Investigating the structure of macromolecular complexes and membrane proteins using transmission electron microscopy

Breanna Sachiko Vollmar

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Tamir Gonen, Chair

Carlos Catalano

Alexey Merz

Program Authorized to Offer Degree:

Department of Biochemistry

©Copyright 2013
Breanna Sachiko Vollmar

University of Washington

Abstract

Investigating the structure of macromolecular complexes and membrane proteins using transmission electron microscopy

Breanna Sachiko Vollmar

Chair of the Supervisory Committee:
Tamir Gonen
Affiliated Associate Professor, Department of Biochemistry
Group Leader, HHMI Janelia Farm Research Campus

Transmission electron microscopy (TEM) has provided atomic resolution details regarding the biological processes that give rise to life. The work presented here focuses on the application of TEM for studying membrane proteins and macromolecular assemblies using electron crystallography and single particle electron cryomicroscopy (cryo-EM). Two membrane proteins, human aquaporin 3 (hAQP3) and galactose permease (GalP) from *Escherichia coli*, were studied using electron crystallography. hAQP3 was successfully expressed in *Pichia pastoris* and purified, but failed to reconstitute to form two-dimensional crystals. Two-dimensional crystals of GalP were optimized to form large, crystalline vesicles and sheets. A projection map of GalP calculated from a cryo-EM image showed that GalP forms trimers in the

two-dimensional crystals. Single particle TEM studies focused on three macromolecular assemblies, αB-crystallin, a designed octahedral protein cage and a mutant HCV IRES-40S ribosome complex. Projection class averages of negatively stained αB-crystallin were combined with NMR and SAXS data to produce a pseudo-atomic model of αB-crystallin and provide a hypothesis for multimerization. A single particle cryo-EM reconstruction of a designed self-assembling octahedral protein oligomer showed that a designed protein assembly forms the predicted symmetry in solution. A 17 Å resolution single particle cryo-EM reconstruction of a mutant HCV IRES bound to the 40S ribosomal subunit revealed that deletion of three nucleotides in the IRES modulate interaction with the 40S ribosomal subunit. The conformational changes induced in the 40S were shown to affect a step after 80S formation, the transition from initiation to elongation. These studies illustrate the application of TEM to study the molecular mechanisms used by both membrane proteins and macromolecular complexes to carry out biological processes.

Table of Contents

List of Figures viii

List of Tables xi

Chapter 1: Biological application of transmission electron microscopy 1

Introduction 1

Electron crystallography 3

Membrane protein expression, solubilization, purification 5

Two-dimensional crystallization 8

Screening two-dimensional crystallization conditions using negative stain TEM 11

Sample preparation, data collection and processing for cryo-EM 12

Biological insight provided by electron crystallography studies 15

Single particle electron cryomicroscopy 18

Samples suitable for single particle cryo-EM studies 20

Cryo-EM grid preparation 21

Data collection 22

Image processing & reconstruction 25

Insight into biological processes provided by single particle cryo-EM 28

Figures 32

Works cited in chapter 1 36

Chapter 2: Investigating the structure & mechanism of glycerol uptake by the human glycerol facilitator, aquaporin 3 44

Summary 44

Introduction 45

Results & discussion 50

Expression & solubilization of recombinant hAQP3 50

Purification of hAQP3 using Ni-affinity & size exclusion chromatography 53

hAQP3 forms pH sensitive oligomers *in vitro* 54

Two-dimensional crystallization trials of hAQP3-6xH 58

Concluding remarks 59

Materials & methods 61

Expression of hAQP3-6xHis in *Pichia pastoris* 61

Solubilization & Purification of hAQP3-6xHis 62

Two-dimensional Crystallization trials of hAQP3-6xHis 63

Negative-stain TEM 63

Acknowledgements 64

Figures 65

Works cited in chapter 2 72

Chapter 3: Two-dimensional crystallization of galactose permease, a prokaryotic sugar transporter 76

Summary 76

Introduction 76

Results & discussion 82
Screening detergents: Protein purification & stability 82
Screening detergents: Reconstitution & lattice formation 84
Negative-stain TEM of GalP-DM two-dimensional crystals 86
Negative-stain TEM of GalP-OG two-dimensional crystals 86
Negative-stain TEM of GalP-LDAO two-dimensional crystals 88
Cryo-EM of GalP two-dimensional crystals 91
Persistent aggregation of GalP during reconstitution experiments 93
Concluding remarks 95
Materials & Methods 98
Protein expression, purification & detergent screening 98
Two-dimensional crystallization 100
Negative-stain TEM 101
Cryo-EM & image processing 101
Acknowledgements 102
Figures 103
Works cited in chapter 3 115

Chapter 4: The N-terminal domain of αB-crystallin provides a conformational switch for multimerization and structural heterogeneity 119
Summary 119
Introduction 120
Results & discussion 122

The N- & C-termini of αB are highly flexible....122
Secondary structure within the N-terminal domain of αB....122
Modeling the N-terminal domain of αB....123
Building a multimer model....124
Heterogeneity of αB....127
Modeling high-order multimers of αB....129
Electron Microscopy and single particle analysis of negatively stained αB....130
Concluding remarks....131
Materials & methods....133
Sample preparation and NMR/SAXS data collection....133
Preparation of mutants....133
Crosslinking....134
Analysis of SAXS data....134
Modeling of multimers....134
Electron microscopy....135
Acknowledgements....136
Figures....137
Works cited in chapter 4....153

Chapter 5: Computation design of self-assembling protein nanomaterials with atomic level accuracy....157
Summary....157
Introduction....158

Results & discussion 159
Design method 159
Evaluation of expressed and purified octahedral and tetrahedral designs 160
Concluding remarks 164
Materials & methods 164
Generating a set of trimeric scaffold proteins 164
Symmetry definition files 165
Symmetrical docking 166
Symmetrical interface design 167
Source Code 169
Amino acid sequences 169
Protein expression, fluorescent labeling, and purification 170
Analytical size exclusion chromatography 172
Analytical ultracentrifugation 173
Negative-stain electron microscopy 174
Electron cryomicroscopy data collection and single particle reconstruction 174
Crystallization of O3-33 in space groups R32 and P4 175
Crystallization of T3-08 176
Crystallization of T3-10 177
Data collection and structure determination 177
Acknowledgements 179
Figures 180
Works cited in chapter 5 193

Chapter 6: HCV IRES manipulates the ribosome to promote the switch from translation initiation to elongation....198
Summary....198
Introduction....199
Results....201
IRES dIIb affects the rate of protein synthesis....201
IRES-40S subunit complex conformation is influenced by dIIb....201
dIIb influences a step after 80S ribosome formation....203
Modeled local interaction between dIIb and rpS5....204
AUG start codon placement is unaffected by dIIb mutations....206
dIIb mutants initiate in the correct reading frame....206
dIIb mutations does not affect peptide bond formation....207
The first translocation even is promoted by dIIb....208
Discussion....209
Materials & methods....213
Plasmid construction and cloning....213
RNA preparation....213
Radiolabeling RNA and primers....214
Ribosome assembly assays....214
Luciferase assays....215
Toeprinting assays....215
Puromycin assays....216

Nuclear magnetic resonance 217
Negative-stain electron microscopy.....217
Electron cryomicroscopy 217
Accession Codes 219
Acknowledgements.....219
Figures.....221
Works cited in chapter 6 228

Chapter 7: Concluding Remarks.....233
Human aquaporin-3 233
E. coli galactose permease 235
αB-crystallin 236
Designed self-assembling protein nanomaterials.....240
Mutant HCV IRES-40S 242
Future of biological TEM 245
Works cited in chapter 7 249

List of Figures

Figure 1.1 Two-dimensional crystal formation 32

Figure 1.2 Examples of possible results during two-dimensional crystallization 33

Figure 1.3 Back-injection method for freezing two-dimensional crystals for cryo-EM.....34

Figure 1.4 Sample preparation for single particle cryo-EM 35

Figure 2.1 The NPA motif and ar/R constriction site of aquaporins 65

Figure 2.2 Expression, alkali membrane treatment and solubilization assay of hAQP3 66

Figure 2.3 Nickel-NTA affinity and Size exclusion purification of hAQP3-6xHis 67

Figure 2.4 pH dependent oligomerization of hAQP3-6xHis 68

Figure 2.5 Human AQP3 homology model showing cysteines and histidines 69

Figure 2.6 Typical results from two-dimensional crystallization trials of hAQP3-6xHis 70

Figure 3.1 Sugar transporter topology and arrangement of helices within the monomer.....103

Figure 3.2 Conformations of sugar transporter structures determined 104

Figure 3.3 Screening GalP homogeneity in different detergents by exchange over SEC 105

Figure 3.4 Example of typical protein-lipid aggregation observed during screening of two-dimensional crystallization conditions by negative-stain TEM.....107

Figure 3.5 Negative-stain TEM of two-dimensional crystals formed with GalP-DM 108

Figure 3.6 Formation of two-dimensional crystals using GalP-OG 109

Figure 3.7 Effect of glycerol and a mixture of detergents on GalP-OG two-dimensional crystal formation.....110

Figure 3.8 GalP-LDAO two-dimensional crystals form with different divalent cations.....111

Figure 3.9 Decreasing $MgCl_2$ concentration and fine screen of LPR improved two-dimensional crystals formed using GalP-LDAO.....112

Figure 3.10 Cryo-EM of GalP-OG two-dimensional crystals shows GalP form trimers 113

Figure 3.11 Cryo-EM of GalP-LDAO two-dimensional crystals 114

Figure 4.1 N-terminal domain of αB-crystallin 137

Figure 4.2 αB-crystallin dimer and arrangement in hexameric unit 138

Figure 4.3 Building a model for the N-terminal domain of αB 139

Figure 4.4 Homology model of αB β1 and β2 140

Figure 4.5 Model of αB 24-mer 141

Figure 4.6 Intermolecular interface of αB multimerization 142

Figure 4.7 Cys cross-linking in αB 143

Figure 4.8 αB 24-mer 144

Figure 4.9 SAXS analysis of αB reveals heterogeneity 145

Figure 4.10 Model of an αB 28-mer 146

Figure 4.11 Electron microscopy and single particle analysis of αB 147

Figure 4.12 Back projections of 24-mer, 26-mer & 28-mer models 148

Figure 4.13 Location of phosphorylation sites in model of αB 24-mer 149

Figure 4.14 Comparison of the archaeal hsp16.5 coordinates (1shs) (A) and human αB coordinates (2klr) (B) fit into the αB density map (EMDataBank entry EMD-1776) 150

Figure 5.1 General approach to designing self-assembling protein nanomaterials 180

Figure 5.2 Experimental characterization of O3-33, T3-08, and T3-10 181

Figure 5.3 Fluorescent SEC chromatograms of (A) O3-33, (B) 3n79-wt, (C) T3-08, and (D) 3ftt-wt in labeled cell lysates 182

Figure 5.4 Analytical ultracentrifugation analysis of O3-33 183

Figure 5.5 Structural characterization of O3-33 184

Figure 5.6 Electron cryomicroscopy of O3-33 185

Figure 5.7 Fluorescent SEC chromatograms of 3n79-wt, O3-33, and various interface point mutants in lysates.....186

Figure 5.8 Structural characterization of T3-08.....187

Figure 5.9 Mutations in T3-10 that stabilize the original designed configuration.....188

Figure 5.10 Structural characterization of T3-10.....189

Figure 6.1 *In vitro* translation analysis of dIIb mutations.....221

Figure 6.2 EM of HCV IRES mutant ΔGCC in complex with the 40S subunit.....222

Figure 6.3 WT and mutant IRES ribosome assembly assays and position of domain IIb.....223

Figure 6.4 Characterization of the structural changes induced by dIIb mutation.....224

Figure 6.5 Biochemical analysis of AUG docking and potential frameshift.....225

Figure 6.6 Puromycin and toeprinting assays with antibiotic.....226

Figure 6.7 Model of the role of dIIb in HCV IRES translation initiation.....227

List of Tables

Table 2.1 Two-dimensional crystallization conditions screened for hAQP3-6xHis 71

Table 3.1 Summary of purification conditions screened and reconstitution results 106

Table 4.1 Distance restraints from solid-state NMR 151

Table 4.2 Similarity statistics.....152

Table 5.1 Mutations and experimental screening results for all designed proteins 190

Table 5.2 Data collection and refinement statistics 192

Acknowledgements

I would like to thank my advisor, Tamir Gonen, for providing guidance throughout my multiple projects, opportunities to collaborate with different groups covering a broad range of topics in biochemistry and support during my transition to a career in pharmacy. I think that over the five years that I have been in Tamir's lab, I have truly grown into a mature scientist and will take everything that I have learned to my new adventure.

Thanks to all the members of the Gonen lab, past and present, for the good conversations, scientific advice and general support. A special thanks to Dan Shi, the EM master, who helped me to become a confident electron microscopist.

I would also like to thank my thesis committee members, Rachel Klevit, Carlos Catalano, Alex Merz, Ning Zheng and Wenqing Xhu, who helped guide me through my research studies with valuable discussions about science and support when exploring different career paths.

This work was funded by the Cell and Molecular Biology Training Grant - Public Health Service, National Research Service Award, T32 GM07270, from the National Institute of General Medical Sciences and the Howard Hughes Medical Institute.

Dedication

To my parents,

Renee Bennett and Mark Vollmar.

Chapter I

Biological application of transmission electron microscopy

INTRODUCTION

Our understanding of the natural world has increased dramatically because of the advancement of technology allowing us to observe wonders beyond those that can be seen with the human eye. Developments in the field of microscopy and our understanding of the nature of light have provided a means to understand the biological processes that give rise to life. The biological unit of all life is the cell, which was first described by Anton van Leeuwenhoek using a compound light microscope in the 17th century. To see detail beyond the diffraction limit of natural light was not possible until the invention of the transmission electron microscope (TEM) by Ernst Ruska in the 20th century. The transmission electron microscope surpassed the resolution barrier of light microscopy by harnessing the power of electron scattering to visualize individual proteins and even single atoms. The resolving power or resolution of a microscope is defined as the minimum separation detectable between two objects and is a function of the wavelength of the energy source. Visible light consists of a spectrum of wavelengths from 390 to 750 nm; the maximum resolution expected using light microscopy is 200 nm. In comparison, the deBroglie wavelength of an electron in a 200 kV TEM is 2.5 pm with a resolution of 1 Å, limited by aberrations of the microscope's optics and recording medium. The atomic resolution of TEM has resulted in the ability to observe the structures of ribosomes, membrane proteins, viruses, organelles & protein complexes providing a view into the unseen structures of the natural world.

An electron microscope is in principal similar to a light microscope, but instead of using electrons at a wide range of wavelengths, a high-energy beam of electrons at the same

wavelength is used. In TEM, electrons emitted from a source (field emission gun, thermionic cathode or crystal) are focused by a set of condenser lenses into a coherent parallel beam that interacts with the specimen. Electrons encountering the specimen are transmitted, during which the phases of the incoming electrons are changed as they pass through the sample. The exiting electrons are then focused by the objective lens and an image is formed that is projected on to a phosphorescent screen or onto a recording medium such as film, a charged coupled device (CCD) or direct detector. Samples for TEM must be extremely thin, only several hundred nanometers thick, so that electrons can be transmitted through the sample. The types of samples that can be observed using TEM is dictated by the thickness of the sample and the type of information that is needed.

To observe biological samples using TEM the sample must be preserved due to the vacuous environment of the TEM column. Cryogens, such as liquid ethane and nitrogen, are used to preserve the hydrated state of biological samples during data collection using TEM. This method of TEM is referred to as electron cryomicroscopy (cryo-EM). In the field of cryo-EM there are three methods for observing biological samples, tomography, single particle imaging and electron crystallography, which vary in the types of samples, data collection method and expected resolution. Tomography is used to observe the molecular architecture of cells, tissues and large protein assemblies. Data for tomography consists of collecting a series of tilted images for a *single* specimen that are then reconstructed to form a three-dimensional model of the imaged specimen. Single particle cryo-EM consists of collecting images of *many* individual protein particles randomly oriented in a thin layer of vitrified ice and then combining the different views of particles to produce a three-dimensional volume. In comparison to tomography and single particle cryo-EM that are exclusively imaging techniques, electron

crystallography uses both the imaging capabilities of TEM as well as the principal of electron diffraction of crystals to determine the structure of proteins. Electron crystallography uses two-dimensional crystals composed of ordered arrays of membrane proteins embedded in a lipid bilayer or ordered arrangements of soluble proteins, such as αβ-tubulin (1). Using a combination of images and electron diffraction patterns, electron crystallography provides information regarding the native structure of membrane proteins in a lipid environment.

Typically tomography is considered a low-resolution technique limited to 4-10 nm, compared to single particle cryo-EM and electron crystallography that can reach atomic-level resolution, allowing the determination of the exact position of atoms within a macromolecular complex or membrane protein. To answer mechanistic questions regarding the specific interactions that contribute to the ability of a given protein or protein complex to carry out biological processes necessary for life, higher resolution information is necessary. Here I will focus on single particle cryo-EM and electron crystallography to investigate the structure and function of macromolecular machines and membrane proteins.

ELECTRON CRYSTALLOGRAPHY

Electron crystallography is a powerful technique to study the structure of membrane proteins embedded in a lipid bilayer. In principal, X-ray and electron crystallography are very similar; both use a beam of high energy to interact with a crystalline sample to produce a diffraction pattern, from which information regarding the structure of a protein can be extracted. The differences between the two techniques are the use of an electron beam compared to an X-ray beam, diffraction of two-dimensional crystals instead of three-dimensional crystals and the ability to collect images of crystals to obtain phase information that is lost during diffraction. The

loss of phase information during diffraction gives rise to the infamous phase problem in X-ray crystallography. Two-dimensional crystals are composed of ordered arrays of membrane protein and lipids, allowing the observation of the structure in a native-like environment. For X-ray crystallography, detergent stabilized membrane proteins are crystallized to form three-dimensional crystals that may not represent the functional state of the protein. Two-dimensional crystals are not suitable for X-ray diffraction because the scattering power of X-rays is not high enough to interact with the thin crystals. The higher scattering power and sensitivity of electrons allows the determination of charged states of residues in the protein, such as visualization of protons, which is not possible with X-rays. Despite the differences between X-ray and electron crystallography both techniques can be used to study atomic level details regarding the mechanism of membrane proteins.

The most significant advantage of electron crystallography is it is not an all or nothing technique. X-ray crystallography requires the formation of three-dimensional crystals that diffract to beyond 10 Å because of the inability to have enough data for indexing and to determine the phases correctly. In electron crystallography, two-dimensional crystals can be imaged obtaining both phases and amplitudes, which allows calculation of projection maps & three-dimensional models from small low-resolution crystals. Even low-resolution projection maps can reveal information regarding oligomeric state and conformation of proteins embedded in the bilayer. These features of electron crystallography make it a unique method to study membrane protein biochemistry.

Studying the structure of membrane proteins using electron crystallography begins with obtaining pure membrane protein to use for screening reconstitution conditions that result in the formation of two-dimensional crystals. Negative-stain TEM is used to analyze reconstitution

conditions for morphology and lattice formation. Two-dimensional crystals are stained using a heavy metal solution such as uranyl formate to increase image contrast and provide an easy method to evaluate crystals. If two-dimensional crystals are observed using negative-stain TEM, then freezing conditions are screened for cryo-EM data collection. A series of images and/or diffraction patterns are collected from many two-dimensional crystals, and then the data is processed to calculate the membrane protein structure. Here I will discuss details of the steps leading to determining the structure of a membrane protein using electron crystallography and how electron crystallography has contributed to our understanding of membrane protein biochemistry.

Membrane protein expression, solubilization and purification

The first step in an electron crystallography study is to determine a source to obtain the membrane protein of interest. Most membrane proteins are expressed in limited amounts where they are naturally found; therefore heterologous expression is necessary in order to obtain milligram quantities of protein needed to adequately screen crystallization conditions. Several options are available for heterologous expression of membrane proteins including bacteria, yeast, insect cells and mammalian cells. The choice of expression system depends on the organism the membrane protein is natively expressed and compatibility of organism to express high amounts of non-native protein. Membrane proteins naturally expressed in prokaryotes are most suitable for heterologous expression in bacteria, most commonly in *Escherichia coli* (2). Membrane proteins of eukaryotic origin may require post-translational modification or chaperones for proper folding and localization that are not available in *E. coli*, thus eukaryotic expression systems maybe needed. Several eukaryotic heterologous expression systems are available for

membrane proteins including yeast, insect cells and mammalian cells. Expressing membrane proteins has become much easier with the development of heterologous expression systems that have reduced proteolytic activity and decreased toxicity of over-expressed proteins (3,4). Finding the optimal expression conditions is key because this determines the amount of starting material for each subsequent experiment leading to reconstitution to form two-dimensional crystals.

Once a reliable source has been identified and expression optimized, the next step is to isolate the membrane protein from the lipid bilayer using detergents. Detergents are amphiphilic molecules that consist of a long aliphatic chain attached to a hydrophilic group. Above the critical micelle concentration (CMC), the detergent forms micelles that can extract the membrane protein from the cell membrane and stabilize the protein in solution (5). Non-ionic detergents, such as n-dodecyl-β-D-maltopyranoside (DDM), n-octyl-β-D-glucopyranoside (OG), n-decyl-β-D-maltopyranoside (DM) and octaethylene glycol monodecyl ether (C12E8), are commonly used as solubilization reagents because they do not perturb the tertiary structure of the protein (5). Some proteins may require certain lipids once solubilized from the membrane to remain folded properly and to function. (6). During solubilization screening, the amount and types of lipids that are solubilized with the protein by the detergent can be determined by using thin-layer chromatography (7). Information about the presence and type of lipids that are solubilized with the protein can help for reproducibility during crystallization screens.

The last step before screening reconstitution conditions for two-dimensional crystal formation is purification of the protein. Membrane proteins are routinely expressed as fusions with N-terminal or C-terminal affinity tags to facilitate purification from the solubilized fraction. Purification using the affinity-based tag is the primary step and the second step is size exclusion chromatography to assess the homogeneity of the purified protein. The most common affinity

purification method is immobilized metal ion affinity chromatography (IMAC) in which a protein is expressed with a terminal tag consisting of a string of at least six histidine residues that binds to either Ni^{2+} or Co^{2+} charged resin. The benefit of the histidine tag is that it is a small tag that does not affect the folding or insertion of membrane proteins (8,9). The addition of a proteolytic cleavage site between the membrane protein and the purification tag is convenient for removing the tag to prevent protein aggregation (10). In some cases it may be necessary to use ion-exchange chromatography to remove proteins that co-purify during the initial purification step. Once purification has been optimized, the next step is assessment of the homogeneity and stability of the protein.

Size exclusion chromatography (SEC) separates proteins based on size and provides an assessment of the homogeneity of the purified membrane protein. In general, SEC is a diagnostic tool to assess your purified protein, but can be used as a purification step to remove proteases used to cleave affinity tags, separate proteins that co-purified during the previous purification steps or to exchange detergents for membrane proteins. The shape of the SEC elution peak provides information regarding the oligomeric state and homogeneity of the purified protein. SEC can also be used to screen detergents for two-dimensional crystallization and protein stability. The detergent that is able to solubilize your protein from the cell membrane may not be the best for two-dimensional crystallization. An example is the use of DDM, which is able to solubilize a number of proteins but the low CMC results in prolonged dialysis times, risking the aggregation of protein before reconstitution can occur. Protein stability is crucial over the course of the reconstitution experiment because if the protein is aggregated before the protein can incorporate with the lipids then reconstitution will not be able to occur.

The steps and consideration for producing pure membrane protein for electron crystallography are essentially the same as other structural biology techniques with the addition of considering the CMC of the particular detergent used in the final purification step. For X-ray crystallography having bulk amounts of detergent present may interfere with the ability of proteins to make sufficient crystal contacts in three-dimensional crystals. In comparison, the process for forming two-dimensional crystals results in the removal of detergent and reconstitution with lipids. The important factors to consider for electron crystallography are the stability of protein during reconstitution as well as the detergent's CMC that controls the kinetics of reconstitution.

Two-dimensional crystallization

Formation of two-dimensional crystals is undoubtedly a bottleneck in electron crystallography. Unlike X-ray crystallography where a complete data set can be collected from a single well-diffracting crystal, electron crystallography requires many crystals because the electron dosage needed to collect diffraction or images damages the crystal after a single exposure. Two-dimensional crystals are formed by combining detergent stabilized membrane protein with detergent solubilized lipids at different lipid to protein ratios (LPR) followed by removal of the detergent using dialysis (Fig. 1.1). As the detergent is removed, the membrane protein will prefer to associate with lipids rather than exposing hydrophobic surfaces to the hydrophilic surrounding solution, which is energetically unfavorable. Reconstitution of membrane proteins with lipids typically results in four different morphologies: vesicles, sheets, tubes or aggregation (Fig. 1.2). The most desired two-dimensional crystal morphology is a single layered sheet representing an ordered array of protein embedded in a single lipid bilayer to avoid

the presence of multiple lattices. For successful two-dimensional crystallization, it is essential to carefully design screens and analyze results for trends. Parameters that are important for two-dimensional crystallization are detergent, lipid, LPR and dialysis conditions.

The choice of detergent for stabilizing the membrane protein and solubilizing lipids is important for protein stability and controlling the kinetics of two-dimensional crystallization (11). During reconstitution, the detergent concentration is decreased to below the CMC through dialysis, allowing incorporation of membrane proteins with the forming lipid bilayer. The kinetics of two-dimensional crystallization is dictated by the CMC of the detergent; higher CMC detergents such as OG (0.5%) fall below the CMC faster than detergents with lower CMCs such as DDM (0.0087%). The longer time for lower CMC detergents to dialyze can have both positive and negative effects depending on protein stability and susceptibility of aggregation during the course of reconstitution. The choice of detergent is a balance between protein stability, reconstitution and crystal formation.

The ability to form two-dimensional crystals depends on specific interactions between the membrane protein and lipids mediated by the removal of detergent. Whether liposomes form then proteins are incorporated or proteins and lipids both associate to form lipid structures is not known. Studies have shown that the lipids are more than just forming a lipidic environment to stabilize membrane proteins, but lipids specifically associate with membrane proteins (12). In the AQP0 structure determined using electron crystallography lipids were observed to conform to the shape of the hydrophobic surfaces of the membrane protein and the phospholipid head groups packed closely to the exposed hydrophilic residues (13). Due to the specific nature of protein-lipid interactions it is important to screen lipids with different alkyl chain lengths as well as different phospholipid head groups to find conditions that yield two-dimensional crystals. Lipids

commonly used for two-dimensional crystallization are DMPC, native lipids and *E. coli* lipid extracts. Researching the lipid composition of where the protein is natively expressed can provide clues as to what lipids maybe helpful for reconstitution. Analysis of co-purified lipids using thin-layer chromatography can also provide information on what types of lipids is best for reconstitution of the specific membrane protein (7).

To form two-dimensional crystals, the membrane protein and lipids are combined in different lipid to protein ratios to find optimal conditions for reconstitution and crystallization. A low LPR will result in aggregation of protein & lipid, where as a high LPR may result in formation of proteoliposomes lacking ordered arrays of protein. For well-ordered two-dimensional crystals there should be enough lipid to form a lipid bilayer so that incorporated proteins will pack tightly to form ordered arrays. The crystal morphology can also change by slightly varying the LPR; sheets tend to form at low LPRs and vesicles at higher LPRs. The LPR may slightly change during screening of conditions so several LPRs should be screened for every condition. Most often LPR refers to the weight ratio and LPR range of 0 to 1.0 are screened with protein concentrations ranging from 0.5 mg/mL to 5 mg/mL. Two-dimensional crystals usually form between LPRs of 0.2 to 0.6 over a wide range of protein concentrations.

The last parameters to screen are dialysis conditions, such as dialysis buffer components (buffer pH, divalent cations, salt concentration, glycerol), dialysis method and temperature. Dialysis conditions are used for fine-tuning the quality of the two-dimensional crystals initially obtained and optimizing the crystal morphology. The specific temperature and dialysis method chosen depends on the CMC of the detergent and transition temperature of the lipids used in the reconstitution mixture. Temperature cycling above and below the transition temperature of the lipids has been used in several conditions to improve lattice formation in two-dimensional

crystals (14,15). Various dialysis methods have been used such as dialysis tubing, slide-a-lyzers and detergent chelators such as methyl-β-cyclodextrin (16) & Bio-Beads (17), but the most successful dialysis method that produced crystals used for high resolution structure determination are dialysis buttons (18). Divalent cations are included in the dialysis buffer because they interact with the zwitterionic or charged lipid head groups and can significantly alter the lipid morphology. Typical divalent cations that are screened for dialysis are $CaCl_2$, $MgCl_2$ and $ZnCl_2$ from 0 to 50 mM. Taking into account all of the conditions that can influence reconstitution and crystallization, carefully screening conditions and analysis of results are important for successful two-dimensional crystallization of a given membrane protein.

Screening two-dimensional crystallization conditions using negative-stain TEM

Two-dimensional crystallization conditions are screened using negative-stain TEM to assess crystal morphology and quality. Negative-stain TEM uses a solution of heavy metals, typically uranyl formate or uranyl acetate, to stain the sample thereby increasing the contrast for imaging and providing protection from radiation damage. The use of negative-stain TEM is only for initial screening because the maximum observable resolution is limited by the size of the heavy metal stain, about 20 Å for uranyl formate, and the quality of the crystals maybe affected by dehydration during the staining process. Screening using low magnifications, ranging from 1,000x to 10,000x, can provide an overview of typical crystal morphologies, the size and number of crystals. Higher magnifications, 20,000x to 50,000x, are needed to assess the size of the crystalline area and order. A Fourier transform of the selected area in a higher magnification image provides information regarding the unit cell dimensions, presence of multiple lattices and the resolution to which the crystals are ordered. A preliminary projection map of the negatively

stained two-dimensional crystals can be calculated using 2dx, a GUI interface for MRC programs for processing images of two-dimensional crystals (19). The low-resolution projection map provides information regarding the oligomeric state and packing of the membrane protein in the lipid bilayer. To get atomic resolution information, two-dimensional crystals must be frozen and data collected using cryo-EM.

Sample preparation, data collection and processing for cryo-EM

To determine the structure of a membrane protein embedded in a lipid bilayer using electron crystallography, the crystals must be cooled to the temperature of liquid nitrogen or helium to preserve high resolution features during data collection. The two methods for preserving two-dimensional crystals for structure determination by TEM are to embed the crystals in sugar or in vitrified ice. Freezing and embedding conditions are screened to find the optimal conditions that preserve the crystal morphology and quality, while protecting the crystals from beam-induced damage.

Embedding the two-dimensional crystals using sugar not only serves as a cryoprotectant but helps to adhere crystals to the carbon coated grids used for data collection and help the crystals to lay flat (20). Sugars typically used for embedding two-dimensional crystals are glucose (21), trehalose (22) and tannic acid (23). Incubation of the two-dimensional crystals with sugar results in dehydration of the crystals, which can help or disrupt the crystalline lattice. Crystals embedded in sugar can be applied to grids using three methods: back-injection, carbon sandwich and direct application. The back-injection method (22) consists of floating a carbon film suspended on an aqueous solution (sugar, buffer or water) onto a molybdenum grid, crystals are then applied to the carbon-free side, incubated for a period of time, blotted crystal side down

onto filter paper and directly frozen in liquid nitrogen (Fig. 1.3). The carbon sandwich method (24) is similar to the back-injection method except a second layer of carbon is applied to the other side of the grid with the crystals between the two layers of carbon. The benefit of the carbon sandwich method is that the crystals lay flat on the grid maximizing the area for data collection and decreasing charging effects. The last method is direct application, in which two-dimensional crystals are applied to a carbon-coated grid, then excess solution is blotted away with filter paper and the grid is frozen in liquid nitrogen. This method is the fastest to screen but the crystals may not lay flat on the grid.

In contrast, direct freezing of two-dimensional crystals using liquid ethane does not require additional reagents. The formation of vitrified ice by direct freezing of the crystals in ethane provides protection from radiation damage and preserves the crystals in the hydrated state. To embed samples in vitrified ice, crystals are applied to a carbon coated molybdenum or copper grid, blotted with filter paper to form a thin film and plunged into liquid ethane. Vitrified samples are prepared using a robot, which allows control over blotting time and fast plunging of the sample into liquid ethane. A drawback to direct freezing is that the crystals may not attach to the grid so crystals maybe loss during the blotting procedure prior to plunging the grid in liquid ethane. The ideal sample preparation conditions should result in crystals lying flat on the carbon support, preservation of the crystal during data collection and a large number of crystals present on the grid.

There are two methods to obtain information regarding the structure of membrane proteins using cryo-EM: electron diffraction and imaging. The advantage of imaging the crystals is that both phase and amplitude information is obtained. The drawback to imaging is that the images are sensitive to aberrations present in the electron microscope, such as drift and charging,

which limit the resolution of structural information that can be obtained. Electron diffraction is not sensitive to the aberrations of the microscope, but phase information is lost as is the same with X-ray diffraction. Diffraction also requires large crystalline areas in excess of 1 μm in order to have enough unit cells to generate high signal to noise intensities in the diffraction pattern. For both types of data, a series of images at different tilt angles must be collected to get structural information in the z-direction, parallel to the beam direction, so that a three dimensional structure can be calculated. Data collection for electron crystallography is time-consuming because a complete data set can not be collected from a single crystal, so conditions that result in a number of crystals per grid is essential to increase efficiency of data collection.

Data processing for the two different types of data sets, electron diffraction and images, is accomplished using well-established protocols in the MRC suite (25). Images are processed by determining the defocus, correcting the CTF, determining the lattice & tilt, unbending the images (correction of lattice distortions), extraction of phases/amplitude information and calculation of a density map. Image processing can be conveniently carried out using 2dx (19,26), a program that provides a GUI interface for the MRC suite (26) of programs as well as access to programs in CCP4 (27). For diffraction data, a similar procedure is carried out beginning with determination of lattice, subtraction of background, integration of intensities observed in the pattern and the merging of diffraction patterns to calculate a three-dimensional density map. The processing of diffraction data can be done using programs such as XDP (28), a GUI interface for MRC programs, and IPLT (29,30). In contrast to images, diffraction data does not contain phase information so phases must be obtained using molecular replacement or from images of two-dimensional crystals. Recently a new method was developed and published that describes using the low-resolution phase information from images for phasing high-resolution electron

diffraction data called phase extension (31). After generating a density map the structure can be built into the density based on the features seen in the map and refined using established procedures similar to those used for X-ray crystallography.

Biological insight provided by electron crystallography studies

Formation of two-dimensional crystals that are large and well ordered has proven to be difficult as seen by the limited number of published atomic resolution structures of membrane proteins by electron crystallography. Despite the difficulty of two-dimensional crystallization, low resolution studies have provided useful details regarding oligomeric state, tertiary structure of membrane proteins and conformations of membrane proteins embedded in a lipid bilayer. There are many examples of low-resolution projection maps and three-dimensional maps including the betaine uptake transporter (32), oxalate transporter (14), copper transporter (33) and connexins (34,35,36). Some studies have been successful in forming large well ordered two-dimensional crystals and have obtained atomic resolution structures of membrane proteins including bacteriorhodopsin (23), several aquaporins (13, 15, 37), the light harvesting complex (38), the acetyl choline receptor (39), glutathione transferase (40) and prostaglandin E2 (41). Besides providing insight into the structure of membrane proteins, electron crystallography also has several applications that are quite unique compared to other structural biology methods.

A key question in membrane protein biochemistry is the nature of the association of lipids with membrane proteins. With electron crystallography there is a unique opportunity to observe phospholipids associated with membrane proteins. Several studies have revealed that lipids form two populations consisting of bulk and annular lipids, which specifically associate with proteins (12,13,23). In the density maps calculated from electron diffraction of AQP0 (13)

and bacteriorhodopsin (23) two-dimensional crystals, both leaflets of the lipid bilayer were able to be visualized with the associated lipids specifically conforming to the hydrophobic transmembrane segments of the proteins and that surprisingly the crystal contacts in two-dimensional crystals are facilitated exclusively by the surrounding lipids. Further questions that remain unanswered are how do the surrounding lipids affect the ability of membrane proteins to undergo large conformational changes and do specific types of lipids associate with membrane proteins.

Most membrane protein structures have been determined using X-ray crystallography, which crystallizes the protein in the presence of detergent; therefore the structure that is observed possibly represents an energetically favorable state for crystallization rather than a true functional representation. Electron crystallography uses crystals consisting of membrane proteins embedded in a lipid bilayer and from the crystals the function can be assessed directly. In the study of AQP1, two-dimensional crystals formed from reconstitution of purified erythrocyte AQP1 with *E. coli* lipids were shown to have high water permeability using carboxyfluoroscein loaded crystalline vesicles (42). The AQP1 crystals were then used to determine the structure of AQP1 demonstrating that the structure of AQP1 in fact represented a functional membrane protein (37). Most recently, studies of the *E. coli* galactose permease have shown that a fluorescent glucose analogue was able to bind specifically to GalP two-dimensional crystals (43). Using electron crystallography we can directly study the structure of functional membrane proteins.

Building on the use of two-dimensional crystals for measuring function, the same crystals used for structure determination can also be used to study conformational changes in response to a stimulus using electron crystallography. An atomizer spray coupled to a rapid freezing device

was used to study the *Torpedo* ray acetylcholine receptor from native membranes to trap open channels in response to a brief exposure to acetylcholine (44). The native membranes formed tubular two-dimensional crystals of ordered acetylcholine receptors after isolation and resuspension in a low salt solution. The acetylcholine receptor structure before and after exposure to acetylcholine could be observed on the millisecond time scale. After exposure of the two-dimensional crystals to acetylcholine the channel opens, undergoing a change in the orientation of the M2 helical segment that results in a conformation change in helices surrounding the pore allowing the conductance of positively charged ions (44). Direct assessment of function and structure using the same materials is only possible with electron crystallography due to the reconstitution of membrane proteins in a lipid bilayer, the native environment for membrane proteins.

The movement of charge is responsible for many physiological processes and is a feature of several membrane proteins such as ion channels, ATP synthases and nutrient transporters. Using electron crystallography, the state of charged residues can be observed due to the higher scattering power and sensitivity of electrons compared to X-rays. The structure of bacteriorhodopsin determined using electron crystallography was able to detect charged residues using difference maps (45). Bacteriorhodopsin is a light driven proton pump that forms purple membranes consisting of ordered protein in *Halobacterium salinarium*. To understand the different conformational states that bacteriorhodopsin undergoes upon absorption of light to transport positively charged protons, the charged states of the residues must be determined. In the analysis of the bacteriorhodopsin structure, Mitsuoka *et al*. identified three negatively charged residues and two positively charged residues that line the pathways for proton transport representing a possible proton pathway (45). The ability to visualize charge distribution has vast

implications for all membrane protein biochemistry, especially processes that involve the movement of protons, such as sugar transport in bacteria where the proton pathway is still not currently understood.

