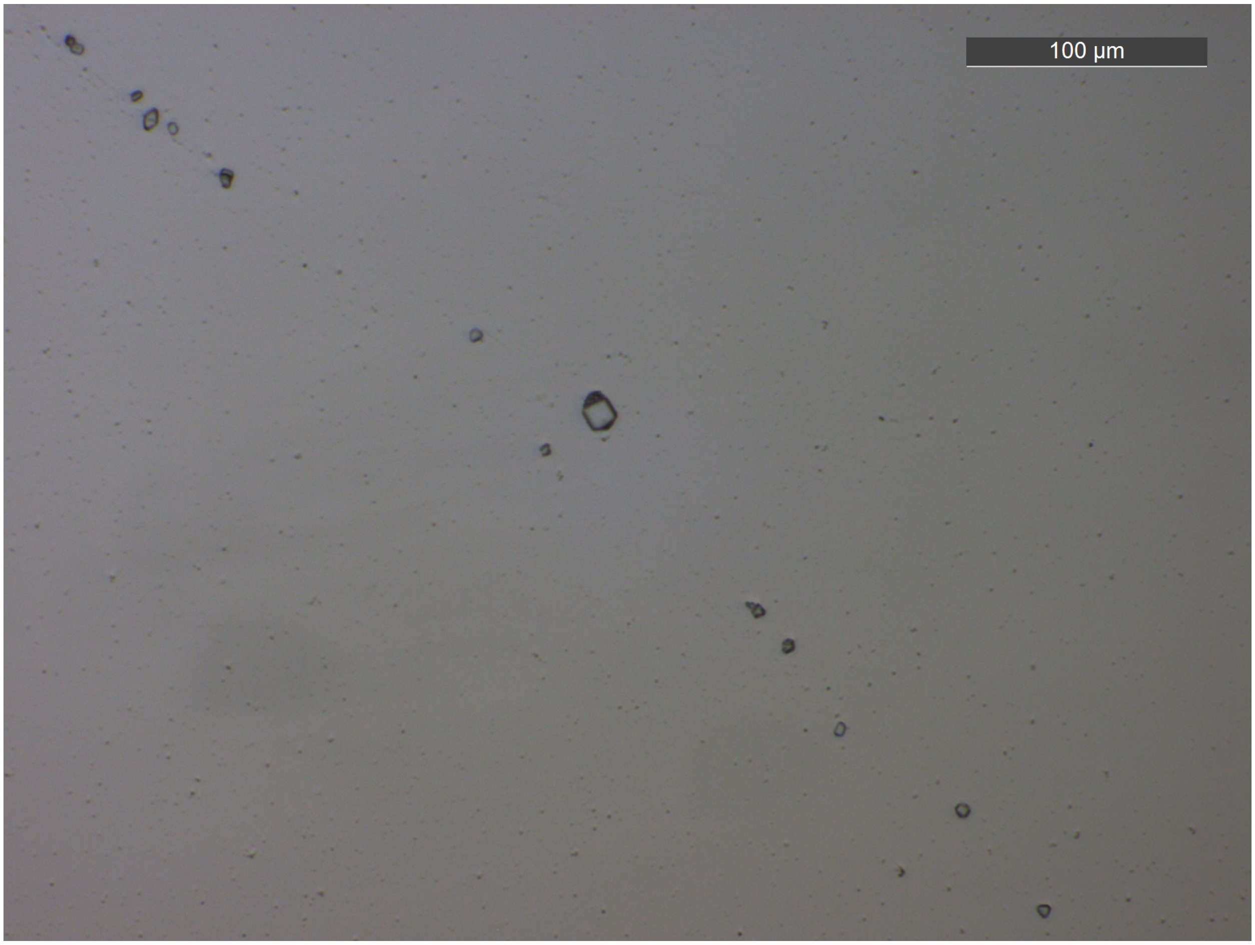


Figure 8.10. Crystals of P7U-7. Crystals were obtained after one round of streak seeding (original drop: seeds crushed and diluted 4x with reservoir solution). Reservoir solution contains 0.15M Ca(OAc)2, 18% PEG-1000 and 0.1M Immidazole, pH 8.0. Drops set up with 10mg/mL PduA-P7, 2.25µL total volume drops, 2:1 (sample:crystallant). Well conditions were the same as those from the original well/drop used to seed this 24-well plate. These crystals will be screened for diffraction in August 2016.

8.6.2 Salt Induced Disruption of Permuted PduA

The decision to utilize ion exchange chromatography for the purification of the PduAp4 redesigns was predicted by the absence of exposed termini in the assembled cage. Given that the new constructs were predicated to form stable cages, it was assumed that the termini would remain luminal in for all of the constructs. Since PduA and PduAp tend to favor alkaline buffer conditions, the protein could be highly purified with tandem IEX purifications. A pH of 9.0 was required to ensure that only PduAp4 mutants predominantly bound the final IEX cation-column. Tightly bound protein resulted, which was not eluted until the NaCl concentration was >0.5M. At no time during the purification of PduAp4 was the protein ever exposed to salt concentrations this high. Additionally, DLS indicated that the cage dissociated into pentamers or monomers when exposed to high NaCl concentrations (chapter 8.1, figure 8.3), a phenomenon common in protein complexes.[132,191,192]

It is possible that once the PduAp4 mutants are in their monomeric form, they cannot reassemble *in vitro*. DLS experiments (chapter 8.1) have suggested that there is interconversion between the cage and pentamer forms depending on the buffer conditions. However, reversible conversion to monomers has not been observed. This could be a result of not yet determining the correct buffer conditions for this, or be a reflection of structural changes in the monomeric subunits. Slight changes to the structure are responsible for the dramatic shift from hexameric tiles to enclosed cages. Similar changes could result in an unforeseen structural conformation that is not capable of forming oligomers.

8.6.3 TEV Coexpression for *In Vivo* Affinity Tag Removal

Traditionally the Pdu shell proteins have been purified with a C-terminal histidine affinity tag.[35,190] This was attempted with the original permutations of PduA, however after initial failures the tag was changed to an N-terminal TEV-cleavable affinity tag. It was thought that the C-terminal tags might interfere with crystallization or assembly of the oligomers due to the close proximity to the oligomer interface. The TEV cleavage site was used in the many of the constructs as it typically results in fewer contaminating *E. coli* proteins, and there is some evidence that removal of histidine tags improves crystallization outcome.[96] However, the presence of the N-terminal affinity tag and TEV site may have significant impact on the protein folding, solubility, structure and crystallization efficieny.[193–197] Additionally, anecdotal evidence suggests that the presence of an N-terminal histidine tag and TEV cleavage site hinders that ability of split GFP to fold (chapter 3). These factors could contribute to the solubility issues found with the P7U constructs. The inability to fully cleave the affinity tag of the P7U-7 suggests that the tag is not fully exposed and may be partially buried in the oligomer, likely affecting the assembly of the oligomer and impacting the potential for crystallization success.

It is a common practice to create multiple expression constructs with affinity tags on either terminus to alleviate the deleterious effects that may occur for one approach. While this is effective when working in large scale, it is not practical for the large numbers of proteins required for protein design efforts. Therefore, a facile approach for screening alternate affinity tag conditions needs to be developed. One way to achieve this would be coexpression with TEV protease. TEV protease has been used for the *in vitro* cleavage of recombinant proteins in numerous studies.[198–203] In the case of designed proteins, one construct would be synthesized and cloned. This would feature a N-terminal histidine tag and TEV cleavage site, or if desired an alternate protease site.[204,205] This construct would be used for initial solubility screening; if the protein were soluble then purification would be performed. If the protein fails to express or is insoluble, the construct could be expressed in a cell line harboring TEV protease on a compatible plasmid.[206] If the *in vivo* removal of the tag results in soluble expression it can be concluded that the affinity tag was responsible for insolubility. The protein could then be purified without an affinity tag, or a C-terminal affinity tag could be attempted.

8.6.4 Structure of Redesigned Permuted PduA Hexamer

The six-fold symmetry of wild type PduA was recapitulated with the permuted PduAp7, designed with Rosetta using the PduAp4 structure. This successfully created a structure that matched the designed model and is nearly identical in structure to wild type PduA. There are no structural variations of consequence with this structure. PduAp7 can now be used for the basis of new fusions, including the PduU beta barrel fusion discussed here. Biological validation of PduAp7's ability to incorporate into Pdu shells and function as a replacement for wild type PduA needs to be determined. These are the next experiments planned for this designed protein.

8.6.5 Interconversion Between Pentamers and Hexamers Has Evolutionary Implications

Ambiguity of the oligomeric state of shell proteins has been observed prior to the PduA permutation work. The ethanolamine utilization microcompartment shell protein EutN was suggested to be a hexamer by the crystal structure (PDB 2X9H). Recently, it was determined to be a pentamer in solution.[207] PduAp4 was the first definitive case of a nearly identical shell protein that folds into the canonical BMC fold[31] existing in two different oligomeric states. In order for a closed shell to form, it needs two components hexamers for the facets and pentamers for the vertices. All known microcompartments, are thoughT to contain proteins in these two symmetries.[24]

The ability for one protein to adopt these two conformations would allow simple microcompartments to form with only one protein required to form the shell. Additional shell proteins could then evolve through gene duplication events. It has been suggested that EutS, a curved shell protein, might assemble into closed structures by itself.[208] It is thought that the curved structure of this protein allows the shell to form, resulting in polymorphic structures of varying sizes. Interconversion between hexamers and pentamers could result in more regular shells, similar to the Pdu MCPs formed with all shell proteins present. This would be especially true if the pentamer form occurs at a low level compared to the hexamer form. If this were the case in early shell proteins and primitive MCPs, then that ability may be restored with alterations to the shell proteins. Permutation of PduA may have resulted in this ability.

It has been difficult to assess the exact oligomeric state of PduAp4 in solution. DLS cannot reliably detect the difference in radii between the two forms. SEC-MALS was performed on a sample of PduAp4, but provided inconclusive results from dilution during the run. Standard SEC purification of PduA permutations resulted in elution volumes consistent with hexamers. However, the physical size differences between the two cyclic oligomers (figure 8.1) may be too small to accurately resolve on an SEC column. Additional experiments should be undertaken to determine the oligomeric state of PduA permutations in differing buffer, similar to what was done with EutN.[207]

8.7 Conclusions

Circular permutation of PduA and subsequent computational redesign of the PduAp4 cage has proven difficult to achieve. Inherent issues computationally designed proteins have been contributing factor in the cage optimizations. Although a stabilized cage has eluded design to date, redesign of the intended PduAp hexamer was successful. Now, we have a new version that can serve as a building block for future designs or used in biological assays of Pdu MCP function. Through iterative rounds of protein design, like what was done to form PduAp7, new versions of MCP shell proteins can now be formed. These can be significantly different from native proteins and offer chances to create highly modified pores compared to simple pore point mutations.[36] Modified pores are applicable to and essential for future efforts to bioengineer MCPs for the encapsulation of non-native enzyme pathways.

8.8 Materials and Methods

8.8.1 Cloning

For all designed sequences, the genes were codon opitmized[95] for expression in *E. coli.* Optimized genes were synthesized as gBlocks (Integrated DNA Technologies) with the terminal sequences GTTTAACTTTAAGAAGGAGATATAC and

TAAAATTCGAGCTCCGTCGACAAGCTTG added to the 5’ and 3’ ends respectively. The genes were then cloned via Gibson assembly[75] into pET22b vector cut with NdeI and EcoRI restriction enzymes and transformed into XL-2 Blue Ultracompetent cells (Agilent Technologies). Once the plasmids were sequence validated, plasmids were transformed in BL21 DE3 expression cells (New England Biolabs)

8.8.2 Protein Expression

All constructs were expressed in LB supplemented ampicillin. 1L LB flasks were inoculated with 10mL of overnight culture and allow to grow for at 37°C until the A_{600nm} optical density reached 0.6-0.8, the temperature was then reduced to 30°C and protein expression was induced with 1mM IPTG and grown for 16 hours.