Electron crystallography is a versatile technique not only used to determine the structure of membrane proteins in a lipid environment but is also used to study structure-function relationships directly using two-dimensional crystals. The use of electron crystallography to investigate the mechanisms used by membrane proteins has provided details regarding protein-lipid interactions, direct observation of the conformational changes of membrane proteins in response to ligand binding, charged states of residues along transport pathways and the structure of several membrane proteins. The limiting factor of producing crystals for structure determination using electron crystallography is mitigated by the fact that meaningful information regarding membrane proteins can still be derived from low-resolution crystals. The versatility of electron crystallography makes it a promising tool for studying the biochemistry of membrane proteins; with further development and research on the formation of two-dimensional crystals the technique will be able to become more widely accessible.

SINGLE PARTICLE ELECTRON CRYOMICROSCOPY

Single particle cryo-EM is a method to study the structural biology of macromolecular complexes. In comparison to other methods, such as X-ray crystallography and NMR, single particle cryo-EM does not require crystallization of the protein and only micrograms of protein are sufficient to collect data. A limitation to both NMR and X-ray crystallography is the inability to investigate large heterogeneous proteins or protein complexes. Single particle cryo-EM fills this gap by being particularly well suited to investigate the structure of these types of

macromolecular complexes that typically can not be recombinantly expressed yielding quantities needed for other methods or form heterogeneous complexes.

The theory of single particle cryo-EM is that when a sample is frozen, the protein particles are randomly oriented in the ice, by imaging the sample different projections or views of the particles are obtained. Through the use of computational algorithms the views of the particles can be reassembled to form a three-dimensional model of the imaged protein particle. The ability to reconstruct a model from different views relies on the projection theorem that states that a two-dimensional Fourier transform of the projection of a three-dimensional object is a central slice through the three-dimensional Fourier transform perpendicular to the projection plane. An inverse Fourier of transfer of the 3D transform then yields a three-dimensional model for the particle (46). As many views as possible are needed to cover the three-dimensional Fourier space and with the low signal to noise present in cryo-EM images, the more images of each particular view the better signal to noise. A reason that single particle cryo-EM has a reputation for low resolution reconstructions is the inability to get broad angular distribution of views and low signal to noise that results in low accuracy of alignment. Although this is seen as a disadvantage of single particle cryo-EM, low-resolution information can still be useful in answering questions regarding the oligomeric state, binding sites and overall architecture of macromolecular machines that cannot be studied by other methods.

The process of reconstructing a model of a particular protein or complex involves sample preparation, imaging, particle selection, classification, determination of initial orientation parameters and reconstruction. At each step of the process, there are several things to consider in order to be confident in the final three-dimensional model and to obtain the highest resolution

possible for a given data set. Here I will give an overview and important considerations at each step of the process leading to generating a reconstruction from single particle cryo-EM images.

Samples suitable for single particle cryo-EM studies

The two most important characteristics for investigation of protein samples by single particle cryo-EM are protein size and sample preparation. The size limit for proteins or protein complexes suitable for single particle cryo-EM is not an upper bound, but is a lower bound of about 300 kDa for sub-nanometer resolution (47). The lower size limitation is due to reduced signal to noise from proteins with lower molecular mass and the tendency of smaller proteins to be more globular with less features that result in inaccurate determination of alignment parameters necessary for high resolution reconstructions. Large protein assemblies and complexes with a molecular mass over 1 MDa have higher signal to noise and more features for alignment resulting in higher accuracy of determining the orientation of each particle in the ice. Smaller proteins on the order of 200 kDa have been reconstructed to resolutions of 10 to 20 Å giving low-resolution details (48) concerning oligomeric state and shape. Currently, the only single particle reconstructions that have obtained at least 5 Å are large symmetrical protein complexes (49,50), the 80S ribosome (51) and large icosahedral viral capsids (52,53,54).

There are key features of sample preparation that hold true across multiple disciplines of structural biology in order to have a high probability of success: the sample must be purified from the source, homogeneous and stable in solution over the course of the experiment. Even though cryo-EM involves selection of protein particles in images for processing, which seems like a purification step, this process is subjective to the proper identification of the particles of interest. Another thing to consider is that a protein particle may orient on the grid in different

orientations giving rise to different views in projection, thus it would be hard to distinguish the particles of interest from contamination. The requirement for homogeneous protein sample (same conformation, oligomeric state or fully-assembled complex) is not as essential as for other structural biology methods because programs have been developed to classify the different projections of protein particles (55,56). Although if multiple states have shared views it maybe difficult to differentiate which views belong to a specific state, therefore a homogeneous protein sample will facilitate the ease of data processing. Stability in solution for cryo-EM is necessary, but on a shorter time scale compared to X-ray crystallography and NMR where samples must be stable for several days. Combining the use of robotics and ethane to freeze protein, samples for cryo-EM can be frozen on the millisecond time scale. Overall, many different types of samples are suitable for single particle cryo-EM which adds to the versatility of the technique to obtain structural information.

Cryo-EM grid preparation

Adapting the use of ethane to freeze biological samples, originally developed by X-ray crystallographers for freezing three-dimensional protein crystals, has facilitated the advancement of structure determination using single particle reconstruction. Freezing biological samples in ethane results in the formation of vitrified ice that bypasses the crystalline ice phase of water forming a thin transparent film. Vitrification serves two purposes: 1) to maintain the hydrated native state of proteins (57) and 2) protects the sample from radiation damage from the high-energy electron beam (58). Samples are prepared by application of a few microliters of sample on to a holey carbon-coated copper grid followed by blotting with filter paper to form a thin aqueous film then plunging into liquid ethane (Fig. 1.4). The use of vitrified ice provides

sufficient contrast in electron scattering of the aqueous background compared to the protein sample for imaging. Within the thin layer of vitrified ice protein particles are randomly oriented but some particles tend to orient preferentially in the ice (Fig. 1.4B). Freezing conditions must be tested for each sample to determine the correct thickness of ice to achieve the proper contrast of the protein sample against the background. The buffer conditions for the protein sample must also be optimized so that the particles do not aggregate. Currently, there are automated vitrification robots (59) that are used to prepare vitrified samples that allow precise control over several parameters, such as blotting time, blotting force and humidity in sample chamber (Fig 1.4A). As long as the sample is kept below 150°C the ice will maintain the vitrified state (60). Vitrified samples can be kept for weeks or months at liquid nitrogen temperature for storage till ready for data collection.

Data collection

For high-resolution structures using single particle cryo-EM it is imperative that the best images possible are collected. The main factors that affect the quality of images and the process of data collection are radiation damage and microscope performance, excluding consideration of factors that can be compensated for during grid preparation. Biological samples are very sensitive to damage from the high-energy electron beams used for imaging, typically 200 kV or 300 kV. To reduce radiation damage, low dose procedures (61) are used for imaging biological samples typically only 10 to 20 $e^-/Å^2$. Low dose imaging consists of three modes, search, focus and exposure. The grid is screened in search mode using a defocused diffraction pattern producing a high contrast image with negligible electron dosage. The user then switches to focus mode, where an area adjacent to the area to be imaged is used for focusing, correction of

astigmatism and setting the defocus value. After focusing, the user switches to exposure mode to collect an image typically using 0.5 sec or 1 sec exposure time. The low dose imaging used for data collection results in low signal to noise, to compensate for this high efficiency detectors are used for imaging and a large number of images are required to have enough signal to recover high resolution features. Complementary metal-oxide-semiconductor (CMOS) detectors have been developed that have lower noise compared to the traditional CCD detectors. The efficiency of collecting images has been improved by implementation of automated data collection programs such as LEGINON (62).

The second factor that affects data collection consists of the aberrations that arise from the contrast transfer function (CTF) that describes the process of image formation in TEM. In TEM, electrons are accelerated from an electron source, collimated to form a coherent beam, pass through the sample to the objective lens system that forms an image composed of both the unscattered and scattered electrons. As electrons pass through the sample a phase shift of the electrons occurs that contributes to image formation by the objective lens described by the CTF. The CTF is a sine wave function that oscillates between -1 and 1, where negative values of the CTF represent no observed signal and positive values represent observed signal. The oscillation of the CTF is seen in the fast Fourier transform (FFT) of an image that displays rings of alternating light and dark regions referred to as Thon rings. As a result of the CTF, images must be collected at values away from the focal point (defocus) in order to have contrast in images. To compensate for the CTF, images are collected over a range of defocus values determined by the relative signal to noise of each particular protein.

Images quality is also determined by factors that modulate the CTF such as astigmatism, drift and charging that arise from the environment or the optics of the microscope. Astigmatism

is a distortion in the magnetic field maintained by the lens of the electron microscope that result in ellipsoid rings in the CTF. During data collection the astigmatism of the objective lens must be corrected and is easily done by the user. In contrast to defocus and astigmatism that can be corrected by the user, drift and charging are affects that can be reduced but cannot be accounted for during data processing. Drift is movement of the sample while in the microscope that results in smearing of high-resolution features of the protein particle. There are two types of drift, thermal drift and mechanical drift. Thermal drift is caused by a temperature differential between the sample and the holder that is outside the microscope. Mechanical drift is the actual movement of the sample due to interference from external noise, magnetic fields and vibrations. Monitoring the room for mechanical sources and using cryo-holders that have a Dewar of nitrogen to keep the sample at a stable temperature can control drift. Whereas drift is a global effect across the entire grid, charging can be a local effect that results in distortion of the image. Charging is the accumulation of charge in the sample that occurs in the column because water is not the best electrical conductor, thus overtime charge may buildup on the sample. Charging, drift and astigmatism can be monitored by observing the Thon rings in the FFT of an image. Each of the modulators of the CTF affect the "roundedness" of the Thon rings, a perfect image would have circular rings radiating from the center equally in all directions. Drift results in the rings disappearing perpendicular to the direction of the drift and an astigmatism results in the ellipsoid shape of the rings. Observation of the behavior of the sample within the microscope during data collection can help to make sure the best possible images are recorded.

Another factor that significantly contributes to loss of high-resolution information is beam-induced movement of protein particles. Until recently rigorous particle selection was used after imaging to remove particles that had smeared features, requiring extensive amounts of time

from the user. A new technique that involves collecting movies of each sample area to correct for sample movement, which has only been made possible by the fast, read out rates of the new direct detectors (63). Using this technique to correct for beam induced movement, Bai *et al*. have been able to reconstruct the 80S ribosome, a large asymmetric protein & RNA complex, to near atomic resolution with only a limited set of 30,000 particles (51). A comparable reconstruction of the 80S ribosome, imaged using film as the recording medium, reached a resolution of only 6 Å and required nearly 1.4 million particles (64). With this new approach to single particle cryo-EM it maybe possible to significantly reduce the amount of individual particles required and improve the routinely attainable resolution of reconstructions (65).

Image processing & reconstruction

There are several steps in the process of going from cryo-EM images to calculating a three-dimensional reconstruction: assessment of image quality, analysis of particles, determining orientation parameters of individual particles and reconstruction of a three-dimensional model. The first step in image processing is to determine the CTF parameters for each image, defocus & astigmatism, which corrects the systematic distortion of the image by the CTF. This is readily accomplished using programs CTFFIND and CTFTILT (for tilted images) (66) that produce a power spectrum for each image that can be used to assess the quality of images. The power spectrum shows the same characteristics Thon rings of the CTF and can be used to judge image quality by looking for distortions in the Thon rings (discussed in detail in the previous section). Careful examination of the raw image will reveal the effects of charging that results in a smearing effect across part of an image compared to drift, which is observed as a smearing across the entire image. Each type of image defect results in a decrease in the resolution of

information that is contained in each image and distortion of the actual particle image affecting the ability to align the particles necessary for reconstruction.

A critical parameter for a successful single particle reconstruction is to have images of protein particles with many different views covering as much of angular space as possible. To assess particle distribution and quality, particles must first be selected from the images either manually by giving the coordinates of each particle in each image (55,67) or by using automated programs to select the particles of interest (56,68). The particles can then be classified using multivariate statistical analysis (69) and aligned with respect to each other in regards to rotation and translation. Projection averages are then calculated representing the average of the particles that have the same orientation. The projection averages can be assessed to determine the different views of particles present in the data set and classes representing contamination or incomplete protein complexes can be identified. Besides serving as a step in the process to calculate a reconstruction, projection averages can provide information regarding different conformations that maybe present, symmetry, oligomeric state and assembly of protein complexes.

The next step is to determine the specific orientation of each of the particles in the data set in order to calculate a three dimensional reconstruction. The orientation of each particle can be represented by five alignment parameters: translational alignment to center the particle (x, y) and the three Euler angles to describe the rotational alignment (α, β, γ). Orientation parameters are determined using an initial reference three-dimensional model generated by angular reconstitution (70,71), random conical tilt (72) or the use of a model from a homologous structure (73,74). Angular reconstitution uses the theory of common lines in real space to reconstruct a three-dimensional volume from a set of projection averages (71). Random conical

tilt records a series of the same area at different angles; the tilt angle (one of the Euler angles) between projections is known and is used to help determine the other Euler angles (72). If a model from a homologous structure is available a low-resolution density can be calculated from the structure and used for projection matching to determine initial orientation parameters (73,74).

After determination of the initial alignment parameters of each particle, a preliminary three-dimensional model can be calculated. Refinement of the reconstruction is then iteration between refinement of particle parameters based on the model using a combination of projection matching and common lines and calculation of a new model based on the new alignment parameters. The iteration between refinement of alignment parameters and reconstruction continues until the three-dimensional model and parameters remain unchanged. Several programs have been developed to carry out refinement and reconstruction such as IMAGIC (56), EMAN (68), SPIDER (55) and FREALIGN (75). FREALIGN (75) is a program that has been developed to reconstruct and refine alignment parameters in Fourier space using the common lines approach. Briefly, FREALIGN searches angular space to find the best alignment of a particular projection with the model, based on a restricted amount of structural information defined by the user. Initially, only low-resolution information is used for alignment of particles and as refinement continues more information is included until there is no improvement in the reconstruction. The FREALIGN algorithms work entirely in Fourier space with the last step as an inverse Fourier transform to produce a three-dimensional model, this results in faster computations than IMAGIC, EMAN or SPIDER (75).

The resolution of reconstructions from single particle cryo-EM data is determined based on the comparison of two independent reconstructions over Fourier space using Fourier shell correlation (FSC) (76). Typically the data is divided into two data sets, reconstruction and

refinement is carried out for each independent of one another. The two data sets are then compared using the FSC, measuring how well the reconstructions compare over increasing resolution. Initially at low resolution (40 Å) the correlation should be close to 1, but with increasing resolution the correlation will drop. In the past, there was a debate about whether to use the 0.5 (77) or 0.143 (78) FSC cutoff to determine the resolution of the reconstruction, but the consensus in the field in the last five years has been to use the more conservative 0.5 criteria. The real testament to the resolution of a three-dimensional reconstruction is the amount of detail that is present in the map. If the structure of a homologous protein or components of the complex are known then comparison of the reconstruction to a map calculated from the known structures can aid in determining the resolution. At sub-nanometer resolution, alpha helices should be appearing as cylinder-like densities and with increasing resolution indentations of helices should be visible. Atomic resolution (< 4 Å) results in the observation of densities characteristic of particular side chains, thus a complete structure could be built into the density. Single particle cryo-EM reconstructions have routinely been able to able to achieve 10 to 20 Å resolution, but larger symmetric particles have been able to achieve near-atomic and atomic level resolution. The development of better methods for alignment will help to push the resolution of smaller asymmetric single particle cryo-EM reconstruction to sub-nanometer resolution.

Insight into biological processes provided by single particle cryo-EM

Biological processes often involve many proteins or many copies of a protein to form macromolecular machines that carry out a variety of functions necessary for the cell. Single particle cryo-EM bridges a gap in the investigation of structural biology with the ability to study large heterogeneous proteins and protein complexes. With the introduction of freezing biological

samples using ethane for TEM imaging, large macromolecular machines could be visualized for the first time. Single particle cryo-EM has significantly contributed to providing a landscape of macromolecular assemblies across a variety of areas in biology, from the translation of proteins to the maturation of viruses.

The highest resolution single particle cryo-EM reconstructions have come from the investigation of the structure of viral capsids that have revealed mechanisms of capsid maturation, and viral entry into the cell. There are now many examples of single particle cryo-EM reconstruction of viral capsids reaching near-atomic resolution (79,80,81,82,83) and a few that have reached atomic resolution (54,84,85). In the structure of the human adenovirus, the extensive network of protein-protein interactions contributing could be mapped as a result of the 3.6-Å resolution achieved (54). The priming mechanism for viral entry of a non-enveloped virus was revealed by a 3.3-Å resolution cryo-EM reconstruction of an aquareovirus (85). In studies where a structure was also determined using X-ray crystallography, the single particle cryo-EM reconstructions have revealed details not observed in the crystal structure, such as the cryo-EM structure of the papillomavirus that showed a C-terminal to extend to a nearby subunit in the capsid but then loop back to the originating subunit (83).

The ability to reach near-atomic resolution has mostly been restricted to viral capsids due to the large molecular mass and the high icosahedral symmetry, giving rise to increased signal to noise with a small number of images. Low-resolution reconstructions still have value depending on the types of questions that are to be addressed, such as large conformational changes and assembly of large protein complexes. Several functional states of the large asymmetric ribosome have been determined using single particle cryo-EM to map binding sites and conformation changes to understand the process of mRNA translation to form proteins (86,87,88,89). The

study of GroEL, an E. coli protein chaperone, has been extensively studied using single particle cryo-EM achieving sub-nanometer resolution to visualize changes that occur with the addition of ligands (90), the apo-GroEL assembly (91) and to understand conformational changes that occur in the presence of ATP (92), the energy needed to drive protein folding. Ubiquitin mediated protein degradation by the proteasome has been the focus of a number of single particle cryo-EM studies that have reached sub-nanometer resolution (93). These examples only represent a small sampling of the macromolecular assemblies studied by single particle cryo-EM, but they demonstrate the potential this particular technique has to provide insight into a broad range of biological processes that cannot be visualized by other structural biology methods.

Currently in the EM Data Bank there are approximately 1400 density maps (94) that have been submitted covering a wide range of cellular biology at varying resolutions determined using single particle cryo-EM. The ability to study large heterogeneous protein complexes with the requirement of crystals or large amounts of purified protein is a significant advantage over other structural biology techniques. Most three-dimensional models determined using single particle cryo-EM achieve only low resolution > 10 Å, but this still provides information to answer questions regarding oligomeric state, assembly of protein complexes and observation of large conformational changes. The use of single particle cryo-EM is limited to large proteins and near-atomic level resolution is only achieved with large data sets or high symmetry. Current developments in the field are working on both of these limitations. As discussed previously, a method has been developed to reduce the amount of data needed for a high-resolution reconstruction by collecting movies to correct for beam-induced movement (51). To address the use of single particle cryo-EM to observe smaller proteins, a method using antibody fragments bound to a specific protein to provide discernible features to improve alignment of the particles

(95). With continuing advancements with TEM, developments of high efficiency imaging devices, data collection and alignment algorithms, single particle cryo-EM will be a versatile method to study the macromolecular machines that carry out many biological processes essential for life.

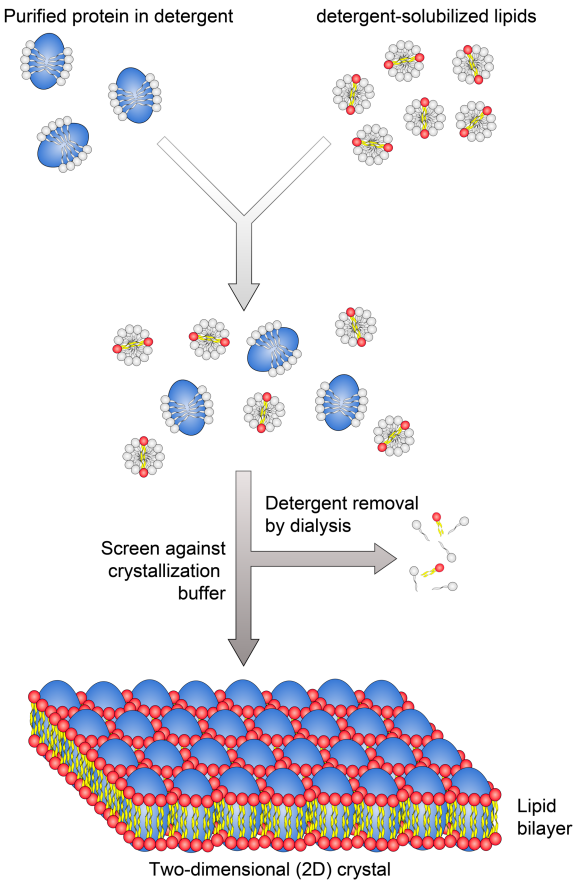

Figure 1.1 Two-dimensional crystal formation. Two-dimensional crystals are produced by combining purified detergent-stabilized membrane protein with detergent solubilized lipids followed by removal of the detergent using dialysis. The membrane proteins then reconstitute with the forming lipid bilayer, which may results in the formation of ordered arrays of membrane protein and lipids (96).

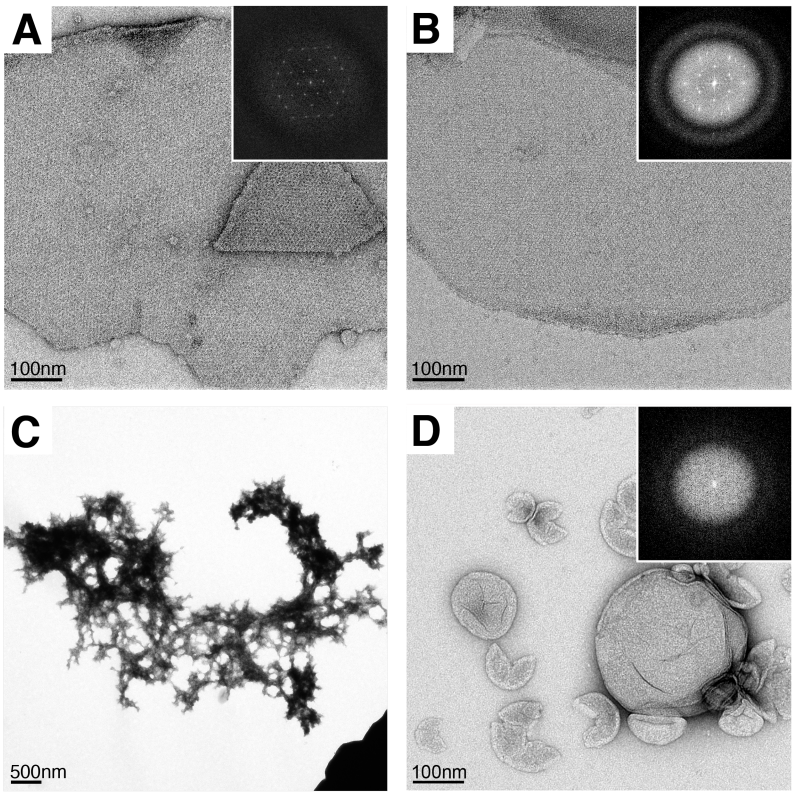

Figure 1.2 Examples of possible results during two-dimensional crystallization. The types of expected results of screening two-dimensional crystallization of membrane proteins are crystalline sheets (**A**), crystalline vesicles (**B**), protein aggregates (**C**) and reconstitution of the membrane protein with lipids forming vesicles without any visible lattice formation (**D**). Fourier transforms of each of the images is shown in the upper right of each panel indicating the lattice formation by the presence of intense spots or the lack there of. The TEM images shown were samples that had been negative stained with uranyl formate.

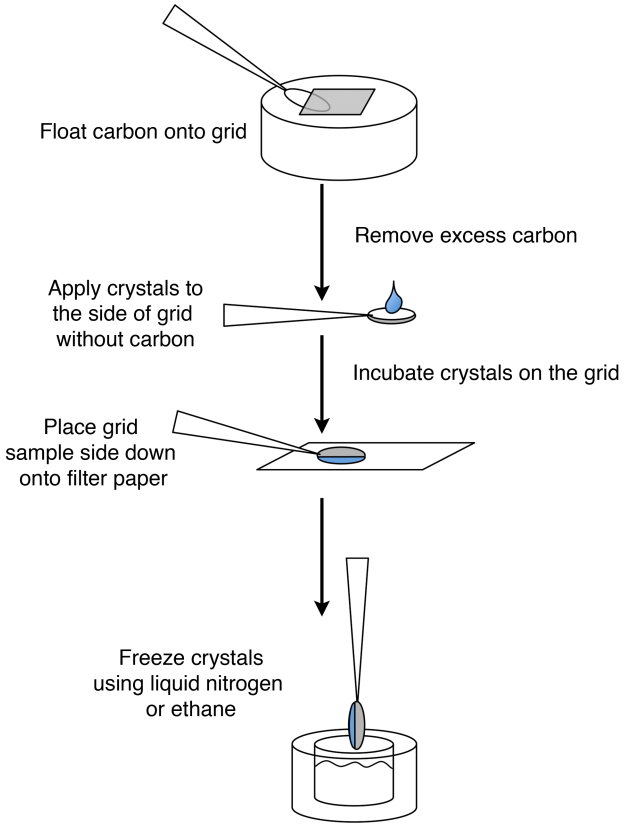

Figure 1.3 Back-injection method for freezing two-dimensional crystals for cryo-EM. The first step to prepare two-dimensional crystals for cryo-EM using back-injection (22) is to float a layer of carbon suspended on top of an aqueous solution onto a copper or molybdenum grid. The excess carbon is removed by gently scraping the opposite side of the grid with a pipette tip. The two-dimensional crystals are applied to the non-carbon coated side of the grid and incubated on the grid for a minute to allow the crystals to stick to the carbon. At this step, embedding solution containing sugar maybe added or the sugar can be added to the solution of crystals directly prior to application on the grid. The grid is then blotted with filter paper by placing the grid sample side down onto filter paper for up to one minute to remove excess solution. Then finally the grid is frozen in liquid nitrogen or ethane and is ready for data collection using TEM.

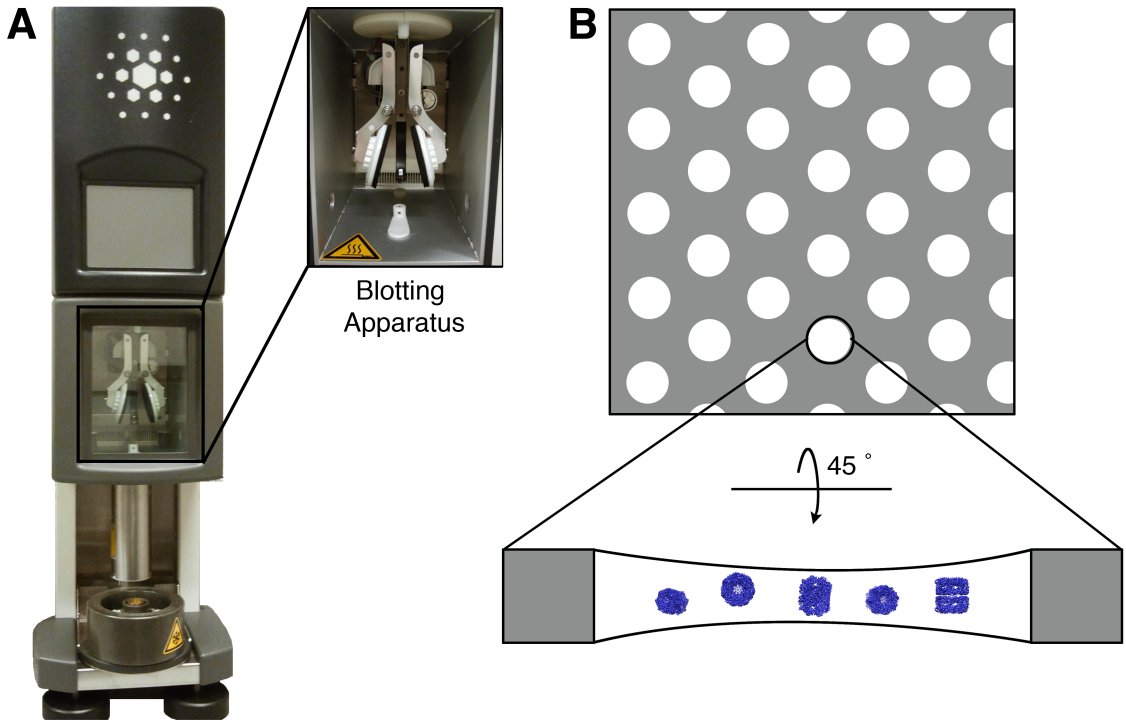

Figure 1.4 Sample preparation for single particle cryo-EM. (**A**) FEI vitrification robot that is used to prepare vitrified samples for single particle cryo-EM. The robot allows control of humidity, blotting time and temperature. A close up of the blotting apparatus contained within the humidity chamber is shown. Tweezers (not shown) hold the delicate grid during the blotting procedure and plunge the grid into the Dewar of liquid ethane to form vitrified ice. (**B**) Holey carbon copper grids are used for single particle cryo-EM. A thin layer of vitrified ice forms across the holes of the carbon-coated grid that contain the protein particles oriented randomly within the ice. Each hole in the grid is then imaged in the TEM, producing images of the protein particles that are processed to produce a single particle reconstruction.

WORKS CITED IN CHAPTER 1

1. Nogales, E., Wolf, S.G. & Downing, K.H. (1998). Structure of the alpha beta tubulin dimer by electron crystallography. *Nature*, **391**, 199-203.

2. Surade, S., Klein, M., Stolt-Bergner, P.C., Muenke, C., Roy, A. & Michel, H. (2006). Comparative analysis and "expression space" coverage of the production of prokaryotic membrane proteins for structural genomics. *Protein Sci.* **15**, 2178-2189.

3. Nannenga, B.L. & Baneyx, F. (2011). Reprogramming chaperone pathways to improve membrane protein expression in *Escherichia coli*. *Protein Sci.* **20**, 1411-1420.

4. Wagner, S. *et al.* (2008). Tuning *Escherichia coli* for membrane protein overexpression. *Proc. Natl. Acad. Sci. USA*, **105**, 14371-14376.

5. Le Maire, M., Champeil, P., Moller, J.V. (2000). Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim. Biophys. Acta*, **1508**, 86-111.

6. Nussberger, S., Dorr, K., Wang, D.N., Kuhlbrandt, W. (1993). Lipid-protein interactions in crystals of plant-light harvesting complex. *J. Mol. Biol*. **234**, 347-356.

7. Hamilton, R.J. & Hamilton, S. (1992). Lipid Analysis-A practical approach, IRL Press at Oxford University Press, Oxford.

8. Lotz, M., Haase, W., Kuhlbrandt, W. & Collinson, I. (2008). Projection structure of yidC: a conserved mediator of membrane protein assembly. *J. Mol. Biol.* **375**, 901-907.

9. Williams, K.A., Geldmacher-Kaufer, U., Padan, E., Schuldiner, S. & Kuhlbrandt, W. (1999). Projection structure of NhaA, a secondary transporter from *Escherichia coli*, at 4.0 Å resolution. *EMBO J.* **18**, 3558-3563.

10. Hefti, M.H., Caroline, J.G., der Toorn V.V., Dixon R. & Vervoort, J. (2001). A novel purification method for histidine-tagged proteins containing a thrombin cleavage site. *Anal Biochem.* **295**, 180–185.

11. Tsai, C. & Ziegler, C. (2005). Structure determination of secondary transport proteins by electron crystallography: Two-dimensional crystallization of the betaine uptake system BetP. *J. Mol. Microbiol. Biotechnol.* **10**, 197-207.

12. Hite, R.K., Li, Z. & Walz, T. (2010). Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2D crystals. *EMBO J.* **29**, 1652-1658.

13. Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S.C. & Walz, T. (2005). Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature*, **438**, 633-638.

14. Heymann, J.A.W., Hirai, T., Shi, D. & Subramaniam, S. (2003). Projection structure of the bacterial oxalate transporter OxlT at 3.4 Å resolution. *J. Struct. Biol.* **144**, 320-326.

15. Hiroaki, Y. *et al.* (2006). Implications of the aquaporin-4 structure on array formation and cell adhesion. *J. Mol. Biol.* **355**, 628-639.

16. Signorell, G.A., Kaufmann, T.C., Kukulski, W., Engel, A. & Remigy, H.W. (2007). Controlled 2D crystallization of membrane proteins using methyl-beta-cyclodextrin. *J. Struct. Biol.* **157**, 321-328.

17. Rigaud, J.L., Mosser, G., Lacaperre, J.J., Olofsson, A., Levy, D. & Ranck, J.L. (1997). Bio-Beads: An efficient strategy for two-dimensional crystallization of membrane proteins. *J. Struct. Biol.* **118**, 226-235.

18. Wisedchaisri, G., Reichow, S.L. & Gonen, T. (2011). Advances in structural and functional analysis of membrane proteins by electron crystallography. *Structure*, **19**, 1381-1393.

19. Gipson, B., Zeng, X., Zhang, Z.Y. & Stahlber, H. (2006). 2dx – User-friendly image processing for 2D crystals. *J. Struct. Biol.* **157**, 64-72.

20. Raunser, S. & Walz, T. (2009). Electron crystallography as a technique to study the structure on membrane proteins in a lipidic environment. *Annu. Rev. Biophys.* **38**, 89-105.

21. Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., Downing, K.H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899-929.

22. Wang, D.N. & Kuhlbrandt, W. (1991). High-resolution electron crystallography of light-harvesting chlorophyll a/b-protein complex in three different media. *J. Mol. Biol.* **217**, 691-699.

23. Kimura, Y. *et al.* (1997). Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature*, **389**, 206-211.

24. Gyobu, N., Tani, K., Hiroaki, Y., Kamegawa, A., Mitsuoka, A. & Fujiyoshi, Y. (2004). Improved specimen preparation for cryo-electron microscopy using a symmetric carbon sandwich technique. *J. Struct. Biol.* **146**, 325-333.

25. Crowther, R.A., Henderson, R. & Smith, J.M. (1996). MRC image processing programs. *J. Struct. Biol.* **116**, 9-16.

26. Gipson, B., Zeng, X. & Stahlberg, H. (2007). 2dx_merge: Data management and merging of 2D crystal images. *J. Struct. Biol.* **160**, 375-384.

27. Collaborative Computational Project, Number 4. (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst. D*, **50**, 760-763.

28. Hirai, T., Murata, K., Mitsuoka, K., Kimura, Y. & Fujiyoshi, Y. (1999). Trehalose embedding technique for high-resolution electron crystallography: application to structural study on bacteriorhodopsin. *J. Electron Microsc. (Tokyo)*, **56**, 131-140.

29. Philippsen, A., Schenk, A.D., Stahlberg, H. & Engel, A. (2003). IPLT – image processing library and toolkit for the electron microscopy community. *J. Struct. Biol.* **144**, 4-12.

30. Philippsen, A., Schenk, A.D., Signorell, G.A., Mariani, V., Berneche, S. & Engel, A. (2007). Collaborative EM image processing with the IPLT image processing library and toolbox. *J. Struct. Biol.* **157**, 28-37.

31. Wisedchaisri, G. & Gonen, T. (2011). Fragment-based phase extension for three-dimensional structure determination of membrane proteins by electron crystallography. *Structure*, **19**, 976-987.

32. Ziegler, C., Morbach, S., Schiller, D., Kramer, R., Tziatzios, C., Schubert, D. & Kuhlbrandt, W. (2004). Projection structure and oligomeric state of the osmoregulated sodium/glycine betaine symporter BetP of *Corynebacterium glutamicum*. *J. Mol. Biol.* **337**, 1137-1147.

33. De Feo, C.J., Aller, S.G., Siluvai, G.S., Blackburn, N.J. & Unger, V.M. (2009). Three-dimensional structure of the human copper transporter hCTR1. *Proc. Natl. Acad. Sci. USA*, **106**, 4237-4242.

34. Unger, V.M., Kumar, N.M. Gilula, N.B. & Yeager, M. (1997). Projection structure of a gap junction membrane channel at 7 Å resolution. *Nat. Struct. Biol.* **4**, 39-43.

35. Oshima, A., Tani, K., Hiroaki, Y., Fujiyoshi, Y. & Sosinsky G.E. (2008). Projection structure of a N-terminal deletion mutant of connexin 26 channel with decreased central pore density. *Cell Commun. Adhes.* **15**, 85-93.

36. Cheng, A., Schweissinger, D., Dawood, F., Kumar, N. & Yeager, M. (2003). Projection structure of full length connexin 43 by electron cryo-crystallography. *Cell Commun. Adhes.* **10**, 187-191.

37. Murata, K. *et al.* (2000). Structural determinants of water permeation through aquaporin-1. *Nature*, **407**, 599-605.

38. Kuhlbrandt, W., Wang, D.N. & Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystallography. *Nature*, **367**, 614-621.

39. Unwin, N. (2005). Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J. Mol. Biol.* **346**, 967-989.

40. Holm, P.J. *et al.* (2006). Structural basis for detoxification and oxidative stress protection in membranes. *J. Mol. Biol.* **360**, 934-945.

41. Jegerschold, C. *et al.* (2008). Structural basis for induced formation of the inflammatory mediator prostaglandin E2. *Proc. Natl. Acad. Sci. USA*, **105**, 11110-11115.

42. Walz, T., Smith, B.L., Zeidel, M.L., Engel, A. & Agre, P. (1994). Biologically active two-dimensional crystals of aquaporin CHIP. *J. Biol. Chem.* **269**, 1583-1586.

43. Zheng, H., Taraska, J., Merz, AJ. & Gonen, T. (2010). The prototypical H^+/galactose symporter GalP assembles into functional trimers. *J. Mol. Biol.* **396**, 593-601.

44. Unwin, N. (1995). Acetylcholine receptor channel imaged in the open state. *Nature*, **373**, 37-43.

45. Mitsuoka, K., Hirai, T., Murata, K., Miyazawa, A., Kidera, A., Kimura, Y. & Fujiyoshi, Y. (1999). The structure of bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: Implication of the charge distribution. *J. Mol. Biol.* **286**, 861-882.

46. Crowther, R.A. (1971). Procedures for 3-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **261**, 221-230.

47. Cheng, Y., Zak, O., Aisen, P., Harrison, S.C. & Walz, T. (2004). Structure of the human transferrin receptor-transferrin complex. *Cell*, **116**, 565-576.

48. Osenkowski, P. *et al.* (2009). Cryoelectron microscopy structure of purified gamma-secretase at 12 A resolution. *J. Mol. Biol*. **385***,* 642–652.

49. Cong, Y. *et al.* (2010). 4.0-A resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement. *Proc. Natl. Acad. Sci. USA*, **107**, 4967-4972.