8.8.3 Ion Exchange Purification of Untagged Constructs

Cells were lysed in a buffer at pH 9.0 with only 50mM NaCl. After the cells were lysed and clarified at 5000 xg for 20 mins, the soluble fraction was diluted 2x in a buffer lacking NaCl. This was then applied to a Q Sepharose cation exchange to remove contaminating *E. coli* proteins (Q Elute). Flow through from this column was then immediately applied to S Sepharose anion-column and washed to remove unbound protein. PduAp mutants were eluted off with 1m NaCl, the protein was immediately diluted to a final NaCl of 100mM. Protein was then concentrated to a volume suitable for further purification by SEC. SEC was performed with a Superpose6 GL 10/300 column to determine is a cage was present, SEC was performed on a s200 GL 10/300 analytical column to determine the oligomeric state of the proteins when it was found cages were not formed.

8.8.4 Purification of Affinity Tagged Proteins

Purification was performed per previously reported methods.[209,210] TEV cleavage performed overnight dialysis in a buffer by dialysis of mixed IMAC purified PduAp constructs was mixed with TEV protease in a 1:50 ratio in a buffer of 10mM Tris pH 8.0, 150mM NaCl, 1mM EDTA, 10% glycerol at 4°C, the TEV protease was removed but a second IMAC purification step prior to SEC.[129] SEC fractions corresponding to hexamers was pooled, concentrated to 10mg/mL, flash frozen and stored for crystallization experiments.

8.8.5 PduAp7 Crystallization

PduAp7 was screened for crystals only in the $AmSO_4$ screen (Qiagen) at both 10mg/mL and 5mg/mL. Small rod crystals were found after 1 week in 2.0M AmSO4, 2.0M NaCl. The Crystals were optimized with the ADDit additive screen (Emerald BioStructures). Diffraction quality crsytals were grown at 5mg/mL in a final condition of 2.0M AmSO4, 2.0M NaCl, 0.6% w/v Xylitol. Crystals were cryo-protected with 25% glycerol prior to flash freezing the crystals in liquid nitrogen.

8.8.6 PduAp7 X-Ray Data Processing and Structure Solution

All data was collected at the Advance Photon Source beam line 24-ID-C. Data was indexed in space group P1 to a resolution of 2.5Å with XDS/XSCALE,[97] phases were determined with Phaser[88] with the computational model serving as the search model. Solution from these programs had refinement attempted with PDB_Redo,[90] Refmac[89] and Phenix,[91] only Buster[211] was able to sufficiently refine the model. Models were manually evaluated in Coot[98] and analyzed with PyMOL.[99]

Table 8.1. Sequences of Rosetta Designed Permuted PduA Icosahedral Constructs

P4_I11
MTVAVIPDPHDDVAAALGGSGGSKALGVVLTVGMTAALAAAAAMVAAANVEMVGIYLIGNGLAA
VVVRGAAGQVIIATAAGAAAAAAR

P4_I12
MTVAVIPNPHDDVAAALGGSGGSKALGAVMTVGLTAALAAAAAMAAAANVEVVGFYVIGNGLVV
VVVRGAAGQVMIATAAGAAAADNR

P4_I13
MTVAVIPNPHDDVAAALGGSGGSKALGVVTTVGLTAALAAAAAMAAAANVEVVGFYVIGNGLVS
VVVRGSIGAVIKATEAGAEAAKNR

P4_I14
MTVAVIPNPHDDVAANLGGSGGSKALGAVTTVGLTAALAAAAAMVAAANVEAVGFYVIGNGLVV
VVVRGSIGAVIKATEAGAEAAKNR

P4_I21
MATHVTPNPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAAAMVAAANVMLVGYEKIGSGLVT
VIVRGSTGAVKAATLAGSLAAATA

P4_I22
MATHVLPNPHTDVEKILGGSGGSEALGMVETKGLTAAIAAAAAMVAAANVMLVGYEKIGSGLVT
VIVRGSIGAVIKATLAGSLAAANA

P4_I23
MAVHVIPNPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAAAMVASANVMLVGYEKIGSGLVT
VIVRGSTGAVIAATLAGSLAAATV

P4_I31
MAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAADAMVQAANVMLVGYEKIGSGLVT
VIVRGDVGAVKAATAAGAAAADSA

P4_I32
MAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAADAMVEAANVMLVGYEKIGSGLVT
VIVRGDVGAVKAATWAGAAAADNA

P4_I33
MAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAADAMVEAANVMLVGYEKIGSGLVT
VIVRGDVGAVKAATWAGALAAITA

P4_I34
MAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLAAAIEAAAAMVAAANVMLVGYEKIGSGLVT
VIVRGDVGAVKAATAAGAAAAAST

P4_I35
MAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAAAMVEAANVMLVGYEKIGSGLVT
VIVRGDVGAVKAATDAGAAAAANA

Table 8.2. Sequences of PduAp4 Icosahedral Constructs From BeAtMuSiC Predictions

PduAp4 – Original Structure
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAADAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVKAATDAGAAAARNVG

DM1 - D43F-K76F
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAFAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVFAATDAGAAAARNVG

DM2 - D43Y-K76F
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAYAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVFAATDAGAAAARNVG

DM3 - D43I-K76W
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAIAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVWAATDAGAAAARNVG

DM4 - D43W-K76Y
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAWAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVYAATDAGAAAARNVG

DM5 - D43I-K76F
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAIAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVFAATDAGAAAARNVG

DM6 - D43R-K76F
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAARAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVFAATDAGAAAARNVG

DM7 - D43W-K76F
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAWAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVFAATDAGAAAARNVG

DM8 - D43F-K76L
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAFAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVLAATDAGAAAARNVG

DM9 - D43I-K76L
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAIAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVLAATDAGAAAARNVG

DM10 - D43R-K76L
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAARAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVLAATDAGAAAARNVG

DM11 - D43W-K76L
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAWAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVLAATDAGAAAARNVG

DM12 - D43Y-K76L
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAYAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVLAATDAGAAAARNVG

DM13 - D43F-K76W
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAFAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVWAATDAGAAAARNVG

DM14 - D43R-K76W
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAARAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVWAATDAGAAAARNVG

DM15 - D43W-K76W
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAWAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVWAATDAGAAAARNVG

DM16 - D43Y-K76W
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAYAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVWAATDAGAAAARNVG

DM17 - D43F-K76Y
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAFAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVYAATDAGAAAARNVG

DM18 - D43I-K76Y
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAIAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVYAATDAGAAAARNVG

DM19 - D43R-K76Y
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAARAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVYAATDAGAAAARNVG

DM20 - D43Y-K76Y
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAAYAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVYAATDAGAAAARNVG

Cysteine Disulfide Mutants
DM21 - D43C-K76C
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAACAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVCAATDAGAAAARNVG

DM22 - D80C-A84C –
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAADAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVKAATCAGACAARNVG

DM23 - R87C –
MKAVHVIPRPHTDVEKILGGSGGSEALGMVETKGLTAAIEAADAMVASANVMLVGYEKIGSGLV
TVIVRGDVGAVKAATDAGAAAACNVG

DM24 - R9C-V72C –
MKAVHVIPCPHTDVEKILGGSGGSEALGMVETKGLTAAIEAADAMVASANVMLVGYEKIGSGLV
TVIVRGDCGAVKAATDAGAAAARNVG

Table 8.3. Sequences of Rosetta Designed PduAp4 Derived Hexamers

PduAp6
MGSDKIHHHHHHENLYFQGAMVYVIPDPHDLVEKILGGSGGSEALGMVETKGLTAAIAAAHAMV
VSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARNVG

PduAp7
MGSDKIHHHHHHENLYFQGAMVYVIPDPHDAVEKILGGSGGSEALGMVETKGLTAAIAAAHAMV
VSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARNVG
CGTGGGTTAACGCGACTTAATTAACTCGTT

PduAp8
MGSDKIHHHHHHENLYFQGKTVYVIPDPHDAVEKILGGSGGSEALGMVETQGLTAAIAAAHAMV
KSANVVLVGYEKIGMGLVTVIVRGDVGAVKAATDAGAAAARNVG

PduAp9
MGSDKIHHHHHHENLYFQGKTVYVIPDPHDAVEKILGGSGGSEALGMVETEGLTAAIAAAHAMV
VSANVVLVGYEKIGMGLVVVIVRGDVGAVKAATDAGAAAARNVG

Table 8.4. Data Collection and Refinement Statistics for the PduAp7 Structure

Resolution range (Å)	56.57 - 2.8 (2.564 - 2.476)
Space group	P 1
Unit cell	39.71 50.27 58.39 102.93 94.68 96.25
Total reflections	50083 (3441)
Unique reflections	14348 (1204)
Multiplicity	3.5 (2.9)
Completeness (%)	92.68 (79.37)
Mean I/sigma(I)	6.91 (1.16)
Wilson B-factor	63.10
R-merge	0.08601 (0.7818)
R-meas	0.1014
CC1/2	0.996 (0.855)
CC*	0.999 (0.96)
R-work	0.3212 (0.4877)
R-free	0.3988 (0.4555)
Number of non-hydrogen atoms	3636
macromolecules	3636
Protein residues	534
RMS(bonds)	0.007
RMS(angles)	1.24
Ramachandran favored (%)	95
Ramachandran outliers (%)	0.57
Clashscore	17.05
Average B-factor	212.40
macromolecules	212.40

Table 8.5. Sequences of PduAp7-PduU Barrel Fusions

P7U-1
MGSDKIHHHHHHENLYFQGQPTTDRMIQEYVPGKQAMVYVIPDPHDAVEKILGGSGGSEALGMV
ETKGLTAAIAAAHAMVVSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARN

P7U-2
MGSDKIHHHHHHENLYFQGQPTTDRMIQEYVPGKQVTMVYVIPDPHDAVEKILGGSGGSEALGM
VETKGLTAAIAAAHAMVVSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARNV

P7U-3
MGSDKIHHHHHHENLYFQGQPTTDRMIQEYVPGKQVTLVYVIPDPHDAVEKILGGSGGSEALGM
VETKGLTAAIAAAHAMVVSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARNV

P7U-4
MGSDKIHHHHHHENLYFQGTPAQAVTTVWVAPGIANATLVIPDPSYAVLATTGGSGGSEALGEV
MTAGMVAANAAAWAMYLSANVVVVGAIIIGAGAVIVIVRGDVGAVKAATDAGAAAAKKV

P7U-5
MGSDKIHHHHHHENLYFQGFPAQAVAVVFVAPGVANLTMVIPDPTDAALAMLGGSGGSEALGMV
TTAGLEAAAAAAEAMAISANVVVVGAVLIGAGVVTVIVRGDVGAVKAATDAGAAAAKKV

P7U-6
MGSDKIHHHHHHENLYFQGTPDQESATHDIAPGVANATWVIPDPVDAVLATTGGSGGSEALGQV
ETSGMAAADAAAWAMALSANVVTVGAVITGSGAVTVIVRGDVGAVKAATDAGAAAAKKV

P7U-7
MGSDKIHHHHHHENLYFQGTPAQTATAVLLSPGVANATWVIPDPSDAALAATGGSGGSEALGMV
TTSGMAAAAAAAAAMAAAANVVNVGYVITGSGAVTVIVRGDVGAVKAATDAGAAAAKKT

P7U-8
MGSDKIHHHHHHENLYFQGTPAQAVTTVWVAPGNAVAAALVIPDPSYATLATTGGSGGSEALGE
VTTAGMSAANAAAEAMYNSANVVVVGAIIIGAGAVTVIVRGDVGAVKAATDAGAAAAKKV

P7U-9
MGSDKIHHHHHHENLYFQGFPAQAVAVVFVAPGQQVAAAMVIPDPSDAVLAMTGGSGGSEALGM
VTTAGLAAAAAAALAMALSANVVVVGAVLIGAGVVTVIVRGDVGAVKAATDAGAAAAKKA

P7U-10
MGSDKIHHHHHHENLYFQGTPDQTSTAVDAAPGNQVSAAWVIPDPSDAALATTGGSGGSEALGQ
VETSGAAAANAAAWAMALSANVVTVGAVITGSGAVTVIVRGDVGAVKAATDAGAAAAKKA

P7U-11
MGSDKIHHHHHHENLYFQGTPAQKAQTHSLAPGAAVASAWVIPDPSDAALAANGGSGGSEALGL
VETSGHSAAAAAAWAMRLAANVVNVGAVITGSGAVAVIVRGDVGAVKAATDAGAAAAKKA