50. Ludtke, S.J., Baker, M.L., Chen, D.H., Song, J.L., Chuang, D.T. & Chiu, W. (2008). De novo backbone trace of GroEL from single particle electron cryomicroscopy. *Structure*, **16**, 441-448.

51. Bai, X., Fernandez, I.S., McMullan, G. & Scheres, S.H.W. (2013). Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife*, **2**, e00461.

52. Zhang, X., Settembre, E., Xu, C., Dormitzer, P.R., Bellamy, R., Harrison, S.C. & Grigorieff, N. (2008). Near-atomic resolution using electron cryomicroscopy and single particle reconstruction. *Proc. Natl. Acad. Sci. USA*, **105**, 1867-1872.

53. Zhang, X., Jin, L., Fang, Q., Hui, W.H. & Zhou, Z.H. (2010). 3.3 Å Cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell*, **141**, 472-482.

54. Liu, H., Jin, L., Koh, S.B.S, Atanasov, I., Schein, S., Wu, L. & Zhou, H.Z. (2010). Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. *Science*, **329**, 1038-1043.

55. Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadi, M. & Leith, A. (1996). SPIDER and WEB: processing and visualization of iamges in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190-199.

56. van Heel, M., Harauz, G. & Orlova, E.V. (1996). A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17-24.

57. Taylor, K. & Glaeser, R.M. (1974). Electron diffraction of frozen, hydrated protein crystals. *Science*, **186**, 1036-1037.

58. Chiu, W. (1986). Electron microscopy of frozen, hydrated biological specimens. *Annu. Rev. Biophys. Chem.* **15**, 237-257.

59. Frederick, P.M., Bomans, P., Franssen, V. & Laeven, P. (2000). A vitrification robot for time resolved cryo-electron microscopy. *Proceedings of 12th European Congress on Electron Microscopy* (S. Cech and R. Janisch, eds.), Vol. I, pp. B383-B384. Reklammi Atelier Kupa Brno.

60. Stewart, M. (1990). Electron microscopy of biological macromolecules. In: *Modern Microscopies* (P.J. Duke & A.G. Michette, eds.) pp.9-40. Plenum Press, New York.

61. Frank, J. (2006). Electron Microscopy of Macromolecular Assemblies, pp. 64-66. Oxford University Press, New York.

62. Suloway, C. *et al.* (2005). Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41-60.

63. Campbell, M.G. *et al*. (2012). Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure*, **20**, 1823–1828.

64. Armache, J.P. *et al.* (2010). Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribsome at 5.5 Å. *Proc. Natl. Acad. Sci. USA*, **107**, 19748-19753.

65. Grigorieff, N. (2013). Direct detection pays off for electron cryo-microscopy. *eLife*, **2**, e00573.

66. Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334-347.

67. Smith, J.M. (1999). Ximdisp--A visualization tool to aid structure determination from electron microscope images. *J. Struct. Biol.* **125**, 223–228.

68. Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I. & Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38-46.

69. van Heel, M. (1984). Multivariate statistical classification of noisy images (randomly oriented biological molecules). *Ultramicroscopy*, **13**, 165-184.

70. Crowther, R.A. (1971). Procedure for three-dimensional reconstruction of spherical ciruses by Fourier synthesis from electron micrographs. *Phil. Trans. R. Soc. Lond. B.* **261**, 221-230.

71. van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, **21**, 111-124.

72. Radermacher, M. (1988). The three-dimensional reconstruction from reconstruction of single paricles from random and non-random tilt series. *J. Electron Microsc. Tech.* **9**, 359-394.

73. Penczek, P., Grassucci, R.A. & Frank, J. (1994). The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryoelectron microscopy of biological particles. *Ultramicroscopy*, **53**, 251-270.

74. Grigorieff, N. (1998). Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 A in ice. *J. Mol. Biol.* **277**, 1033-1046.

75. Grigorieff, N. (2007). FREALIGN: High-resolution refinement of single particle structures. *J. Struct. Biol.* **157**, 117-125.

76. Harauz, G. & van Heel, M. (1986). Exact filters for general geometry 3-dimensional reconstruction. *Optik*, **73**, 146-156.

77. Bottcher, B., Wynne, S.A. & Crowther, R.A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*, **386**, 88-91.

78. Rosenthal, P.B. & Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721-745.

79. Jiang, W., Baker, M.L., Jakana, J., Weigele, P.R., King, J. & Chiu, W. (2008). Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. *Nature*, **451**, 1130-1134.

80. Zhang, X. *et al.* (2008). Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl. Acad. Sci. USA*, **105**, 1867-1872.

81. Chen, J.Z. *et al.* (2009). Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. *Proc. Natl. Acad. Sci. USA*, **106**, 10644-10648.

82. Cheng, L., Zhu, J., Hui, W.H., Zhang, X., Honig, B., Fang, Q. & Zhou, Z.H. (2010). Backbone model of an aquareovirus virion by cryo-electron microscopy and bioinformatics. *J. Mol. Biol.* **397**, 852-863.

83. Wolf, M., Garcea, R.L., Grigorieff, N. & Harrison, S.C. (2010). Subunit interactions in bovine papillomavirus. *Proc. Natl. Acad. Sci. USA*, **107**, 6298-6303.

84. Yu, X., Jin, L., Zhou, Z.H. (2008). 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature*, **453**, 415-419.

85. Zhang, X., Jin, L., Fang, Q., Hui, W.H. & Zhou, Z.H. (2010). 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell*, **141**, 472-482.

86. Connell, S.R. *et al.* (2007). Structural basis for interaction of the ribsoome with the switch regions of GTP-bound elongation factors. *Mol. Cell*, **25**, 751-764.

87. Julian, P. *et al.* (2008). Structure of ratcheted ribsoomes with tRNAs in hybrid states. *Proc. Natl. Acad. Sci. USA*, **105**, 16924-16927.

88. Fu, J. *et al.* (2010). Visualizing the transfer-messenger RNA as the ribosome resumes translation. *EMBO J.* **29**, 3819-3825.

89. Strunk, B.S. *et al.* (2011). Ribosome assembly factors prevent premature translation initiation by 40S assembly intermediates. *Science*, **333**, 1449-1453.

90. Clare, D.K., Bakkes, P.J., van Heerikhuizen, H., van der Vies, S.M. & Saibil, H.R. (2009). Chaperonin complex with a newly folded protein encapsulated in the folding chamber. *Nature,* **457**, 107-110.

91. Ludtke, S.J., Baker, M.L., Chen, D.H., Song, J.L., Chuang, D.T. & Chiu, W. (2008). De novo backbone trace of GroEL from single particle electron cryomicroscopy. *Structure*, **16**, 441-448.

92. Clare, D.K. *et al.* (2012). ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell*, **149**, 113-123.

93. Lander, G.C., Estrin, E., Matyskiela, M.E., Bashore, C., Nogales, E. & Martin, A. (2012). Complete subunit architecture of the proteasome regulatory particle. *Nature*, **482**, 186-191.

94. EM Data Bank, www.emdatabank.org, accessed 03/01/2013.

95. Wu, S. *et al.* (2012). Fabs enable single particle cryoEM studies of small proteins.

Structure, **20**, 582-592.

96. Nannenga, B.L., Iadanza, M.G., Vollmar, B.S., Gonen, T. (2013). Overview of electron crystallography of membrane proteins: crystallization and screening strategies using negative stain electron microscopy. *Curr. Protoc. Protein Sci.* **72**, 17.15.1-17.15.11.

Chapter II

Investigating the structure & mechanism of glycerol uptake by the human glycerol facilitator, aquaporin 3

SUMMARY

Aquaporins are membrane proteins that facilitate the diffusion of water across the cell membrane. Aquaglyceroporins are a subset of aquaporins that not only facilitate the diffusion of water, but also small solutes such as glycerol and urea. Aquaporin 3, an aquaglyceroporin, has been shown to be involved in urine concentration in the kidney (1), glycerol transport in red blood cells (2), necessary for wound healing (3) and hydration of the skin (4). Recently, hAQP3 has been shown to be over-expressed in skin carcinomas (5). Moreover, AQP3 knockout mice are protected from tumor formation (5). Therefore, determining the structure and regulation of AQP3 would be beneficial to guide future studies of the particular role AQP3 plays in tumorigenesis. Currently, there are two structures of aquaglyceroporins, GlpF (6) from *Escherichia coli* and PfAQP (7) from *Plasmodium falciparum*. These structures provided details regarding the mechanism of glycerol and water diffusion through aquaglyceroporins, but differ significantly in substrate regulation and rates of substrate diffusion compared to mammalian aquaglyceroporins. Questions still remain as to what is the structure of a mammalian aquaglyceroporin, what characteristics of the pore facilitate both water and glycerol diffusion, and how is diffusion regulated. To provide insight into the above questions, electron crystallography can be used to determine the structure of a mammalian aquaglyceroporin in a lipid environment and to assess the regulation of hAQP3. Here I present results that show the heterologous over-expression of human aquaporin-3 (hAQP3) in *Pichia pastoris*, purification of

hAQP3 and two-dimensional crystallization trials for structure determination using electron crystallography.

INTRODUCTION

Water is an essential molecule for all physiological processes necessary for life, in both plants and animals. Water can diffuse freely across the lipid bilayer, but at very slow rates. Peter Agre's group discovered a membrane protein in red blood cells that was able to facilitate the diffusion of water across the membrane at high rates (8,9). This discovery led to the identification of several water channels that are expressed in a wide variety of tissues referred to as aquaporins. Members of the aquaporin family provide an extremely fast and highly regulated pathway for water permeation.

There are currently 13 known human aquaporins that are expressed in a variety of tissues and organs, such as the skin, red blood cells, the eyes, the intestinal tract, the lung, the kidneys and the brain (10). The aquaporin family is divided into two classes based on channel selectivity. Aquaporins that are only selective to water include AQP0, AQP1, AQP2, AQP4 and AQP5. The aquaglyceroporins, AQP3, AQP7, AQP9 and AQP10, are less selective and allow water, glycerol and urea to permeate the pore. Two aquaporins appear to be ion-channel like: AQP6 is permeable to small anions and AQP8 is permeable to ammonia, which do not fall into the traditional categories (11). AQP11 and AQP12 have recently been identified but have not been extensively characterized yet (11).

The first structure determined of an aquaporin was AQP1 by electron crystallography in 2000 (12). Since then, several structures of mammalian (AQP0 (13,14,15), AQP1 (16), AQP4 (17,18) & AQP5 (19)), plant (SoPIP2;1 (20)), yeast (21), bacterial (GlpF (6) & AqpZ (22)) and

archaea (AqpM (23)) aquaporins have been determined, revealing characteristics of the aquaporin family. Aquaporins are composed of six transmembrane helices that form a barrel through which solutes pass. Four aquaporin monomers come together to form a tetramer in which each monomer forms a permeable pore (24).

An important characteristic of aquaporins is that they exclude protons from diffusing through the channel. This seems surprising because it is expected that if water is able to travel through the channel close to the rate of diffusion that protons would be able to as well. From structural and mutagenesis studies, two features of the channel are responsible for the exclusion of protons and cations, as well as regulating solute permeability: the asparagine-proline-alanine (NPA) motif and the aromatic-arginine (ar/R) constriction site (Fig. 2.1). The NPA region is located at the center of the pore where two short helices meet, HB and HE. Helices HB and HE are oriented with their N-termini meeting at the center of the pore. The asparagine of the NPA motif is thought to bind to the oxygen of a water molecule orienting the hydrogens to be perpendicular to the channel axis rendering them unable to make hydrogen bonds with neighboring waters in the channel. This specific orientation of the water molecule results in the inability to form a continuous hydrogen-bonding network thereby blocking proton diffusion. This mechanism for inhibition of proton diffusion was first proposed by Murata *et al*. describing the first atomic resolution structure of AQP1 determined using electron crystallography and is referred to as the proton exclusion mechanism (12). Recent mutagenesis and molecular dynamics simulations have provided support for this hypothesis as well as showing that both the ar/R constriction site and NPA motif participate in inhibition of protons and cations (25,26).

Members of the aquaglyceroporin family can facilitate the diffusion of a wide variety of small solutes including water, glycerol, urea, anions and ammonia. The ar/R constriction site

governs the solute selectivity in aquaporins. The ar/R constriction site consists of an arginine residue surrounded by several aromatic residues and is located in the extracellular vestibule of the pore immediate to the NPA motif. The individual residues that make up the aromatic region and the diameter of the pore control what solutes are able to diffuse through the channel. In the structure of bovine AQP1 (16), the ar/R constriction site consists of residues phenylalanine-58, histidine-182, cysteine-191 & arginine-197 that form a narrow pore measuring approximately 3 Å, which is only slightly larger than water. The substitution of histidine-182 in bovine AQP1 (16) to glycine-191 in GlpF (6), cysteine-191 to phenylalanine-200 and phenylalanine-58 to trptophan-48 results in the formation of a larger pore measuring approximately 3.8 Å in the structure of GlpF. The larger pore in GlpF allows glycerol, a molecule larger than water, to pass through the pore in an extended conformation (6). In addition, the substitutions in the ar/R constriction site result in a more hydrophobic pore (6), able to accommodate the carbon backbone of glycerol. Some aquaporins contain a cysteine residue in the ar/R constriction site, cysteine-191 in bovine AQP1, which binds mercury, blocking water permeation through the pore (26,27). Solute selectivity is an important feature of aquaporins that defines the differences among family members that at the level of secondary structure seem almost identical but small changes in residues lining the channel can have a dramatic effect.

Diffusion of solutes through aquaporins is regulated through a variety of ways at both the level of the protein and at the level of transcription. For example, in the plant aquaporin SoPIP2;1 phosphorylation of two serine residues in cytosolic loops opens a gate, unblocking the pore (19). Plant aquaporins also have a conserved histidine residue located on the cytosolic face of the protein that when protonated during times of flooding closes the channel (29). The lens aquaporin, AQP0, is regulated by a combination of phosphorylation and calmodulin that is

involved in lens protein homeostasis. Phosphorylation of a serine residue located in a calmodulin binding domain in the C-terminus of AQP0 results in the inability of calmodulin to bind the C-terminus preventing closure of the channel (30,31). Members of the aquaporin family can be regulated using multiple methods, such as AQP0, which is also regulated by pH (32), calmodulin (32,33), phosphorylation (30,34) and by proteolytic cleavage of the C-terminus (35). Other mechanisms of regulation of different members of the aquaporin family include sensitivity to cations (36,37), localization to different membranes (38) and expression of aquaporins at different times during cellular processes (39). Aquaporins have adapted many modes of regulation in response to the needs of the cell and the cellular processes that are taking place in a specific tissue. Regulation of aquaglyceroporins is not as well understood due to a lack of structural and mechanistic studies.

Currently a structure of a mammalian aquaglyceroporin has not been determined. Some insight has been provided from the structures of aquaglyceroporins, GlpF (6) from *E. coli* and PfAQP (7) from *P. falciparum*. Each contains the canonical six transmembrane helices that form a barrel through which water diffuses and four monomers assemble into a tetramer. From their structures, it was suggested that glycerol is able to pass through the pore as an extended linear molecule because the pore-lining residues create an amphipathic channel. The ar/R constriction site in both structures is altered to be more hydrophobic and the diameter of the constriction site has increased to accommodate the larger glycerol molecule. Despite these differences, aquaglyceroporins maintain the proton exclusion mechanism. Although these two structures provided details as to how aquaglyceroporins accommodate glycerol and water, there are many differences between GlpF and PfAQP, and the mammalian aquaglyceroporins, including solute diffusion rates, selectivity and regulation. To provide insight into the structure and mechanism of

diffusion through mammalian aquaglyceroporins, I focused my research on human aquaporin 3, a glycerol facilitator.

Human aquaporin 3 (hAQP3) is an aquaglyceroporin is permeable to water, glycerol & urea. The protein is found in a wide distribution of tissues in the body, including the eye, skin, kidney, red blood cells, the brain, respiratory tract and digestive tract (11). Its expression is regulated by epinephrine (39), insulin (40), epidermal growth factor (41) and p73 (42), a member of the p53 family. hAQP3 is expressed most commonly in cells in the basolateral tissue layer and along with many water-selective aquaporins (11), but is usually the only aquaglyceroporin present. The expression pattern maybe related to the general function of hAQP3 as a channel to transfer solutes between tissues or as a way to target specific effectors to the basolateral tissue layer.

The specific role that hAQP3 plays in individual cellular processes is not clear. AQP3 knockout mice do not have an observable phenotype except increased urination, suggesting hAQP3 is involved in urine concentration in the kidneys (1). Insight into the particular cellular processes that hAQP3 maybe involved have come from studies where AQP3 knockout mice have been observed under times of stress. Studies of AQP3 knockout mice have revealed that without AQP3 wound healing is delayed (4) and tumorigenesis is inhibited (5). The underlying process that was found to be affected was a decrease in cell migration and cell proliferation indicating that somehow AQP3 maybe involved or necessary for these processes (4,5). Questions remain as to how AQP3 is specifically involved in cell migration and cell proliferation; is AQP3 acting as a way to target other proteins or is somehow glycerol or water transport involved. From these and other studies the role of AQP3 and aquaglyceroporins are just being understood.

Although the specific cellular processes that hAQP3 participates in are not thoroughly understood, mechanistic studies of hAQP3 can shed light on how the protein maybe involved in these processes. Biochemical studies of the regulation of hAQP3 permeability have revealed that water and glycerol transport are sensitive to acidic pH (43), nickel (36) and copper (37). Mutagenesis studies have revealed that serine-152, histidine-241 and histidine-154 are involved in hAQP3 sensitivity to nickel, copper and acidic pH (36,37). Histidine-154 when mutated to alanine completely abolished water permeability indicating that it plays an important role in channel permeability and regulation. These residues are hypothesized to be located in extracellular loops close to the extracellular vestibule of the pore, but how these residues participate in regulating solute permeation through the pore is unclear.

Using electron crystallography to study hAQP3 we can investigate the structure and regulation within a lipid environment. We have shown that in the presence of detergent, hAQP3 undergoes a pH dependent oligomerization. In acidic conditions, hAQP3 exists as a tetramer and under basic conditions there is an equilibrium between the tetramer and a species consistent with associating tetramers using size exclusion chromatography and negative-stain TEM. This association of tetramers could be disrupted by the use of reducing agents suggesting that cysteine residues mediate the interaction. The biological significance of this interaction is not known but maybe important for hAQP3's role in different signaling pathways.

RESULTS AND DISCUSSION

Expression and solubilization of recombinant hAQP3

For protein structure determination studies large amounts of purified protein are required and the first step is to find a source from which to obtain the protein. *P. pastoris* is a

methyltrophic yeast that has been used for heterologous over-expression of mammalian membrane proteins with yields of >1 mg/L of culture (44,45,46). The cDNA encoding hAQP3 was purchased from OpenBiosystems and subcloned into the vector pPICZA (Invitrogen) for expression in *P. pastoris* with either an N-terminal or C-terminal 6xHis-tag. hAQP3 his-tagged constructs were integrated into the AOX1 locus in the *P. pastoris* genome using homologous recombination. Transformants with multiple copy gene integration were selected for on YPDS plates containing high concentrations of Zeocin (Invitrogen). Three colonies were selected for each of the his-tagged constructs for expression analysis. Expression of his-tagged hAQP3 constructs over the course of five days of methanol induction was analyzed using Western blot with HisProbe-HRP. Expression of the C-terminal his-tagged construct was detectable after two days of methanol induction and remained stable over the course of the expression analysis (Fig. 2.2A). The N-terminal 6xHis-tag construct of hAQP3 failed to grow after induction with methanol. Therefore the following studies proceeded with the C-terminal his-tagged construct of hAQP3 (hAQP3-6xHis).

For large-scale production of hAQP3-6xHis in *P. pastoris* typically six liters of culture were grown at 30°C and 300 rpm in baffled flasks. Protein expression was induced with 0.5% methanol and cells were harvested after 72 hours of induction. After 72 hours of induction the OD600 ranged between 16-20 and yielded approximately 25 grams of cells per liter of culture. Cells were lysed using a microfluidizer and crude membranes were pelleted using ultracentrifugation. Crude membranes were prepared at 20 mg/mL total membrane protein concentration assayed using bicinchoninic acid (BCA) protein assay (Pierce) and stored at -80°C. Crude membranes prepared from *P. pastoris* expressing hAQP3-6xHis were subjected to alkaline washes to determine whether the protein was properly folded and inserted into the membrane.

Alkaline wash treatment of crude membranes removes peripheral membrane proteins, leaving the integral membrane proteins intact (47). After washing crude membranes at 20 mg/mL with urea/sodium hydroxide, hAQP3 was detected in the membrane fraction using Western blot analysis with HisProbe-HRP (Fig. 2.2B). Treatment of crude membranes using sodium carbonate buffer (48), a milder treatment compared to sodium hydroxide, produced similar results to the urea/sodium hydroxide treatment (Fig. 2.2C). Following alkaline wash treatment with either urea/hydroxide or sodium carbonate, membranes were resuspended to 10 mg/mL total membrane protein concentration and stored at -80°C. Using two different protocols for alkaline wash resistance, hAQP3-6xHis was found to remain in the membrane fraction indicating that the protein when expressed in *P. pastoris* is integrated into the membrane.

To purify membrane proteins, the protein must first be extracted from the membrane and stabilized in solution using detergents. Solubilization of hAQP3-6xHis was tested using a variety of detergents including decyl maltopyranoside (DM), dodecyl maltopyranoside (DDM), octylglucoside (OG), phosphcholine-10 (FC-10), nonylmaltopyranoside (NM) and Triton X-100. Briefly hAQP3-6xHis crude membranes at 10 mg/mL prepared using an alkaline wash step were incubated with detergent over a range of temperatures (4°C, 25°C and 37°C) and length of times (30 min, 1 hr and 2 hrs). The membrane suspension was then centrifuged at 100,000 x g, separating the insoluble membrane fraction in the pellet and solubilized protein in the supernatant. The fractions after solubilization were analyzed using Western blot analysis with HisProbe-HRP. Initial solubilization attempts using membranes treated with urea/sodium hydroxide did not yield significant amounts of solubilized hAQP3. Solubilization with crude membranes treated with sodium carbonate yielded better results and were used for further testing of solubilization conditions. hAQP3-6xHis was solubilized in DM, DDM and NM within one

hour at 4°C, but failed to be solubilized by FC-10, OG and TritonX100. The best solubilization results were obtained using 1% DDM at 4°C for one hour with gentle stirring (Fig. 2.2D).

Purification of hAQP3 using Ni-affinity and Size-exclusion chromatography

hAQP3-6xHis was purified using a combination of nickel affinity chromatography and size exclusion chromatography (SEC) at 4°C. Briefly, hAQP3-6xHis solubilized with 1% DDM for 1 hour at 4°C with 20 mM imidazole pH 8 was mixed with Ni-NTA resin for 2 hours, then washed with 25 mM Tris-HCl pH8, 100 mM NaCl, 20 mM imidazole and 0.3% DM, followed by another wash step with the same buffer with 40 mM imidazole to remove proteins bound non-specifically to the resin. hAQP3-6xHis was eluted from the resin with 25 mM Tris-HCl pH 8, 100 mM NaCl, 0.3% DM, 300 mM imidazole and analyzed using SDS-PAGE. The detergent was exchanged from DDM to DM during the nickel affinity purification step because DM has a higher critical micelle concentration compared to DDM, which is better for two-dimensional crystallization trials. Fractions from the nickel affinity purification visualized by SDS-PAGE revealed four major bands at the apparent molecular weight of 100 kDa, 75 kDa, 50 kDa and 25 kDa, which corresponded to the approximate molecular weight of a tetramer, trimer, dimer and monomer of hAQP3-6xHis (Fig. 2.3A). The identity of the bands was confirmed using Western blot analysis with HisProbe-HRP. In the literature, aquaglyceroporins even under reducing conditions show the characteristic tetramer, trimer, dimer and monomer bands so this is not surprising that these bands were observed. The monomer ran faster than the calculated molecular weight of 31.5 kDa this is not unusual for membrane proteins, likely due to increased binding of SDS (49). From SDS-PAGE it was observed that hAQP3-6xHis was the most abundant component in the elution from the Ni-NTA resin (Fig. 2.3A).

Following elution from the Ni-NTA resin, hAQP3-6xHis was further purified using a S200 SEC column (GE Healthcare) with 10 mM Tris-HCl pH 8, 100 mM NaCl and 0.3% DM. Initially, hAQP3-6xHis eluted as two peaks at 12 mL and 14.7 mL from the gel filtration. The addition of 1 mM DTT and 10% glycerol during both the nickel affinity and size exclusion chromatography reduced hAQP3-6xHis to a single peak eluting at 14.7 mL corresponding to a molecular weight of 150 kDa (Fig. 2.3B). This matches the molecular weight of the tetramer of hAQP3 (126 kDa) plus the DM detergent micelle (MW ~ 35 kDa). Fractions from size exclusion were run on SDS-PAGE to assess the purity of hAQP3 peak fractions (Fig. 2.3B,inset). The identity of the peak fraction was confirmed using Western blot analysis using HisProbe-HRP. Protein corresponding to the peak fraction was examined using negative-stain TEM and was found to consist of homogeneous particles with a diameter of ~ 10 nm (Fig. 2.3C), which agrees with previously published data on aquaporins (50,51). Results from SDS-PAGE, western blot analysis and TEM single particle analysis indicate that hAQP3-6xHis was purified as a homogeneous tetramer. The typical yield from one liter of *P. pastoris* culture is about 1 mg of purified hAQP3 determined using BCA protein concentration assay (Pierce).

hAQP3 forms pH sensitive oligomers *in vitro*

An interesting result came when purifying hAQP3-6xHis using SEC, two peaks were observed, corresponding to ~150 kDa and ~400 kDa (Fig. 2.4A, blue). Both peaks correspond to hAQP3-6xHis confirmed using SDS-PAGE and western blot analysis with HisProbe-HRP. The two peaks from the chromatograms were then examined using negative-stain TEM. In the electron micrographs, the ~150 kDa protein was a square shaped particle (Fig. 2.4C,E), which resembles the tetramer (50,51) and the ~400 kDa protein appeared to be two 150 kDa particles

stacked together (Fig. 2.4B,D). This association could be disrupted by 1 mM DTT as indicated by the presence of only a single peak on SEC when DTT was added to the running buffer (Fig. 2.3B). In addition, 10% glycerol was also required in order to stabilize the tetrameric form of hAQP3-6xHis and prevent association of the tetramers.

To determine what type of interactions might be facilitating the association several additives to the purification buffer were screened including acidic pH, EDTA, lipids and detergent during SEC. When hAQP3-6xHis was analyzed using SEC with running buffer containing 25 mM sodium acetate pH 5.0, 150 mM NaCl and 0.3% DM, the protein eluted as a single peak at 14.5 mL approximately the size expected for the tetramer (Fig. 2.4A, red). The sample was confirmed to be homogeneous hAQP3-6xHis using SDS-PAGE and Western blot analysis. Negative-stain TEM analysis of the peak fraction revealed monodisperse particles similar to those in the tetrameric fraction of hAQP3-6xHis purified at pH 8 (Fig. 2.4C). Other conditions tested to disrupt the association include the addition of 5 mM EDTA/EGTA, 1 M imidazole pH 8, 0.01% DMPC and exchanging the protein into other detergents, but none of these conditions disrupted the interaction as observed using SEC. Disruption of the associating tetramers by 1 mM DTT or acidic pH suggest that the residues involved in the interaction maybe histidine or cysteine.

To assess the possibility that the interaction is facilitated by histidines or cysteines, a homology model of human AQP3 was calculated using the program ModWeb (52) based on the structure of GlpF (6), an *E. coli* glycerol facilitator. Careful analysis of the sequence and homology model of hAQP3 revealed that there are six histidines and seven cysteines (Fig. 2.5). In addition to the six native histidines there are six histidines on the C-terminus of hAQP3 that comprise the his-tag used for Ni-NTA affinity purification. The acidic proton of the imidazole

ring of histidine has a pKa around 6.0, which correlates to the pH dependent shift of oligomerization of hAQP3-6xHis observed by SEC. Histidine when de-protonated can undergo weak pi-stacking interactions that would occur when the pH is above the pKa of the imino proton. The pi-stacking interaction would not exist when histidine is protonated because the side chain becomes positively charged. This is consistent with the observed result that at pH 5 hAQP3-6xHis runs as a tetramer on SEC but at pH 8 there is a second peak that corresponds to the association of the tetramers (Fig. 2.4A). Although this provides a reasonable explanation of the SEC analysis, this is not consistent with results that show the addition of DTT results in a single peak. In addition, imidazole can act as a competitor for the histidine pi-stacking interaction thereby disrupting the intermolecular interaction but two peaks were observed during SEC with 1 M imidazole pH 8. Thus pH-dependent histidine pi-stacking is not consistent with the biochemical data as the type of interactions that might be facilitating the association of hAQP3-6xHis tetramers.

Another amino acid that has a pKa near physiological pH is cysteine. Cysteine is not an obvious choice to explain the association of hAQP3-6xHis tetramers because in solution the sulfhydryl proton has a pKa of 8.5. But in proteins the pKa of a particular side chain is determined by neighboring amino acids and potential hydrogen bonding interactions could change the pKa of a hydrolyzable proton. Human creatine kinase is an example in which cysteine-282 in the active site has a pKa that is shifted to 5.6 due to hydrogen bonding interactions with serine-284 and an amide NH-group (53). Cysteine maybe facilitating the association of the tetramers through disulfide bond formation. Disulfide bond formation susceptibility is influenced by pH because the thiolate form is more reactive than the thiol. Thus the observation of two populations of hAQP3-6xHis at pH 8 maybe a result of cysteine and the

formation of a disulfide bond between tetramers. In fact this hypothesis is consistent with the result that when DTT is present during SEC a single peak is observed corresponding to the tetramer.

Examination of the homology model for the structure of hAQP3 reveals that there are seven cysteines and of the seven cysteines only three are predicted to be solvent exposed (Fig. 2.5). The other cysteine residues are buried in the interior of the protein or exist in regions of transmembrane helices that would be protected by detergent. The three cysteines are located on the intracellular side of hAQP3 in the homology model: cysteine-11, cysteine-91 and cysteine-267. According to the homology model and sequence analysis there are neighboring residues that could significantly affect the pKa of cysteine, such as serine and protonated histidines that could participate in hydrogen bonding with cysteine. For cysteine-11, located in the N-terminus of hAQP3 there is a serine residue at position 9 and for cysteine-267 there is a neighboring histidine residue. Cystein-91 does not have a histidine or serine residue in close proximity in the homology model. Cysteines 11 or 267 could be candidates for intermolecular disulfide bond formation to explain the association of tetramers of hAQP3-6xHis by SEC.

The next question is does the interaction of the hAQP3-6xHis tetramers relate to the biological role that hAQP3 could have in the cell. The cysteine residues are located on the intracellular surface of hAQP3, which excludes the possibility for formation of junctions between cells. It is possible that the cysteine residues are involved in regulation of the channel as an internal sensor of pH but this has not been demonstrated. Thus, it is unlikely that this interaction is biologically significant although the importance for the cysteines for water and glycerol permeability has not been determined or the possibility that the residues are mediating protein-protein interactions important in signaling for cellular processes, such as cell migration.

To answer questions regarding how the cysteines may contribute to the structure or function of hAQP3, we need to determine the structure of hAQP3.

Two-dimensional crystallization trials of hAQP3-6xHis

Electron crystallography requires two-dimensional crystals that are comprised of ordered arrays of membrane protein and lipids. To form two-dimensional crystals purified membrane protein solubilized with detergent is combined with lipids and through dialysis crystals in the form of vesicles, sheets or tubes can be formed. Conditions conducive for reconstitution of the membrane protein into vesicles and eventual lattice formation must be screened. Two-dimensional crystallization conditions for hAQP3-6xHis were screened using protein purified from SEC with 10 mM Tris-HCl pH8, 150 mM NaCl, 0.3% DM, 1 mM DTT, 10% glycerol and concentrated to 1 mg/mL. hAQP3-6xHis at 1 mg/mL was incubated with lipids at a range of lipid to protein ratios (LPR) for 30 minutes and assembled into dialysis buttons that were allowed to dialyze for up to four weeks. Two-dimensional crystallization conditions screened include LPR, magnesium chloride concentration, temperature, buffer pH, dialysis rate, different lipids and detergents (Table 2.1).

After two to four weeks of dialysis, dialysis buttons were opened, the sample was collected and stored at 4°C. Reconstitution experiments were screened using negative-stain TEM analysis and observation of the transparency of the button before removing sample was noted. Successful reconstitution experiments in dialysis buttons result in cloudiness throughout the solution and indications of the presence of precipitant on the membrane surface can suggest that the protein is unstable during the reconstitution process. After extensively screening dialysis conditions as well as different lipids and detergents, reconstitution of hAQP3-6xHis to form

crystalline vesicles was unsuccessful. Typical results from selected conditions are depicted in Figure 2.6. Most conditions resulted in protein aggregation (Fig. 2.6A), lipid-protein aggregation (Fig. 2.6B), lipid vesicles or small vesicles containing protein without any order (Fig. 2.6C). The dialysis buttons during the reconstitution experiments had a significant amount of white precipitant on the dialysis membrane indicating that the protein was not stable throughout the reconstitution experiment. Instability of hAQP3-6xHis could be due to a number of factors including the intact his-tag. His-tagged proteins have been shown to be prone to aggregation (50) so removal of the tag may help stability of hAQP3-6xHis. To improve the stability of hAQP3 for reconstitution experiments new constructs must be sub-cloned.

CONCLUDING REMARKS

The vast majority of membrane proteins whose structure has been determined come from bacteria. Currently there are only ten structures of mammalian membrane proteins listed in the Protein Data Bank out of ~62,000 other structures (soluble & membrane) that have been determined. Determining the structure of human AQP3 would be the first mammalian aquaglyceroporin structure and would provide insight into the mechanism by which water and glycerol are able to diffuse through the pores of mammalian aquaglyceroporins. In the work presented here we show that hAQP3-6xHis can be heterologously over-expressed in *P. pastoris* and purified yielding 1 mg per liter of culture. An observation during SEC purification showed that cysteine disulfide bond formation gave rise to dimers of tetramers confirmed by negative-stain TEM analysis. Reconstitution of hAQP3-6xHis to form two-dimensional crystals was unsuccessful after exhaustive screening possibly due to protein aggregation facilitated by the his-tagged construct. Efforts to remove the his-tag of purified hAQP3-6xHis using limited

proteolysis with trypsin resulted in significant digestion of the protein. In order to continue studies in pursuit of a structure of a mammalian aquaglyceroporin more constructs need to subcloned and tested for stability during reconstitution experiments.

Investigating the molecular mechanism that aquaglyceroporins use to transport both water and glycerol is a first step to try to understand these special aquaporins. From studies of the structure of GlpF from *E. coli* and PfAQP from *P. falciparum* we have learned how aquaglyceroporins are able to accommodate larger and more hydrophobic substrates. By studying the structure of hAQP3 we hope to not only understand the diffusion mechanism of substrates through a mammalian aquaglyceroporin, but to provide a model to understand the regulation of hAQP3 by pH and cationic metals. After determining the structure of hAQP3, the next experiments would focus on mutagenesis studies of residues lining the pore to determine which residues contribute the most to water and glycerol transport, in addition to the pH, nickel and copper sensitivities that have been observed. An interesting experiment would be to investigate the effect of pH on wound healing and skin hydration specifically looking at AQP3 function. Mutagenesis studies could be conducted to alter AQP3 pH insensitive and then see if wound healing or skin hydration is effected in the same way. Future studies should be focused on connecting molecular mechanism of regulation to the physiological processes that hAQP3 influences and controls.

The importance of aquaglyceroporins and in particular AQP3 in physiological processes is only emerging with targeted knockout studies. Currently, there are no diseases associated with AQP3, but an emerging role in tumorigenesis is being investigated. Studies by Alan Verkman's group have found that in AQP3 knockout mice tumorigenesis is inhibited, suggesting that AQP3 maybe a potential cancer therapy target (4). Investigation into how AQP3 promotes

tumorigenesis may provide information on its role in related processes, such as cell migration & wound healing. Connecting the information regarding the mechanistic studies of aquaglyceroporins to the physiological processes in which they are involved in is essential to understanding the importance that this small class of membrane proteins has in cell biology.

MATERIALS AND METHODS

Expression of hAQP3-6xHis in *Pichia pastoris*

The cDNA encoding the human AQP3 gene was purchased from OpenBiosystems and subcloned into the vector pPICZA (Invitrogen) using restriction sites EcoRI and SalI. Two hAQP3 constructs were subcloned to be expressed as fusion with a C-terminal or N-terminal 6xHis-tag in *Pichia pastoris* under the control of a methanol inducible promoter. Each construct was transformed into *P. pastoris* and integrated into the genome using homologous recombination into the AOX1 loci. Successful genome integration was indicated by growth of colonies on YPDS plates containing between 100 to 1000 μg/mL of the antibiotic, Zeocin (Invitrogen). The C-terminal his-tagged hAQP3 strain was grown in BMGY (buffered glycerol-complex media) till an OD 600 of 0.6 was reached then expression was induced by transferring cells to BMMY (buffered methanol-complex media) containing 0.5% methanol. Cultures were grown at 30°C, 300 rpm and 0.5% methanol was added every 24 hours. Cells were harvested after 72 hours of induction with a final OD 600 of 16, pelleted at 3000xg for 10 minutes, washed with breaking buffer (50mM sodium phosphate pH 7.4, 1 mM EDTA, 5% glycerol) and resuspended in breaking buffer plus protease inhibitors, 1 mM PMSF and 0.3 μg/mL of pepstatin. Cells were lysed after 3 passes using a microfluidizer (Microfluidics) at 25,000 psi. Cell lysate was centrifuged at 3000xg followed by 10,000xg to remove cell debris. Crude

membranes were pelleted at 130,000xg, washed with membrane preparation buffer (25 mM Tris-HCl pH8, 150 mM NaCl, 10% glycerol), resuspended to a total membrane protein concentration of 20 mg/mL and stored at -80°C.

Membranes were then washed with 4M urea followed by 20 mM sodium hydroxide or with 20 mM sodium carbonate pH ~11. The membranes were then washed with membrane preparation buffer to neutralize the pH of the membranes. Membranes were then frozen at -80°C at 10 mg/mL or purification steps immediately proceeded.

Solubilization & Purification of hAQP3-6xHis

Solubilization assays of hAQP3-6xHis were conducted by adding an equal of volume of detergent at ten times the critical micelle concentration to membranes at 10 mg/ml and incubated at 4°C, 25°C or 37°C for a length of time ranging from 30 minutes to 2 hours. The mixture was then centrifuged at 130,000 xg. The resulting pellet and supernatant were analyzed using Western blot probed with HisProbe-HRP (Pierce).

hAQP3-6xHis was purified using nickel affinity chromatography and size exclusion chromatography. Membranes solubilized with 1% DDM at 4°C for 1 hour were incubated with 2 mL of Ni-NTA resin (Qiagen) in the presence of 20 mM imidazole pH 8 for 2 hours at 4°C with gently stirring. The resin was then washed with 10 column volumes of 25 mM Tris-HCl pH 8, 150 mM NaCl, 20 mM imidazole, 0.3% DM, 0.5 mM DTT, 10% glycerol and 10 column volumes of 25 mM Tris-HCl pH 8, 150 mM NaCl, 40 mM imidazole, 0.3% DM, 0.5 mM DTT, 10% glycerol. hAQP3-6xHis was eluted from the Ni-NTA resin with five column volumes of 25 mM Tris-HCl pH 8, 150 mM NaCl, 300 mM imidazole, 0.3% DM, 0.5 mM DTT and 10% glycerol. The elution was then loaded onto a S200 size exclusion column equilibrated with 25

mM Tris-HCl pH 8, 150 mM NaCl, 0.3% DM, 0.5 mM DTT, 10% glycerol and ran at 0.5 mL/min. For purification of hAQP3-6xHis at pH5, 500 mM sodium acetate buffer pH 5 was added to the Ni-NTA elution to get a final concentration of 50 mM sodium acetate and the pH of the solution was adjusted to pH 5 using 12 M hydrochloric acid. The pH adjusted Ni-NTA elution was then ran over SEC with the following buffer: 25 mM sodium acetate pH 5, 150 mM NaCl, 1 mM EDTA, 0.3% DM, at 0.5 ml/min. Fractions pertaining to hAQP3-6xHis were concentrated using a Vivaspin6 (GE Healthcare) 50 kDa molecular weight cut-off centrifugal concentrator to 1 mg/mL. Protein concentrations were assayed using BCA protein assay (Pierce).

Two-dimensional Crystallization trials of hAQP3-6xHis

Purified hAQP3-6xHis at 1 mg/mL was used to setup two-dimensional crystallization trials. Lipids used for two-dimensional crystallization screens were prepared at a concentration of 10 mg/mL in 1% DM and 2% OG. Purified hAQP3-6xHis, lipids and buffer were combined and incubated at room temperature for 30 minutes prior to assembling buttons. 50 μL dialysis buttons (Hampton Research) were used for two-dimensional crystallization trials with 12-14 kDa molecular weight cut-off dialysis membranes. Assembled dialysis buttons were incubated in 100 mL of dialysis buffer for two to four weeks at room temperature, changing the buffer everyday. After four weeks the dialysis buttons were opened and screened using negative stain electron microscopy.

Negative-stain TEM

Briefly, 2 μL of sample was applied to freshly glow discharged carbon coated 400 mesh copper grids, blotted with filter, washed with water and stained with 0.75% uranyl formate. The

grids were then viewed using a 100 kV transmission electron microscope (Morgagni M268, FEI). Images were collected using a 4k x 2k Gatan charge-coupled device (CCD) over a wide range of magnifications. Single particle images of hAQP3-6xHis purified at pH 8 with 1 mM DTT and 10% glycerol were imaged using a Tecnai Spirit operating at 120 kV and a 4k x 4k Tietz CMOS detector.

ACKNOWLEDGEMENTS

B.S.V. was supported in part by Public Health Service, National Research Service, T32 GM07270, from the National Institute of General Medical Sciences.

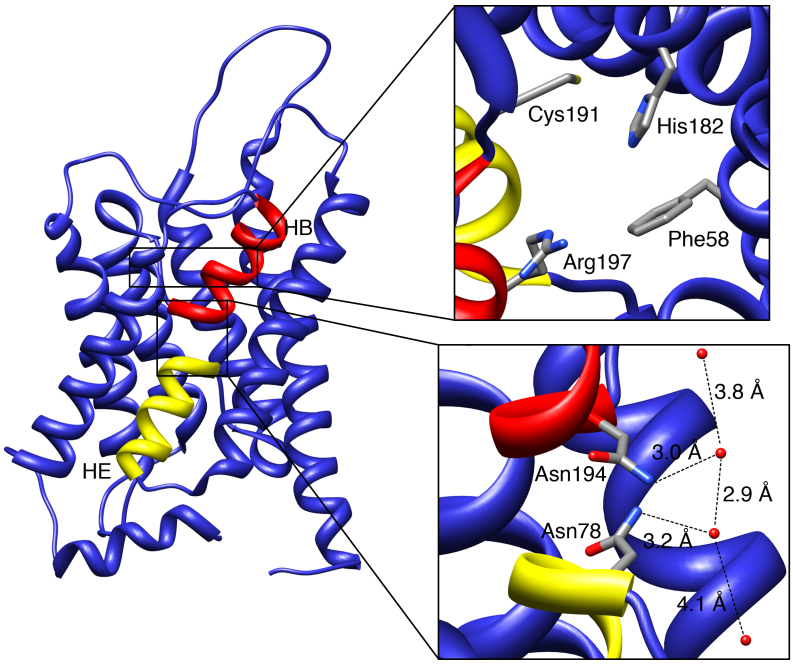

Figure 2.1 The NPA motif and ar/R constriction site of aquaporins. The monomeric structure of AQP1 (PDB 1J4N) (16) is depicted highlighting the two half helices, HB and HE, that contain the NPA motifs responsible for water conductance and inhibit proton diffusion through the channel. The ar/R constriction site of AQP1 is formed by cysteine-191, histidine-182, arginine-197 and phenylalanine-58 and is responsible for pore selectivity (right, top). Two asparagines from the two NPA motifs orient water molecules in the pore such that hydrogen bonding is prevented between water molecules, inhibiting proton diffusion through the pore (right, bottom). In the structure of bovine AQP1 (16), 4 water molecules were observed in the crystal structure, two are participating in hydrogen bonds with asparagine-194 and asparagine-78, while the other two, positioned above and below the central waters, are not positioned close enough to be hydrogen bonded.

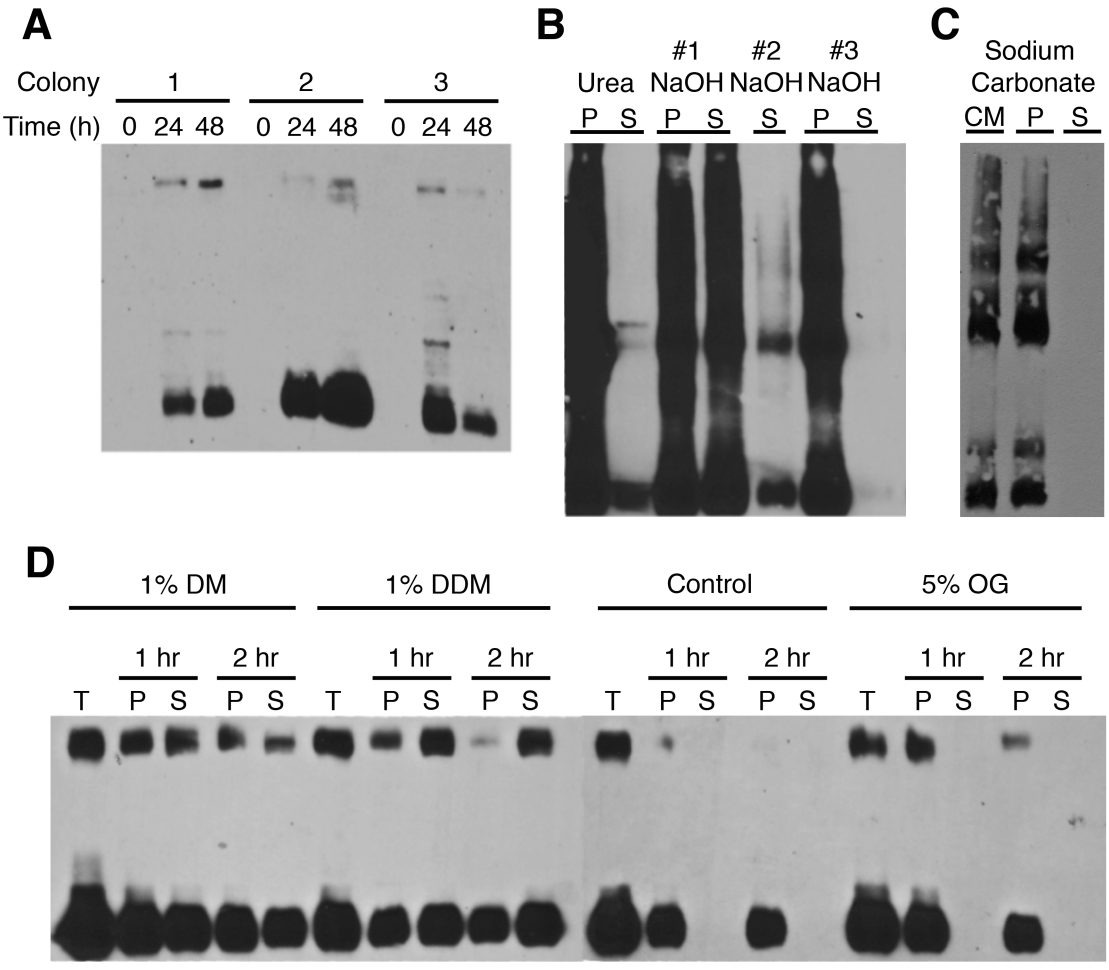

Figure 2.2 Expression, alkali membrane treatment and solubilization assay of hAQP3. (A) Western blot analysis of C-terminal his-tagged hAQP3 expression in *P. pastoris* monitored over the course of two days of overexpression induction with 0.5% methanol. Expression was monitored for a total of five days and remained constant after the first 48 hours. (**B**) Alkali treatment of hAQP3-6xHis crude membranes with 4 M urea followed by three successive washes with 20 mM sodium hydroxide. Western blot analysis of both the pellet and supernatant after each wash is shown. hAQP3-6xHis remains in the membrane fraction (pellet) after alkali treatment indicating that hAQP3-6xHis is integrated into the membrane. (**C**) Alkali treatment of hAQP3-6xHis membranes with 20 mM sodium carbonate, producing similar results to the use of urea/sodium hydroxide. (**D**) Solubilization assay of sodium carbonate treated membranes of hAQP3-6xHis carried out at 4°C and monitored using western blot analysis. All western blots were probed against the C-terminal his-tag of hAQP3 using HisProbe-HRP (Pierce). CM = crude membrane, P = pellet, S = supernatant.

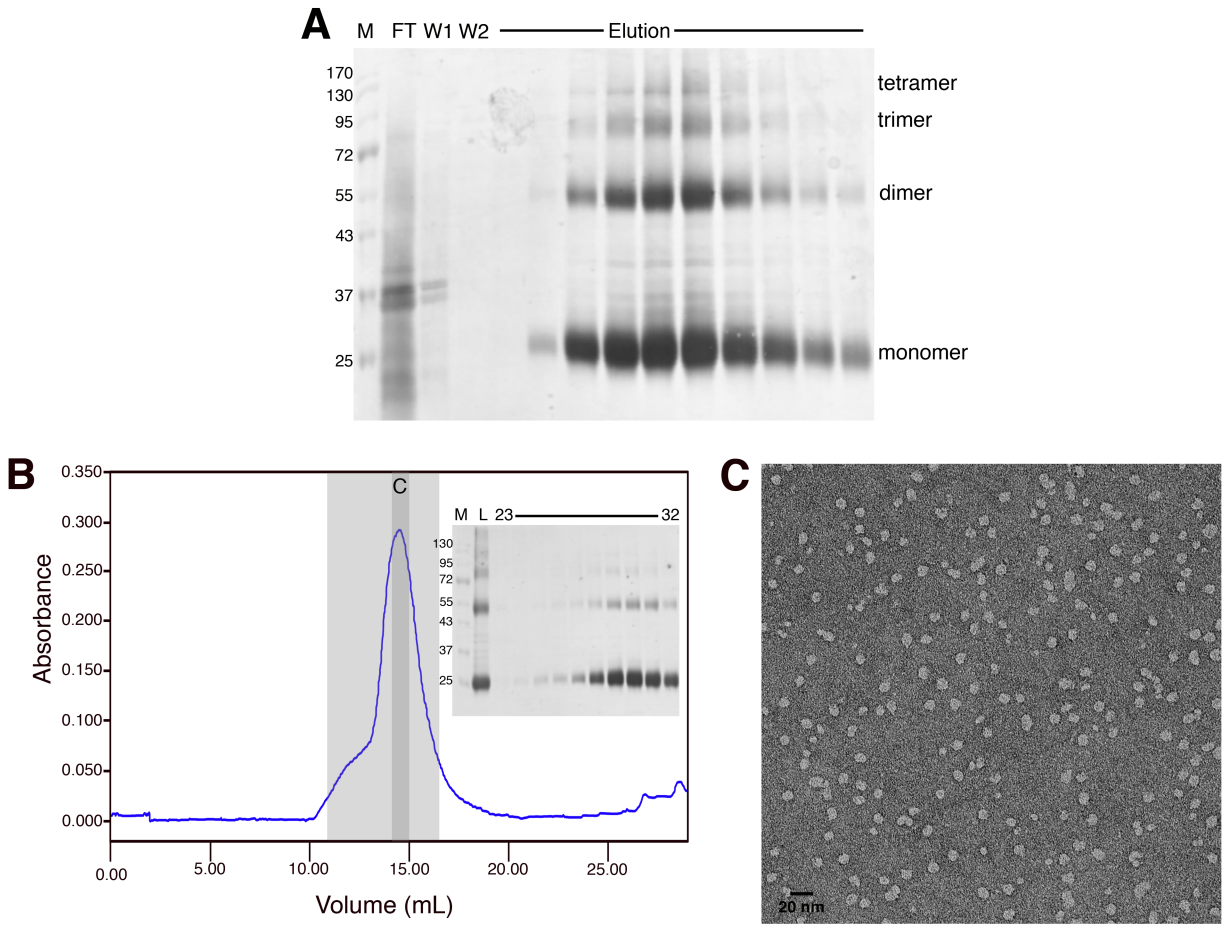

Figure 2.3 Nickel-NTA affinity and Size exclusion purification of hAQP3-6xHis. (**A**) hAQP3-6xHis was purified first using nickel affinity chromatography from membranes solubilized with 1% DDM. After two hours of incubation with the resin, the resin was washed with Ni Binding Buffer pH8 (W1), followed by Ni Wash Buffer (W2) and finally eluted with 300 mM imidazole Elution Buffer. (**B**) Size exclusion chromatography was used as the second step in purification of hAQP3-6xHis using a S200 SEC column. hAQP3-6xHis eluted at 14.7 mL as a single peak in the presence of 1 mM DTT and 10% glycerol. Inset, fractions from SEC were run on SDS-PAGE to observe the purity of the fractions; fractions on gel correspond to the light gray box in the SEC chromatogram. (**C**) The peak fraction (dark grey box in B) from SEC was then applied to a freshly glow dis-charged grid, stained with uranyl formate and observed using TEM. Particles were homogeneous with a diameter of approximately 10 nm.

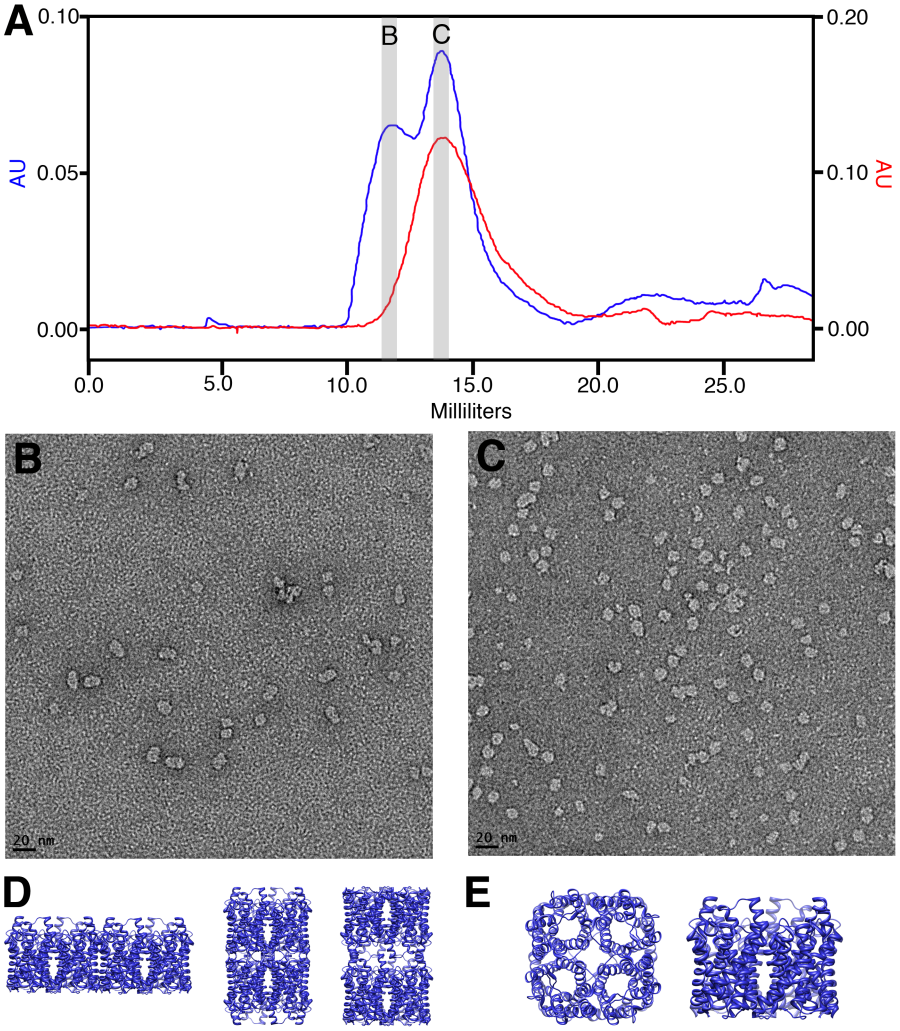

Figure 2.4 pH dependent oligomerization of hAQP3-6xHis. hAQP3-6xHis purified using size exclusion chromatography at pH 5 and pH 8 show a shift in oligomerization state (**A**). At pH 5 (A, red) hAQP3-6xHis elutes as a single peak that when observed using negative-stain TEM adopts a tetrameric state with particles about 10 nm in diameter as shown in part C. At pH 8 (A, blue) hAQP3-6xHis elutes as two peaks corresponding to 150 kDa and 400 kDa that when each is observed using negative-stain TEM the larger complex resembles two associating tetramers (**B**) while the smaller peak fraction resembles the tetramer (**C**). The 400 kDa peak could correspond to the association of two tetramers; possible binding arrangements are shown (**D**), left is interaction of the transmembrane regions, middle is interaction of the intracellular domains and right is interaction of extracellular domains (PDB 1FX8) (6). (**E**) Representation of possible views of the tetramer seen in the negative TEM image in **C** (PDB 1FX8) (6).

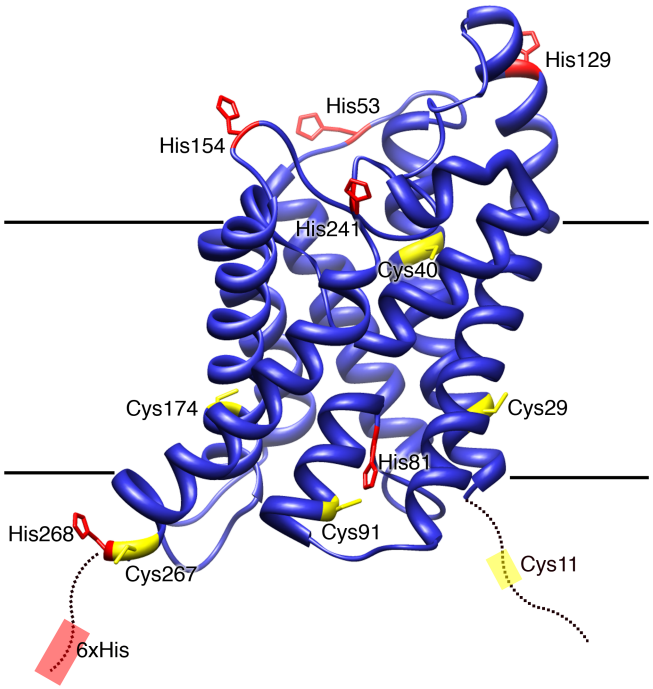

Figure 2.5 Human AQP3 homology model showing cysteines and histidines. Homology model of hAQP3-6xHis calculated using ModWeb (52) based on the structure of GlpF (6). The homology model shows the positions of cysteines and histidines that maybe involved in the pH dependent oligomerization observed using size exclusion chromatography and negative-stain TEM.

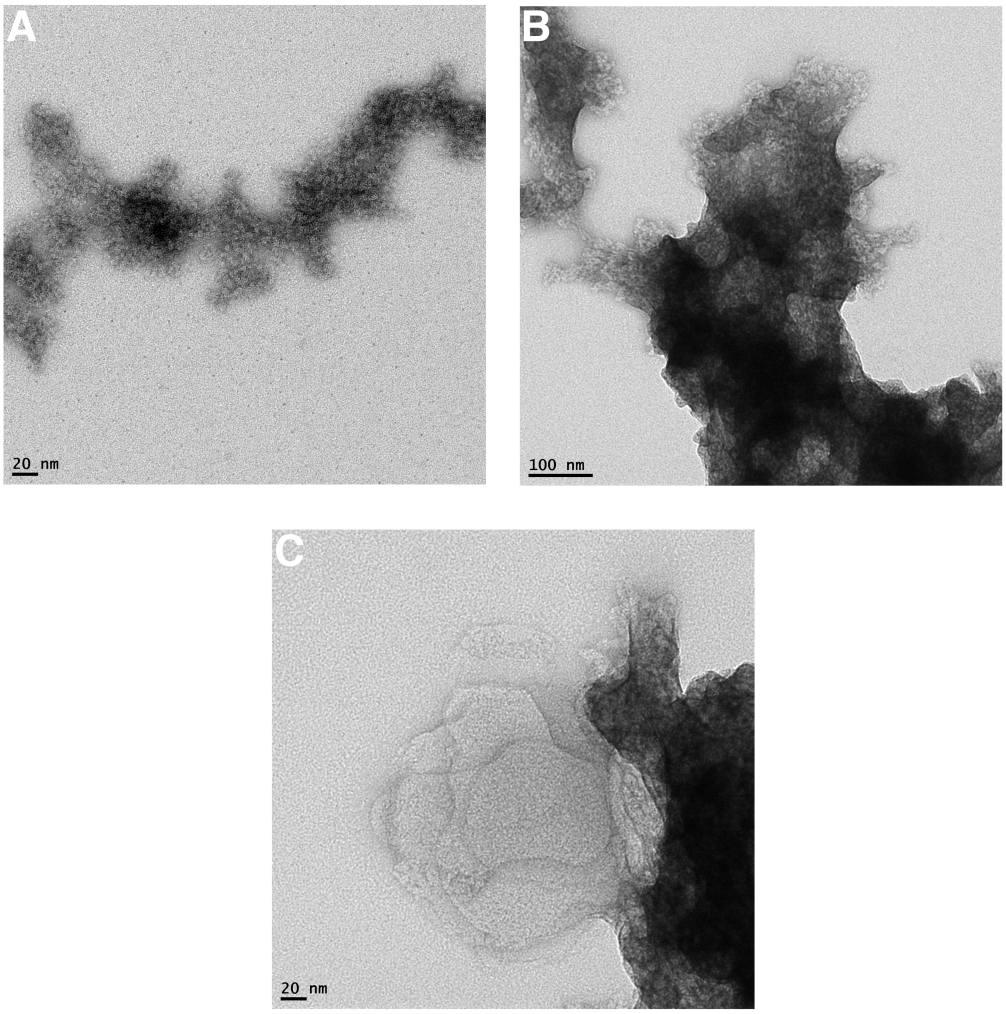

Figure 2.6 Typical results from two-dimensional crystallization trials of hAQP3-6xHis. Two-dimensional crystallization trials of hAQP3-6xHis typically yielded protein aggregation (**A**), protein-lipid aggregation (**B**) and reconstituted protein into lipid vesicles that did not have observable lattice formation (**C**).

Table 2.1 Two-dimensional crystallization conditions screened for hAQP3-6xHis

Lipids	LPR	pH	Mg_2Cl	Temperature	Glycerol
DMPC	0.25-1.5	4	50 mM	20 ˚C	10%
E. coli Polar Lipid Extract	1.0-10.0	5	200 mM	4 ˚C	20%
POPC		6		20 ˚C - 37 ˚C - 20 ˚C	
POPE		7			
DOPC		8			
DOPE					

WORKS CITED IN CHAPTER 2

1. Zhao, D., Bankir, L., Qian, L., Yang, D. & Yang, B. (2006). Urea and urine concentrating ability in mice lacking AQP1 and AQP3. *Am. J. Physiol. Renal Physiol.* **291**, F429-F438.

2. Roudier, N., Ripoche, P., Gane, P., Le Pennec, P., Daniels, G., Cartron, J-P. & Bailly, P. (2002). AQP3 Deficiency in human and the molecular basis of a novel blood group system, GIL. *J. Biol. Chem.* **277**, 45854-45859.

3. Hara-Chikuma, M. & Verkman, A.S. (2008). Aquaporin-3 facilitates epidermal cell migration and proliferation during wound healing. *J. Mol. Med.* **86**, 221-31.

4. Hara-Chikuma, M. & Verkman, A.S. (2003). Glycerol replacement corrects defective hydration, elasticity, and barrier function in aquaporin-3 deficient mice. *Proc. Natl. Acad. Sci. USA*, **86**, 221-31.

5. Hara-Chikuma, M. & Verkman, A. (2008). Prevention of skin tumorigenesis and impairment of epidermal cell proliferation by targeted aquaporin-3 gene disruption. *Mol. Cell Biol.* **28**, 326-332.

6. Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J. & Stroud, R.M. (2000). Structure of a glycerol-conducting channel and the basis for its selectivity. *Science*, **290**, 481-486.

7. Newby, Z.E.R, O'Connell, J., Robles-Colmenares, Y., Khademi, S. Miercke, L.J. & Stroud, R.M. (2008). Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite *Plasmodium falciparum. Nat. Struct. Mol. Biol.* **15**, 619-625.

8. Agre, P., Saboori, A.M., Asimos, A. & Smith, B.L. (1987). Purification and partial characterization of the Mr 30,000 integral membrane protein associated with the erythrocyte Rh(D) antigen. *J. Biol. Chem.* **262**, 17497-17503.

9. Preston, G.M., Carroll, T.P. Guggino, W.B. & Agre, P. (1992). Appearance of water channel in Xenopus oocytes expressing red cell CHIP28 protein. *Science*, **256**, 385-387.

10. King, L.S., Kozono, D. & Agre, P. (2004). From structure to disease: The evolving tale of aquaporin biology. *Nat. Rev. Mol. Cell Biol*. **5**, 687-697.

11. Rojek, A., Praetorius, J., Frokiaer, J., Nielsen, S. & Fenton, R. (2008). A current view of the mammalian aquaglyceroporins. *Annu. Rev. Physiol.* **70**, 301-327.

12. Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J.B., Engel, A., Fujiyoshi, Y. (2000). Structural determinants of water permeation through aquaporin-1. *Nature*, **407**, 599-605.

13. Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, F. Harrison, S.C. & Walz, T. (2005). Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature*, **438**, 633-638.

14. Harries, W.E.C., Akhavan, D., Miercke, L.J., Khademi, S., Stroud, R.M. (2004). The channel architecture of aquaporin-0 at 2.2 Å resolution. *Proc. Natl. Acad. Sci. USA*, **101**, 14045-14050.

15. Hite, R.K., Li, Z & Walz, T. (2010). Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2D crystals. *EMBO J.* **29**, 1652-1658.

16. Sui, H., Han, B.G., Lee, J.K., Walian, P. & Jap, B.K. (2001). Structural basis of water specific transport through the AQP1 water channel. *Nature*, **414**, 872-878.

17. Hiroaki, Y. *et al.* (2005). Implications of the Aquaporin-4 Structure on Array Formation and Cell Adhesion. *J.Mol.Biol.* **355**, 628-639.

18. Ho, J.D. *et al.* (2009). Crystal structure of human aquaporin 4 at 1.8 Å and its mechanism of conductance. *Proc. Natl. Acad. Sci. USA*, **106**, 7437-7442.

19. Horsefield, R. *et al*. (2008). High resolution X-ray structure of human aquaporin 5. *Proc. Natl. Acad. Sci. USA*, **105**, 13327-13332.

20. Tornroth-Horsefield, S. *et al*. (2006). Structural mechanism of plant aquaporin gating. *Nature*, **439**, 688-694.

21. Fisher, G. *et al.* (2009). Crystal structure of a yeast aquaporin at 1.15 angstrom reveals a novel gating mechanism. *PLoS Biol.* **7**, e1000130.

22. Savage, D.F., Egea, P.F., Robles-Colmenares, Y., O'Connell III, J.D. & Stroud, R.M. (2003). Architecture and Selectivity in Aquaporins: 2.5 Å X-ray Structure of Aquaporin Z. *PloS Biology*, **1**, 334-340.

23. Lee, J.K. (2005). Structural basis for conductance by the archael aquaporin AqpM at 1.68 Å. *Proc. Natl. Acad. Sci. USA*, **102**, 18932-18937.

24. Gonen, T. & Walz, T. (2006). The structure of aquaporins. *Quart. Rev. Biophys.* **39**, 361-396.

25. Groot, B.L. & Grubmeller, H. (2005). The dynamics and energetics of water permeation and proton exclusion in aquaporins. *Curr. Opin. Struct. Biol.* **15**, 176-183.

26. Wu, B., Steinbronn, C., Alsterfjord, M., Zeuthen, T. & Beitz, E. (2009). Concerted action of two cation filters in the aquaporin water channel. *EMBO J.* **28**, 2188-2194.

27. Preston, G.M., Jung, J.S., Guggino, W.B. & Agre, P. (1993). The mercury sensitive residue at cysteine 189 in the CHIP28 water channel. *J. Biol. Chem.* **268**, 17-20.

28. Zhang, R. van Hoek, A.N., Biwersi, J & Verkman, A.S. (1993). A point mutation at cysteine 189 blocks the water permeability of rat kidney water channel CHIP28k. *Biochemistry*, **32**, 2938-2941.

29. Johansson, I., Karlsson, M., Shukia, V.K., Chrispeels, M.J., Larsson, C. & Kjeilbom, P. (1998). Water transport activity of the plasma membrane aquaporin PM28A is regulated by phosphorylation. *Plant Cell*, **10**, 451-459.

30. Reichow, S.L. & Gonen, T. (2008). Noncanonical binding of calmodulin to aquaporin-0: Implications for channel regulation. *Structure*, **16**, 1389-1398.

31. Gold, M.G. *et al*. (2012). AKAP2 anchors PKA with aquaporin-0 to support ocular lens transparency. *EMBO Mol. Med.* **4**, 15-26.

32. Nemeth-Cahalan, K.L., Kalman, K. & Hall, J.E. (2004). Molecular basis of pH and Ca2+ regulation of aquaporin water permeability. *J. Gen. Physiol.* **123**, 573-580.

33. Varadaraj, K., Sumari, S., Shiels, A & Mathias, R.T. (2005). Regulation of water permeability in the lens. *Invest. Ophthalmol. Vis. Sci.* **46**, 1393-1402.

34. Ball, L.E., Garland, D.L., Crouch, R.K. & Schey, K.L. (2004). Post-translational modifications of aquaporin-0 (AQP0) in the normal human lens: spatial and temporal occurrence. *Biochemistry*, **43**, 9856-9865.

35. Gonen, T., Cheng, Y., Kistler, J. & Walz, T. (2004). Aquaporin-0 membrane junctions form upon proteolytic cleavage. *J. Mol. Biol.* **342**, 1337-1345.

36. Zelenina, M., Bondar, A.A., Zelenin, S. & Aperia, A. (2003). Nickel and extracellular acidification inhibit the water permeability of human aquaporin-3 in lung epithelial cells. *J. Biol. Chem*. **278**, 30037-30042.

37. Zelenina, M., Tritto, S., Bondar, A.A., Zelenin, S. & Aperia, A. (2004). Copper inhibits the water and glycerol permeability of aquaporin-3. *J. Biol. Chem*. **279**, 51939-51943.

38. Christensen, B.M., Zelenina, M., Aperia, A. & Nielsen, S. (2000). Localization and regulation of PKA-phosphorylated AQP2 in response to V_2-receptor agonist/antagonist treatment. *Am. J. Physiol. Renal Physiol.* **278**, F29-F42.

39. Yasui, H., Kubota, M., Iguchi, K., Usui, S., Kiho, T. & Hirano, K. (2008). Membrane trafficking of aquaporin 3 induced by epinephrine. *Biochem. Biophys. Res. Comm.* **373**, 613-617.

40. Higuchi, S., Kubota, K., Iguchi, S., Usui, T., Kiho, K. & Hirano, K. (2007). Transcriptional regulation of aquaporin 3 by insulin. *J. Cell. Biochem.* **102**, 1051-1058.

41. Cao, C. *et al*. (2006). EGFR-mediated expression of aquaporin-3 is involved in human skin fibroblast migration. *Biochem. J.* **400**, 225-234.

42. Zheng, X. & Chen, X. (2001). Aquaporin 3, a glycerol and water transporter, is regulated by p73 of the p53 family. *FEBS Lett.* **489**, 4-7.

43. Zeuthen, T. & Klaerke, D.A. (1999). Transport of water and glycerol in aquaporin 3 is gated by H+. *J. Biol. Chem.* **274**, 21631-21636.

44. Chloupkov, M. *et al.* (2007). Expression of 25 human ABC transporters in the yeast *Pichia pastoris* and characterization of purified ABCC3 ATPase activity. *Biochemistry*, **46**, 7992-8003.

45. Wetterholm, A., Molina, D.M., Nordlund, P., Eshaghib, S. & Haeggström, J.Z. (2008). High-level expression, purification, and crystallization of recombinant rat leukotriene C4 synthase from the yeast Pichia pastoris. *Protein Expr. & Purif*. **60**, 1-6.

46. Nico-Farber, K., Harder, W., Ab, G., and Veenhuis, M. (1995). Methylotrophic Yeasts as Factories for the Production of Foreign Proteins. *Yeast*, **11**, 1331-1344.

47. Fotiadis, D. *et al.* (2001) Structural Characterization of Two Aquaporins Isolated from Native Spinach Leaf Plasma Membranes. *J. Biol. Chem.* **276**, 1707-1714.

48. Fujiki, Y., Hubbard, A.L., Fowler, S. & Lazarow, P.B. (1982). Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. *J. Cell Biol.* **93**, 97-102.

49. Rath, A., Glibowicka, M., Nadeau, V.G., Chen, G. & Deber, C.M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Natl. Acad. Sci. USA,* **106**, 1760-1765.

50. Viadiu, H., Gonen, T. & Walz, T. (2007) Projection of aquaporin-9 at 7Å resolution. *J. Mol. Biol.* **367**, 80-88.

51. Walz, T., Smith, B.L., Agre, P. & Engel, A. (1994) The three-dimensional structure of human erythrocyte aquaporin CHIP. *The EMBO J.* **13**, 2985-2993.

52. MODWEB server, http://salilab.org/modweb

53. Naor, M.M. & Jensen, J.H. (2004). Determinants of cysteine pKa values in creatine kinase and α1-antitrypsin. *Proteins*, **57**, 799-803.

Chapter III

Two-dimensional crystallization of galactose permease, a prokaryotic sugar transporter

SUMMARY

Glucose homeostasis and sugar uptake in the human body is facilitated by a family of membrane proteins called sugar transporters. Efforts to determine the structure of members of the facilitative glucose transporter family have not been successful due to the inability to recombinantly express, purify and crystallize the proteins. To gain insight into the mechanism used by this family of sugar transporters, homologous prokaryotic sugar transporters are studied due to the increased stability in solution and large quantities that can be obtained using simple expression systems such as *Escherichia coli*. Here we present two-dimensional crystallization of galactose permease (GalP) from *E. coli* and a electron cryomicroscopy (cryo-EM) projection map at 18 Å showing that GalP forms trimers when embedded in a phospholipid bilayer.

INTRODUCTION

Glucose is essential for the production of ATP via aerobic respiration, anaerobic respiration and fermentation. It is also an important metabolite used for the synthesis of wide variety of molecules including triglycerides, glycogen, protein modifications, amino acids and vitamin C. Glucose is ingested and taken up by cells lining the intestine via transporters that couple the movement of sodium down the electrochemical gradient, maintained by ion pumps in the cell, to the transport of glucose against its concentration gradient. Once glucose enters cells lining the small intestine it is then transported into the blood stream via proteins that facilitate the

diffusion of glucose down its concentration gradient. Both of these processes are mediated by a group of membrane proteins called sugar transporters.

Sugar transporters are members of the major facilitator superfamily consisting of membrane proteins that transport a wide variety of substrates including amino acids, sugars, neurotransmitters, ions and peptides. Sugar transporters are divided into two groups, the sodium-coupled glucose transporters (sGLT) and the facilitative glucose transporters (GLUT), based on energy requirement for transport, predicted secondary structure and physiological function (1). The GLUT family of proteins exhibits a conserved fold of 12 transmembrane helices while the sGLT family is predicted to have 14 transmembrane helices (2). The sGLTs use the sodium gradient generated by the Na/K ATPase as energy to drive the transport of glucose against its concentration gradient (3), while GLUTs do not require energy from a secondary substrate to drive transport of glucose (4). Physiologically, sGLTs are responsible for absorption of sugar in the intestine and re-absorption in the kidney compared to GLUTs that are essential for glucose homeostasis in the body (1).

There are 14 members of the GLUT family that have been identified in the human body with different affinities for sugars and tissue localization. There are three classes within the GLUT family defined by sequence homology: i) GLUT1-4, ii) GLUT 5,7,9 & 11 and iii) GLUT 6,8,10,12 & the proton coupled myo-inositol transporter (HMIT) (2,5). All proteins within these classes exhibit a similar fold of 12 transmembrane helices with the N and C termini located on the intracellular side of the membrane and a single N-linked glycosylation site. The classical transporters, GLUT1-4, include the erythrocyte sugar transporter, GLUT1 (6), and the transporter implicated in diabetes, GLUT4 (7). Class ii is composed of the odd transporters that

transport fructose and class iii are composed of the even transporters that contain the glycosylation site in loop 9 as opposed to loop 1 for classes i and ii (8).