P7U-12
MGSDKIHHHHHHENLYFQGTPAQAVTTVWVAPGIAAMVYVIPDPHDAVEKILGGSGGSEALGMV
ETKGLTAAIAAAHAMVVSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARNV

P7U-13
MGSDKIHHHHHHENLYFQGFPAQAVAVVFVAPGVAAMVYVIPDPHDAVEKILGGSGGSEALGMV
ETKGLTAAIAAAHAMVVSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARNV

P7U-14
MGSDKIHHHHHHENLYFQGTPDQESATHDIAPGVAAMVYVIPDPHDAVEKILGGSGGSEALGMV
ETKGLTAAIAAAHAMVVSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARNV

P7U-15
MGSDKIHHHHHHENLYFQGTPAQTATAVLLSPGVAAMVYVIPDPHDAVEKILGGSGGSEALGMV
ETKGLTAAIAAAHAMVVSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARNV

P7U-16
MGSDKIHHHHHHENLYFQGTPAQAVTTVWVAPGNAVAAVYVIPDPHDAVEKILGGSGGSEALGM
VETKGLTAAIAAAHAMVVSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARNV

P7U-17
MGSDKIHHHHHHENLYFQGFPAQAVAVVFVAPGQQVAAVYVIPDPHDAVEKILGGSGGSEALGM
VETKGLTAAIAAAHAMVVSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARN

P7U-18
MGSDKIHHHHHHENLYFQGTPDQTSTAVDAAPGNQVSAVYVIPDPHDAVEKILGGSGGSEALGM
VETKGLTAAIAAAHAMVVSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARNV

P7U-19
MGSDKIHHHHHHENLYFQGTPAQKAQTHSLAPGAAVASVYVIPDPHDAVEKILGGSGGSEALGM
VETKGLTAAIAAAHAMVVSANVVLVGYEKIGDGLVTVIVRGDVGAVKAATDAGAAAARNV
AAAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAG

CHAPTER 9

Structural Studies of the PduA – PduD Helix Interaction

9.1 Introduction

Bacterial microcompartments encapsulate diverse and complex metabolic pathways.[24,25,29] Although the pathways found in these can be quite divergent, they are thought to have evolved to seclude volatile or toxic intermediates away from the cytoplasm of the bacterium. There is a growing desire to utilize these shells for the creation of bioreactors.[28] In order to achieve this, non-native enzymes must be encapsulated by the shells. This mechanism of this encapsulation remains ambiguous for many of these shells.

9.1.1 Microcompartment Enzyme Encapsulation

Carboxysomes were the first MCP discovered and are the simplest, containing only he enzymes RuBisCO and carbonic anhydrase for CO_2 fixation.[26,27,212,213] Enzyme encapsulation in the carboxysome is facilitated by a preformed enzyme core recruiting the shell proteins to the core through auxiliary proteins.[37,214] Enzyme encapsulation in the more complex MCPs, including the Pdu MCP, is not well understood. It has been well established that the N-terminus of many of the encapsulated enzymes contains an amino acid sequence, thought to be helical in character, that targets the enzymes to the MCPs through interactions with the luminal face of shell proteins.[39,40,208,215–217] These terminal sequences, or tails, target both native and heterologous proteins to the lumen of the shells. The global arrangements of the enzymes, in the shells remain a mystery. Emerging evidence suggests there are in fact interactions between the tail segments of the enzymes forming larger assemblies in the lumen of the MCP, but there has yet to be evidence of fully preformed enzyme cores as the case with the carboxysome.[218–220]

In order to better understand the protein-protein interaction between the targeting tails and the shell proteins, efforts were undertaken to solve the x-ray crystal structure of the interaction. The PduA-

PduD tail interaction was selected for study. The PduD tail has been found to target both PduD and heterologous enzymes to the lumen of the Pdu MCP.[215] The PduA C-terminus and PduD tail interaction was also predicted through computational docking studies[24] but not experimentally validated. Only the interaction between the PduA C-terminus and PduP tail has been conclusively shown.[40] Interestingly, the PduD and PduP tails are thought to bind the same cleft formed by the C-terminal helix of PduA except in the exact opposite orientation. The PduD tail is predicted to have the N-terminus pointing towards the edge of the PduA hexamer and the C-terminus towards the pore. Fortuitously, this orientation allows fusion of the C-terminus of PduA to the N-terminus of the PduD tail (PduAD). Similar fusions are not possible with the PduP tail due to its predicted orientation.[24]

9.1.2 Summary of Previous Crystallographic Efforts

Dr. Dan McNamara performed preliminary crystallographic studies. For these experiments he attempted to determine the structure of the interaction between PduA and synthesized tails of both PduD and PduP. An edge mutant of PduA K26A was used for these studies as it crystallizes in a form (PDB 4P2S, space group C2) with large solvent channels providing access to the C-terminus of PduA. Apo PduA crystals in this form were soaked with the peptides and co-crystallization experiments with PduA and the peptides were also attempted. In several instances the resolution of the diffracting crystals was poor compared to the apo PduA crystals. In all instances no peptides could be seen in the electron density.

Fusions of PduA K26A residues 1-89 and PduD residues 2-16 or 2-18 were made with a short 'GSG' linker between the two segments. These constructs readily crystallized in conditions contained $AmSO_4$ as the precipitant, and formed hexagonal crystals just as PduA typically crystallizes. These fusions crystallized in space group P6 2 2, structures were solved and refined. In all instances, only the electron density of PduA was visible, the PduD fusion was completely disordered. We hypothesized that the fusions may have failed as a result of the short linker used.

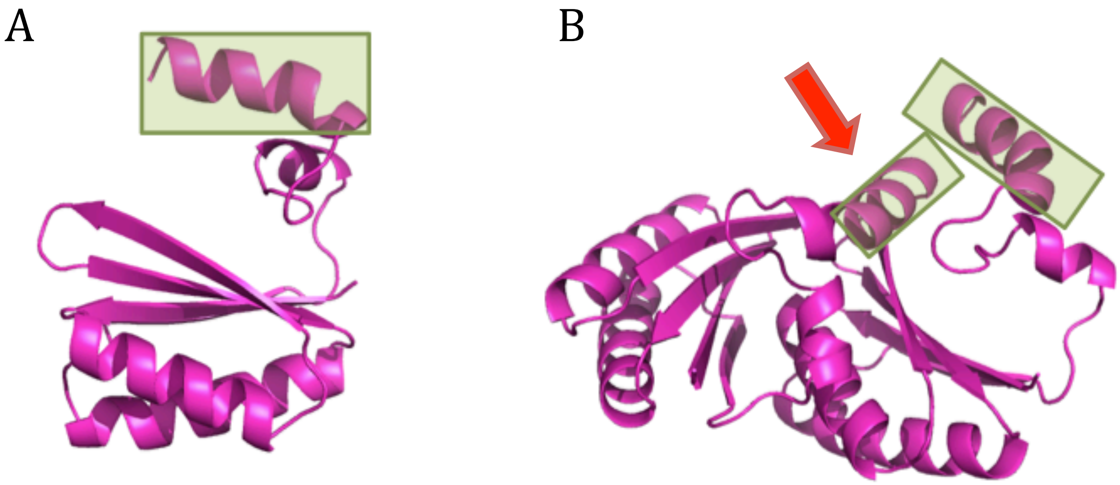


Figure 9.1. Rosetta Opimized Fusion Model of a PduA-PduD N-terminal Fusion. (A) The monomer of the PduAD model has the PduD tail located above the PduA C-terminal tail (green box). The structure of the PduA BMC domain is not altered. (B) When the PduA hexamer forms the PduD tail is predicted to sit on the interface of two monomers (red arrow).

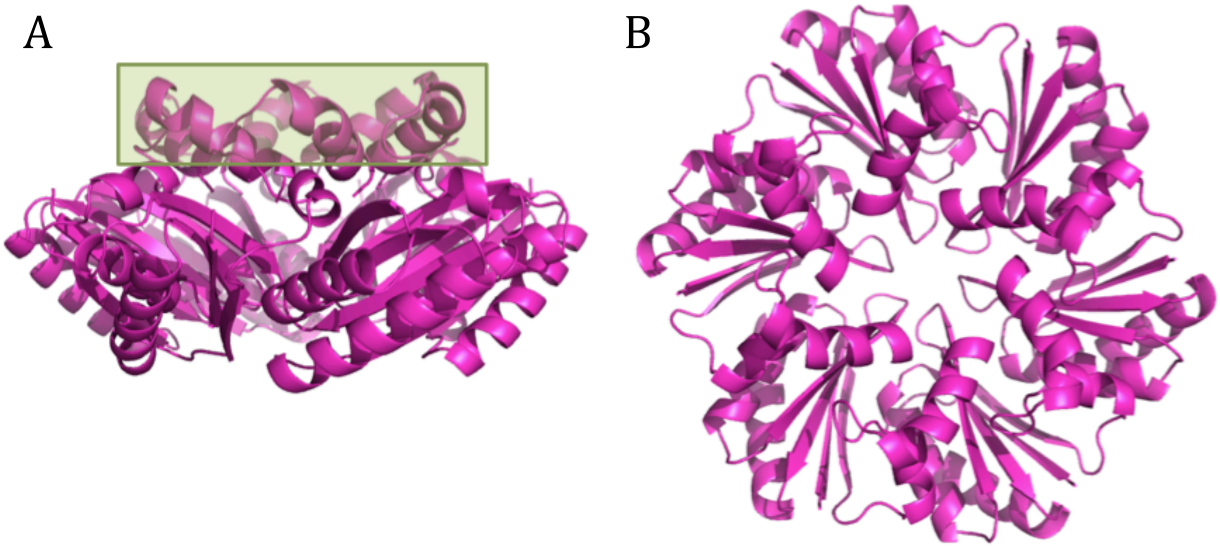
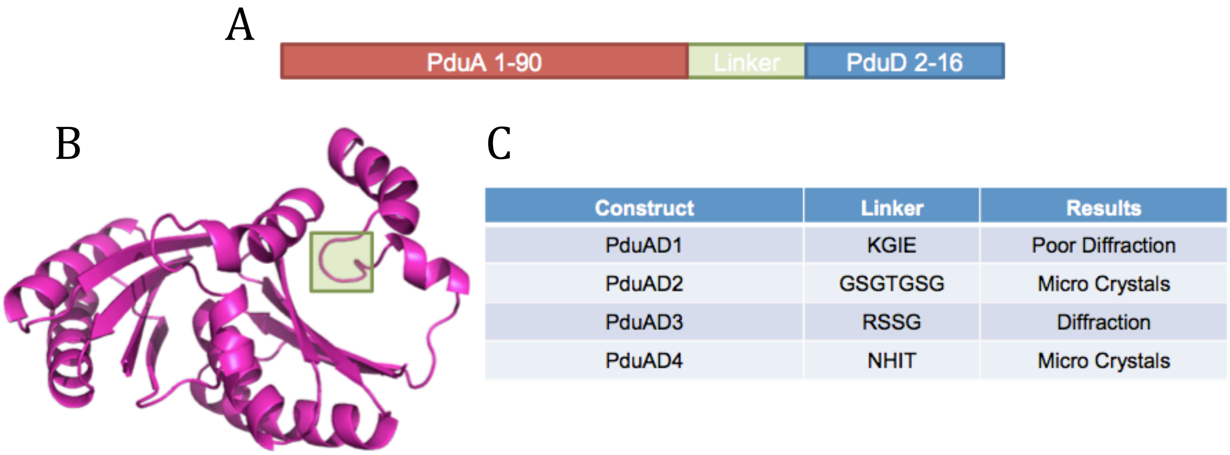


Figure 9.2. Hexametric Model of the PduAD Rosetta Design. (A) In the computational model sixe PduAD fusions could pack as a hexamer with the PduD tails pack tightly against the PduA C-terminal helices (green box). (B) Viewed down the six-fold axis the PduAD fusions do not disrupt the hexamer or central pore of PduA.



Construct	Linker	Results
PduAD1	KGIE	Poor Diffraction
PduAD2	GSGTGSG	Micro Crystals
PduAD3	RSSG	Diffraction
PduAD4	NHIT	Micro Crystals

Figure 9.3. PduAD Construct Designs. (A) PduAD featured the first 90 residues, a variable linker and PduD residues 2-16. (B) A turn (green box) is required to set the correct geometry to allow the tail fold on to the PduA putative binding site. (C) Four versions of the loop were attempted, PduAD1 is the Rosetta optimized sequence, PduAD2 is a long flexible linker, PduAD3 and PduAD4 are sequences with a high propensity to form beta turns.

9.2 Results

9.2.1 PduAD Fusion Constructs

A model of the PduAD fusion was generated with Rosetta[16] (figure 9.1). This model indicated that the PduA could accommodate the six PduD tail fusions and still form the native hexamer (figure 9.2). Key to this model is a turn that forms the linker (figure 9.3), from Rosetta this turn was designed as ‘KGIE’ (‘GSG’ was used in preliminary studies). This designed constructs as well as three others with differing linkers were synthesized and expressed. It was hypothesized that the previous ‘GSG’ linker was either too short or too flexible. To test this a longer flexible ‘GSGTGSG’ linker replaced the designed ‘KGIE’. This designed loop was also replaced with one of two sequences (‘RSSG’ and ‘NHIT’) with a high propensity to form a β-turn.[221] In total four constructs were expressed, purified and crystallized (table 9.1, figure 9.3).

Purification of all four constructs was performed and each construct ran as a predicted hexamer by SEC purification. Hexagonal crystals rapidly grew in conditions containing $AmSO_4$ as a precipitant for all four proteins. PduAD2 (‘GSGTGS’ linker) and PduAD4 (‘NHIT’ linker) crystals were extremely fragile with rounded corners. These crystals could not be harvested for diffraction experiments and could not be optimized. The crystals of PduAD1 (‘KGIE’ linker) and PduAD3 (‘RSSG’ linker) were more regular hexagons with well-defined edges and corners. Crystals from both constructs diffracted. The diffraction of PduAD1 was extremely anisotropic with poorly resolved reflections, resulting in difficulty processing the data and solving the structure. PduAD3 was the only construct where a crystal structure could be solved.

9.2.2 PduAD3 Crystal Structures

The first crystal form of PduAD3 grew as hexagonal crystals that diffracted to 2.5Å and were indexed in space group P1. Phases were determined by molecular replacement, which was only successful with the PduA hexamer serving as the search model and contained two full hexamers in the

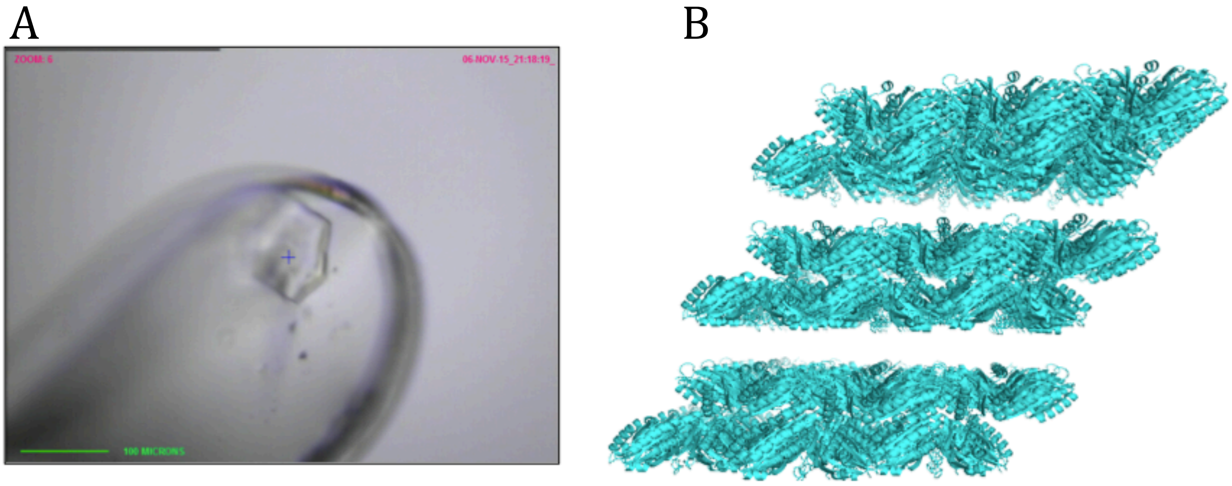


Figure 9.4. Crystal Structure of the PduAD3 P1 Crystal Form. (A) Crystals in the P1 crystal form grew as well-defined hexagons. (B) The crystal structure was found to have layers of the PduA hexamers with a large space between the layers. Although the PpduAD model would full this space, no density was observed besides that of the PduA BMC core.

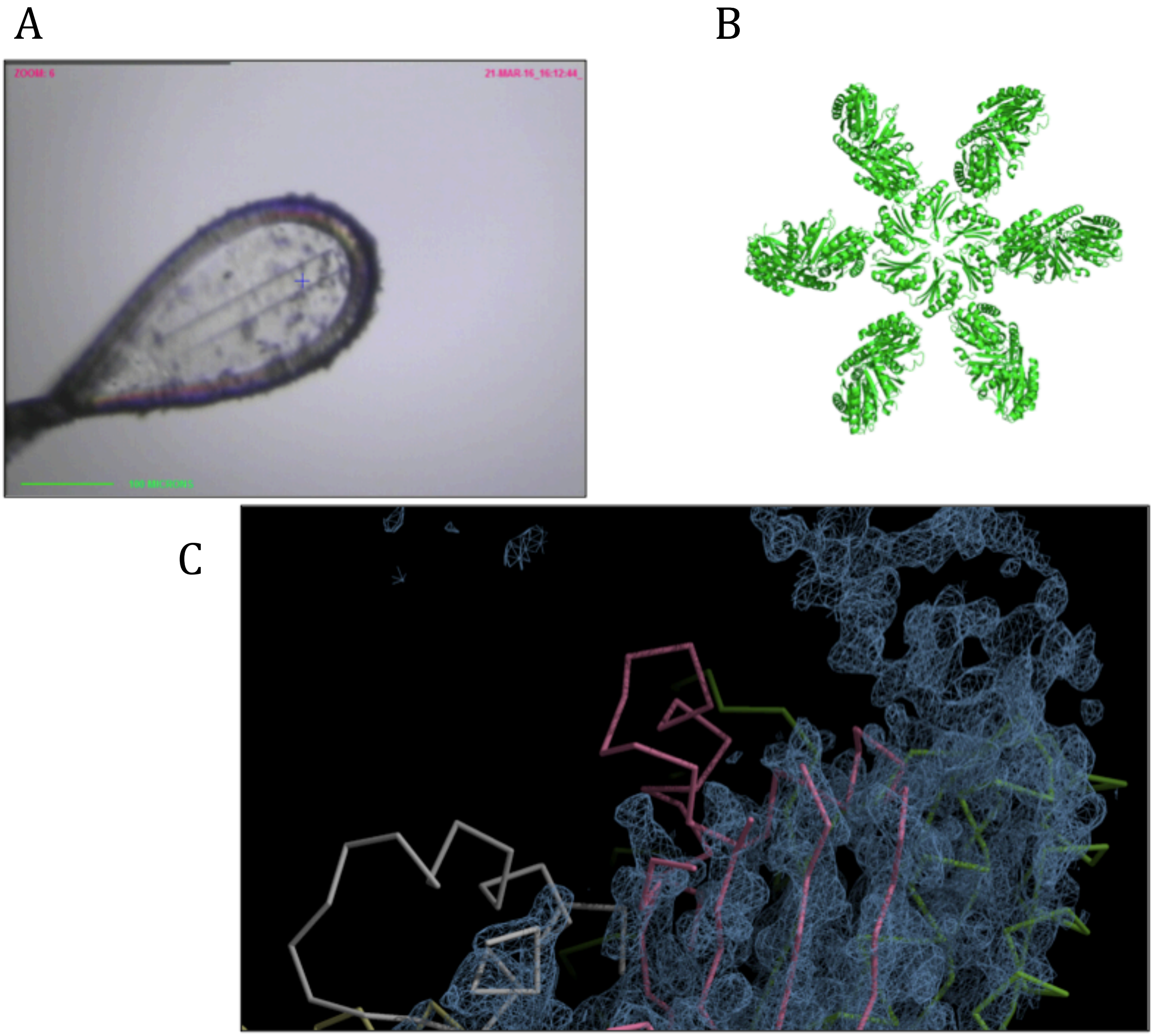


Figure 9.5. Crystal Structure of the PduAD3 P4$_3$22 Crystal Form. (A) Crystals in the P4$_3$22 were found to be large rods. (B) A single hexamer was in the asymmetric unit (central hexamer) with six hexamers forming crystal contacts, one on each edge of the central hexamer. (C) PduAD model is placed into the density of the structure. No electron density is present for the PduD tail helix.

asymmetric unit. The two hexamers packed with the convex surfaces forming a crystal contact with near two-fold symmetry relating the two hexamers. The hexamers tessellated into layers, with a large empty space between the layers (figure 9.4). The PduAD was modeled into the structure and the empty space could be accommodated by the PduD helices. Significant refinement was attempted but electron density for the PduD tails could not been seen and the R-factors for the structure remained high (>40%). The broken six-fold symmetry of this crystal form suggested there might be interactions between the PduA and the PduD tails. Since the density for the tails was completely absent, new crystal forms were sought.

In order obtain a new crystal forms of PduAD3 the original crystallization condition was subject to additive screening.[222–224] Additive screening resulted in large rod crystals in one condition. These crystals diffracted well (2.0Å) and were found to be in space group $P4_3 2 2$. Again only the PduA hexamer could serve as the molecular replacement search model for successful phase determination. In this crystal form, one hexamer comprised the asymmetric unit (figure 9.5). The electron density for the PduA portion of fusion was complete and well defined. Again the electron density ended at the C-terminus of PduA and with no density for the PduD tail. Refinement was performed with Phenix,[91] PDB_Redo,[90] or Buster[211] and did not reveal density for the PduD tail. Finally feature enhanced maps[225] and non crystallographic symmetry[226] maps were generated, but these too failed to reveal density for the tails.

9.3 Discussion

A crystal structure of the PduA-PduD interaction has remained elusive. Multiple factors may contribute to this, including the symmetry of PduA and shell proteins in general. Flaws in the current model of the PduA and its tail may also be factors in the difficulties thus far.

9.3.1 Obscured Electron Density From Six-Fold Symmetry

The six-fold symmetry could be detrimental for crystal structures with bound tails if there is not a 1:1 stoichiometry of PduA and bound tail. It is possible that only one of the tails may bind to PduA with the fusion constructs and the previous peptide soaking and co-crystallization experiments. If this were the case the density of the bound peptide would be obscured by the six-fold symmetry. In the case where one tail is bound, it can bind into any of the six equivalent position of PduA. Further, the PduA hexamer can pack in any orientation since each monomer can form equivalent crystals contacts. If this occurs, electron density of a single bound tail would be reduced to one-sixth from the averaging of the six monomers. This would fully obscure the density of the tail, which would further be affected if not all hexamers had a bound tail. The changes in crystal morphologies, resolution and space groups of the fusion proteins and peptide soaks suggest binding events are occurring to some extent, even though they can not be visualized.

9.3.2 PduAD Crystals

Broken symmetry in the hexagonal crystal form of PduAD3 suggests structural changes are occurring in the crystal. Previous experiments crystallized in similar morphologies but retained crystallographic six-fold symmetry. Breaking this symmetry is an indication that the PduA hexamers are being distorted in some manner. The poorly formed crystals of the other three PduAD constructs are another indication of structural changes possibly induced by tail binding. These structural changes are likely a cause for the poorly ordered crystals and other defects (fragile crystals, high anisotropy and poorly defined reflections).

PduA is highly tolerant to C-terminal extensions. In one study, the researchers neglected to includes a stop codon for their construct. This resulted in fusion of twenty amino acids from the vector to the PduA construct.[227] The resulting structure showed no density for the vector residues, similar to the PduAD3 structure, but no crystal defects were observed. Taken as a whole, these new crystals with defects and the previous structures with terminal fusion and no structural alterations, the targeting tails may be binding to some extent.