The classical transporters, GLUT1-4, are the most well studied class and are associated with several diseases. GLUT1 is found ubiquitously in the body but is known as the erythrocyte sugar transporter comprising up to 20% of membrane proteins in erythrocytes (9). GLUT1 deficiency syndrome is the result of a variety of genetic mutations that lead to developmental delay, abnormal brain development, seizures and decreased coordination of movement due to decreased rates of glucose transport across the blood brain barrier (10). GLUT2 is found in the insulin producing cells of the pancreas and mutations in the gene result in glycogen storage diseases, stunted growth, sugar intolerance and hypoglycemia (8). GLUT4 is implicated in the development of type-2 (adult onset) diabetes and is caused by improper translocation of GLUT4 from intracellular storage vesicles to the plasma membrane in response to insulin in skeletal and cardiac muscle, as well as brown and white adipose tissue (7). Diabetes affects approximately 26 million people with 2 million new cases each year in the United States and leads to severe complications when left untreated, such as kidney failure, heart disease, amputation and blindness (11). The prevalence of diabetes highlights the importance to understanding the mechanisms of glucose homeostasis in order to treat the multitude of diseases associated with sugar transporters.

Despite efforts over the past forty years to understand the mechanism of sugar transport used by GLUTs, details regarding the structure and mechanism are still not thoroughly understood. The lack of information regarding the mechanism is attributed to the difficulty of expressing and purifying large amounts of mammalian membrane proteins necessary for

structural studies. To study the mechanism of sugar transport, the field has turned to studying bacterial homologues of the mammalian sugar transporters.

Several structures of *E. coli* sugar transporters have been studied using X-ray crystallography including lactose permease (LacY) (12), fucose permease (FucP) (13) and most recently the xylose permease (XylE), a bacterial homologue of the classic GLUTs (14). These studies revealed that the proteins form monomers composed of 12 transmembrane helices with the N and C termini located on the intracellular side of the membrane. The protein can be divided into two domains, the N and C terminal domains, composed of six helices that are related by pseudo two-fold symmetry, connected by a long cytoplasmic loop (Fig. 3.1 top). The helices are oriented such that helices 1,2,4,5,7,8,10 & 11 form a central cavity and helices 3,6,9,12 are peripheral to the cavity (Fig. 3.1 bottom). The substrate binding sites are formed at the interface of the N and C terminal domains. The specific residues that make up the substrate-binding site slightly vary between the structures determined according to the substrate transported. LacY and FucP both contain aspartate and glutamate residues implicated in binding sugar and proton induced translocation (12,13). XylE, in contrast to previous structures, does not contain any charged residues participating in the binding of xylose, instead the binding site contains several asparagines and glutamines (14). With the absence of acidic residues near the sugar binding pockets, it is not understood how proton and sugar translocation are coupled by XylE, possible residues are also conserved among GLUTs, which are not driven by a proton gradient (14).

All MFS proteins are thought to work via a similar mechanism using an alternating access model (15). In this model, substrate binds to the outward-facing conformation of the protein triggering a conformational change to the inward-facing conformation; release of the substrate then triggers return of the protein to the outward-facing conformation ready to bind

substrate. The structures of sugar transporters determined thus far have provided snap shots of the proteins in different conformations that largely agree with this proposed model. The structure of LacY was determined in the inward-open ligand-bound state (12) (Fig. 3.2 left); in comparison FucP was determined in outward-open conformation (13) (Fig. 3.2 middle). Most recently, XylE was determined in a ligand-bound outward conformation that was partially occluded to the extracellular environment (14) (Fig. 3.2 right). Taking all these structures into account it is speculated that simple rigid body movement of the two domains would produce a conformational change from outward to inward facing structures (13,14).

XylE shares approximately 50% sequence similarity with GLUT1-4 (14), but there are still significant differences in the transport mechanisms. GLUT1, the most widely studied of all facilitative transporters, shows very distinct biochemical characteristics compared to the prokaryotic homologues for which the structures have been determined (16). GLUT1 transports glucose as well as other hexose sugars without the use of a proton electrochemical gradient and the alternating access model is insufficient to account for the multi-phasic transport kinetics of sugar transport by GLUT1 (16). The most significant difference between the prokaryotic sugar transporters and GLUT1 is that the structures of FucP, LacY and XylE indicated that the proteins form a monomer in the crystal (12,13,14) while human GLUT1 has been shown to exist as a dimer (17). The most well studied of the *E. coli* sugar transporters, LacY, has been shown to function as exclusively as a monomer (18,19,20). In comparison, human GLUT1 has been shown to function as a dimer and in some conditions form a tetramer (17,21,22). Kinetic studies of GLUT1 have also shown that cooperativity exists between the GLUT1 dimers (16). Questions remain as to what is the structure of an oligomeric sugar transporter and how does it compare to the structure of monomeric prokaryotic sugar transporters determined thus far.

The structures of the prokaryotic sugar transporters determined used X-ray crystallography in which the structure is determined in the presence of detergents (12,13,14). The environment of detergent micelles is in stark contrast to the lipid bilayer that membrane proteins are embedded in the cell. The difference in environment may account for discrepancy of the oligomeric states between the structures determined and GLUT1. How the presence of a lipid bilayer affects the structure of sugar transporters is a remaining question in the field. According to the proposed transport mechanism, the proteins undergo large conformational changes and the affect that lipids have on this process is currently unknown. To investigate questions regarding oligomerization and the structure of a sugar transporter embedded in a lipid bilayer we used electron crystallography. Electron crystallography uses two-dimensional crystals, an array of protein embedded in a lipid bilayer, diffraction data is then collected from these crystals using a transmission electron microscope. The structure of the membrane protein is determined within the context of a lipid bilayer environment providing more biologically relevant information relating to how the protein is structured in the cell and how lipids associate with the protein.

Because of the difficulty of working with human membrane proteins such as GLUT1, we chose to study the structure of *E. coli* galactose permease (GalP). GalP is a prokaryotic sugar transporter that uses a proton electrochemical gradient to drive the transport of galactose into the cell (23,24). We chose to study GalP due to similarities to human GLUT1 that are contrasted from the prokaryotic sugar transporter protein structures so far studied. GalP has been shown to transport glucose without the use of an electrochemical proton gradient (23,25,26) and is inhibited by cytochalasin B (26) and forskolin (27,28), which are potent inhibitors of GLUT1 (22,29,30). Recently in our lab we have shown that GalP reconstitutes as an oligomer in two-dimensional crystals and that in these crystals GalP is functional (31). Thus GalP represents a

suitable model to study the structure of the oligomeric sugar transporter, GLUT1, embedded in a lipid bilayer.

Results presented in this study build upon previously published work that established that GalP forms trimers as single particles and when reconstituted in a lipid bilayer (31). In this study, purification of GalP and two-dimensional crystallization conditions were optimized in order to obtain large two-dimensional crystals that were suitable for data collection using cryo-EM. Two-dimensional crystals were embedded in trehalose and a low-resolution cryo-EM projection map was calculated from a single image indicating GalP forms trimers. Optimization of two-dimensional crystals and embedding conditions is still needed to obtain a high-resolution structure of GalP in a membrane.

RESULTS AND DISCUSSION

Screening detergents: protein purification & stability

Detergents have a large effect on the stability of the protein post-purification and during reconstitution experiments; therefore multiple detergents were screened during purification to optimize protein stability. The general purification scheme for screening detergents is outlined briefly and every step was carried out at 4˚C. Membranes prepared from BL21 C43 *E. coli* cells over-expressing N-terminal his-tagged GalP (6xHis-GalP) were solubilized using 1% n-decyl-β-D-maltopyranoside (DM) for one hour followed by nickel ion affinity purification using a 5 mL His-Trap HP column (GE Healthcare). The N-terminal his-tag was then removed using human α-thrombin (Enzyme Research Laboratories) digestion overnight. The detergent was exchanged from DM to the detergent of interest over a Superdex 200 size exclusion column (GE Healthcare) equilibrated with detergent plus 25 mM Tris pH 8, 150 mM NaCl, 1 mM EDTA

(SEC buffer). The elution profile of the size exclusion chromatogram was analyzed for each detergent screened as well as stability before and after centrifugal concentration.

Five different detergents were screened by size exclusion chromatography (SEC) covering different alkyl chain lengths, sugar moiety and charge, as listed in Figure 3.3A. In addition to screening types of detergents, additives were screened including lipids, glycerol and mixtures of detergents with different critical micelle concentrations (CMC). Based on the SEC elution profile of GalP after detergent exchange, the most mono-dispersed, Gaussian shaped peak was observed after exchange in to SEC buffer containing 0.2% lauryldimethylamine oxide (LDAO) (Fig. 3.3B). After exchange of GalP into LDAO, the protein was able to be concentrated to > 30 mg/mL and was stable at 4˚C for up to one week at 10 mg/mL. The maltoside detergents tested, n-nonyl-β-D-maltopyranoside (NM) and DM, resulted in single peaks that were less Gaussian in shape (Fig. 3.3C). In both of the maltoside detergents, GalP was able to be concentrated to 10 mg/mL and was stable in solution up to 3 days at 4˚C. Exchange of GalP into the glucopyranoside detergents, n-nonyl-β-D-glucopyranoside (NG) & n-octyl-β-D-glucopyranoside (OG) resulted in a single peak for OG and two peaks for NG in the SEC elution profile (Fig. 3.3D). After elution from the SEC column, GalP in both NG and OG precipitated when concentrated beyond 1 mg/mL and stability was limited to hours when stored at 4˚C. The addition of 10% glycerol or *E. coli* polar lipid extract during detergent exchange using SEC did not improve the elution profile and the stability of GalP purified with DM or OG was only marginally improved (data not shown).

The ability of a particular detergent to stabilize a membrane protein in solution depends on the affinity of the alkyl chain to bind to exposed hydrophobic regions and compatibility of the head group of the detergent with the hydrophilic surface of the membrane protein (32). The

reason that LDAO was able to stabilize GalP is not clear but maybe due to the long alkyl chain, 12 carbons, in LDAO. Of all the detergents tested, LDAO has the longest alkyl chain as well as the smallest head group, a simple amine oxide compared to the larger sugar moiety head groups of the maltopyranoside and glucopyranoside detergents. The combination of longer chain length and smaller head group might allow LDAO to bind and conform to the hydrophobic regions of GalP better than the other detergents tested. Previously it (33) was shown that GalP is stable in n-dodecyl-β-D-maltopyranoside (DDM), a maltopyranoside detergent with alkyl chain length of 12 carbons, suggesting that longer chain length is important for GalP stabilization in solution. DDM was not included in this screen because the low CMC (0.0087%) results in longer dialysis times.

Screening detergents: reconstitution & lattice formation

The second aim of screening detergents during purification of GalP was to find reconstitution conditions that yielded large two-dimensional crystals suitable for data collection using cryo-EM. Despite the instability of GalP in some detergents, two-dimensional crystallization conditions were screened for each of the conditions tested. Two-dimensional crystals are formed by combining purified membrane protein stabilized with detergent with lipids solubilized in the same detergent, followed by dialysis to remove the detergent. As the detergent is removed from the mixture, the membrane protein reconstitutes into the forming lipid bilayer, possibly forming ordered arrays of protein and lipid in the form of sheets, vesicles or tubes. Single layer sheets are the preferred two-dimensional crystal form but tubes or vesicles can also be used for structure determination. Conditions that form multi-layered sheets or stacks of vesicles should be avoided because these will have multiple lattices that prevent accurate indexing during structure determination.

The procedure for reconstitution of GalP to form two-dimensional crystals is briefly described. GalP was purified, concentrated to 1 mg/mL and combined with 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) solubilized with the same detergent as used for protein purification at different lipid to protein ratios (LPR), ranging from 0.1 to 0.6. The GalP-DMPC-detergent mixtures were then incubated at 4˚C for 30 minutes prior to the start of dialysis. Dialysis was carried out using 50 µL dialysis buttons with 12-14 kDa molecular weight cutoff dialysis membrane and incubated with 100 mL of 10 mM Tris pH, 150 mM NaCl, 0.2% NaN3 with 5 mM EGTA or 50 mM $MgCl_2$ or 50 mM $CaCl_2$ at room temperature. Dialysis was carried out over three weeks with daily buffer changes the first week followed by once per week for the remaining time. Reconstitution and two-dimensional crystal formation was evaluated using transmission electron microscopy of samples stained with 0.75% uranyl formate.

After three weeks of dialysis all conditions showed complete removal of detergent determined by observing water droplet spread during preparation of negative stained grids. Prior to removing sample from the dialysis buttons, it was observed that all buttons had a thin white layer of precipitation on the bottom of the button that did not resuspend to form a homogeneous solution. A summary of the purification conditions and reconstitution results are presented in Table 3.1. GalP purified in OG (GalP-OG) formed crystalline vesicles with 50 mM $CaCl_2$ in the dialysis buffer despite the observation that GalP is not stable in OG. Reconstitution with GalP purified in LDAO (GalP-LDAO) produced crystalline vesicles in all buffers tested and crystalline sheets with dialysis buffer containing 50 mM $CaCl_2$. Reconstitution of GalP purified in DM (GalP-DM) produced vesicles with small crystalline patches measuring 100 to 200 nm in diameter. Other detergent conditions screened produced either vesicles lacking order or protein-lipid aggregation. An example of typical protein-lipid aggregation results is shown in Figure 3.4.

From the detergent screen, purification of GalP in DM, OG, LDAO and reconstitution with DMPC produced two-dimensional crystals. Optimization of two-dimensional crystals of GalP-DM, -OG and -LDAO was conducted next to produce crystals suitable for structure determination.

Negative-stain TEM of GalP-DM two-dimensional crystals

Previous work on determining two-dimensional crystallization conditions showed that GalP formed crystalline vesicles when purified with DM and reconstituted with DMPC (31). These results are reproduced here with the dialysis buffer 10 mM Tris pH 8, 150 mM NaCl, 0.2% NaN_3 and 50 mM $MgCl_2$ at an LPR of 0.6. GalP-DM two-dimensional crystals formed small vesicles measuring 200 nm to 300 nm that had P3 symmetry and only first or second order reflections were observed in Fourier transforms of negative-stain TEM images (Fig. 3.5). The crystalline vesicles appeared to be mostly clustered and none of the vesicles were larger than 300 nm. Attempts to optimize the initial two-dimensional crystals by screening additives such as glycerol, $MgCl_2$ concentration and temperature did not result in improvement of crystals (data not shown). Two-dimensional crystallization of GalP-DM was not further pursued due to the limited stability of GalP in DM and the low resolution of the crystals produced as observed by negative-stain TEM.

Negative-stain TEM of GalP-OG two-dimensional crystals

GalP purified in OG was unstable following detergent exchange but yielded two-dimensional crystals when dialyzed in 10 mM Tris pH 8, 150 mM NaCl, 50 mM $CaCl_2$, 0.2% NaN_3 at room temperature for one week. Initial two-dimensional crystallization conditions of

GalP-OG yielded large crystalline vesicles measuring several microns in diameter (Fig. 3.6). The crystals exhibited P3 symmetry with unit cell dimensions of $a = b = 77$ Å and $\gamma = 120^{\circ}$. Most crystals exhibited several orders of reflections in the image Fourier transform, the highest resolution spots visible were 20 Å, the resolution of negative-stain TEM. Two-dimensional crystals of GalP obtained at LPR of 0.2 had the largest coherent areas measuring more than 500 nm in diameter with sharp spots present in the Fourier transform, but the crystalline vesicles tended to aggregate and fold on themselves as observed in negative stain (Fig. 3.6B). Adjusting the LPR from 0.2 to 0.3 yielded crystals that were larger and appeared to be less prone to aggregation but there was a decrease in the size of the coherent areas of the crystal with some vesicles only having coherent areas measuring 300 nm to 400 nm in diameter (Fig. 3.6A). On the edges of the coherent areas of the crystal obtained at LPR 0.3 (Fig. 3.6A) there is disordered reconstituted protein in contrast to crystals at LPR 0.2 (Fig. 3.6B) that show coherence throughout the entire imaged area.

Reconstitution conditions of GalP-OG were screened to try to improve crystal morphology and to increase the size of coherent areas. Conditions that were assayed included increasing concentration of OG to 2% during purification, temperature, reducing agents, $CaCl_2$ concentration, dialysis rate, dialysis method and glycerol. Most conditions resulted in similar conditions to those obtained previously or protein-lipid aggregation. The addition of 10% glycerol during purification and dialysis seemed to improve coherence over larger area measuring about 700 nm by 400 nm with four order of reflections visible in the image Fourier transform (Fig. 3.7A). Unfortunately the crystals seemed to grow out of large protein-lipid aggregates and crystals clustered together.

To try to improve coherence and aggregation, a combination of DM and OG was used during purification of GalP to stabilize GalP during reconstitution. Both 0.2% DM/ 1% OG and 0.1% DM/ 1% OG were used to purify GalP to screen for both protein stability and reconstitution. The hypothesis was to include only the minimum amount of DM to stabilize the protein during reconstitution, which may lead to crystals with larger coherent areas and less clustering of the crystals. After reconstitution with GalP exchanged from DM to 0.1% DM / 1% OG over SEC, large crystalline vesicles in excess of 3 μm were observed with dialysis with buffer containing 50 mM $MgCl_2$ at an LPR of 0.6 (Fig. 3.7B). Negative-stain TEM revealed some separated vesicles present but most of the vesicles tended to aggregate.

The ability of GalP-OG to reconstitute to form two-dimensional crystals despite showing limited instability beyond 1 mg/mL may have actually helped the reconstitution process by promoting the association of GalP with DMPC rather than the detergent. During screening of the crystallization conditions, there was not an abundance of protein in the background indicating that either the protein aggregated forming large clusters that tend to not stick to the sample grids or that most of the protein was reconstituted. The results obtained with GalP-OG were significantly improved over GalP-DM because the two-dimensional crystals had large coherent areas, which resulted in reflections in the image Fourier transforms up to the resolution of negative stain (Fig. 3.6). The actual resolution of the GalP-OG two-dimensional crystals will have to be assessed under cryogenic conditions as opposed to negative stain because then the resolution is limited by sample prep and the coherence of the crystals.

Negative-stain TEM of GalP-LDAO two-dimensional crystals

In contrast to GalP-OG, GalP-LDAO exhibited prolonged stability before reconstitution and during concentration steps. Initial reconstitution screens of GalP-LDAO with DMPC produced two-dimensional crystals in all three of the dialysis buffers tested, 10 mM Tris pH8, 150 mM NaCl, 0.2% NaN_3 with 50 mM $MgCl_2$ or 50 mM $CaCl_2$ or 5 mM EGTA with different crystal morphologies and size of coherent areas (Fig. 3.8). Using dialysis buffer with 50 mM $MgCl_2$ at an LPR of 0.2 yielded small aggregated vesicles and sheets, approximately 500 nm in diameter. Most of the areas that could be imaged had coherent areas measuring 200 nm by 200 nm with several orders of reflections visible in the Fourier transform (Fig. 3.8A). Dialysis buffer containing 50 mM $CaCl_2$ and an LPR of 0.5 produced two-dimensional crystals that were large sheets measuring up to 10 μm in diameter (Fig. 3.8B). The sheets appeared to be multi-layered as observed by two sets of lattices in the Fourier transform of the image (Fig. 3.8B). It was not possible to distinguish whether the sheets observed in negative stain were multi-layered sheets or were perhaps large vesicles that broke along the edges of the vesicle during the negative staining procedure. In Figure 3.8B, the crystal measures 1.8 μm in diameter with several orders of reflections present in the Fourier transform of the image but the higher resolution reflections were weak in intensity indicating that the crystal was not completely coherent. The third condition tested was dialysis with buffer containing 5 mM EGTA at an LPR of 0.2, which resulted in large crystalline vesicles measuring in up to 5 μm in diameter (Fig. 3.8C). The vesicles tended to be highly aggregated as observed in the negative-stain TEM image in Figure 3.8C. The coherent regions of the crystals tended to be small, similar to the crystals obtained in 50 mM $MgCl_2$, measuring up to 400 nm in diameter and a Fourier transform of the image shows multiple lattices with several orders of reflections. Although dialysis with buffer containing 5 mM EGTA or 50 mM $CaCl_2$ yielded larger crystalline vesicles and sheets, the resolution of the

small patches of the $MgCl_2$ was higher than the larger coherent areas of the other dialysis conditions tests.

Since the crystals of GalP-LDAO produced from dialysis with $MgCl_2$ exhibited higher resolution reflections these crystals were chosen for optimization for aggregation and size of coherent crystalline areas. A broad screen of $MgCl_2$ concentration was conducted during reconstitution including 10 mM $MgCl_2$, 25 mM $MgCl_2$ or 50 mM $MgCl_2$ in the standard dialysis buffer of 10 mM Tris pH 8, 150 mM NaCl, 0.2% NaN_3. The best results were obtained with 10 mM $MgCl_2$ at an LPR of 0.2 yielding large crystalline vesicles measuring several microns in diameter with coherent areas measuring 400 nm in diameter (Fig. 3.9A), although still not coherent throughout the entire area that can be imaged.

The degree of coherence in the two-dimensional crystals could be related to the amount of lipid present. If there are too many lipids present in the crystals the proteins may not pack as tightly resulting in low coherence and decreased resolution of crystals. To improve coherence a fine screen of LPR was conducted with dialysis buffer containing 10 mM $MgCl_2$. The LPR range screened was 0 to 0.34 in increments of 0.02, with the best crystals obtained at LPR of 0.24 (Fig. 3.9B). The crystals formed large vesicles and sheets measuring several microns in diameter with coherent areas measuring 500 nm in diameter. Fourier transform of coherent areas of the crystals produced three to four orders of reflections with the highest resolution reflections visible at 25 Å. These crystals as well as others produced using $MgCl_2$ in the dialysis buffer were clustered reducing the area available to image.

Aggregation of the crystals could be a result of interactions between the lipids, exposed surfaces of the protein or a result of protein-lipid interactions. The following reagents were incubated with the crystals to try to disrupt aggregation of the crystals, 10 mM dithiothreitol, 3 M

urea, 1 M NaCl, 0.5 M NaOH, 0.5 M acetic acid and 1% trehalose. Reagents were incubated with the crystals at room temperature overnight and samples were stained with 0.75% uranyl formate. None of the conditions tested were able to separate the crystals (data not shown). Sonication of the crystals was also tested to try to separate the crystals using mechanical force. Crystals were sonicated in a water bath at room temperature for 15 minutes then negative stain samples were prepared. Again, crystalline vesicles still remained in clusters although more of the clusters were observed on the grid, possibly freeing the clusters bound to precipitated protein aggregates.

GalP-LDAO two-dimensional crystals showed an increase in the size of crystalline vesicles compared to GalP-OG crystals and the formation of sheets with dialysis buffer containing 50 mM $CaCl_2$. Despite the large crystals produced using GalP-LDAO the size of coherent crystalline area was not as large as with GalP-OG, further optimization of GalP-LDAO crystals with respect to concentration of $CaCl_2$ and LPR may increase the coherence of the crystals. The coherence of the crystals may also be affected by dehydration of the crystals during preparation of negatively stained samples, thus the ultimate test to see the quality of the crystals is to screen freezing conditions for cryo-EM.

Cryo-EM of GalP two-dimensional crystals

Both GalP-OG and GalP-LDAO produced two-dimensional crystals that had regions that were ordered to 20 Å resolution as observed by negative-stain TEM (Figs. 3.6 & 3.8). The aggregation that was observed consistently throughout screening of two-dimensional crystals maybe due to the dehydration of the crystals during the staining procedure. To assess the resolution of the crystals, freezing conditions were screened with the best crystals obtained for

GalP-OG and GalP-LDAO for imaging using cryo-EM where resolution is only limited by the crystals and aberrations of the microscope.

GalP-OG crystals from reconstitution with DMPC at an LPR of 0.20 dialysis with 10 mM Tris pH 8, 150 mM NaCl, 0.2% NaN_3 and 50 mM $CaCl_2$ were screened using slow freezing with a variety of cryoprotectants: glucose, trehalose and tannic acid, ranging from 2.5% to 10%. The best results were from crystals incubated overnight at 4˚C with 10% trehalose, grids were prepared using back-injection method and then cooled to liquid nitrogen temperature after insertion of the holder into the microscope. Preservation with trehalose resulted in visible reflections to 27 Å in the Fourier transform of the image and after processing reflections with weak intensity was visible at 18 Å (Fig. 3.10A). The embedded crystals exhibited P3 symmetry with unit cell dimensions of $a = b = 80$ Å and $\gamma = 120^{\circ}$. A projection map at 18 Å was calculated from a single cryo-EM image using 2dx showing densities with three-fold symmetry indicating that GalP forms a trimer (Fig. 3.10B), as was seen previously seen in negative-stain TEM (31). In the crystal, there are no protein-to-protein contacts so the lipids surrounding each GalP trimer mediate crystallization. Aggregation of crystalline vesicles was still an issue interfering with the ability to find areas that consisted of a single crystal to be imaged. Breaks in the crystalline lattice were also observed in the cryo-EM image indicating that the crystal is dry and that more embedding conditions must be screened.

Two-dimensional crystals of GalP-LDAO reconstituted with DMPC at LPR of 0.5 with dialysis buffer of 10 mM Tris pH 8, 150 mM NaCl, 0.2% NaN_3 and 50 mM $CaCl_2$ had large crystalline sheets several microns in diameter compared to the smaller crystalline vesicles obtained from GalP-OG. Freezing conditions of GalP-LDAO were screened using 1% trehalose with a variety of grid preparation and freezing methods. The best results were from GalP-LDAO

crystals incubated for 1 min with 1% trehalose, then blotted for 5 seconds and plunged into liquid nitrogen to freeze the crystals (Fig. 3.11). A Fourier transform of the image shown in Figure 3.11 of GalP-LDAO crystals exhibited P3 symmetry with unit cell dimensions of $a = b =$ 85 Å and $\gamma = 120^{\circ}$. The highest resolution reflections visible in the Fourier transform of the image were at 27 Å and appeared as a smeared ring. The smearing of higher order reflections is probably a result of a mosaic crystal lattice.

With the cryogenic conditions screened here for crystals of GalP-OG and GalP-LDAO both seemed to only have low-resolution reflections present in the Fourier transform of the images. The low-resolution observed maybe a result of non-optimal freezing conditions and embedding of the crystals or the coherence of the crystals themselves. Aggregation was a pronounced issue seen under cryogenic conditions that inhibited further data collection. Despite the improvement in the two-dimensional crystallization of GalP presented here, aggregation remained a large factor that hindered data collection to determine the structure of GalP embedded in a phospholipid bilayer.

Persistent aggregation of GalP during reconstitution experiments

An observation during all of the reconstitution experiments with GalP-OG or GalP-LDAO is that during the initial two days of dialysis a white layer was formed at the bottom of the dialysis button. After the completion of dialysis, the white particulates were observed using light microscopy and appeared to be similar in morphology to dialysis of GalP without DMPC indicating that the white particulates may be protein precipitation. This observation was more pronounce with GalP-LDAO as there was 10 times more protein used for dialysis than with GalP-OG, 5 mg/mL versus 0.5 mg/mL, respectively. All reconstitution experiments were

initially setup at 4˚C then dialysis was carried out at the desired temperature but when GalP-LDAO was mixed with lipids and allowed to incubate at room temperature within 30 minutes white precipitation formed. The amount of white precipitation present was correlated to the amount of DMPC that was added, higher LPRs exhibited more pronounced precipitation than lower LPRs. After overnight incubation this white precipitation was observed using light microscopy, the white particulates appeared to be protein precipitation as observed previously.

Aggregation of GalP-LDAO after addition of DMPC was quite a surprising observation because neither the protein nor the lipid controls prepared under the same temperature and buffer conditions produced any observable precipitation. To find conditions that prevented the precipitation of GalP upon addition of lipids several experiments were conducted to assay the effect of NaCl concentration, protein concentration, glycerol and lipids. Salt concentrations from 150 mM NaCl to 1M NaCl were screened revealing that an increase in salt concentration resulted in prolonged time before precipitation was visible. After 24 hours the GalP & DMPC mixture with 1 M NaCl did not appear to have any precipitation present. The presence of 20% glycerol in the 10 mg/mL stock of GalP before the addition of lipid did not prevent precipitation. GalP concentrations from 0.5 mg/mL to 5 mg/mL were screened with DMPC at an LPR of 0.3 that resulted in a similar observation as with LPR, higher protein concentrations exhibited larger amounts of precipitation. The observed effect of protein concentration maybe a result of the simple fact that if more protein is present than more is available to be precipitated by the lipid.

To test whether this effect was lipid dependent a screen of different lipids including 1,2-dimyristoyl-*sn*-glycero-3-phosphoethanolamine (DMPE), 1,2-dipalmitoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (DPPG), L-α-phosphatidylserine, *E. coli* polar lipid extracts, 1:4 DMPE:DMPC and DMPC (control) was setup at an LPR of 0.3 and 0.5 with GalP-LDAO at 5 mg/mL. Different

lipids induced precipitation of GalP within varying amounts of time. DPPG had the most pronounced induction of precipitation, within a couple of minutes a cloud of white was observed, while a 1:4 DMPC:DMPE reduced the amount of white precipitation observed after 30 minutes. The lipids tested thus far all resulted in varying degrees of precipitation when added to GalP-LDAO.

A possible explanation for this observation is that GalP prefers certain lipids that were stripped during the solubilization and purification steps. From the detergent screens it appears that GalP preferred longer chained detergents, thus longer chained lipids may be better for reconstitution such as palmityl (16 carbon chain) or stearyl lipids (18 carbon chain). Analysis of the lipid composition of *E. coli* strain K12 has shown that the majority (~80%) of phospholipids are phosphatidylethanolamines and that a majority of fatty acid moieties of the phospholipids are 16:0, 16:1 and 18:1, comprising about 80% to 90% of phospholipids (34,35). Reconstitution of GalP-LDAO in lipids having these characteristics might help the initial precipitation of GalP that is observed, as well as the aggregation of two-dimensional crystals.

CONCLUDING REMARKS

Membrane proteins are encoded by approximately 30% of the human genome (37) yet the number of structures determined of human membrane proteins is only a small fraction compared to the abundance of structures of soluble proteins. Due to the difficulty in working with eukaryotic, especially human, membrane proteins, we have turned to investigating homologous prokaryotic membrane proteins to gain insight into the mechanisms of membrane proteins essential in human physiological processes. There are still significant differences in the specific details of the mechanisms used by prokaryotic versus eukaryotic membrane proteins

such as the regulation or gating of transport, but determining the structure of homologous prokaryotic membrane proteins allows investigators to make hypotheses that can be tested either *in vitro* or *in vivo*.

The mechanism used by the human glucose transporter family has been elusive even though GLUT1 was identified over 20 years ago (37). Over 25 years of research has shown that GLUT1 exhibits asymmetric glucose transport (38,39), but the mechanism has not been determined because there is no structure of GLUT1. To gain insight into sugar transporter structure the field has turned to investigating sugar transporters in prokaryotes. Currently, there are three structures of *E. coli* sugar transporters that have been determined, LacY (12), FucP (13) and XylE (14), which transport sugar using a proton electrochemical gradient. These structures provided details regarding the overall architecture, the substrate binding pocket and a proposed model for sugar transport. The recently determined structure of XylE showed an intracellular four helix bundle formed by helices in the large cytoplasmic loop connecting transmembrane helices 6 and 7 and the C-terminus that was seen for the first time (14). Interaction of the four helix bundle and the intracellular surface of the transmembrane domains is mediated by conserved sugar porter family amino acid sequence motifs that are conserved in XylE as well as the classical glucose transporter family members, GLUT1-4 (14). Mutation of residues in this domain resulted in significantly reduced proton driven sugar transport of XylE, suggesting that this bundle plays an integral role in the transport of sugar (14). How this intracellular bundle of helices modulates sugar transport is currently unknown but provides a model to assess function.

All of the structures determined thus far were determined using X-ray crystallography, which uses crystals of the protein stabilized by detergent. The native environment of a membrane protein is embedded in a phospholipid bilayer and detergents are not equal substitutes for a

membrane. Studies of the influence of lipids on the structure and mechanism of LacY have shown that the presence of phosphatidylethanolamine lipids is required for not only proper insertion of LacY into the membrane (40), but also transport of lactose (41). Determination of the structure of a membrane protein in a detergent environment is only a small part of understanding the mechanism; the protein-lipid interactions have a tremendous influence on the structure and function.

GLUT1, the most studied of the classical human glucose transporters has been shown to function as a dimer and under certain conditions as a tetramer (17). The structures of prokaryotic sugar transporters determined have all been monomers, thus there are still questions remaining about how sugar transporters form oligomers and what is the structure of each of the subunits in the oligomeric state. LacY that crystallized as a monomer has been shown to function as a monomer (12), but LacY shows limited homology to the human GLUTs (25). GalP is about 50% similar in sequence to the classical glucose transporters (25) and we have shown that GalP embedded in a phospholipid bilayer forms trimers. The two-dimensional crystals of GalP in DMPC have been shown to be able to bind substrate as published previously (31). The particular interactions that contribute to the formation of trimers are not currently understood because of the limited resolution of information obtained from imaging the two-dimensional crystals. The ability of GalP to oligomerize as well as the ability to transport glucose without the use of an electrochemical gradient suggests that determining the structure of GalP in a phospholipid bilayer could provide information to help understand the complex mechanisms of sugar transport used by GLUT1.

Here, we presented an optimized purification, two-dimensional crystallization conditions of GalP and for the first time a projection map of GalP obtained using cryo-EM. The crystals of

GalP-LDAO reconstituted with DMPC with 50 mM $CaCl_2$ included in the dialysis buffer exhibited large sheets that if were coherent would be suitable for diffraction where atomic resolution structural information can be obtained (Fig. 3.8B). These conditions were improved over those previously published (31) but more optimization still remains to be able to form two-dimensional crystals that will yield atomic resolution information. Future efforts for optimization of GalP two-dimensional crystals will focus on screening lipids to help stabilize GalP during reconstitution, which may increase the coherence of the crystals.

MATERIALS AND METHODS

Protein expression, purification & detergent screening

The *E. coli* galactose permease gene was cloned into PET15b vector for over-expression in BL21 DE3 C43 *E. coli* cells with an N-terminal histidine tag followed by a thrombin cleavage site. Large scale expression of GalP (12 L) was carried using the following method: two aliquots of 75 mL LB with 100 μg/mL ampicillin were inoculated with colonies from streak plates of GalP-PET15b *E. coli* from glycerol stock. The starter cultures were grown overnight at 37˚C, shaking at 250 rpm. Starter cultures were used to inoculate 12 L of LB with 100 μg/mL ampicillin, approximately 10 mL culture per liter of LB. The 1L cultures were grown at 37˚C, shaking at 250 rpm till an OD600 of ~ 0.5 was reached, then 0.3 mM IPTG was added to each liter of LB and then temperature of the shaker reduced to 18˚C. GalP expression was carried out overnight at 18˚C shaking at 250 rpm.

Cells were harvested at 3000 x g, resuspended in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM PMSF (lysis buffer) to a density of 10 g per 100 mL. Cells were lysed with 3 passes using a Microfluidics microfluidizer operating at 15,000 psi. Cell lysate was then centrifuged at

10,000 x g to pellet cell debris and membranes were isolated by centrifugation of supernatant at 120,000 x g. Membranes were washed with lysis buffer and resuspended to a final volume of 120 mL (10 mL per liter of culture); 10 mL aliquots were frozen and stored at -80˚C.

GalP was purified as published previously with some minor changes (31). Membranes were solubilized by adding 1.1 mL of 10% decyl maltopyranoside (DM) to a 10 mL aliquot of membranes and rotating for 1 hour at 4˚C, then insoluble membrane was pelleted at 120,000 x g. Supernatant containing solubilized GalP was then applied to a 5 mL His-Trap HP pre-packed column (GE Healthcare) equilibrated with 25 mM Tris-HCl pH 8.0, 500 mM NaCl, 30 mM imidazole and 0.3% DM for nickel ion affinity purification using a Bio-Rad FPLC. The concentration of NaCl and imidazole in the solubilization supernatant was adjusted to correspond to the HisTrap equilibration buffer. The flow rate during binding was 1 ml/min to allow sufficient time for protein to bind resin, for subsequent wash and elution steps the flow rate was increased to 5 ml/min. The resin was then washed with 10 column volumes (CV) of 25 mM Tris-HCl pH8.0, 500 mM NaCl, 0.3% DM, 30 mM imidazole, followed by 5 CV of buffer containing 60 mM imidazole and GalP was eluted using 10 CV of buffer containing 500 mM imidazole. Fractions from the elution peak were collected and the N-terminal his-tag was cleaved using human α-thrombin (ERL), 10 units per mg of GalP, overnight at 4˚C. GalP was then concentrated using a Vivaspin 6 50 kDa MWCO centrifugal concentrator at 2,000 x g at 4˚C to 1 mL. Concentrated and cleaved GalP was then loaded on to a 24 mL Superdex 200 10/300 GL (GE Healthcare) equilibrated with 25 mM Tris-HCl pH 8.0, 150 mM NaCl, detergent and 1 mM EDTA and ran at 0.5 ml/min for purification using size exclusion chromatography. For screening of detergents for purification and subsequent two-dimensional crystallization trials, the detergent was exchanged from DM during the size exclusion chromatography step using buffer containing

at least two CMCs of the new detergent, with the exception of LDAO; 0.2% LDAO was used for SEC. All detergents were purchased from Affymetrix and were of the highest grade, anagrade. Fractions corresponding to GalP were then concentrated using a Vivapsin6 30 kDa MWCO (LDAO) or 50 kDa MWCO (all other detergents) centrifugal concentrator to 10 mg/ml or 1mg/mL depending on the detergent and assessed using BCA protein concentration assay (Pierce).

Two-dimensional crystallization

GalP was purified and concentrated to 1 mg/mL or 10 mg/mL depending on stability in detergent was used for reconstitution to form two-dimensional crystals. Lipids for reconstitution experiments were prepared corresponding to the particular detergent used to purify GalP. Lipids used to produce two-dimensional crystals of GalP-OG were prepared as follows: 20 mg of DMPC were solubilized with 200 μL of 20% OG then diluted to 2 mL using 20 mM Tris pH8, 150 mM NaCl for a final concentration of 2% OG. For crystals produced using GalP-LDAO, 20 mg of DMPC was solubilized with 200 μL of 10% LDAO and diluted to 2 mL using 20 mM Tris pH8, 150 mM NaCl, for a final concentration of 1% LDAO. 200 μL aliquots of lipids were frozen at -80˚C. For LDAO, the lipids were diluted two-fold with 20 mM Tris pH8, 150 mM NaCl to lower the LDAO concentration to 0.5%. All the lipids used were purchased from Avanti Polar Lipids in powder form. Mixtures of GalP and lipid containing the appropriate lipid-to-protein ratios were incubated at 4˚C for 30 min prior to assembling the dialysis buttons. For removal of the detergent by dialysis, 50 μL dialysis buttons (Hampton Research) were used with 12-14 kDa molecular weight cut-off dialysis membranes (SpectraPor). Dialysis was carried out against 100 mL of detergent free buffer in 250 mL sterile bottles (Corning) at the desired

temperature with buffer changed daily during the first two weeks of dialysis and once a week during the last two weeks of dialysis. After four weeks of dialysis buttons were opened and screened using negative stain electron microscopy.