9.3.3 Fusion Protein Creates Incorrect Stoichiometry

The Rosetta model suggests that the PduA hexamer can accommodate up to six bound tails. However, this model does not account for any structural changes in the hexamer upon tail binding. Encapsulated Pdu enzymes are quite large (PduP molecular weight is 50 kDa) and it is very unlikely that six enzymes will bind to the PduA hexamer. If this were the case there would be severe clashes between the enzymes bound to each hexamer and adjacent hexamers. It is simply not possible for this to occur *in vivo.* Emerging evidences suggests this is the case. It has been proposed that the tails can self associate in addition to binding to the shells. It is though that enzyme complexes of unknown size assemble, and one component of this complex can then bind the shell.[38] Assembly in this manner could accommodate structural changes in the PduA hexamer when a tail binds, something that can not be accommodated with the PduAD fusion since all six PduA monomers are fused to a PduD tail. Again, if only one tail was bound the electron density would likely be obscured in the crystal structure.

9.3.4 Crystallization May Select For Unbound Structures

Significant structural changes induced by tail binding may perturb the PduA to the point where is cannot readily crystallize. Therefore, crystallization may select for PduA hexamers without bound tails. Additionally the crystallization conditions may interfere with the PduA – tail interactions. PduA readily crystallizes in conditions composed of 1.6-2.0M $AmSO_4$ and high NaCl concentrations (up to 2.0M) and alkaline pHs (pH 8.0 -9.0). These conditions may disrupt the protein interaction between PduA and the tail, especially if the interaction is weak and only occurs transiently during initial MCP formation.[192,228–231] Limited screening of crystallization conditions was performed for the PduAD fusions because of the ease in which PduA derivatives crystalize in $AmSO_4$ conditions. Additional broad condition screening should be performed to determine if other crystallization conditions are available. It is possible that new conditions could favor the PduA-tail interaction if the high salt, alkaline conditions are interfering with the tail binding.

9.3.5 Ambiguity in the Binding Partners and Orientations of Targeting Tails

The only published direct PduA-tail interaction is that of PduA C-terminus to the N-terminal tail of PduP. Computational models suggest that other tails bind PduA but so far there is limited data to confirm this. The PduAD fusion was pursued because the computational modeling suggested the PduD tail bound in a favorable orientation for terminal fusions. If this model were wrong then it would be impossible for the fusions to work. This is a very real possibility given the prediction that the PduP tail binds in the opposite orientation. It remains a possibility that the binding model is drastically wrong as well and the predicted binding cleft along the C-terminal helix of PduA is not the location of binding. The true binding site may be on another location of the PduA hexamer or may occur between multiple shell hexamers. The latter example has yet to be explored experimentally or computationally.

9.3.5.1 Preferential Binding of Tails to Specific Shells

Preliminary binding assays involving Pdu shell proteins and tail peptides have been performed. From these assays, graduate student Sunny Chun has observed a 10-fold weaker affinity of the PduD tail to PduA than the PduP tail (unpublished work). Conversely, PduD has a better affinity for the PduA paralog PduJ over the PduP tail, suggesting that future crystallographic experiments should explore the PduJ interaction with the PduD tail, either through peptide soaks, co-crystallization or PduJ-PduD tail fusions.

PduJ has recently had its crystal structure solved (unpublished). The sequences of PduA and PduJ are remarkably similar (17 amino acids are different) and both form the same predicted binding cleft for the tail helices. The only differences between the putative binding sites are three amino acids under the C-terminal helix. To reconcile the tail binding affinity differences with the similar binding sites, the terminal helix would have to be displaced upon tail binding. If this is true, then major structural changes are likely occurring during the tail binding. These changes would affect the crystal lattice stability and render the PduAD fusion impractical for determining the structure of the interaction.

9.3.6 Crystallographic Rescue Through Broken Symmetry Scaffolds

The discussed issues with crystallization of the shell-tail interaction require new approaches to determine the crystal structure of this interaction. An approach that would remedy all of the issues would be to utilize a large scaffolding protein. With this approach a large protein, whether monomer or oligomer, would serve to form a crystal lattice instead of PduA. To this the tail sequences could be terminally fused on earthier end of the protein. PduA, or another shell protein could then be soaked into crystals if suitable solvent channels are present or added during crystallization at varying concentrations. Under this scenario, PduA could bind to the tails in any orientation at the true binding site and accommodate structural changes in PduA, as PduA structural changes would not perturb the crystal lattice. With this approach, a single binding to the shell could be resolved since there is no six-fold symmetry to average out and obscure the density of a single bound tail.

9.4 Conclusions

A structure revealing the PduA interaction with the targeting tails has remained elusive. Numerous factors may contribute to this. Changes to the crystal forms with the new PduAD fusions suggest binding events may be occurring. However, the exact binding partners to PduA have not been fully validated experimentally and new evidence suggest the PduD tail is the preferred partner for PduA. With the uncertainty in binding partners and binding sites, novel approaches must be attempted to solve the bound structure.

9.5 Materials and Methods

9.5.1 Cloning

For all designed sequences, the gnes were codon opitmized[95] for expression in *E. coli.* Optimized genes were synthesized as gBlocks (Integrated DNA Technologies) with the terminal sequences GTTTAACTTTAAGAAGGAGATATAC and

TAAAATTCGAGCTCCGTCGACAAGCTTG added to the 5’ and 3’ ends respectively. The genes were then cloned via Gibson assembly[75] into pET22b vector cut with NdeI and EcoRI restriction enzymes and transformed into XL-2 Blue Ultracompetent cells (Agilent Technologies). Once the plasmids were sequence validated, plasmids were transformed in BL21 DE3 expression cells (New England Biolabs)

9.5.2 Protein Expression

All constructs were expressed in LB supplemented ampicillin. 1L LB flasks were inoculated with 10mL of overnight culture and allow to grow for at 37°C until the A_{600nm} optical density reached 0.6-0.8, the temperature was then reduced to 30°C and protein expression was induced with 1mM IPTG and grown for 16 hours.

9.5.3 Protein Purification

Purification was performed per previously reported methods.[209,210] SEC was performed in a buffer composed of 30mM Tris pH 9.0, 50mM NaCl fractions corresponding to hexamers was pooled, concentrated to 10mg/mL, flash frozen and stored for crystallization experiments.

9.5.4 PduAD Crystallization

PduAD constructs were screened for crystals only in the $AmSO_4$ screen (Qiagen) at both 10mg/mL and 5mg/mL. Hexagonal rystals grew with in 48 hours for all four constructs in various conditions. Diffraction quality crystals from PduAD3 were grown in a condition of 0.1M Bicine pH 9.0, 1.6M $AmSO_4$. Crystals were cryo-protected with 25% glycerol prior to flash freezing the crystals in liquid nitrogen.

9.5.4.1 PduAD3 Optimization

Additive screening was performed with protein at both 10mg/mL and 5mg/mL concentrations. The well solution was composed of 0.1M Bicine pH 9.0, 1.6M $AmSO_4$. The final condition in which rod crystals were grown was composed of 0.1M Bicine pH 9.0, 1.6M $AmSO_4$, 0.033% w/v 2,7-

naphthalenedisulfonic acid disodium saltm 0.033% w/v azelaic acid, 0.033% trans-Cinnamic acid, 0.002M Hepes pH 6.8. Crystals were cryo-protected with 25% glycerol prior to flash freezing the crystals in liquid nitrogen.

9.5.5 X-ray Data Processing and Structure Solution

All data was collected at the Advance Photon Source beam line 24-ID-C. Data was indexed with XDS/XSCALE,[97] phases were determined with Phaser[88] with the a performed with Phenix[91], Models were manually evaluated in Coot[98] and analyzed with PyMOL.[99]

9.5.5.1 PduAD3 P1 Crystal Form

Data was indexed in space group P1 at a resolution of 2.5Å with a unit cell of a=67.920 b=67.920 c=75.250 α=99.39 β=105.99 γ=120.27. Refinement was attempted with the PduAD Rosetta model and no density could be found for the PduD tail helices.

9.5.5.2 PduAD3 $P4_322$ Crystal Form

Data was indexed in space group $P4_322$ at a resolution of 2.0Å, with a unit cell of a=b=96.640 c=175.18 α=β=γ=90. AgLike with the P1 form refinement was carried out with both PduA hexamer and PduAD Rosetta model and no density could be found for the PduD tail helices. FEM maps were generated with Phenix[91] and NCS maps generates with Coot.[98]

Table 9.1. Sequences of PduAD Fusion Constructs

PduAD1

MHHHHHHQQEALGMVETKGLTAAIEAADAMVASANVMLVGYEKIGSGLVTVIVRGDVGAVKAATDAGAAAARNVGEVKAVHVIPRPHTDVEKILGKGIEINEKLLRQIIEDVL

PduAD2

MHHHHHHQQEALGMVETKGLTAAIEAADAMVASANVMLVGYEKIGSGLVTVIVRGDVGAVKAATDAGAAAARNVGEVKAVHVIPRPHTDVEKILGKGSGTGSGEINEKLLRQIIEDVL

PduAD3

MHHHHHHQQEALGMVETKGLTAAIEAADAMVASANVMLVGYEKIGSGLVTVIVRGDVGAVKAATDAGAAAARNVGEVKAVHVIPRPHTDVEKILGSRSSGSINEKLLRQIIEDVL

PduAD4

MHHHHHHQQEALGMVETKGLTAAIEAADAMVASANVMLVGYEKIGSGLVTVIVRGDVGAVKAATDAGAAAARNVGEVKAVHVIPRPHTDVEKILGSNHITGINEKLLRQIIEDVL

REFERENCES

1. Stacy, R. *et al.* Structural genomics of infectious disease drug targets: the SSGCID. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **67,** 979–984 (2011).
2. Bernhard Rupp. in *Structural Genomics and High Throughput Structural Biology* 61–104 (CRC Press, 2005).
3. Walter, T. S. *et al.* Lysine methylation as a routine rescue strategy for protein crystallization. *Struct. Lond. Engl. 1993* **14,** 1617–1622 (2006).
4. Cooper, D. R. *et al.* Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr. D Biol. Crystallogr.* **63,** 636–645 (2007).
5. Kobe, B., Ve, T. & Williams, S. J. Fusion-protein-assisted protein crystallization. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **71,** 861–869 (2015).
6. Banatao, D. R. *et al.* An approach to crystallizing proteins by synthetic symmetrization. *Proc. Natl. Acad. Sci.* **103,** 16230–16235 (2006).
7. Wukovitz, S. W. & Yeates, T. O. Why protein crystals favour some space-groups over others. *Nat. Struct. Mol. Biol.* **2,** 1062–1067 (1995).
8. Laganowsky, A. *et al.* An approach to crystallizing proteins by metal-mediated synthetic symmetrization. *Protein Sci. Publ. Protein Soc.* **20,** 1876–1890 (2011).
9. Salgado, E. N. *et al.* Metal templated design of protein interfaces. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 1827–1832 (2010).
10. Salgado, E. N., Radford, R. J. & Tezcan, F. A. Metal-directed protein self-assembly. *Acc. Chem. Res.* **43,** 661–672 (2010).

11. Forse, G. J. *et al.* Synthetic symmetrization in the crystallization and structure determination of CelA from Thermotoga maritima. *Protein Sci. Publ. Protein Soc.* **20,** 168–178 (2011).
12. Cabantous, S. & Waldo, G. S. In vivo and in vitro protein solubility assays using split GFP. *Nat. Methods* **3,** 845–854 (2006).
13. Nguyen, H. B., Hung, L.-W., Yeates, T. O., Terwilliger, T. C. & Waldo, G. S. Split green fluorescent protein as a modular binding partner for protein crystallization. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 2513–2523 (2013).
14. Radford, R. J., Nguyen, P. C., Ditri, T. B., Figueroa, J. S. & Tezcan, F. A. Controlled protein dimerization through hybrid coordination motifs. *Inorg. Chem.* **49,** 4362–4369 (2010).
15. 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).
16. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487,** 545–574 (2011).
17. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510,** 103–108 (2014).
18. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336,** 1171–1174 (2012).
19. Cabantous, S. *et al.* A New Protein-Protein Interaction Sensor Based on Tripartite Split-GFP Association. *Sci. Rep.* **3,** (2013).
20. Avnir, D., Braun, S., Lev, O. & Ottolenghi, M. Enzymes and Other Proteins Entrapped in Sol-Gel Materials. *Chem. Mater.* **6,** 1605–1614 (1994).