Negative-stain TEM

Grids were prepared from the reconstitution experiments as follows for screening by negative-stain TEM. 3 μL of sample was applied to a negative glow-discharged carbon coated 400 mesh copper grid (Ted Pella) and incubated for 30 seconds. Excess sample was blotted with filter paper and stained using 0.75% uranyl formate. Grids were viewed using FEI Tecnai Spirit transmission electron microscope operating at 120kV and images of samples were recorded on a 4k x 4k Tietz CMOS detector over a range of magnifications.

Cryo-EM and image processing

For GalP-OG crystals, the crystals were incubated overnight with 10% trehalose and then applied to a molybdenum grid using the back injection technique. Back injection consists of floating carbon on a water surface, scooping the carbon with a sample grid, application of the sample to the side of the grid without carbon, then blotting by placing the grid sample side down on filter paper. GalP-OG crystals embedded in 10% trehalose were then loaded onto a Gatan 626 cryo-holder and inserted into the microscope. Once grid quality was assessed the holder was then cooled to -175˚C. Grids were viewed using an FEI Tecnai F20 TEM operating at 200 kV and images were collected using a Gatan 4k x 4k CCD at a magnification of 50,000x yielding a pixel size of 2.16 Å/pixel.

For GalP-LDAO two-dimensional crystals, the samples were prepared using fast freezing in comparison to the slow freezing method used with the GalP-OG crystals. GalP-LDAO crystals were aliquoted onto a carbon coated copper grid (Quantifoil 4/1, Ted Pella) with 1% trehalose for 1 minute, then blotted with filter paper for 30 seconds and plunged into liquid nitrogen. The grid was then loaded onto a liquid nitrogen cooled Gatan 626 cryo-holder and inserted into an FEI Tecnai F20 TEM operating at 200kV. Images were collected at a magnification of 62,000x yielding a pixel size of 1.33 Å/pixel using a Tietz CMOS detector.

Images that yielded several orders of reflections in the image Fourier transforms were processed using 2dx (42,43), a graphical user interface for MRC programs (44), to calculate projection maps of GalP.

ACKNOWLEDGEMENTS

We would like to thank the Murdock Charitable Trust and the Washington Research Foundation for support of the cryo-EM facility at the University of Washington – Department of Biochemistry and the Howard Hughes Medical Institute for support of the cryo-EM at Janelia Farm Research Institute. Research presented here was funded by Howard Hughes Medical Institute and B.S.V. was funded in part by NIGMS 2T32 GM007270.

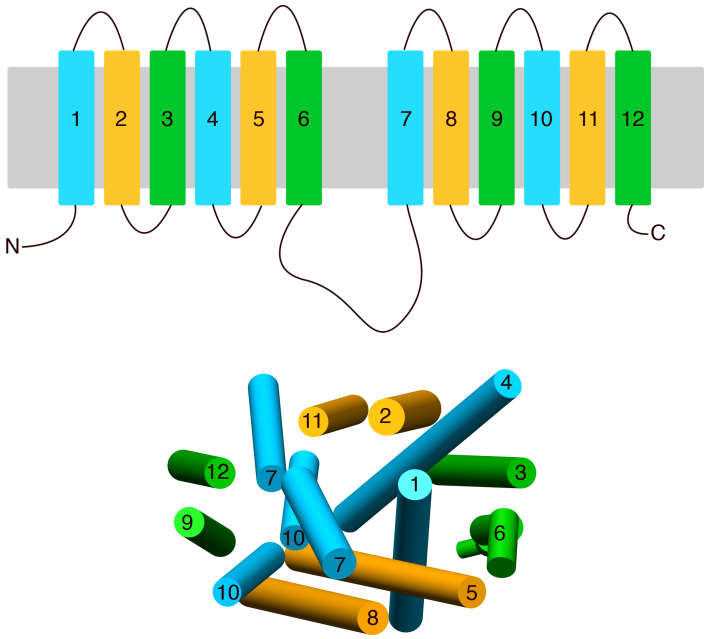

Figure 3.1 Sugar transporter topology and arrangement of helices within the monomer. (Top) Sugar transporters exhibit the typical MFS topology with 12 transmembrane helices that can be separated in to the N-terminal and C terminal domain. The two domains have pseudo two-fold symmetry and are connected via a long intracellular loop. (Bottom) Typical arrangement of the helices in the structures of sugar transporters determined with the binding cavity composed of helices 1,2, 4, 7, 5, 8, 10 and 11, surrounded by peripheral helices, 3, 6, 9 and 12. Bottom figure produced using the XylE structure (PDB 4GBY) (14) and the program UCSF chimera (45).

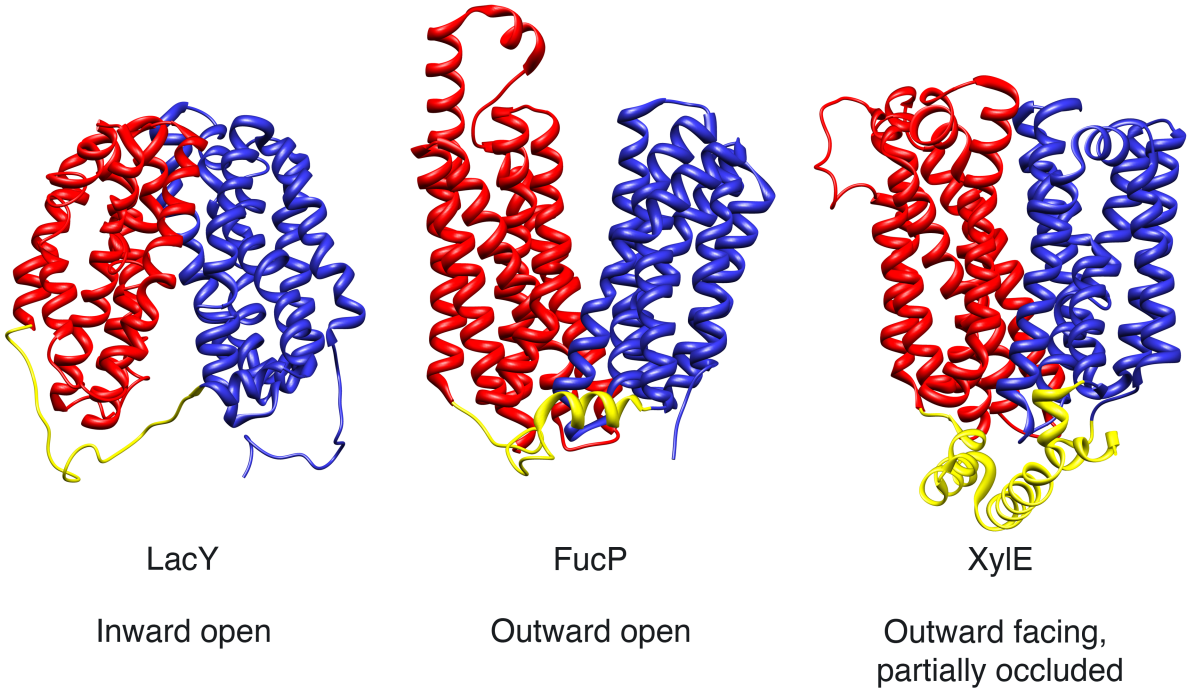

Figure 3.2 Conformations of sugar transporter structures determined. Structures of the three sugar transporters determined using X-ray crystallography in different conformations. LacY was determined in the inward open conformation with (12) and without substrate bound (46), the structure without substrate is shown (PDB 2V8N) (46). The structure of FucP was determined in the outward open state (PDB 3O7Q) (13). The GLUT homologue, XylE, was determined with substrate bound in the outward facing partially occluded state (PDB 4GBY) (14).

A

Detergent	CMC
Lauryldimethylamine-N-oxide	0.023%
n-Decyl-β-D-maltopyranoside	0.087%
n-Nonyl-β-D-maltopyranoside	0.28%
n-Octyl-β-D-glucopyranoside	0.53%
n-Nonyl-β-D-glucopyranoside	0.20%

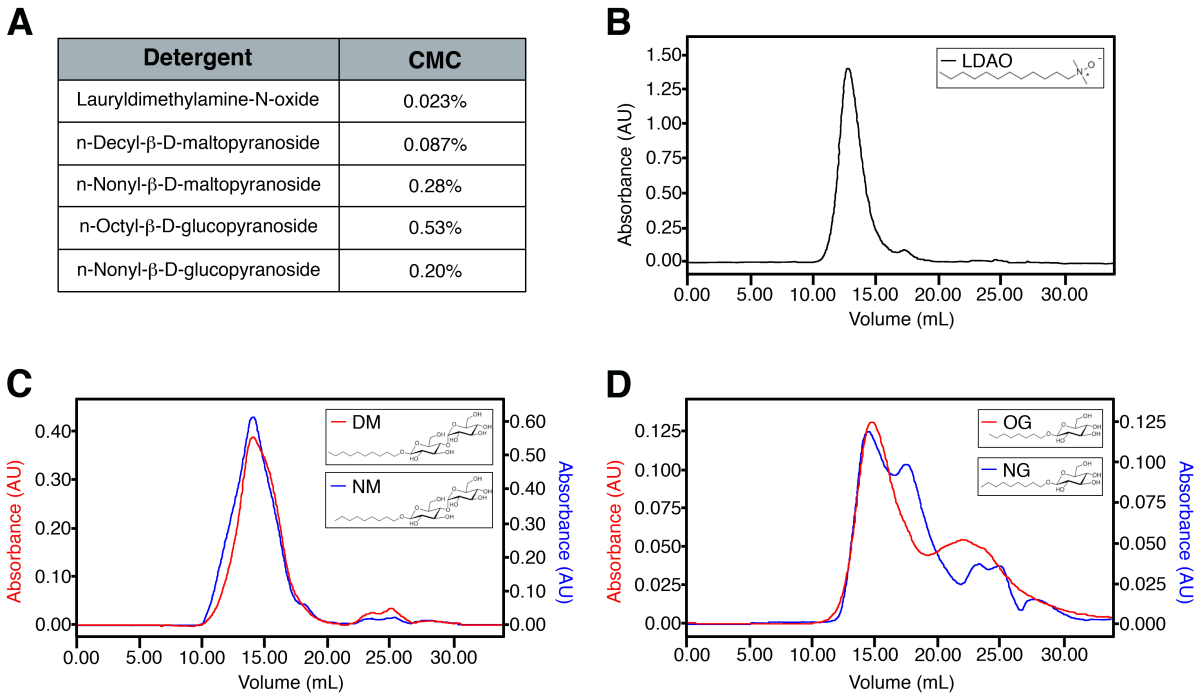

Figure 3.3 Screening GalP homogeneity in different detergents by exchange over SEC. (**A**) Table of the five different detergents and corresponding CMC of each detergent screened by exchange from DM using SEC. (**B-D**) SEC chromatograms of GalP exchanged into different detergents showing that GalP-LDAO (**B**) had the most Gaussian shaped elution peak. GalP purified in either of the maltoside detergents (**C**), NM & DM, resulted in a single broad peak that was less Gaussian shaped. Purification in the glucopyranoside detergents (**D**), OG & NG, resulted in two peaks for NG and a broad single peak in OG.

Table 3.1 Summary of purification conditions screened and reconstitution results.

Detergent	Stability	Reconstitution (Lattice Formation)		
		$MgCl_2$	$CaCl_2$	EGTA
0.3% DM	Concentrate	V (yes)	PLA	PLA
0.3% DM + ECPL	Concentrate	V (no)	PLA	–
0.3% DM + 10% Glycerol	Concentrate	V (no)	PLA	–
1% OG	Limited	V (no)	V (yes)	PLA
1% OG + ECPL	Limited	PLA	PLA	–
1% OG + 10% Glycerol	Limited	PLA	V (yes)	–
2% OG	Limited	V (no)	V (yes)	–
0.2% DM / 1% OG	Limited	V (yes)	PLA	–
0.1% DM / 1% OG	Limited	V (yes)	PLA	–
0.7% NM	Concentrate	PLA	PLA	PLA
0.4% NG	Limited	V (no)	V (no)	PLA
0.2% LDAO	Concentrate	V (yes)	V/S (yes)	V (yes)

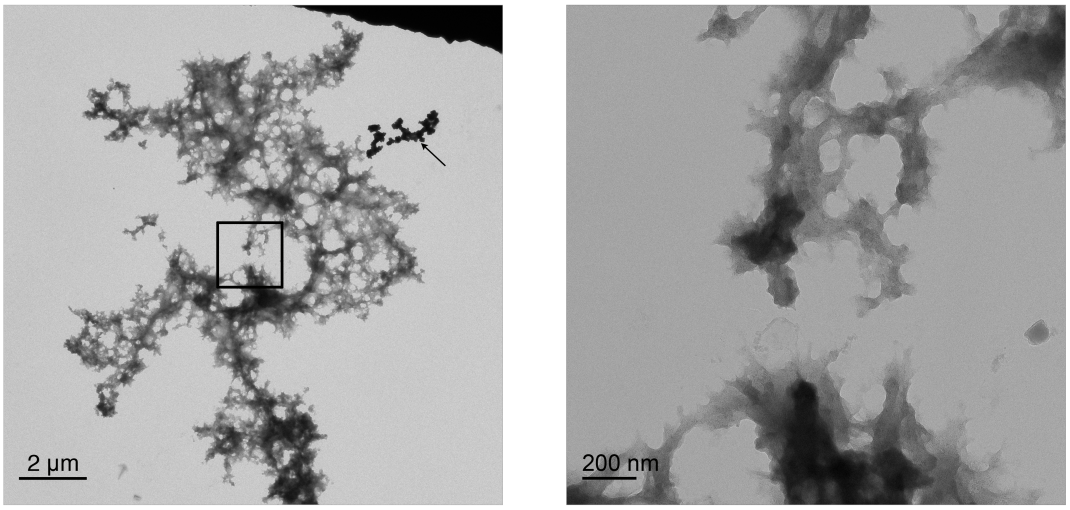

Figure 3.4 Example of typical protein-lipid aggregation observed during screening of two-dimensional crystallization conditions by negative-stain TEM. Several purification conditions screened for two-dimensional crystallization resulted in protein-lipid aggregates as observed by negative-stain TEM. Shown on the left is a micrograph of a protein-lipid aggregate observed after GalP-NM reconstitution with DMPC at an LPR of 0.4 with dialysis buffer containing 50 mM $CaCl_2$. The image on the right was taken at a higher magnification of the boxed region of the left micrograph showing the branched web-like structures present in the protein-lipid aggregates. The arrow in the micrograph on the left indicates a structure representing protein aggregation with no association with lipids.

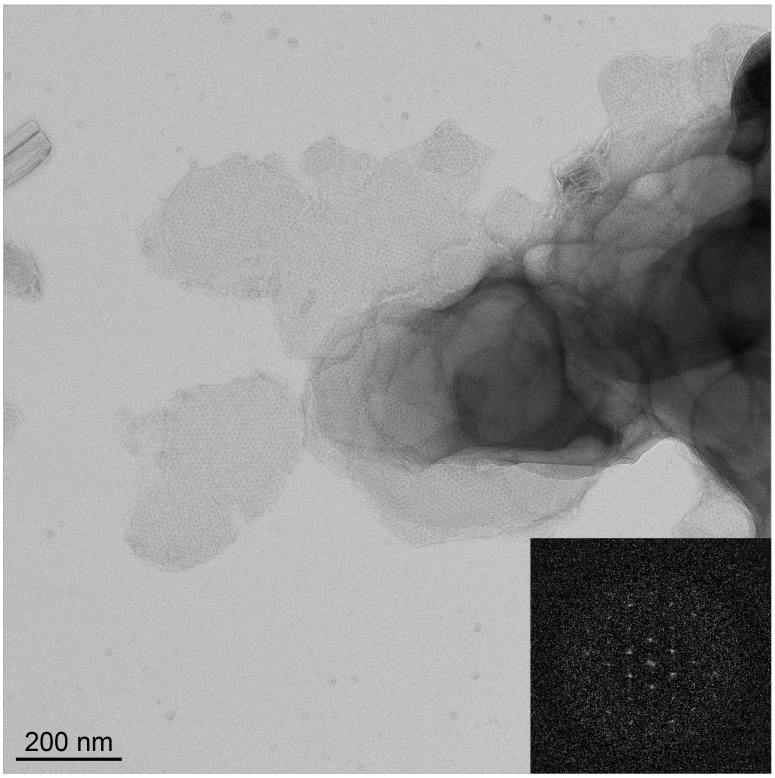

Figure 3.5 Negative-stain TEM of two-dimensional crystals formed with GalP-DM. GalP-DM formed small crystalline vesicles measuring about 200 to 300 nm in diameter. Fourier transforms of small regions revealed second or third order reflections (inset). The GalP-DM vesicles tended to be highly clustered as seen in the right portion of the micrograph.

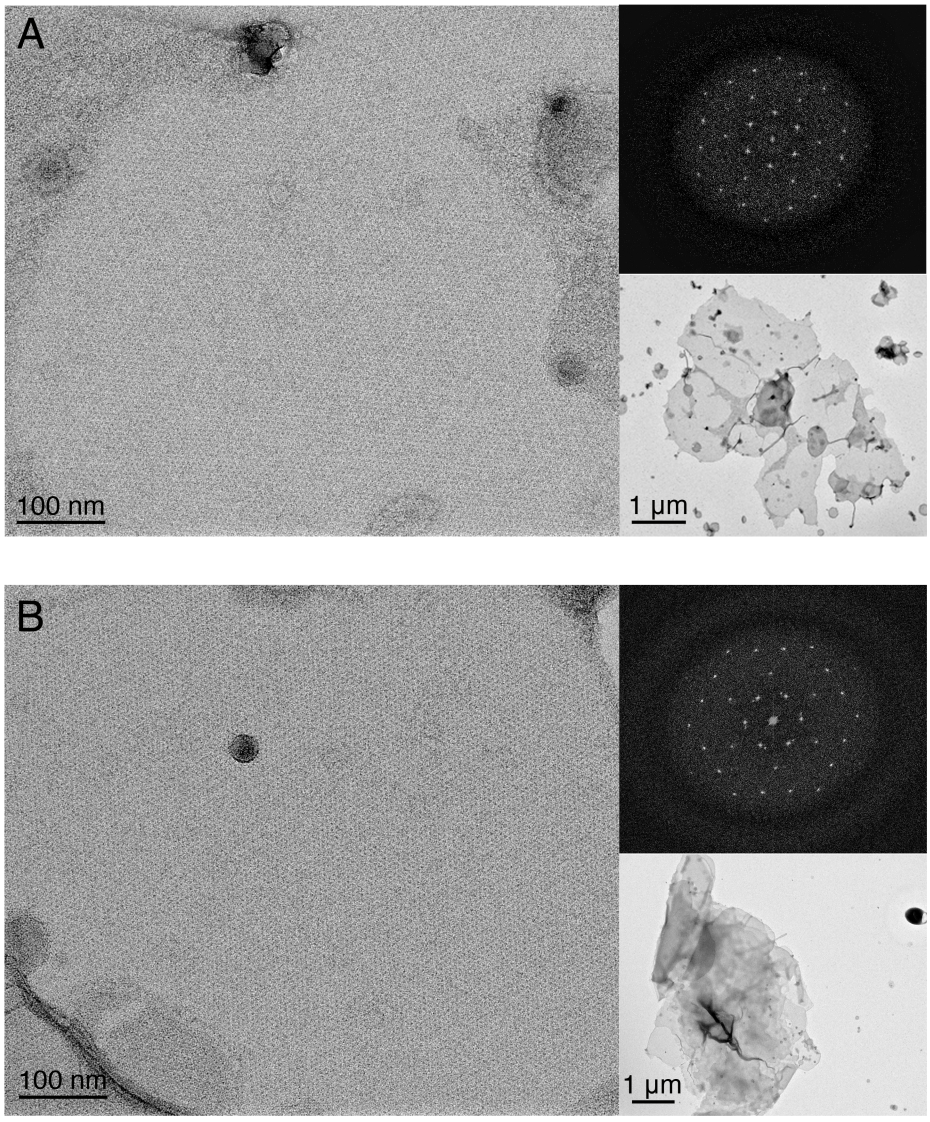

Figure 3.6 Formation of two-dimensional crystals using GalP-OG. (**A**) Reconstitution of GalP-OG with DMPC at an LPR of 0.3 using dialysis buffer containing 50 mM $CaCl_2$ resulted in large sheet-like vesicles measuring several microns in diameter (**A, bottom inset**). Crystalline patches measured 400 nm in diameter with disordered reconstituted protein visible on the edges of the crystalline patches. Fourier transform of the micrograph shows some forth order reflections, which is at 20 Å, the resolution of negative stain (**A, top inset**). Decreasing the LPR to 0.2 resulted in larger crystalline areas measuring more than 500 nm in diameter and was coherent throughout the imaged area (**B**). In the Fourier transform of the image, sharp reflections to the resolution of negative stain are observed (**B, top inset**) but the crystals tended to be highly clustered (**B, bottom inset**).

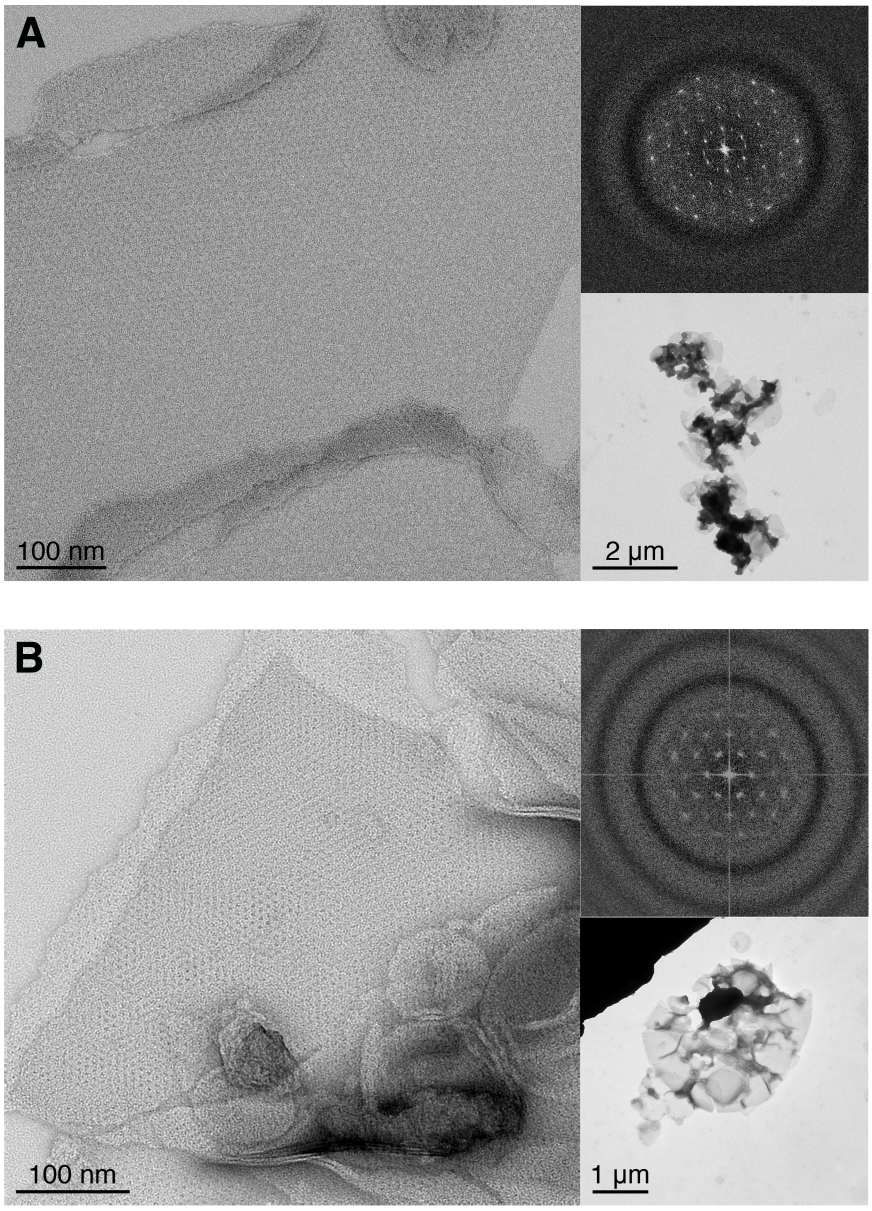

Figure 3.7 Effect of glycerol and a mixture of detergents on GalP-OG two-dimensional crystal formation. The addition of 10% glycerol during purification and dialysis resulted in large vesicle crystals with the coherent region shown measuring 800 nm by 400 nm (**A**). A Fourier transform of the micrograph shows several forth order reflections to the resolution limit of negative stain (**A, top inset**) but the crystals seemed to form out of protein-lipid aggregates (**A, bottom inset**). Purification of GalP using a combination of 0.1% DM and 1% OG and reconstitution with DMPC resulted in crystalline vesicles measuring 3 µm in diameter (**B, bottom inset**) but clustering of vesicles resulted in limited amount of areas that could be imaged. Visible areas were coherent but measured only 300 nm in diameter due to the clustering of nearby vesicles (**B**). Fourier transform of a region of the crystal showed three orders of reflections with the highest resolution reflections at 25 Å (**B, top inset**).

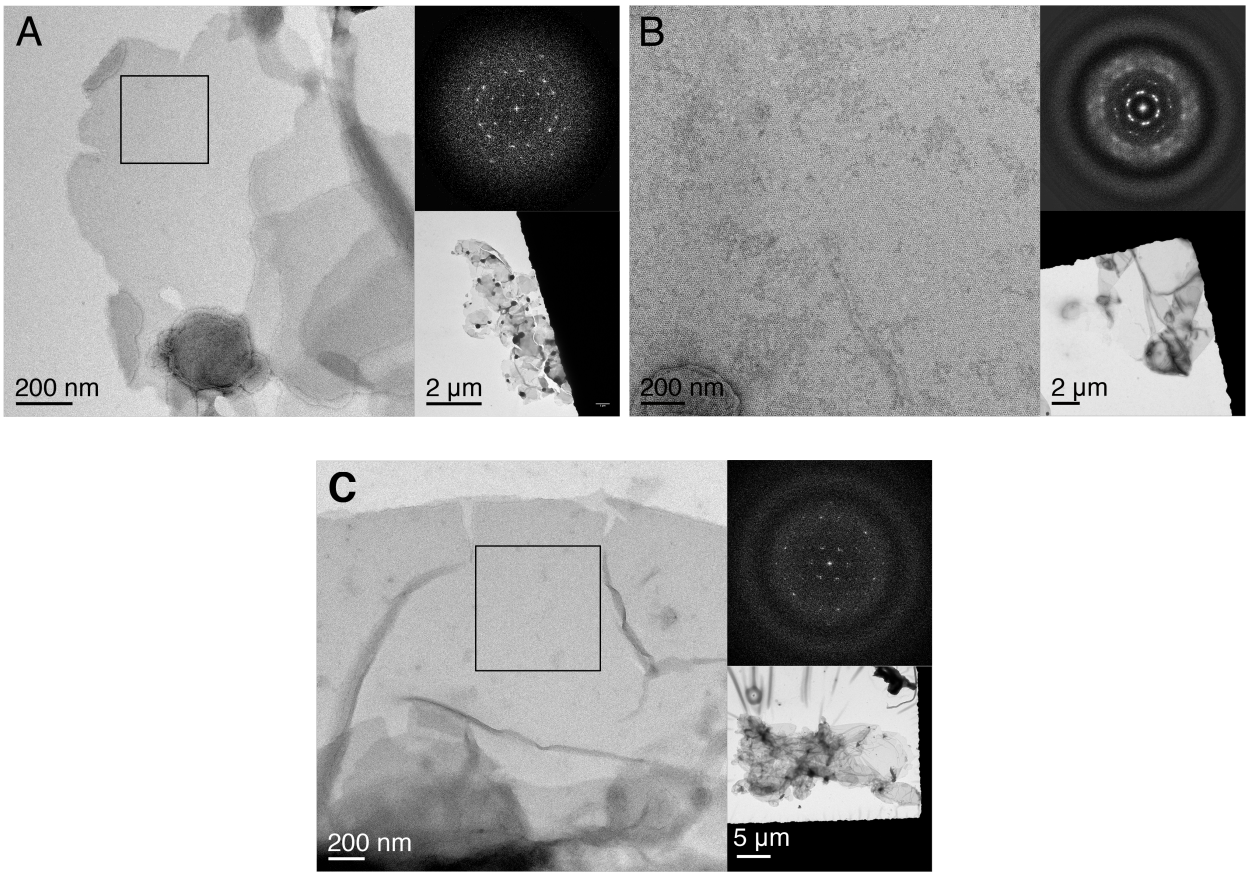

Figure 3.8 GalP-LDAO two-dimensional crystals form with different divalent cations. Two-dimensional crystals of GalP-LDAO formed aggregated crystalline vesicles in the presence of 50 mM $MgCl_2$ in the dialysis buffer (**A, bottom inset**). The crystalline vesicles formed measured about 500 nm in diameter with coherent areas measuring 200 nm (**A**). Fourier transform of a small region of the crystals formed in the presence of $MgCl_2$ showed forth order reflections at a resolution of 20 Å (**A, top inset**). In contrast, dialysis buffer containing 50 mM $CaCl_2$ formed large sheets measuring 10 µm (**B, bottom inset**) and crystalline areas measured at least 2 µm in diameter (**B**). A Fourier transform of the image showed blurred high-resolution spots indicating that the crystal is not coherent throughout the area imaged (**B, top inset**). Large crystalline vesicles measuring more than 5 µm in diameter were obtained at an LPR of 0.2 with dialysis buffer containing 5 mM EGTA (**C, bottom inset**). Vesicles were crystalline throughout but coherent areas measured 400 nm in diameter (**C**) with only weak third order reflections visible in the Fourier transform of the region shown in panel C (**C, top inset**).

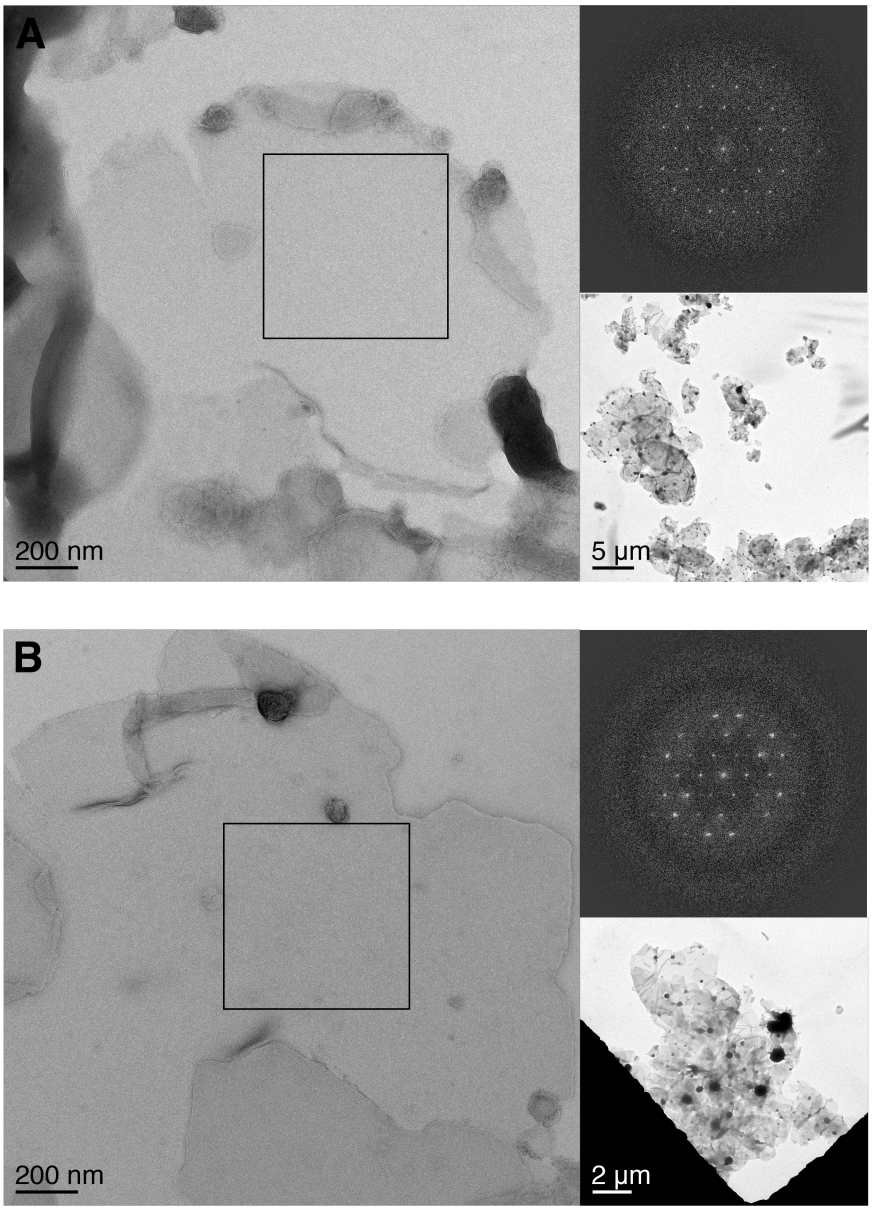

Figure 3.9 Decreasing $MgCl_2$ concentration and fine screen of LPR improved two-dimensional crystals formed using GalP-LDAO. Two-dimensional crystallization of GalP-LDAO using 10 mM $MgCl_2$ in the dialysis buffer significantly improved the crystal morphology producing crystalline vesicles measuring 800 nm in diameter and coherent areas measuring 400 nm (**A**). Fourier transform of a region in the micrograph shown in panel **A** shows several forth order reflections (**A, top inset**) but clustering of vesicles was still an issue (**A, bottom inset**) that limited the amount of areas to image and the crystals were not coherent throughout the area imaged. A fine screen of LPR produced large sheets or vesicle-like sheets at an LPR of 0.24 that had crystalline patches measuring 400 nm in diameter (**B**). Fourier transform of the region shown in panel B revealed only third order reflections indicating that the crystals are not as coherent (**B, top inset**). As with all two-dimensional crystals, clustering of the crystals was observed (**B, bottom inset**).

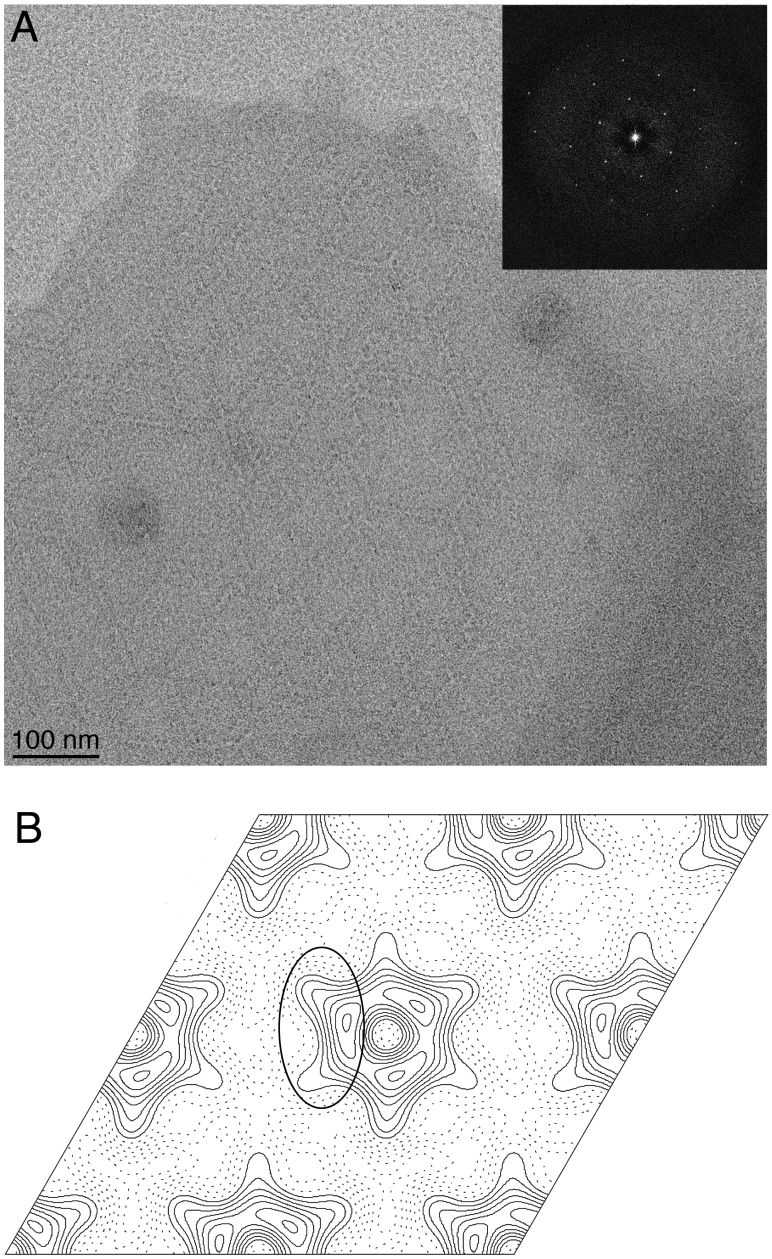

Figure 3.10 Cryo-EM of GalP-OG two-dimensional crystals shows GalP form trimers. GalP-OG crystals formed with DMPC at an LPR of 0.20 with 50 mM $CaCl_2$ in the dialysis buffer were embedded in 10% trehalose and frozen in the microscope (**A**). Due to the clustering of GalP-OG crystals, even under cryogenic conditions, there were only a limited amount of areas that could be imaged. The Fourier transform of the image in panel **A** only shows reflections to 18 Å after unbending procedures carried out using 2dx (42) (**A, inset**). A projection map calculated from the micrograph in panel **A** shows that GalP forms a trimer surrounded by lipids (**B**) with the unit cell dimensions $a = b = 80$ Å and $\gamma = 120^{\circ}$.