21. Ellerby, L. M. *et al.* Encapsulation of proteins in transparent porous silicate glasses prepared by the sol-gel method. *Science* **255,** 1113–1115 (1992).
22. Wang, X. *et al.* Sol-gel encapsulation of biomolecules and cells for medicinal applications. *Curr. Top. Med. Chem.* **15,** 223–244 (2015).
23. Hu, J., Zhang, G. & Liu, S. Enzyme-responsive polymeric assemblies, nanoparticles and hydrogels. *Chem. Soc. Rev.* **41,** 5933–5949 (2012).
24. Jorda, J., Lopez, D., Wheatley, N. M. & Yeates, T. O. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci.* **22,** 179–195 (2013).
25. Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. & Bobik, T. A. Diverse bacterial microcompartment organelles. *Microbiol. Mol. Biol. Rev. MMBR* **78,** 438–468 (2014).
26. Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C. & Shively, J. M. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.* **6,** 681–691 (2008).
27. Rae, B. D., Long, B. M., Badger, M. R. & Price, G. D. Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO2 fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev. MMBR* **77,** 357–379 (2013).
28. Frank, S., Lawrence, A. D., Prentice, M. B. & Warren, M. J. Bacterial microcompartments moving into a synthetic biological world. *J. Biotechnol.* **163,** 273–279 (2013).
29. Tsai, S. J. & Yeates, T. O. Bacterial microcompartments insights into the structure, mechanism, and engineering applications. *Prog. Mol. Biol. Transl. Sci.* **103,** 1–20 (2010).

30. Lawrence, A. D. *et al.* Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. *ACS Synth. Biol.* **3,** 454–465 (2014).

31. Yeates, T. O., Jorda, J. & Bobik, T. A. The shells of BMC-type microcompartment organelles in bacteria. *J. Mol. Microbiol. Biotechnol.* **23,** 290–299 (2013).

32. Yeates, T. O., Thompson, M. C. & Bobik, T. A. The protein shells of bacterial microcompartment organelles. *Curr. Opin. Struct. Biol.* **21,** 223–231 (2011).

33. Yeates, T. O., Crowley, C. S. & Tanaka, S. Bacterial microcompartment organelles: protein shell structure and evolution. *Annu. Rev. Biophys.* **39,** 185–205 (2010).

34. Havemann, G. D., Sampson, E. M. & Bobik, T. A. PduA is a shell protein of polyhedral organelles involved in coenzyme B(12)-dependent degradation of 1,2-propanediol in Salmonella enterica serovar typhimurium LT2. *J. Bacteriol.* **184,** 1253–1261 (2002).

35. Crowley, C. S. *et al.* Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. *J. Biol. Chem.* **285,** 37838–37846 (2010).

36. Chowdhury, C. *et al.* Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 2990–2995 (2015).

37. Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A. & Polka, J. K. The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. *PLoS One* **8,** e76127 (2013).

38. Aussignargues, C., Paasch, B. C., Gonzalez-Esquer, R., Erbilgin, O. & Kerfeld, C. A. Bacterial microcompartment assembly: The key role of encapsulation peptides. *Commun. Integr. Biol.* **8,** e1039755 (2015).

39. Fan, C. *et al.* Short N-terminal sequences package proteins into bacterial microcompartments. *Proc. Natl. Acad. Sci.* **107,** 7509–7514 (2010).

40. Fan, C., Cheng, S., Sinha, S. & Bobik, T. A. Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 14995–15000 (2012).

41. Jorda, J., Liu, Y., Bobik, T. A. & Yeates, T. O. Exploring bacterial organelle interactomes: a model of the protein-protein interaction network in the Pdu microcompartment. *PLoS Comput Biol* **11,** e1004067 (2015).

42. Leibly, D. J. *et al.* A Suite of Engineered GFP Molecules for Oligomeric Scaffolding. *Structure* **23,** 1754–1768 (2015).

43. Chun, W., Waldo, G. S. & Johnson, G. V. Split GFP complementation assay for quantitative measurement of tau aggregation in situ. *Methods Mol Biol* **670,** 109–123 (2011).

44. Chun, W., Waldo, G. S. & Johnson, G. V. Split GFP complementation assay: a novel approach to quantitatively measure aggregation of tau in situ: effects of GSK3β activation and caspase 3 cleavage. *J. Neurochem.* **103,** 2529–2539 (2007).

45. Kaddoum, L., Magdeleine, E., Waldo, G. S., Joly, E. & Cabantous, S. One-step split GFP staining for sensitive protein detection and localization in mammalian cells. *Biotechniques* **49,** 727–8 (2010).

46. Listwan, P., Terwilliger, T. C. & Waldo, G. S. Automated, high-throughput platform for protein solubility screening using a split-GFP system. *J. Struct. Funct. Genomics* **10,** 47–55 (2009).

47. Waldo, G. S. & Cabantous, S. *Protein subcellular localization assays using split fluorescent proteins*. (2009).

48. Judge, R. A., Forsythe, E. L. & Pusey, M. L. The effect of protein impurities on lysozyme crystal growth. *Biotechnol. Bioeng.* **59,** 776–785 (1998).

49. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88 (2006).

50. Parekh, B. S., Schwimmbeck, P. W. & Buchmeier, M. J. High efficiency immunoaffinity purification of anti-peptide antibodies on thiopropyl sepharose immunoadsorbants. *Pept. Res.* **2,** 249–252 (1989).

51. Makrides, S. C. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiol. Rev.* **60,** 512–538 (1996).

52. Denoncin, K. & Collet, J.-F. Disulfide bond formation in the bacterial periplasm: major achievements and challenges ahead. *Antioxid. Redox Signal.* **19,** 63–71 (2013).

53. de Marco, A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microb. Cell Factories* **8,** 26 (2009).

54. Jeong, H. *et al.* Genome sequences of Escherichia coli B strains REL606 and BL21(DE3). *J. Mol. Biol.* **394,** 644–652 (2009).

55. Ojo, K. K. *et al.* Structure determination of glycogen synthase kinase-3 from Leishmania major and comparative inhibitor structure-activity relationships with Trypanosoma brucei GSK-3. *Mol. Biochem. Parasitol.* **176,** 98–108 (2011).

56. Tanaka, Y. *et al.* Structural implications for heavy metal-induced reversible assembly and aggregation of a protein: the case of Pyrococcus horikoshii CutA1. *FEBS Lett.* **556,** 167–174 (2004).

57. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).

58. Yamaguchi, H. & Miyazaki, M. Refolding Techniques for Recovering Biologically Active Recombinant Proteins from Inclusion Bodies. *Biomolecules* **4,** 235–251 (2014).

59. Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. & Dougherty, W. G. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal. Biochem.* **216,** 413–417 (1994).

60. Oganesyan, N., Kim, S.-H. & Kim, R. On-column protein refolding for crystallization. *J. Struct. Funct. Genomics* **6,** 177–182 (2005).

61. Mogridge, J. Using light scattering to determine the stoichiometry of protein complexes. *Methods Mol. Biol. Clifton NJ* **1278,** 233–238 (2015).

62. Folta-Stogniew, E. J. in *eLS* (John Wiley & Sons, Ltd, 2001).

63. Berne, B. J. & Pecora, R. *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics*. (Courier Corporation, 1976).

64. Slabinski, L. *et al.* XtalPred: a web server for prediction of protein crystallizability. *Bioinforma. Oxf. Engl.* **23,** 3403–3405 (2007).

65. Pearse, B. M. Clathrin: a unique protein associated with intracellular transfer of membrane by coated vesicles. *Proc. Natl. Acad. Sci. U. S. A.* **73,** 1255–1259 (1976).

66. Blatch, G. L. & Lässle, M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **21,** 932–939 (1999).

67. Doyle, L. *et al.* Rational design of α-helical tandem repeat proteins with closed architectures. *Nature* **528,** 585–588 (2015).

68. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.* **5,** (2014).

69. Torres, J. Z. STARD9/Kif16a is a novel mitotic kinesin and antimitotic target. *Bioarchitecture* **2,** 19–22 (2012).

70. Torres, J. Z. *et al.* The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly. *Cell* **147,** 1309–1323 (2011).

71. von Stetten, D., Noirclerc-Savoye, M., Goedhart, J., Gadella, T. W. J. & Royant, A. Structure of a fluorescent protein from Aequorea victoria bearing the obligate-monomer mutation A206K. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **68,** 878–882 (2012).

72. Sousa, R. Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. *Acta Crystallogr. D Biol. Crystallogr.* **51,** 271–277 (1995).

73. Trakhanov, S. & Quiocho, F. A. Influence of divalent cations in protein crystallization. *Protein Sci. Publ. Protein Soc.* **4,** 1914–1919 (1995).

74. Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr. D Biol. Crystallogr.* **50,** 414–423 (1994).
75. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
76. 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).
77. Lai, Y.-T., Tsai, K.-L., Sawaya, M. R., Asturias, F. J. & Yeates, T. O. Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. *J. Am. Chem. Soc.* **135,** 7738–7743 (2013).
78. Lai, Y.-T., King, N. P. & Yeates, T. O. Principles for designing ordered protein assemblies. *Trends Cell Biol.* **22,** 653–661 (2012).
79. Lai, Y.-T., Cascio, D. & Yeates, T. Crystal structure of a 16 nm, half-megadalton protein cage designed by fusing symmetric oligomeric domains. in *Protein Science* **21,** 115–116 (WILEY-BLACKWELL 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2012).
80. Lai, Y.-T., Cascio, D. & Yeates, T. O. Structure of a 16-nm cage designed by using protein oligomers. *Science* **336,** 1129–1129 (2012).
81. Lai, Y.-T., Jiang, L., Chen, W. & Yeates, T. O. On the predictability of the orientation of protein domains joined by a spanning alpha-helical linker. *Protein Eng. Des. Sel.* gzv035 (2015).

82. Zhang, G., Gurtu, V. & Kain, S. R. An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochem. Biophys. Res. Commun.* **227,** 707–711 (1996).
83. Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **67,** 509–544 (1998).
84. Stepanenko, O. V., Stepanenko, O. V., Kuznetsova, I. M., Verkhusha, V. V. & Turoverov, K. K. Beta-Barrel Scaffold of Fluorescent Proteins: Folding, Stability and Role in Chromophore Formation. *Int. Rev. Cell Mol. Biol.* **302,** 221–278 (2013).
85. Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A. & Getzoff, E. D. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc. Natl. Acad. Sci.* **100,** 12111–12116 (2003).
86. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33,** 491–497 (1968).
87. Kantardjieff, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci. Publ. Protein Soc.* **12,** 1865–1871 (2003).
88. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **63,** 32–41 (2006).
89. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242 (2011).
90. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1,** 213–220 (2014).
91. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).
92. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

93. Nagai, T., Sawano, A., Park, E. S. & Miyawaki, A. Circularly permuted green fluorescent proteins engineered to sense Ca2+. *Proc. Natl. Acad. Sci.* **98,** 3197–3202 (2001).

94. Topell, S., Hennecke, J. & Glockshuber, R. Circularly permuted variants of the green fluorescent protein. *FEBS Lett.* **457,** 283–289 (1999).

95. Puigbò, P., Guzmán, E., Romeu, A. & Garcia-Vallvé, S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res.* **35,** W126–W131 (2007).

96. Choi, R. *et al.* Immobilized metal-affinity chromatography protein-recovery screening is predictive of crystallographic structure success. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **67,** 998–1005 (2011).

97. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 (2010).

98. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).

99. *The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.*

100. Chun, E. *et al.* Fusion Partner Toolchest for the Stabilization and Crystallization of G Protein-Coupled Receptors. *Struct. England1993* **20,** 967–976 (2012).

101. Zou, Y., Weis, W. I. & Kobilka, B. K. N-Terminal T4 Lysozyme Fusion Facilitates Crystallization of a G Protein Coupled Receptor. *PLOS ONE* **7,** e46039 (2012).

102. Rosenbaum, D. M. *et al.* GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **318,** 1266–1273 (2007).

103. Cherezov, V. *et al.* High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318,** 1258–1265 (2007).

104. Baase, W. A., Liu, L., Tronrud, D. E. & Matthews, B. W. Lessons from the lysozyme of phage T4. *Protein Sci. Publ. Protein Soc.* **19,** 631–641 (2010).