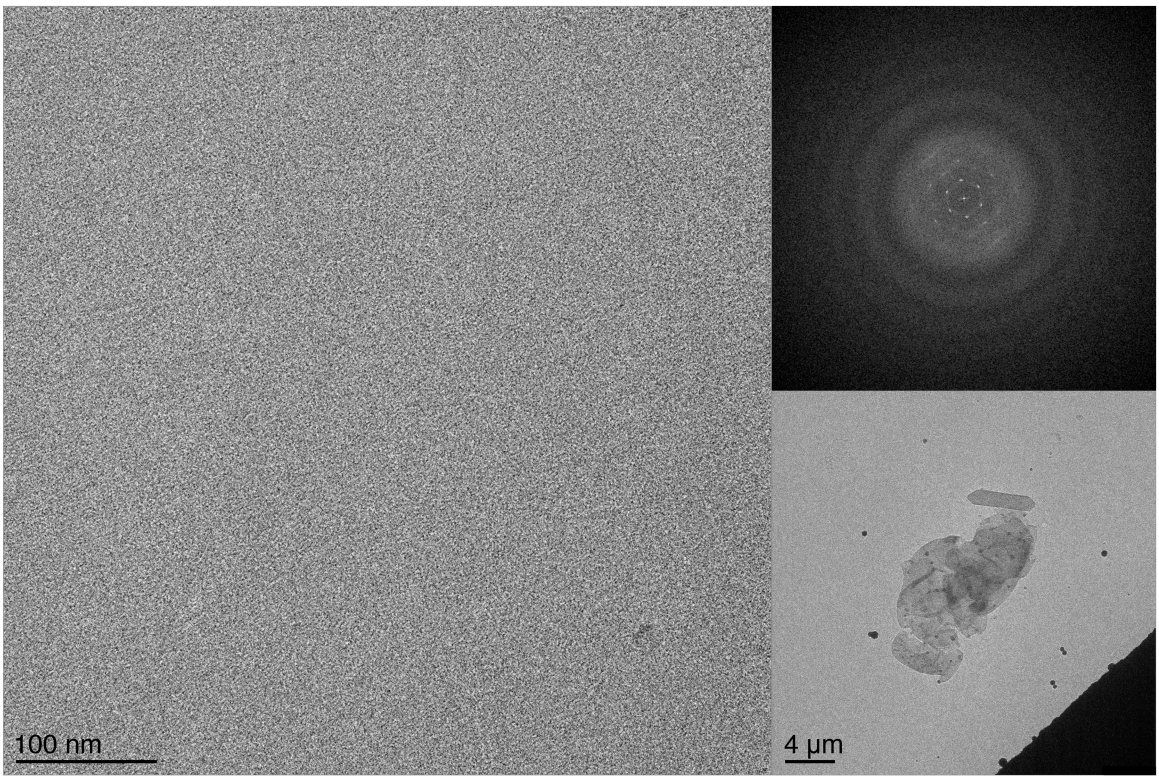

Figure 3.11 Cryo-EM of GalP-LDAO two-dimensional crystals. Two-dimensional crystals of GalP-LDAO with DMPC at an LPR of 0.5 with 50 mM $CaCl_2$ were embedded in 1% trehalose and frozen in liquid nitrogen prior to insertion into the microscope. The crystals formed large vesicles measuring 4 µm in diameter with crystalline patches measuring 500 nm. Fourier transform of the cryo-EM image revealed three orders of reflections to a resolution of 25 Å but the highest resolution spots were smeared indicating that the crystals are either not coherent or that the embedding conditions need to be optimized.

WORKS CITED IN CHAPTER 3

1. Zhao, F. & Keating, A.F. (2007). Functional properties and genomics of glucose transporters. *Curr. Genomics*, **8**, 113-128.

2. Joost, H.G. & Thorens, B. (2001). The extended GLUT-family of sugar/polyol transport facilitators: nomenclature, sequence characteristics and potential function of its novel members (review). *Mol. Membr. Biol.* **18**, 247-256.

3. Wright, E.M. & Turk, E.M. (2004). The sodium/glucose cotransport family SLC5. *Pflugers Arch.* **447**, 510-518.

4. Widdas, W.F. (1988). Old and new concepts of the membrane transport for glucose in cells. *Biochim. Biophys. Acta*, **947**, 385-404.

5. Joost, H.G. *et al*. (2002). Nomenclature of the GLUT/SLC2A family of sugar/polyol transport facilitators. *Am. J. Physiol. Endocrinol. Metab.* **282**, E974-E976.

6. Kasahara, M. & Hinkle, P.C. (1977). Reconstitution and purification of the D-glucose transporter from human erythrocytes. *J. Biol. Chem.* **253**, 7384-7390.

7. Huang, S. & Czech, M.P. (2007). The GLUT4 glucose transporter. *Cell Metab.* **5**, 237-252.

8. Augustin, R. (2010). The protein family of glucose transport facilitators: Its not only about glucose after all. *IUBMB Life,* **62**, 315-333.

9. Gorga, F.R. & Lienhard, G.E. (1982). Changes in the intrinsic fluorescence of the human erythrocyte monosaccharide transporter upon ligand binding. *Biochemistry*, **21**, 1905-1908.

10. Brockmann, K. (2009). The expanding phenotype of GLUT1-deficiency syndrome. *Brain Dev.* **31**, 545-552.

11. Centers for Disease Control and Prevention. (2011). National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

12. Abramson, J. *et al.* (2003). Structure and mechanism of the lactose permease of *Escherichia coli*. *Science*, **301**, 610-615.

13. Dang, S. *et al.* (2010). Structure of a fucose transporter in an outward-open conformation. *Nature*, **467**, 734-738.

14. Sun, L. *et al.* (2012). Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature*, **490**, 361-366.

15. Radestock, S. & Forrest, L.R. (2011). The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *J. Mol. Biol.* **407**, 698-715.

16. Carruthers, A., DeZutter, J., Ganguly, A. & Devaskar, S.U. (2009). Will the original glucose transporter isoform please stand up. *Am. J. Physiol. Endocrinol. Metab.* **297**, E836-E848.

17. Graybill, C., van Hoek, A.N., Desai, D., Carruthers, A.M. & Carruthers, A. (2006). Ultrastructure of human erythrocyte GLUT1. *Biochemistry*, **45**, 8096-8107.

18. Costello, M.J., Escaig, J., Matsushita, K., Viitanen, P.K., Menick, D.R. & Kaback, H.R. (1987). Purified lac permease and cytochrome o oxidase are functional as monomers. *J. Biol. Chem.* **262**, 17072-17082.

19. Dornmair, K., Corin, A.F., Wright, J.K. & Jähnig, F. (1985). The size of the lactose permease derived from rotational diffusion measurements. *EMBO J.* **4**, 3633-3638.

20. Sahin-Toth, M., Lawrence, M.C. & Kaback, H.R. (1994). Properties of permease dimer, a fusion protein containing two lactose permease molecules from Escherichia coli. *Proc. Natl. Acad. Sci. USA*, **91**, 5421-5425.

21. Blodgett, D.M., Graybill, C. & Carruthers, A. (2008). Analysis of glucose topology and structural dynamics. *J. Biol. Chem.* **283**, 36416-36424.

22. Robichaud, T., Appleyard, A.N., Herbert, R.B., Henderson, P.J.F. & Carruthers, A. (2011). Determinants of ligand binding affinity and cooperativity at the GLUT1 endofacial site. *Biochemistry*, **50**, 3137-3148.

23. Henderson, P.J., Giddens, R.A. & Jones-Mortimer, M.C. (1977). Transport of galactose, glucose and their molecular analogues in *Escherichia coli* K12. *Biochem. J.* **162**, 309-320.

24. Macpherson, A.J.S., Jones-Mortimer, M.C., Horne, P. & Henderson, P.J.F. (1983). Identification of the GalP galactose transport protein of *Escherichia coli*. *J. Biol. Chem.* **258**, 4390-4396.

25. Henderson, P.J., Roberts, P.E., Martin, G.E., Seamon, K.B., Walmsley, A.R., Rutherford, N.G. *et al.* (1993). Homologous sugar-transport proteins in microbes and man. *Biochem. Soc. Trans.* **21**, 1002-1006.

26. McDonald, T.P., Walmsley, A.R. & Henderson, P.J.F. (1997). Asparagine 394 in putative helix 11 of the galactose-H^+ symport protein (GalP) from *Eserichia coli* is associated

with the internal binding site for cytochalasin B and sugar. *J. Biol. Chem.* **272**, 15189-15199.

27. Martin, G.E.M, Seamon, K.B., Brown, F.M., Shanahan, M.F., Roberts, P.E. & Henderson, P.J.F. (1994). Forskolin specifically inhibits the bacterial galactose-H^+ transport protein, GalP. *J. Biol. Chem.* **269**, 24870-24877.

28. Martin, G.E.M, Rutherford, N.G., Henderson, P.J.F. & Walmsley, A.R. (1995). Kinetics and thermodynamics of the binding of forskolin to the galactose-H^+ transport protein, GalP, of *Escherichia coli*. *Biochem. J.* **308**, 261-268.

29. Bloch, R. (1973). Inhibition of sugar transport in the human erythrocyte by cytochalasin B. *Biochemistry*, **12**, 4799-4801.

30. Sergeant, S. & Kim, H.D. (1985). Inhibition of 3-O-methylglucose transport in human erythrocytes by forskolin. *J. Biol. Chem.* **260**, 14677-14682.

31. Zheng, H., Taraska, J., Merz, AJ. & Gonen, T. (2010). The prototypical H^+/galactose symporter GalP assembles into functional trimers. *J. Mol. Biol.* **396**, 593-601.

32. le Maire, M., Champeil, P. & Moller, J.V. (2000). Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim. Biophys. Acta*, **1508**, 86-111.

33. Kaufmann, T.C., Engel, A. & Remigy, H.W. (2006). A novel method for detergent concentration. *Biophys. J.* **90**, 310-317.

34. Lugtenberg, E.J.J. & Peters, R. (1976). Distribution of lipids in cytoplasmic and outer membranes of Escherichia coli K12. *Biochim. Biophys. Acta*, **441**, 38-47.

35. Oursei, D., Loutelier-Bourhis, C., Orange, N., Chevalier, S., Norris, V. & Lange, C.M. (2007). Lipid composition of membranes of Escherichia coli by liquid chromatography/tandem mass spectrometry using negative electrospray ionization. *Rapid Commu. Mass Spectrom.* **21**, 1721-1728.

36. Almen, M.S., Nordstrom, K.J.V., Frederiksson, R. & Schioth, H.B. (2009). Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biology*, **7**, 50.

37. Mueckler, M. *et al.* (1985). Sequence and structure of a human glucose transporter. *Science*, **229**, 941-945.

38. Lowe, A.G. & Walmsley, A.R. (1986). The kinetics of glucose transport in human red blood cells. *Biochim. Biophys. Acta*, **857**, 146-154.

39. Bloch, R. (1973). Human erythrocyte sugar transporter: kinetic evidence for an asymmetric carrier. *J. Biol. Chem.* **249**, 3543-3550.

40. Bogdanov, M., Xie, J., Heacock, P. & Dowhan, W. (2008). To flip or not to flip: lipid-protein charge interactions are a determinant of final membrane protein topology. *J. Cell Biol.*, **182**, 925-935.

41. Vitrac, H., Bogdanov, M. & Dowhan, W. (2013). Proper fatty acid composition rather than an ionizable lipid amine is required for full transport function of lactose permease from *Escherichia coli*. *J. Biol. Chem.* **288**, 5873-5885.

42. Gipson, B., Zeng, X., Zhang, Z.Y. & Stahlber, H. (2006). 2dx – User-friendly image processing for 2D crystals. *J. Struct. Biol.* **157**, 64-72.

43. Gipson, B., Zeng, X. & Stahlberg, H. (2007). 2dx_merge: Data management and merging of 2D crystal images. *J. Struct. Biol.* **160**, 375-384.

44. Crowther, R.A., Henderson, R. & Smith, J.M. (1996). MRC image processing programs. *J. Struct. Biol.* **116**, 9-16.

45. Pettersen, E. F. *et al*. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-12.

46. Guan, L., Mirza, O., Verner, G., Iwata, S. & Kaback, H.R. (2007). Structural determination of wild-type lactose permease. *Proc. Natl. Acad. Sci. USA*, **104**, 15294-15298.

Chapter IV

The N-terminal domain of αB-crystallin provides a conformational switch for multimerization and structural heterogeneity

Stefan Jehle[1], **Breanna Vollmar[1]**, Benjamin Bardiaux[3], Katja K. Dove[1], Ponni Rajagopal[1], Tamir Gonen[1,2], Hartmut Oschkinat[3,4], & Rachel E. Klevit[1]

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA
[2]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA
[3]Leibnizinstitut Für Molekulare Pharmakologie, Robert-Rössle-Straße 10, 13125 Berlin, Germany 4Freie Universität Berlin, Takustraße 3, 14195 Berlin, Germany

This work was published in the journal of the Proceedings of the National Academy of Science USA (2011), volume 108, pages 6409-6414.

SUMMARY

The small heat shock protein (sHSP) αB-crystallin (αB) plays a key role in the cellular protection system against stress. For decades, high-resolution structural studies on heterogeneous sHSPs have been confounded by the polydisperse nature of αB oligomers. We present an atomic-level model of full-length αB as a symmetric 24-subunit multimer based on solid-state NMR, small-angle X-ray scattering (SAXS), and electron microscopy (EM) data. The model builds on our recently reported structure of the homodimeric α-crystallin domain (ACD) and C-terminal IXI-motif in the context of the multimer. A hierarchy of interactions contributes to build multimers of varying sizes: interactions between two ACDs define a dimer, three dimers connected by their C-terminal regions define a hexameric unit, and variable interactions involving the N-terminal region define higher-order multimers. Within a multimer, N-terminal regions exist in multiple environments, contributing to the heterogeneity observed by NMR. Analysis of SAXS data allows determination of a heterogeneity parameter for this type of system. A mechanism of multimerization into higher order asymmetric oligomers via the

addition of up to six dimeric units to a 24-mer is proposed. The proposed asymmetric multimers explain the homogeneous appearance of αB in negative-stain EM images and the known dynamic exchange of αB subunits. The model of αB provides a structural basis on which to understand known disease-associated missense mutations and makes predictions concerning substrate binding and the reported fibrilogenesis of αB.

INTRODUCTION

Small heat shock proteins (sHSP) help to maintain protein homeostasis by interacting with partly folded substrates to prevent cell damage (1-3). The ATP-independent chaperone αB-crystallin (αB, 20 kDa, 175 residues) is an archetypal example (4). Discovered as a highly abundant protein in the eye lens that plays a critical role in maintenance of lens transparency, the known biological roles of αB continue to expand. The protein is expressed in many tissue types, notably muscle and brain, in a stress-inducible manner, where it presumably serves as a chaperone for misfolded cellular proteins. Consistent with such a role, αB is implicated in a growing number of diseases that includes cardiac myopathies and neurodegenerative diseases such as Alexander's and Alzheimer's disease (5-7). Furthermore, αB has been shown to play a protective role and can reverse symptoms of multiple sclerosis (8). Thus, a full structural description of αB is an important step towards understanding its mode(s) of action. Past models of **α**B have been based on biochemical data (9); however recent advances in structural biology of sHSPs (9-11) allow for a more detailed understanding of the assembly of αB multimers.

As for all sHSPs, αB is organized in three domains (Fig. 4.1A): (*i*) an N-terminal domain of approximately 60 residues, (*ii*) a central α-crystallin domain (ACD) of about 90 residues involved in dimerization (Fig. 4.2A), and (*iii*) a C-terminal domain of 25 residues containing the

IXI-motif, usually comprised of two Ile residues separated by an intervening residue, that is highly conserved in sHSPs. Recent studies using solid-state NMR and X-ray crystallography have yielded the first atomic-level structures of ACDs from polydisperse mammalian sHSPs, which have been refractory to structure determination (10-12). We reported a structure determined from solid-state NMR measurements on full-length αB in which a highly curved homodimer comprised of two ACDs forms the basic building block of multimers (10). A recent EM structure of negatively stained αB showed a tetrahedrally symmetric oligomer representing the shape of a 24-mer (13). This arrangement could be recapitulated from the curved dimer structure solved by solid-state NMR and an approximate oligomer shape from small-angle X-ray scattering (SAXS) (10). αB is a highly dynamic species in which multimers ranging from 24 to 32 subunits coexist and exchange rapidly (14), and although the aforementioned structures represent a significant advance in αB structural biology, they do not address the hallmark heterogeneity of αB and other sHSPs. Moreover, structural aspects of the N-terminal domain of αB remain poorly understood.

Here we present a model of the N-terminal region (residues 1-65) based on solid-state NMR restraints and similar fragments from proteins in the protein data bank (PDB). Our analysis identifies two β-strands that exist in multiple structural environments. We present a model for full-length αB as a tetrahedrally symmetrical 24mer based on the published EM map of αB (13). Next we use SAXS to measure αB heterogeneity in solution and together with new EM analysis we propose a model in which additional dimeric units can fill existing openings in a 24-mer to create higher-order multimers that look alike in negatively-stained preparations in electron microscopy.

RESULTS AND DISCUSSION

The N- and C-termini of αB are highly flexible

The domain organization of αB is shown schematically in FIG. 1*A*. Solution-state NMR measurements on αB oligomers have revealed that the first 5 N-terminal and last 10-12 C-terminal residues are flexible (15, 16). The methyl resonance region of a 2D ^{1}H-^{13}C INEPT MAS solid-state NMR spectrum of an αB preparation precipitated with PEG8000 contains a number of intense, sharp peaks that arise from flexible residues (Fig. 4.1B). Based on chemical shift similarity with the solution-state NMR spectra, we assign the resonances to Met1 and the methyl-containing side chains at the extreme C-terminus of αB (Ala168, Val169, Thr170, Ala171, Ala172). This region of the INEPT spectrum also contains distinct but weaker signals in positions consistent with isoleucines, likely from the extreme N-terminal region (Ile3, Ile5, and potentially Ile10). These observations indicate that the highly flexible termini observed in solution are also flexible in αB preparations precipitated with PEG8000.

Secondary Structure within the N-terminal domain of αB

Compared to the well-defined ACD, most residues in the N-terminal domain have low signal intensity and chemical shift dispersion in solid-state NMR spectra. Multiple chemical shift sets for individual residues complicated the assignment and detection of structural restraints. Nevertheless, we were able to assign resonances for a majority of residues in the N-terminal domain (Biological Magnetic Resonance Bank entry 16391) (17, 18). Backbone resonance chemical shifts analyzed using TALOS (19) predict β-strand structure for residues Leu44-Tyr48 and Ser59-Thr63. Distance restraints observed between the ^{13}Cα resonances of Tyr48 and Thr63 and between the ^{13}Cα resonance of Leu49 and the ^{13}Cα resonance of Asp62 and Thr63 and ^{13}Cβ

resonance of Phe61 further corroborate the prediction and indicate an antiparallel orientation between the two strands (Fig. 4.3A and Table 4.1). Although the chemical shift analysis did not yield a high confidence prediction of other regular secondary structure in the N-terminal region, eight distance restraints observed in 3D NCACX and NCOCX spectra indicate that residues 14-17 and 27-32 adopt helical conformations, as typical (*i*, *i*+3) and (*i*, *i*+4) correlations could be identified for these residues (illustrated in Figure 4.3A and summarized in Table 4.1).

Modeling the N-terminal domain of αB

To augment the sparse experimental restraints observed for the N-terminal domain, we performed a sequence similarity search for αB residues 1-65 against proteins in the PDB to detect fragments with known structures. Three significant matches were identified (Fig. 4.3B and Table 4.2). The longest match was for αB residues 12-66 with residues 12-62 of acetyl xylan esterase from *Thermotoga maritima* (PDB 1vlq, 47% similarity, Table 4.2). Notably, the esterase structure contains β-strands that align with the two predicted strands in αB. In the esterase, the strands form a two-stranded antiparallel sheet connected by a long loop. Esterase residues 23-37 form an α-helix and align with αB residues 23-37, which give helical distance restraints. Two shorter αB sequences gave significant similarity scores: with residues 5-27 of 2'-specific/double-stranded RNA-activated interferon-induced antiviral protein 2'-5'-oligoadenylate synthetase (PDB 1px5, 65% similarity, Table 4.2) and with residues 25-48 of methyltransferase-fold protein from *Erwinia carotovora atroseptica* (PDB 2p7h, 54% similarity, Table 4.2). Helical secondary structure is predicted for αB residues 19-38 based on the alignment with a fragment of the synthetase, corroborating the prediction based on the esterase. αB residues 2-25 have 54% similarity with N-terminal residues of the methyltransferase fold protein. Taken together, the

solid-state NMR observations and sequence alignments are consistent with the N-terminal domain containing two helical segments and an antiparallel β1-loop-β2 structure comprised of residues 44-65.

The heterogeneity of NMR signals observed for the N-terminal region indicates that the structures described above do not all exist simultaneously in the same environment in all copies of αB subunits in all multimers. For simplicity, a model of the N-terminal region that includes all these structural features was generated by fusing the relevant fragments of the esterase and the methyltransferase-fold protein, as shown in Figure 4.3C. A fragment (residues 1-20) that contains helix α1 and a fragment containing α2, β1, and β2 (residues 21-65) were modeled based on 2p7h and 1vlq (Fig. 4.4), respectively. The resulting fragments were connected and energy minimized using the solid-state NMR restraints with Discovery Studio (Accelrys).

Building a multimer model

Our previous NMR studies defined the ACD homodimer as the basic building block of αB oligomers. Three dimers form a triangular array on the surface of a multimer, each connected via its IXI-motif bound in the groove between the β4 and β8 strands of a neighboring dimer (Figs. 4.5A, 4.2B and C) (10). A similar arrangement is observed in the octahedrally symmetrical oligomer of sHSP 16.5 from *Methanococcus janashii* (Fig. 4.2D) (20), suggesting that a triangular arrangement of dimers is a conserved structural motif among sHSPs.

Four copies of triangular hexamers comprised of the ACD and the C-terminal region were fitted into the published EM density of αB (EM Databank entry EMD-1776) (13), which exhibits tetrahedral symmetry (Fig. 4.5A). This arrangement results in four threefold axes and three twofold axes (Fig. 4.5A). Views onto the threefold and twofold axes each revealed density

in the EM map that was not yet accounted for that likely originates from N-terminal domain residues (boxed region and indicated with arrows, respectively, in Fig. 4.5A). In the esterase structure (1vlq), the loop connecting β1 and β2 contacts another subunit's loop, forming a ring-like assembly within the multimeric structure (Fig. 4.4). αB constructs lacking the N-terminal domain (e.g. αB 69-175) fail to form large multimers, consistent with involvement of N-terminal residues in multimer formation (21, 22). We therefore positioned the β1-loop-β2 model comprised of αB residues Ser45-Leu65 in an orientation analogous to the esterase into empty EM density observed in the twofold axis (Fig. 4.5B).

Placement of the β1/β2 loops into the 24-mer as described above results in two environments for this structural element (shown schematically in Figure 4.6A). β1/β2 of one subunit within a homodimer contacts the β3 strand of its own ACD while the other β1/β2 crosses a two-fold axis; its β2 does not contact a β3. We distinguish these as $\beta2_{intra}$ and $\beta2_{free}$, respectively (Fig. 4.5B). A third environment for β1/β2, called $\beta2_{inter}$, arises in multimers with greater than 24 subunits (described below). We refer to β2 as originally defined from X-ray structures (11, 12). However, our NMR restraints and the sequence similarities with other known structures suggest that $\beta2_{intra}$ is longer (i.e. spanning residues 60-69) while $\beta2_{free}$ and $\beta2_{inter}$ are shorter, (spanning residues 60-63) with residues 66-72 serving as linker between β2 and β3. Met68 and adjacent residues that are part of the linker will therefore exist in variable structural environments, consistent with the heterogeneity observed for Met68 resonances in solid-state NMR spectra of human αB (17). Heteronuclear NOEs measured in solution on a truncated ACD dimer indicated flexibility for these residues (17) and the β2 strand is present in only one subunit of a dimer in an X-ray structure of a truncated ACD dimer (11). In published crystallographic studies of αB ACDs, protein constructs had to be truncated at Met68 to obtain crystals (11, 12).

Variability in β2 is also a feature in other sHSP structures: the ACD from *Methanococcus janashii* sHSP16.5 has a β2 and a β1 strand (20); sHSP16.9 from wheat lacks a β1 strand (23); and Tsp36 from tapeworm has shorter β2 strands compared to sHSP16.5 and sHSP16.9 (24). In summary, this segment appears to adopt multiple structures in multiple environments, thereby contributing to αB's inherent heterogeneity.

The 24-mer model positions two β1/β2 loops in close proximity (Fig. 4.5B). To test the model's validity, single cysteine residues were substituted at positions within the β1-loop-β2 sequence to ascertain whether disulfides can form within oligomeric αB samples (αB contains no Cys residues in its sequence). Crosslinks are formed within multimers comprised of αB mutants with Cys substitution at Ala57, Ser59, or Tyr63, but not at a position within the ACD, Val145 (Fig. 4.7). This finding corroborates the proximity of two β1/β2 loops within an αB multimer connecting two dimers with each other, comparable to the esterase structure used to model these loops.

Positioning of the β1/β2 loops as described above left unoccupied density in the EM map into which residues 1-44 were fitted. The density is observed when looking onto the threefold axis defined by a hexamer, but viewed from the far side (Fig. 4.5A, center). The density is low, suggesting occupancy by loosely packed and/or flexible residues. N-terminal deletion constructs of the closely related αA-crystallin that lack the first 35 residues form multimers of a size comparable to the full-length protein (25). Together with the observation that αB constructs lacking the entire N-terminal domain do not form large multimers (21,22), we assign residues in the second half of the N-terminal domain of αB as important for the formation of multimers. Two fragments, each consisting of α1 and α2 (see Figure 4.3), were docked to each other assigning hydrophobic residues in the more C-terminal helix (Leu32, Leu33, Leu37, and Phe38)

as interacting residues in HADDOCK. Three copies of the resulting structure, each comprised of two copies of residues 10-39 with intermolecular interactions involving residues in α2, were modeled into the empty density on the threefold axis in the 24-mer (Fig. 4.5C). The arrangement places the hydrophilic segment Thr40-Ser41-Thr42-Ser43 on the surface of the 24-mer, with α2 residing in the shell of the EM map. Residues 1-44 are likely to be dynamic due to exchange of subunits among the ensemble of multimers and the model should be viewed as a transient rather than static model.

Finally, flexible residues of the N and C termini identified by NMR were included in the model (Fig. 4.5D,E). Because of different orientations of the ACDs, and therefore the IXI motifs bound to them, half the subunits in a 24-mer have their C-terminal residues 166-175 on the surface and half have them pointing to the interior (Fig. 4.5D). Residues 1-10, including parts of putative α1, were loosely packed in the interior to produce a model that fits radius of gyration measurements by SAXS (see below). Figures 4.5E and 4.8C shows a stereoview of an αB 24-mer comprised of full-length subunits. A central hollow with an approximately 4 nm diameter is surrounded by flexible residues from the N and C termini.

Heterogeneity of αB

SAXS data measured at pH 7.5 were used to assess the 24-mer model. The experimental curve was compared to the curve calculated for the 24-mer (Fig. 4.9A). The calculated SAXS curve is typical for a hollow particle, consistent with the model that has an inner hollow diameter of 4 nm and a 4 nm-wide ring occupied by flexible and/or disordered N and C termini (Fig. 4.9B). The radius of gyration measured by SAXS is consistent with the model. The hollow feature observed in two previous EM studies (13, 26) is not recapitulated in the experimental

SAXS curve. However, Haley et al later postulated variable density in the center of the αB oligomer (27).

The SAXS curve calculated for the 24-mer model fits the experimental curve with a χ^2 of 10.8. Heterogeneity of a particle's diameter and shape can modulate the observed SAXS profile of an ensemble of particles. To generate a better fit to the data, we used an algorithm that allows determination of heterogeneity of an ensemble of colloidal particles whose radii vary from SAXS curves. The algorithm essentially modulates the SAXS curve predicted for a homogeneous species by assuming a Schulz distribution for the particles' radii (28). Introduction of a "degree of polydispersity" of 16% produced a fit to the experimental αB SAXS curve with a χ^2 of 1.0. In the case of αB, this heterogeneity parameter cannot be interpreted directly as a distribution of the particles' radii per se, because both size and shape polydispersity are likely to be present simultaneously. Notably, shape polydispersity has been shown to result in overestimation of size polydispersity (29). Nevertheless, we have included the estimated Schulz distribution of the outer particle's radius in Figure 4.8D. The Schulz distribution of the radius (5.5 - 9.5 nm) is in good agreement with the radius determined by cryo-EM (5.5 - 8 nm; 26) taking the 3.6 nm resolution of the cryo-EM data into account. The inferred variable radii of αB multimers could result from an expansion due to incorporation of additional subunits and/or from dynamic C- and N-terminal extensions in the interior of the oligomer. The N- and C-terminal residues can undergo dynamic fluctuations to produce a variable shell thickness within multimers, resulting in a "breathing" of the spherical particle (represented in Figure 4.9B between the inner and dashed circle). Internal flexibility of dimers within a multimer and the exchange of subunits can also alter the apparent size of the particle, with the SAXS data thus producing a static "average" image of a heterogeneous population. A cavity with *ca*. 8 nm

diameter was determined from EM density maps calculated using single particle reconstruction (13, 26). In this process, particles are averaged to reconstruct a structure, so density from flexible or disordered regions is averaged out, producing an apparently larger cavity.

Modeling higher-order multimers of αB

It is well established that αB exists as a heterogeneous population of multimers with a variable number of subunits (14). A 24-mer is the simplest species to model due to its symmetry characteristics, but it represents only about 5% of multimeric species as detected by mass-spectrometry (14). Therefore, a realistic view of αB assemblies must include higher-order multimers. Given our earlier results showing a dimer to be the basic building block, formation of higher order multimers likely occurs via incorporation of dimers into existing openings in the shell of the 24-mer. The EM density contains six gaps with dimensions that can accommodate a dimer at the edges of the three two-fold axes. Thus, a 26-mer would contain one additional dimer in any one of these openings; a 28-mer would contain two dimers, distributed randomly (Fig. 4.10A), and so forth. Figures 4.10B and C show a 28-mer with two dimers (green and yellow chains) residing in the edge openings of one twofold axis (indicated with a dotted line and arrows at the edges), viewed onto that twofold axis. Viewed along this twofold axis, the same 28-mer is seen to have its additional dimers in the central openings, one at the front and one at the back of this view (Fig. 4.10D,E). We previously reported inter-subunit contacts between β2 residues Ser59-Phe61 and a β3 strand in the ACD, observed by NMR (10). This inter-subunit β2 contact is not satisfied in the 24-mer model and must therefore exist in other species. Accordingly, two $\beta2_{free}$ strands from two dimer units in a 24-mer crossing an unoccupied twofold axis become $\beta2_{inter}$ strands by contacting β3 strands of a new ACD dimer, integrating the new dimeric unit

into a larger multimer (Fig. 4.6B). A dimer integrated in this manner can participate in additional intermolecular interactions with its N-terminal and C-terminal domain (Fig. 4.10E). This mode of multimerization allows the formation of multimers of up to 36 subunits, with 24-mers serving as a conserved scaffold into which six additional dimers can be randomly distributed. Given the dynamic behavior of the regions involved in forming higher-order multimers, it is likely that αB multimers undergo exchange so that dimers and/or monomers that comprise the 24-mer scaffold can be replaced by dimers in a twofold axis and vice versa.

Electron microscopy and single particle analysis of negatively-stained αB

At the surface, the heterogeneity of αB observed in our solid-state NMR and SAXS data appears at odds with recent 3D reconstructions derived from negative-stain EM studies (13). We therefore analyzed our heterogeneous preparation of αB by negative-stain electron microscopy. In a representative image of the preparation, particles appear homogeneous in size and are evenly dispersed on the grid (Fig. 4.11A). Projection averages were calculated from 4,398 selected particles in SPIDER (30). Six highly populated class averages are presented as insets. The particles measure approximately 13.6 nm in diameter, with some averages displaying a cavity approximately 4nm in diameter. These projection averages are comparable to recently published data (Fig. 4.11B) (13). Back projections calculated from the published map of αB (EMD-1776) (13) appear comparable to class averages calculated from our data (compare Figure 4.11A inset with 4.11C). Back projections were also calculated for our models of αB multimers containing 24, 26, and 28 subunits (Fig. 4.11C and Fig. 4.12A-C). The calculated projections are similar to each other and to those calculated from the published EM density. Our analysis indicates that single particle reconstructions of negatively-stained preparations cannot unambiguously

distinguish among several species that are likely to coexist in a heterogeneous population of αB. Thus, although the heterogeneity observed in solid-state NMR data and SAXS data of αB appear to be in conflict with the EM studies [both ours and previously published (13)], it is clear that the size deviations of αB multimers calculated from SAXS data (Fig. 4.9 and Fig. 4.8D) fall below the approximately 20 Å-resolution limit of negative staining.

CONCLUDING REMARKS

The model based on NMR observations reveals four regions of αB that are involved in its multimerization. First, two β6+7 strands within the ACD align to form a homodimer, the smallest building block (Fig. 4.2A) (10). Second, the conserved IXI motif from the C terminus of one dimer binds to hydrophobic pockets formed by the β4 and β8 strands on the edge of another dimer (Fig. 4.2B). This interaction creates a triangular array of three dimers that defines the threefold axis (Fig. 4.2C). Third, the β2 strand is involved in variable interactions: β2 can contact β3 intramolecularly or intermolecularly (Figs. 4.5 and 4.10). Fourth, N-terminal residues interact at the edge of the three-fold axis. Thus, the ACD is responsible for dimer formation, the ACD dimer and the C-terminal regions are responsible for defining hexameric units, and the N-terminal region and ACD are responsible for higher-order multimerization. The regions involved in multimerization in the model are in good agreement with sequences identified by peptide-binding assays as involved in αB self-assembly: Leu37-Pro58, Phe75-Lys82 (β3), Leu131-Ser139 (β8), Gly141-Pro148 (β9), and Pro158-Glu165 (containing the IXI motif) (31).

The ACD dimer has a twisted conformation, which results in two orientations of the six subunits within a hexameric unit. The β4/β8 substrate-binding groove at the edge of each ACD, occluded by the C-terminal IXI motif, faces toward the outside in half the subunits and to the

inside of a multimer in the other half (Fig. 4.8A,B). Dissociation of an IXI motif from the groove in response to activation of αB by pH (10) would presumably occur primarily in the outward-facing copies, leaving half the interactions in place to retain the hexameric unit. Virtually all residues of the protein are solvent accessible somewhere in the multimer, with the exception of the extreme N-terminus and helices α1 and α2, which are packed in the density at the edge of a threefold axis or tucked on the inside. In this context, it is notable that the N-terminal domain itself forms disordered aggregates (22, 25), so sequestration of this region in multimers may be important to prevent stochastic oligomerization. The observation that all other regions of αB are accessible in a multimer may explain why almost all regions have been implicated in the binding of one or more client proteins (31, 32). Of note, the high abundance of known phosphorylation sites (33) at one edge of the three-fold axis suggests the presence of a binding interface for kinases in this region (Fig. 4.13).

The model provides structural context for known mutations in αB associated with its dysfunction and/or disease. The best characterized of these is R120G-αB (5); Arg120 from one subunit interacts with Asp109 from the other subunit across the ACD dimer interface so its substitution may affect αB structure at the level of the dimer. The D140N mutation (34) could affect a putative intermolecular salt-bridge with Arg56 in the N-terminal region, thereby affecting higher-order structure (35). The early truncation mutant Q151X (36) lacks the IXI motif, which plays a role in the formation of hexamers and the extreme C terminus that is responsible for the high solubility of αB oligomers. Furthermore, variable intra- and intermolecular interactions of β-strands, a hallmark of amyloid-forming proteins (37, 38), provide an explanation for the propensity of fibril formation of αB (16).

In conclusion, αB is a highly dynamic assembly with flexible termini that contribute

substantially to its heterogeneous appearance. The model we present here suggests a simple mechanism by which αB dimeric units serve as building blocks that can add to a symmetric 24-mer assembly to form higher-order multimers with minimal conformational change. This model for assembly therefore also allows for facile exchange of subunits with minimal energy. The interactions that define higher-order multimers are all the same and are essentially the same as those that define the 24-mer, consistent with the broad distribution of species observed and indicative of equivalent stabilities. Although the resulting ensemble of multimers cannot be distinguished at the level of EM analysis on negatively stained preparations, the presence of multimers with varying symmetries presumably contributes to the prevention of crystal- and high-molecular weight aggregate formation, an essential property for biological functions of αB which occurs in high concentrations in the eye lens.

MATERIALS AND METHODS

Sample Preparation and NMR/SAXS data collection

αB was expressed, purified and solid-state NMR spectra were recorded as described elsewhere (10, 17). SAXS data were collected at the European Molecular Biology Laboratory SAXS X33 beamline, Hamburg, Germany.

Preparation of mutants

Single cysteine mutations were introduced using QuikChange™ Site-Directed Mutagenesis (Strategene). Mutations were confirmed by DNA sequencing.

Crosslinking

Samples containing 20μM of WT, A57C, S59C, T63C, or V145C αB in 50mM sodium phosphate buffer (pH7.5) with 100mM sodium chloride were incubated for 30 min on ice. Subsequently, samples were visualized by non-reducing 12% SDS-PAGE. Iodoacetamide (150 mM; Sigma) was added to the SDS-PAGE loading buffer to inhibit disulfide formation during sample preparation for gel electrophoresis.

Analysis of SAXS data

Fits to the experimental SAXS curve were performed using CRYSOL from the ATSAS program suite (39). The χ^2 values quoted represent the goodness of fit. Modulation of the theoretical SAXS curve according to the algorithm of Mittelbach and Glatter (28) was performed by using an in-house written python script that calculates an I(q) function modulated by a user defined amount of heterogeneity from an I(q) function.

Modeling of multimers

Three dimers connected by their C-terminal IXI motif were fitted in the EM-density (EMD-1776) using CHIMERA (40). A hexamer was built of model 7 from PDB entry 2klr using non-crystallographic symmetry restraints given in the PDB file. Model 7 was chosen because it is closest to the average structure. Dimers were connected by their C-terminal domain containing the IXI-motif and the linker (residues Q151-P155) was energy minimized (10). The $\beta2_{inter}$-strand was docked to the β3 strand using HADDOCK (41) by defining the residues in β3, (R120, Y122, D109, R107) and in $\beta2_{inter}$ as residues, which interact with each other. The $\beta1/\beta2_{free}$ structural segment was placed into the EM density to fulfill the crosslinks for A57, S59, and T63, similar to

structure 1vlq. The α2 helices of two monomers were docked defining residues Leu32, Leu33, Leu37 and Phe38 as interacting residues using HADDOCK. Results from HADDOCK were selected according to a low rmsd and for the feasibility to connect the docked complexes with their respective N-terminal partners using the residues T40 through S43 as flexible linkers. The flexible C- and N-terminal residues including parts of helix α1 were placed in the inside of the 24-mer to fit the radius of gyration measured by SAXS in an iterative process. All fragments were connected using Discovery Studio 1.6 (Accelrys) and the linkers were energy minimized. After the complete assembly of the 24-mer, a local energy minimization with Discovery Studio (Accelrys) was performed to remove intermolecular clashes. In Fig. 14, we compare the fit of an archael ACD dimer (1shs) and the curved human ACD dimer (2klr) in the EM map of human **α**B. Coordinates for the model of the 24-mer are available at http://faculty.washington.edu/klevit/Protein_Portraits.html.