105. Mooers, B. H. M., Tronrud, D. E. & Matthews, B. W. Evaluation at atomic resolution of the role of strain in destabilizing the temperature-sensitive T4 lysozyme mutant Arg 96 → His. *Protein Sci.* **18,** 863–870 (2009).

106. Moon, A. F., Mueller, G. A., Zhong, X. & Pedersen, L. C. A synergistic approach to protein crystallization: Combination of a fixed-arm carrier with surface entropy reduction. *Protein Sci. Publ. Protein Soc.* **19,** 901–913 (2010).

107. Hurst, H. C. Transcription factors 1: bZIP proteins. *Protein Profile* **2,** 101–168 (1995).

108. Herdegen, T. & Leah, J. D. Inducible and constitutive transcription factors in the mammalian nervous system: control of gene expression by Jun, Fos and Krox, and CREB/ATF proteins. *Brain Res. Brain Res. Rev.* **28,** 370–490 (1998).

109. Glover, J. N. & Harrison, S. C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **373,** 257–261 (1995).

110. O'Shea, E. K., Rutkowski, R., Stafford, W. F. & Kim, P. S. Preferential heterodimer formation by isolated leucine zippers from fos and jun. *Science* **245,** 646–648 (1989).

111. O'Shea, E. K., Lumb, K. J. & Kim, P. S. Peptide 'Velcro': design of a heterodimeric coiled coil. *Curr. Biol. CB* **3,** 658–667 (1993).

112. Arai, R., Ueda, H., Kitayama, A., Kamiya, N. & Nagamune, T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng.* **14,** 529–532 (2001).

113. Chen, X., Zaro, J. & Shen, W.-C. Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug Deliv. Rev.* **65,** 1357–1369 (2013).

114. Reddy Chichili, V. P., Kumar, V. & Sivaraman, J. Linkers in the structural biology of protein–protein interactions. *Protein Sci. Publ. Protein Soc.* **22,** 153–167 (2013).

115. Uversky, V. N. The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini. *FEBS Lett.* **587,** 1891–1901 (2013).

116. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,** 845–858 (2015).

117. Walavalkar, N. M., Gordon, N. & Williams, D. C. Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine binding domain 2 (MBD2) homologues and GATA zinc finger domain containing 2A (GATAD2A/p66α). *J. Biol. Chem.* **288,** 3419–3427 (2013).

118. 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. App.* **715,** 307–329 (1998).

119. Lindhout, D. A., Litowski, J. R., Mercier, P., Hodges, R. S. & Sykes, B. D. NMR solution structure of a highly stable de novo heterodimeric coiled-coil. *Biopolymers* **75,** 367–375 (2004).

120. Litowski, J. R. & Hodges, R. S. in *Peptides for the New Millennium* (eds. Fields, G. B., Tam, J. P. & Barany, G.) 285–286 (Springer Netherlands, 2002).

121. Hirano, T. & Mitchison, T. J. A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. *Cell* **79,** 449–458 (1994).

122. Ren, G. & Bardwell, J. C. A. Engineered Pathways for Correct Disulfide Bond Oxidation. *Antioxid. Redox Signal.* **14,** 2399–2412 (2011).

123. Kaplan, A. *et al.* Small molecule-induced oxidation of protein disulfide isomerase is neuroprotective. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E2245–2252 (2015).

124. Bennett, K. L. *et al.* Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping--a novel approach to assess intermolecular protein contacts. *Protein Sci. Publ. Protein Soc.* **9,** 1503–1518 (2000).

125. del Rosario, R. B., Wahl, R. L., Brocchini, S. J., Lawton, R. G. & Smith, R. H. Sulfhydryl site-specific cross-linking and labeling of monoclonal antibodies by a fluorescent equilibrium transfer alkylation cross-link reagent. *Bioconjug. Chem.* **1,** 51–59 (1990).

126. Hwang, Y.-J., Granelli, J. & Lyubovitsky, J. Effects of zero-length and non-zero-length cross-linking reagents on the optical spectral properties and structures of collagen hydrogels. *ACS Appl. Mater. Interfaces* **4,** 261–267 (2012).

127. El-Shafey, A. *et al.* 'Zero-length' cross-linking in solid state as an approach for analysis of protein-protein interactions. *Protein Sci. Publ. Protein Soc.* **15,** 429–440 (2006).

128. Geer, L. Y. *et al.* The NCBI BioSystems database. *Nucleic Acids Res.* **38,** D492–496 (2010).

129. Bryan, C. M. *et al.* High-throughput protein production and purification at the Seattle Structural Genomics Center for Infectious Disease. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **67,** 1010–1014 (2011).

130. Johnson, J. E. & Speir, J. A. Quasi-equivalent viruses: a paradigm for protein assemblies. *J. Mol. Biol.* **269,** 665–675 (1997).

131. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 105–153 (2000).

132. Levy, E. D. & Teichmann, S. Structural, evolutionary, and assembly principles of protein oligomerization. *Prog. Mol. Biol. Transl. Sci.* **117,** 25–51 (2013).

133. Douglas, T., Allen, M. & Young, M. in *Biopolymers Online* (Wiley-VCH Verlag GmbH & Co. KGaA, 2005).

134. Luo, Y., Wang, X., Du, D. & Lin, Y. Hyaluronic acid-conjugated apoferritin nanocages for lung cancer targeted drug delivery. *Biomater. Sci.* **3,** 1386–1394 (2015).

135. Chruszcz, M. *et al.* Analysis of solvent content and oligomeric states in protein crystals--does symmetry matter? *Protein Sci. Publ. Protein Soc.* **17,** 623–632 (2008).

136. Flenniken, M. L. *et al.* A library of protein cage architectures as nanomaterials. *Curr. Top. Microbiol. Immunol.* **327,** 71–93 (2009).

137. 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).

138. Lai, Y.-T. *et al.* Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem.* **6,** 1065–1071 (2014).

139. Chayen, N. E. A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *J. Appl. Crystallogr.* **30,** 198–202 (1997).

140. Huo, L. *et al.* The CC1-FHA tandem as a central hub for controlling the dimerization and activation of kinesin-3 KIF1A. *Struct. Lond. Engl. 1993* **20,** 1550–1561 (2012).

141. Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. The importance of alignment accuracy for molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 1229–1236 (2004).

142. Keegan, R. M. & Winn, M. D. MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **64,** 119–124 (2008).

143. Bunkóczi, G. *et al.* Phaser.MRage: automated molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 2276–2286 (2013).

144. Booth, D. S., Avila-Sakar, A. & Cheng, Y. Visualizing Proteins and Macromolecular Complexes by Negative Stain EM: from Grid Preparation to Image Acquisition. *J. Vis. Exp. JoVE* (2011). doi:10.3791/3227

145. Rames, M., Yu, Y. & Ren, G. Optimized negative staining: a high-throughput protocol for examining small and asymmetric protein structure by electron microscopy. *J. Vis. Exp. JoVE* e51087 (2014). doi:10.3791/51087

146. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative Staining and Image Classification – Powerful Tools in Modern Electron Microscopy. *Biol. Proced. Online* **6,** 23–34 (2004).

147. Boekema, E. J., Folea, M. & Kouřil, R. Single particle electron microscopy. *Photosynth. Res.* **102,** 189–196 (2009).

148. Mertens, H. D. T. & Svergun, D. I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172,** 128–141 (2010).

149. Skou, S., Gillilan, R. E. & Ando, N. Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nat. Protoc.* **9,** 1727–1739 (2014).

150. Zhao, Z. *et al.* Nanocaged enzymes with enhanced catalytic activity and increased stability against protease digestion. *Nat. Commun.* **7,** 10619 (2016).

151. Jeong, W. H. *et al.* Connecting two proteins using a fusion alpha helix stabilized by a chemical cross linker. *Nat. Commun.* **7,** 11031 (2016).

152. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G. & Galas, D. J. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res.* **17,** 6545–6551 (1989).

153. Kunkel, J. & Asuri, P. Function, Structure, and Stability of Enzymes Confined in Agarose Gels. *PLOS ONE* **9,** e86785 (2014).

154. Phelps, E. A. *et al.* Maleimide cross-linked bioactive PEG hydrogel exhibits improved reaction kinetics and cross-linking for cell encapsulation and in situ delivery. *Adv. Mater. Deerfield Beach Fla* **24,** 64–70, 2 (2012).

155. Asthana, A. *et al.* Bromo-oxidation reaction in enzyme-entrapped alginate hollow microfibers. *Biomicrofluidics* **5,** (2011).

156. Cao, L. *Carrier-bound Immobilized Enzymes: Principles, Application and Design*. (John Wiley & Sons, 2006).

157. Macario, A., Moliner, M., Corma, A. & Giordano, G. Increasing stability and productivity of lipase enzyme by encapsulation in a porous organic–inorganic system. *Microporous Mesoporous Mater.* **118,** 334–340 (2009).

158. Polshettiwar, V. & Asefa, T. *Nanocatalysis: Synthesis and Applications*. (John Wiley & Sons, 2013).

159. Spivey, H. O. & Ovádi, J. Substrate channeling. *Methods San Diego Calif* **19,** 306–321 (1999).

160. Matthaei, J. F. *et al.* Designing Two-Dimensional Protein Arrays through Fusion of Multimers and Interface Mutations. *Nano Lett.* **15,** 5235–5239 (2015).

161. Suzuki, Y. *et al.* Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533,** 369–373 (2016).

162. Gonen, S., DiMaio, F., Gonen, T. & Baker, D. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **348,** 1365–1368 (2015).

163. Hecht, H. J., Sobek, H., Haag, T., Pfeifer, O. & van Pée, K. H. The metal-ion-free oxidoreductase from Streptomyces aureofaciens has an alpha/beta hydrolase fold. *Nat. Struct. Biol.* **1,** 532–537 (1994).

164. Weng, M., Pfeifer, O., Krauss, S., Lingens, F. & van Pée, K. H. Purification, characterization and comparison of two non-haem bromoperoxidases from Streptomyces aureofaciens ATCC 10762. *J. Gen. Microbiol.* **137,** 2539–2546 (1991).

165. Di Toro, S., Zanaroli, G. & Fava, F. Intensification of the aerobic bioremediation of an actual site soil historically contaminated by polychlorinated biphenyls (PCBs) through bioaugmentation with a non acclimated, complex source of microorganisms. *Microb. Cell Factories* **5,** 11 (2006).

166. Michel, F. C., Quensen, J. & Reddy, C. A. Bioremediation of a PCB-Contaminated Soil Via Composting. *Compost Sci. Util.* **9,** 274–284 (2001).

167. Passatore, L., Rossetti, S., Juwarkar, A. A. & Massacci, A. Phytoremediation and bioremediation of polychlorinated biphenyls (PCBs): State of knowledge and research perspectives. *J. Hazard. Mater.* **278,** 189–202 (2014).

168. Singer, A. C., Gilbert, E. S., Luepromchai, E. & Crowley, D. E. Bioremediation of polychlorinated biphenyl-contaminated soil using carvone and surfactant-grown bacteria. *Appl. Microbiol. Biotechnol.* **54,** 838–843 (2000).

169. Vasilyeva, G. K. & Strijakova, E. R. Bioremediation of soils and sediments contaminated by polychlorinated biphenyls. *Microbiology* **76,** 639–653 (2007).

170. Kumar, P. *et al.* Structural insight into the expanded PCB-degrading abilities of a biphenyl dioxygenase obtained by directed evolution. *J. Mol. Biol.* **405,** 531–547 (2011).

171. Wagner, C., Molitor, I. M. & König, G. M. Critical view on the monochlorodimedone assay utilized to detect haloperoxidase activity. *Phytochemistry* **69,** 323–332 (2008).

172. Itoh, N., Morinaga, N. & Nomura, A. A variety of catalases and bromoperoxidases in genus Pseudomonas and their characterization. *Biochim. Biophys. Acta* **1122,** 189–195 (1992).

173. Pée, K. H. van & Lingens, F. Purification of bromoperoxidase from Pseudomonas aureofaciens. *J. Bacteriol.* **161,** 1171–1175 (1985).

174. van Pée, K.-H. & Lingens, F. Detection of a bromoperoxidase in Streptomyces phaeochromogenes. *FEBS Lett.* **173,** 5–8 (1984).

175. Wiesner, W., van Pee, K.-H. & Lingens, F. Detection of a new chloroperoxidase in Pseudomonas pyrrocinia. *FEBS Lett.* **209,** 321–324 (1986).

176. Hülsmeyer, M. *et al.* Crystal structure of cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase from a PCB degrader at 2.0 A resolution. *Protein Sci. Publ. Protein Soc.* **7,** 1286–1293 (1998).

177. Dai, S. *et al.* Identification and analysis of a bottleneck in PCB biodegradation. *Nat. Struct. Biol.* **9,** 934–939 (2002).

178. Horsman, G. P. *et al.* Kinetic and structural insight into the mechanism of BphD, a C-C bond hydrolase from the biphenyl degradation pathway. *Biochemistry (Mosc.)* **45,** 11071–11086 (2006).

179. Nandhagopal, N. *et al.* Crystal structure of 2-hydroxyl-6-oxo-6-phenylhexa-2,4-dienoic acid (HPDA) hydrolase (BphD enzyme) from the Rhodococcus sp. strain RHA1 of the PCB degradation pathway. *J. Mol. Biol.* **309,** 1139–1151 (2001).

180. Qu, Y. *et al.* Catalytic properties of 2,3-dihydroxybiphenyl 1,2-dioxygenase from Dyella Ginsengisoli LA-4 immobilized on mesoporous silica SBA-15. *J. Mol. Catal. B Enzym.* **99,** 136–142 (2014).

181. Stayner, R. S., Min, D.-J., Kiser, P. F. & Stewart, R. J. Site-specific cross-linking of proteins through tyrosine hexahistidine tags. *Bioconjug. Chem.* **16,** 1617–1623 (2005).

182. Minamihata, K., Goto, M. & Kamiya, N. Site-Specific Protein Cross-Linking by Peroxidase-Catalyzed Activation of a Tyrosine-Containing Peptide Tag. *Bioconjug. Chem.* **22,** 74–81 (2011).

183. Umeda, A., Thibodeaux, G. N., Zhu, J., Lee, Y. & Zhang, Z. J. Site-specific Protein Cross-Linking with Genetically Incorporated 3,4-Dihydroxy-L-Phenylalanine. *ChemBioChem* **10,** 1302–1304 (2009).

184. Kamiya, N., Takazawa, T., Tanaka, T., Ueda, H. & Nagamune, T. Site-specific cross-linking of functional proteins by transglutamination. *Enzyme Microb. Technol.* **33,** 492–496 (2003).

185. Behrens, C. R. & Liu, B. Methods for site-specific drug conjugation to antibodies. *mAbs* **6,** 46–53 (2014).

186. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U. S. A.* **109,** E690–697 (2012).

187. Ohashi, K. & Mizuno, K. A novel pair of split venus fragments to detect protein-protein interactions by in vitro and in vivo bimolecular fluorescence complementation assays. *Methods Mol. Biol. Clifton NJ* **1174,** 247–262 (2014).

188. Rush, C. *et al.* Purification, crystallisation and preliminary X-ray analysis of the vanadium-dependent haloperoxidase from Corallina officinalis. *FEBS Lett.* **359,** 244–246 (1995).

189. Dehouck, Y., Kwasigroch, J. M., Rooman, M. & Gilis, D. BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations. *Nucleic Acids Res.* **41,** W333–339 (2013).

190. Crowley, C. S., Sawaya, M. R., Bobik, T. A. & Yeates, T. O. Structure of the PduU Shell Protein from the Pdu Microcompartment of Salmonella. *Structure* **16,** 1324–1332 (2008).

191. Hall, Z., Hernández, H., Marsh, J. A., Teichmann, S. A. & Robinson, C. V. The role of salt bridges, charge density, and subunit flexibility in determining disassembly routes of protein complexes. *Struct. Lond. Engl. 1993* **21,** 1325–1337 (2013).

192. Dumetz, A. C., Snellinger-O'Brien, A. M., Kaler, E. W. & Lenhoff, A. M. Patterns of protein–protein interactions in salt solutions and implications for protein crystallization. *Protein Sci. Publ. Protein Soc.* **16,** 1867–1877 (2007).

193. Bucher, M. H., Evdokimov, A. G. & Waugh, D. S. Differential effects of short affinity tags on the crystallization of Pyrococcus furiosus maltodextrin-binding protein. *Acta Crystallogr. D Biol. Crystallogr.* **58,** 392–397 (2002).

194. Carson, M., Johnson, D. H., McDonald, H., Brouillette, C. & Delucas, L. J. His-tag impact on structure. *Acta Crystallogr. D Biol. Crystallogr.* **63,** 295–301 (2007).

195. Perron-Savard, P., De Crescenzo, G. & Le Moual, H. Dimerization and DNA binding of the Salmonella enterica PhoP response regulator are phosphorylation independent. *Microbiol. Read. Engl.* **151,** 3979–3987 (2005).

196. Ledent, P. *et al.* Unexpected influence of a C-terminal-fused His-tag on the processing of an enzyme and on the kinetic and folding parameters. *FEBS Lett.* **413,** 194–196 (1997).

197. Klose, J. *et al.* Hexa-histidin tag position influences disulfide structure but not binding behavior of in vitro folded N-terminal domain of rat corticotropin-releasing factor receptor type 2a. *Protein Sci. Publ. Protein Soc.* **13,** 2470–2475 (2004).

198. Woodall, N. B., Yin, Y. & Bowie, J. U. Dual-topology insertion of a dual-topology membrane protein. *Nat. Commun.* **6,** 8099 (2015).

199. Kapust, R. B. & Waugh, D. S. Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr. Purif.* **19,** 312–318 (2000).

200. Uhlmann, F., Wernic, D., Poupart, M.-A., Koonin, E. V. & Nasmyth, K. Cleavage of Cohesin by the CD Clan Protease Separin Triggers Anaphase in Yeast. *Cell* **103,** 375–386 (2000).

201. Wernimont, A. & Edwards, A. In Situ Proteolysis to Generate Crystals for Structure Determination: An Update. *PLOS ONE* **4,** e5094 (2009).

202. Satoh, A. & Warren, G. In situ cleavage of the acidic domain from the p115 tether inhibits exocytic transport. *Traffic Cph. Den.* **9,** 1522–1529 (2008).

203. Shih, Y.-P., Wu, H.-C., Hu, S.-M., Wang, T.-F. & Wang, A. H.-J. Self-cleavage of fusion protein in vivo using TEV protease to yield native protein. *Protein Sci. Publ. Protein Soc.* **14,** 936–941 (2005).

204. Waugh, D. S. An Overview of Enzymatic Reagents for the Removal of Affinity Tags. *Protein Expr. Purif.* **80,** 283–293 (2011).

205. Zhao, X. *et al.* Several Affinity Tags Commonly Used in Chromatographic Purification, Several Affinity Tags Commonly Used in Chromatographic Purification. *J. Anal. Methods Chem. J. Anal. Methods Chem.* **2013, 2013,** e581093 (2013).

206. Novick, R. P. Plasmid incompatibility. *Microbiol. Rev.* **51,** 381–395 (1987).

207. Wheatley, N. M., Gidaniyan, S. D., Liu, Y., Cascio, D. & Yeates, T. O. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. *Protein Sci.* **22,** 660–665 (2013).

208. Choudhary, S., Quin, M. B., Sanders, M. A., Johnson, E. T. & Schmidt-Dannert, C. Engineered Protein Nano-Compartments for Targeted Enzyme Localization. *PLOS ONE* **7,** e33342 (2012).

209. Jorda, J., Leibly, D. J., Thompson, M. C. & Yeates, T. O. Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein. *Chem. Commun.* **52,** 5041–5044 (2016).

210. Sinha, S. *et al.* Alanine Scanning Mutagenesis Identifies an Asparagine–Arginine–Lysine Triad Essential to Assembly of the Shell of the Pdu Microcompartment. *J. Mol. Biol.* **426,** 2328–2345 (2014).

211. *BUSTER version 2.10.0.* (Cambridge, United Kingdom: Global Phasing Ltd., 2016).

212. Shively, J. M. & English, R. S. The carboxysome, a prokaryotic organelle: a mini-review. *Can. J. Bot.* **69,** 957–962 (1991).

213. Cannon, G. C. *et al.* Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra. *Appl. Environ. Microbiol.* **67,** 5351–5361 (2001).

214. Chaijarasphong, T. *et al.* Programmed Ribosomal Frameshifting Mediates Expression of the α-Carboxysome. *J. Mol. Biol.* **428,** 153–164 (2016).

215. Fan, C. & Bobik, T. A. The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment. *J. Bacteriol.* **193,** 5623–5628 (2011).

216. Dahlgren, K., Bloch, S., Perdue, S. & Schmidt-Dannert, C. Identification of Specific Cargo and Shell Protein Binding Partners in Eut BMCs. (2015).

217. Held, M. *et al.* Engineering formation of multiple recombinant Eut protein nanocompartments in E. coli. *Sci. Rep.* **6,** (2016).

218. Akita, K. *et al.* Purification and some properties of wild-type and N-terminal-truncated ethanolamine ammonia-lyase of Escherichia coli. *J. Biochem. (Tokyo)* **147,** 83–93 (2010).

219. Tobimatsu, T., Kawata, M. & Toraya, T. The N-terminal regions of beta and gamma subunits lower the solubility of adenosylcobalamin-dependent diol dehydratase. *Biosci. Biotechnol. Biochem.* **69,** 455–462 (2005).

220. Aussignargues, C., Paasch, B. C., Gonzalez-Esquer, R., Erbilgin, O. & Kerfeld, C. A. Bacterial microcompartment assembly: The key role of encapsulation peptides. *Commun. Integr. Biol.* **8,** (2015).

221. Marcelino, A. M. C. & Gierasch, L. M. Roles of beta-turns in protein folding: from peptide models to protein engineering. *Biopolymers* **89,** 380–391 (2008).

222. McPherson, A. & Cudney, B. Searching for silver bullets: an alternative strategy for crystallizing macromolecules. *J. Struct. Biol.* **156,** 387–406 (2006).

223. Larson, S. B., Day, J. S., Cudney, R. & McPherson, A. A new crystal form of bovine pancreatic RNase A in complex with 2'-deoxyguanosine-5'-monophosphate. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **63,** 728–733 (2007).

224. Chruszcz, M. *et al.* Function-biased choice of additives for optimization of protein crystallization – the case of the putative thioesterase PA5185 from Pseudomonas aeruginosa PAO1. *Cryst. Growth Des.* **8,** 4054–4061 (2008).

225. Afonine, P. V. *et al.* FEM: feature-enhanced map. *Acta Crystallogr. D Biol. Crystallogr.* **71,** 646–666 (2015).

226. Terwilliger, T. C. Finding non-crystallographic symmetry in density maps of macromolecular structures. *J. Struct. Funct. Genomics* **14,** 91–95 (2013).

227. Pang, A., Frank, S., Brown, I., Warren, M. J. & Pickersgill, R. W. Structural insights into higher order assembly and function of the bacterial microcompartment protein PduA. *J. Biol. Chem.* **289,** 22377–22384 (2014).

228. Tsumoto, K., Ejima, D., Senczuk, A. M., Kita, Y. & Arakawa, T. Effects of salts on protein-surface interactions: applications for column chromatography. *J. Pharm. Sci.* **96,** 1677–1690 (2007).

229. Moon, Y. U., Curtis, R. A., Anderson, C. O., Blanch, H. W. & Prausnitz, J. M. Protein—Protein Interactions in Aqueous Ammonium Sulfate Solutions. Lysozyme and Bovine Serum Albumin (BSA). *J. Solut. Chem.* **29,** 699–718 (2000).

230. Ugwu, S. O., Apte, S. P., Ugwu, S. O., Iv, O. B. & Apte, S. P. The effect of buffers on protein conformational stability. *Pharm Technol* 86–113

231. Mao, Y.-J., Sheng, X.-R. & Pan, X.-M. The effects of NaCl concentration and pH on the stability of hyperthermophilic protein Ssh10b. *BMC Biochem.* **8,** 28 (2007).

232. Thompson, M. C., Cascio, D., Leibly, D. J. & Yeates, T. O. An allosteric model for control of pore opening by substrate binding in the EutL microcompartment shell protein. *Protein Sci.* **24,** 956–975 (2015).