Electron microscopy

A 2μl drop of purified αB at pH 7.5 in 50 mM sodium phosphate, 100 mM NaCl was applied to freshly glow-discharged electron microscopy grid and stained with 2% uranyl acetate. The preparation was viewed on a 100kV transmission electron microscope (Morgagni M268, FEI, Hillsboro, OR) and images recorded at a nominal magnification of 67,000x at the specimen level using a bottom-mount 4k x 2k Gatan charge coupled device camera, corresponding to a pixel size of 1.34Å. A total of 4,398 particles were selected in WEB and processed in SPIDER (30). Class averages were generated after 8 cycles of reference free multivariant statistical analysis in SPIDER (30). Back projections of the different oligomeric models of αB and of the published EM-density map (13) were calculated using SPIDER (30).

ACKNOWLEDGEMENTS

This work was funded by NIH grant 1R01 EY017370 (to R. E. K.). SAXS data collection was supported by European Community - EMBL Hamburg Outstation; DESY Hamburg X33 beamline. Support for the electron cryomicroscopy laboratory is from Murdock Charitable Trust and Washington Research Foundation. B.V. is supported in part by NIGMS 2T32 GM007270. T.G. is supported by American Diabetes Association (Award # 1-09-CD-05), NIH (R01 GM079233), and by Howard Hughes Medical Institute Early Career Scientist award. The authors thank Profs. J. Buchner and S. Weinkauf, and Dr. N. Braun for providing their EM map prior to deposition. The authors declare no competing financial interests.

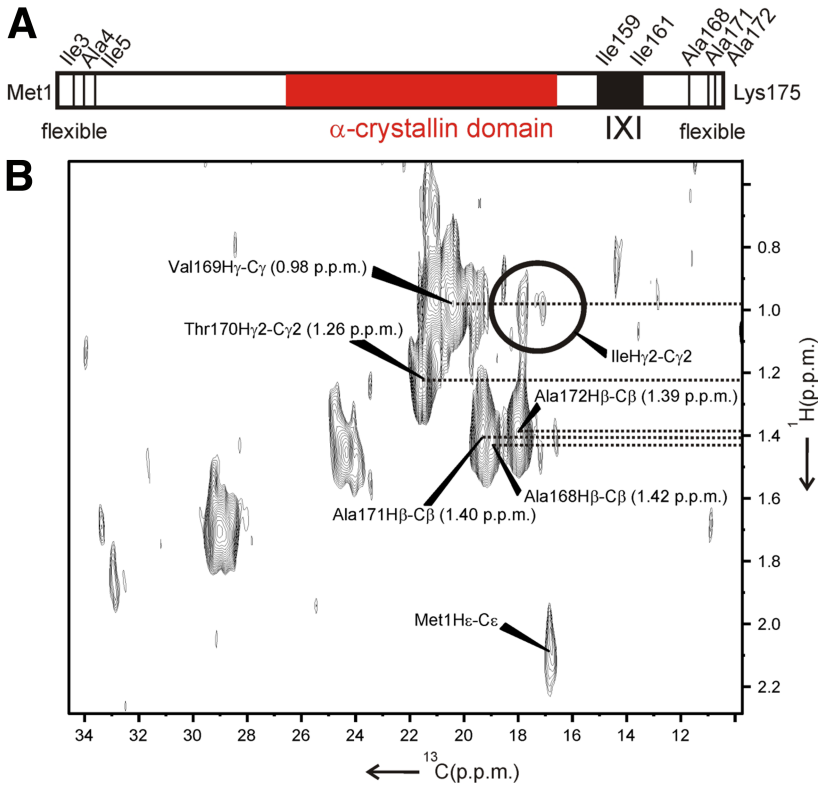

Figure 4.1 N-terminal domain of αB crystallin. (**A**) Domain architecture of αB: an N-terminal domain essential for oligomerization, an α-crystallin domain involved in dimerization, and a C-terminal domain containing the conserved IXI-motif. (**B**) 2D ^{1}H-^{13}C INEPT MAS solid-state NMR spectrum show flexible residues in N and C termini; assignments taken from Carver et al. (15) are indicated with dotted lines.

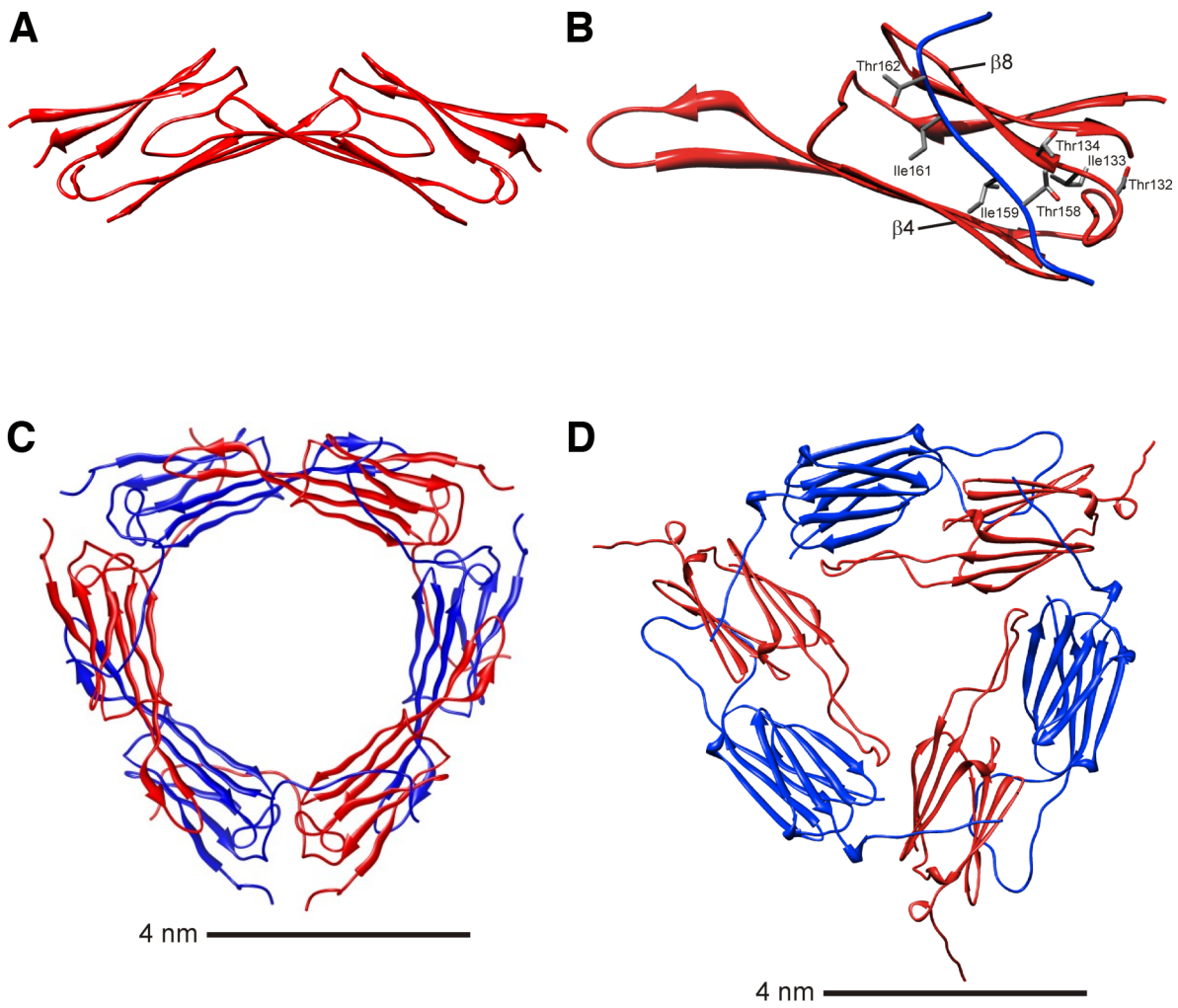

Figure 4.2 αB crystallin dimer and arrangement in hexameric unit. (**A**) Side view of αB ACD dimer, showing its curved shape [solid-state NMR, pH 7.5, PDB ID code 2klr]. (**B**) Intermolecular interactions of C-terminal domain (blue) including the conserved IXI motif (Ile159-Pro160-Ile161) and binding groove between β4 and β8 at the edge of an αB ACD (red). (**C**) Triangular arrangement of three αB dimers. Neighboring dimers are connected by the interaction illustrated in B. (**D**) Triangular arrangement of three ACD dimers in the threefold axis of the octahedrally symmetrical oligomer of *Methanococcus janashii* sHSP16.5.

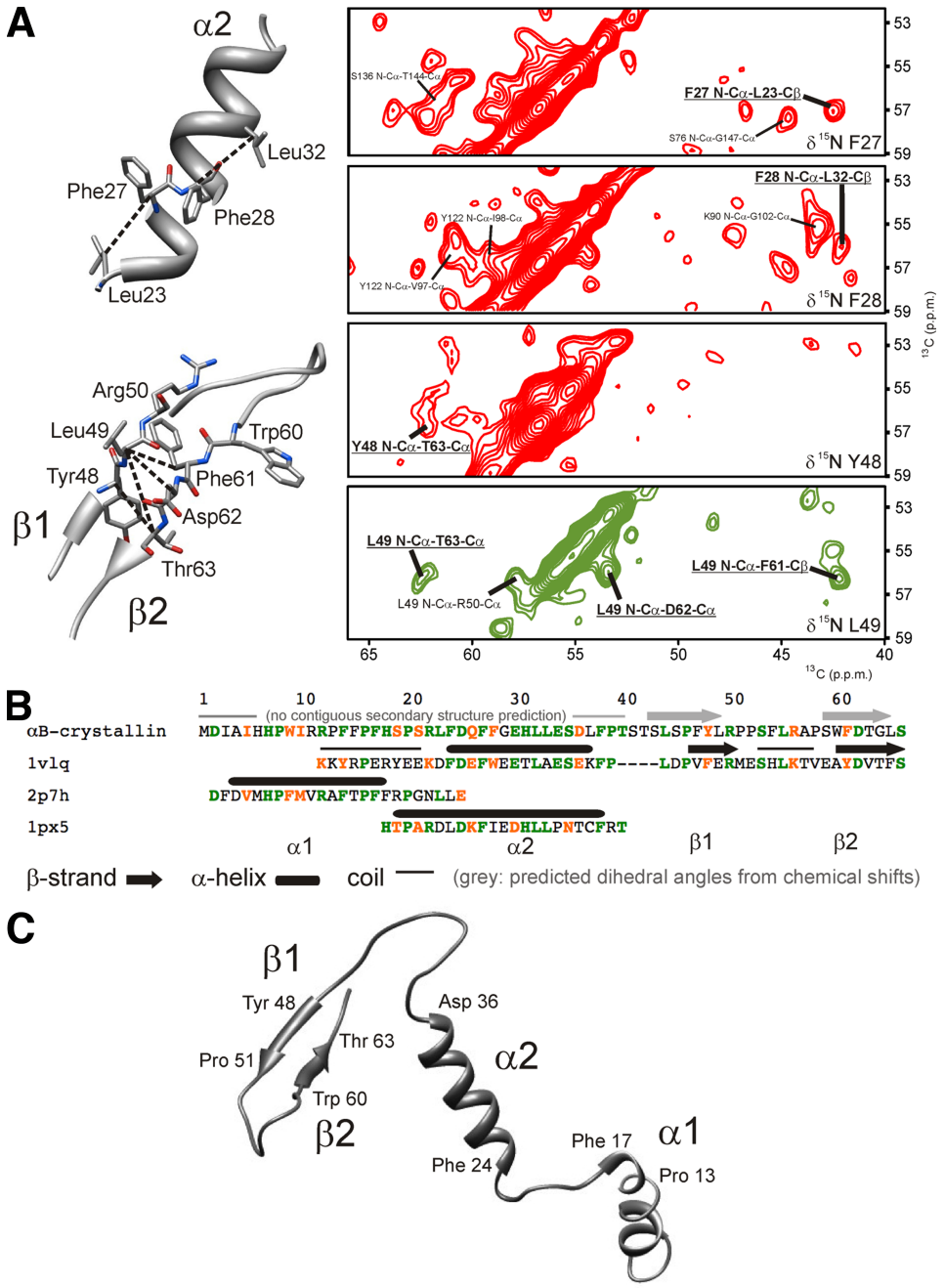

Figure 4.3 Building a model for the N-terminal domain of αB. (**A**) Contour plots from 3D NCACX spectra measured by solid-state NMR recorded using 500 ms spin-diffusion mixing. Distance restraints support secondary structure for the N terminus. Data were collected from αB expressed with either 1,3-[^{13}C]-glycerol (green contours) or 2-[^{13}C]-glycerol (red contours). Correlations that define the relevant distant restraints are underlined. (**B**) Sequence alignment of the N terminus of αB with similar protein fragments. The residue numbering is for αB. (**C**) Model of the N-terminal domain of αB residues 1-65 based on solid-state NMR restraints, dihedral angles from chemical shifts and similarity to 1vlq (β1, β2, α2) and 2p7h (α1). Residue numbers indicate the first and last residues of secondary structure. Pro13 and Phe17 (indicated) flank the region where NMR restraints are observed.

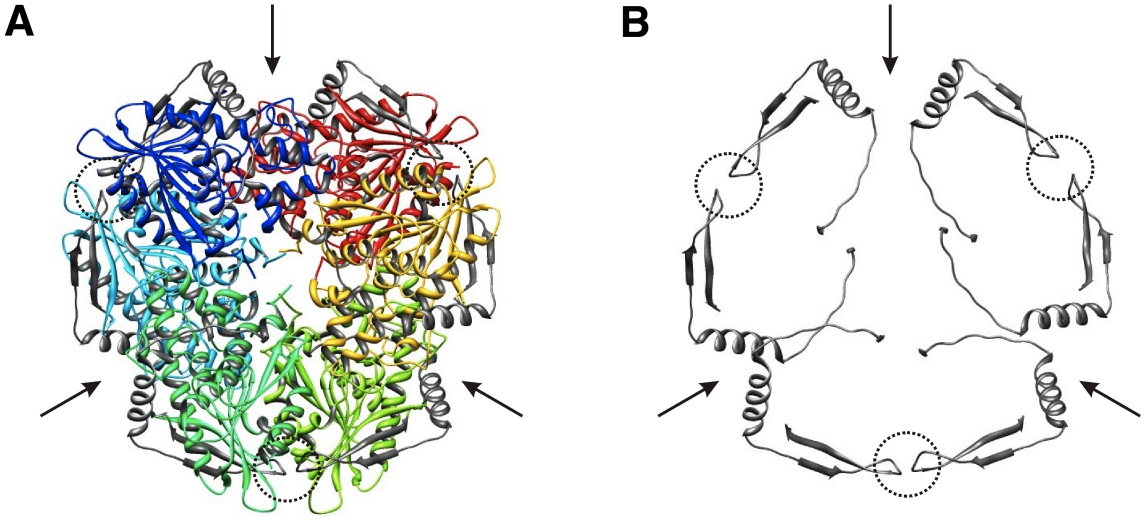

Figure 4.4 Homology model of αB β1 and β2. (**A**) Acetyl xylan esterase (PDB ID code 1vlq) from *Thermotoga maritima* showing all subunits of the hexamer (PDB ID code 1vlq). (**B**) Residues 12–62 of 1vlq are homologous to αB residues 12–66. The residues in the homologous structure (1vlq) show intermolecular contacts between the tip of the loops that connect β1 and β2 (indicated with dotted circles). The arrows guide the reader to follow from A to B.

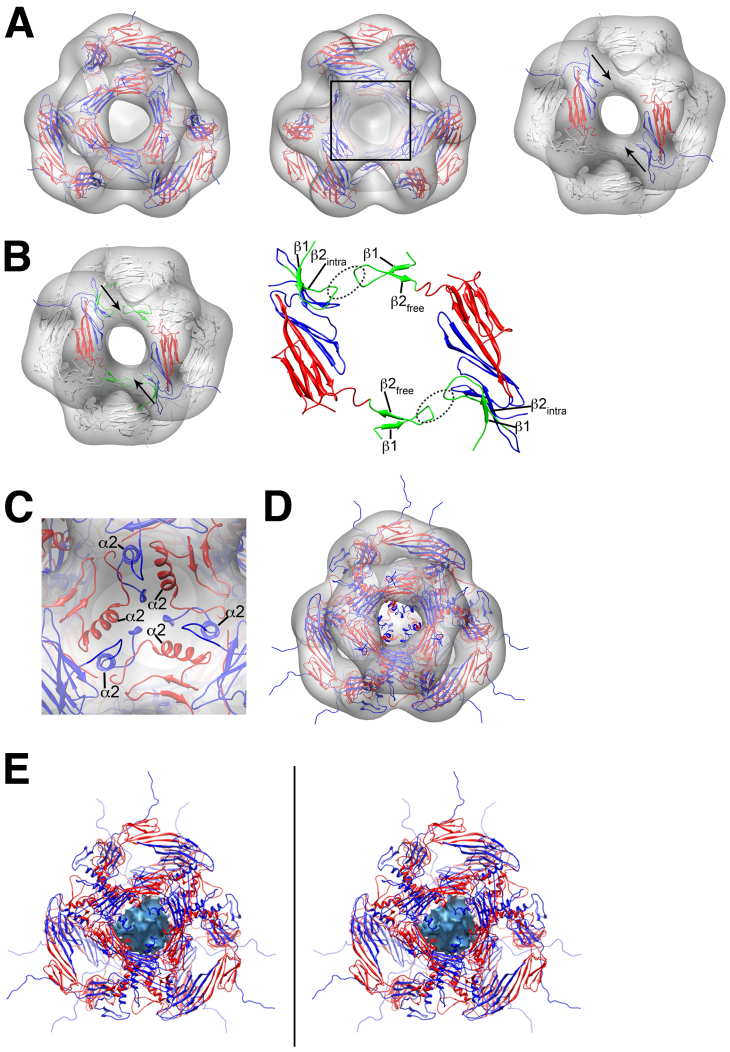

Figure 4.5 Model of αB 24-mer. (**A**) Triangular arrangement of three ACD dimers (PDB 2klr, residues 69-151) connected via the IXI motif of a neighboring dimer (residues 152-165) fit into EM density (EMD-1776) (13) (see also Fig. 14). (Left and center) A three-fold axis is shown viewed from the near and far side, respectively; (Right) view onto a two-fold axis. Unfilled density is indicated by the square (center) and arrows (right). (**B**) View onto a two-fold axis. (Left) N-terminal residues 45-68 (β1-loop-$β2_{free}$) fitted into unfilled EM-density marked by arrows; (Right) same view, with EM-density removed to illustrate different environments of β1/β2 loops. The tips of two loops (indicated by dashed circles) are close to each other and result in crosslinking of single cysteine mutants. (**C**) Enlargement of the area denoted by the square in A, Center, showing N-terminal residues 10-39 placed into the three-fold axes of the 24-mer (see text for details). (**D**) Flexible residues 166-175 on the surface for half the subunits and point to the inside of the 24-mer in the other half. Residues 1-10 reside on the inside. (E) Stereo view of 24-mer illustrating the small inner hollow in the center (central blue sphere) surrounded by flexible residues from the N and C termini.

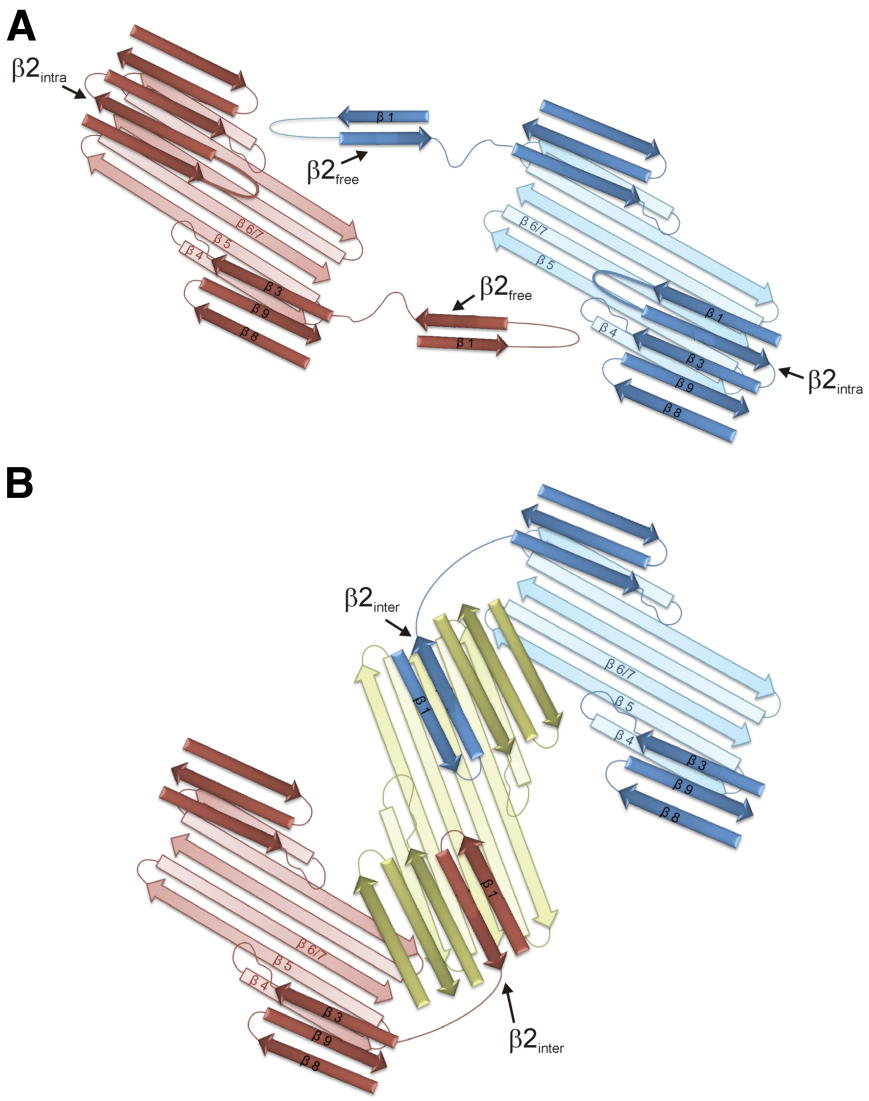

Figure 4.6 Intermolecular interface of αB multimerization. (**A**) View onto a twofold axis of a 24-mer shown schematically to highlight the arrangements of β2intra and β2free. (**B**) Schematic view onto a twofold axis occupied by an αB dimer in a multimer with greater than 24 subunits, showing the arrangement of the intermolecular interactions of β2inter. See main text for details.

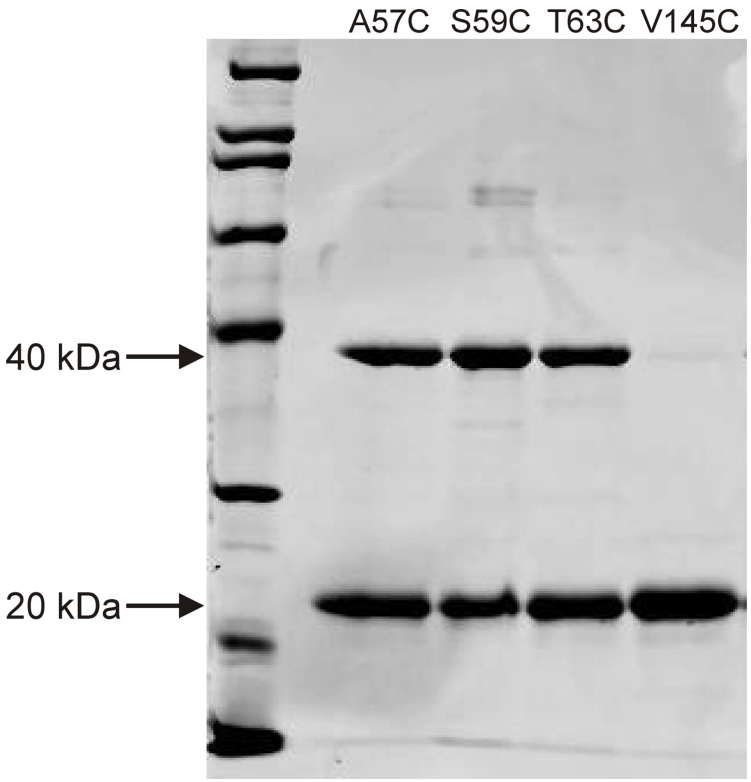

Figure 4.7 Cys cross-linking in αB. SDS-PAGE analysis performed in the absence of reducing agent shows disulfide-linked dimers are formed in multimers comprised of αB mutants A57C, S59C, and T63C but not mutant V145C.

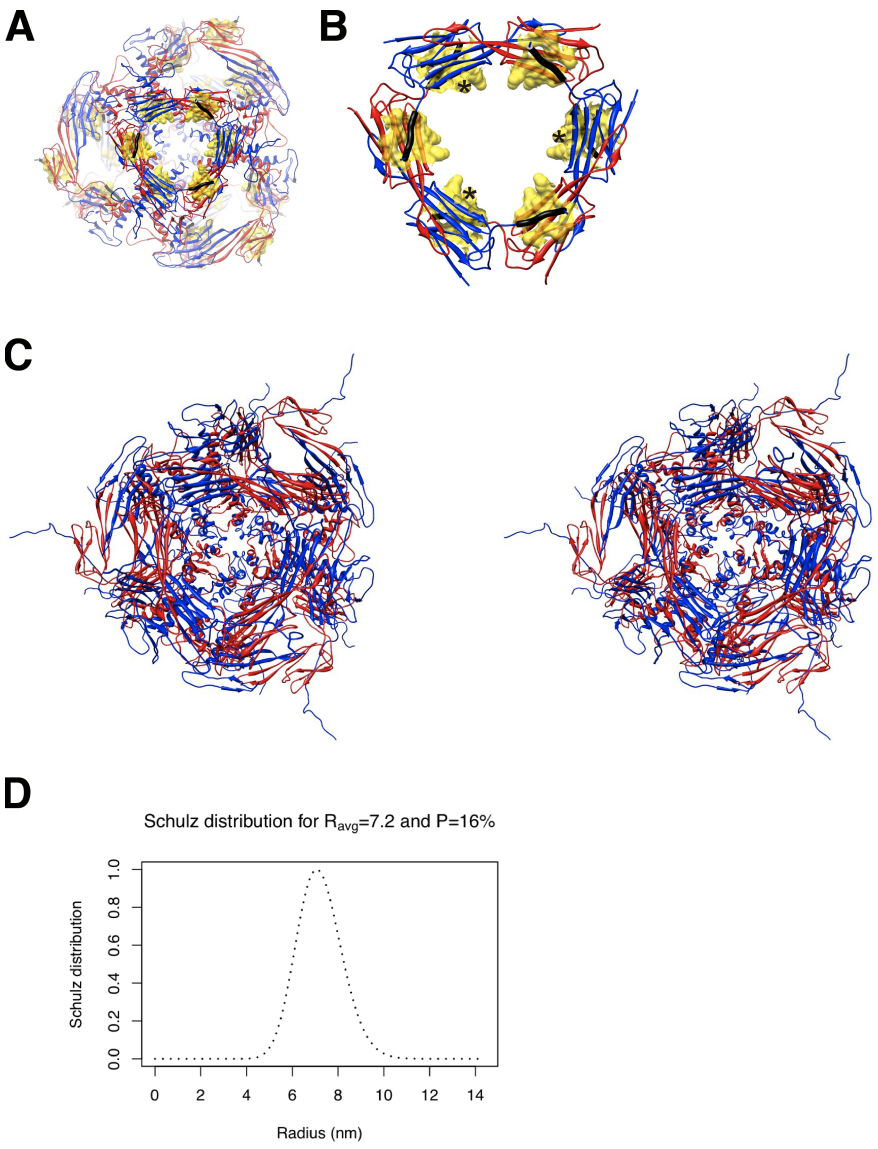

Figure 4.8 αB 24-mer. (**A**) Putative client-binding sites in αB 24-mer are shown in yellow, covered by the C-terminal IXI motif (black). (**B**) Hexameric unit from A, showing three client-binding sites face the surface, with the other three sites pointing to the interior of the multimer behind the plane of projection (indicated with asterisks). (**C**) Stereo representation of αB 24-mer comprised of full-length αB protomers (residues 1–175). (**D**) Schulz distribution of the particles' radius assuming a heterogeneity parameter of 16% and an average outer radius of 7.2 nm for the particle (according to Mittlebach and Glatter) (28). The Schulz distribution with 16% heterogeneity results in a standard deviation of 9.3 Å for a mean outer radius of 72 Å. Accordingly, 68% of the population is between the average radius fl⁄− SD, i.e., between 62.6 and 81.4 Å for the outer radius.

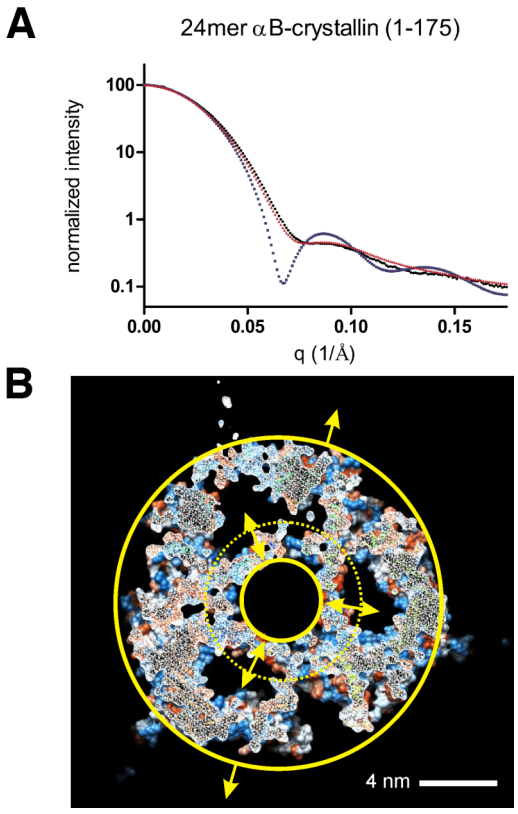

Figure 4.9 SAXS analysis of αB reveals heterogeneity. (**A**) SAXS data (black) were compared with the theoretical SAXS curve (blue) calculated for the 24-mer shown in Fig. 5D. A theoretical SAXS curve assuming 16% heterogeneity (28) is shown (red). (**B**) Cut-away view of the inner hollow occupied by flexible N-terminal residues and half the flexible C termini. The inner solid circle defines the inner cavity and the dashed circle defines the region occupied by flexible residues, placed to be in agreement with the radius of gyration measured by SAXS.

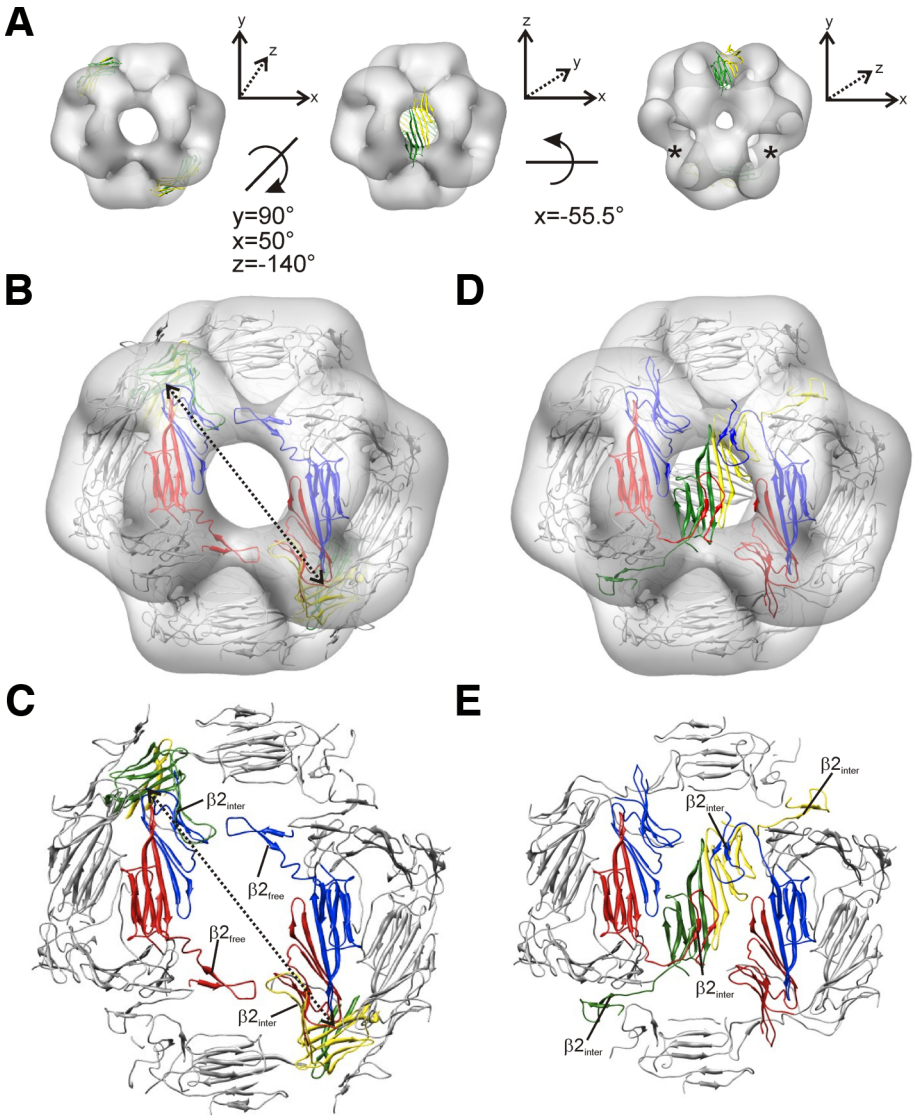

Figure 4.10 Model of an αB 28-mer. The EM density of αB has six openings on the edges of the three twofold axes. Each opening is large enough to accommodate an additional dimer. (**A**) Three views of EM density (EMD-1776) showing two additional dimers (each colored green and yellow) in openings at the edges of a twofold axis. (Left) View onto a twofold axis with its edge openings filled; (Center) view along a different twofold axis, showing the two dimers as placed in the left view; (Right) view along a threefold axis, again showing the two dimers as placed in the left-hand panel. Cartesian axes and the axis of rotation are shown for the three perspectives; the same views are shown in B–E. Asterisks (Right) indicate openings without an integrated dimer. (**B**) View as in A, Left, showing a 24-mer model (red, blue, and gray) and two additional dimers. (**C**) Cropped view of B with EM density removed to illustrate the arrangement of β-strands around the opening of the shell in one of the twofold axes. The two additional dimers of the 28-mer are covered in B and C by the othermolecules in the multimer. (**D**) The same oligomer as in B is viewed as in A, Center. The two dimers that break the symmetry of the multimer are shown in green/yellow and gray (on the far side). (**E**) A cropped view of D with EM density removed to highlight structural rearrangements of the β2free to β2inter strand.

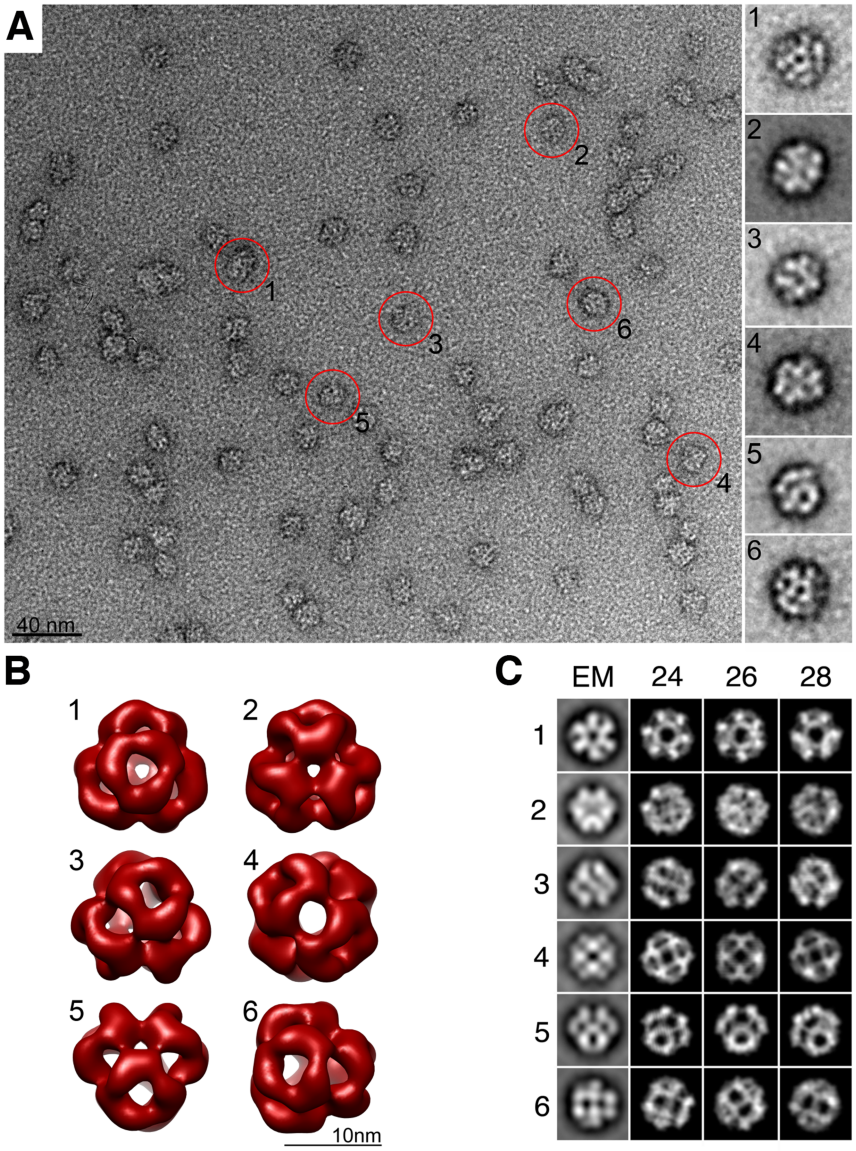

Figure 4.11 Electron microscopy and single particle analysis of αB. (**A**) Representative micrograph of negatively stained αB. The particles appear homogeneous in size. (Inset) Examples of six prominent class averages calculated using reference-free multivariant statistical analysis procedures in SPIDER (30). Red circles identify examples of the six class averages. The width of the box of the projection averages is 27 nm. (**B**) Different views of the αB density map (EMD -1776) at 20-Å resolution (13). The six views presented correspond to the six class averages shown in A, Inset. (**C**) Calculated back projection of the density map presented in B, and from models of αB 24-, 26-, and 28-mers. The six chosen back projections correspond to the six class averages in A and to the views in B.

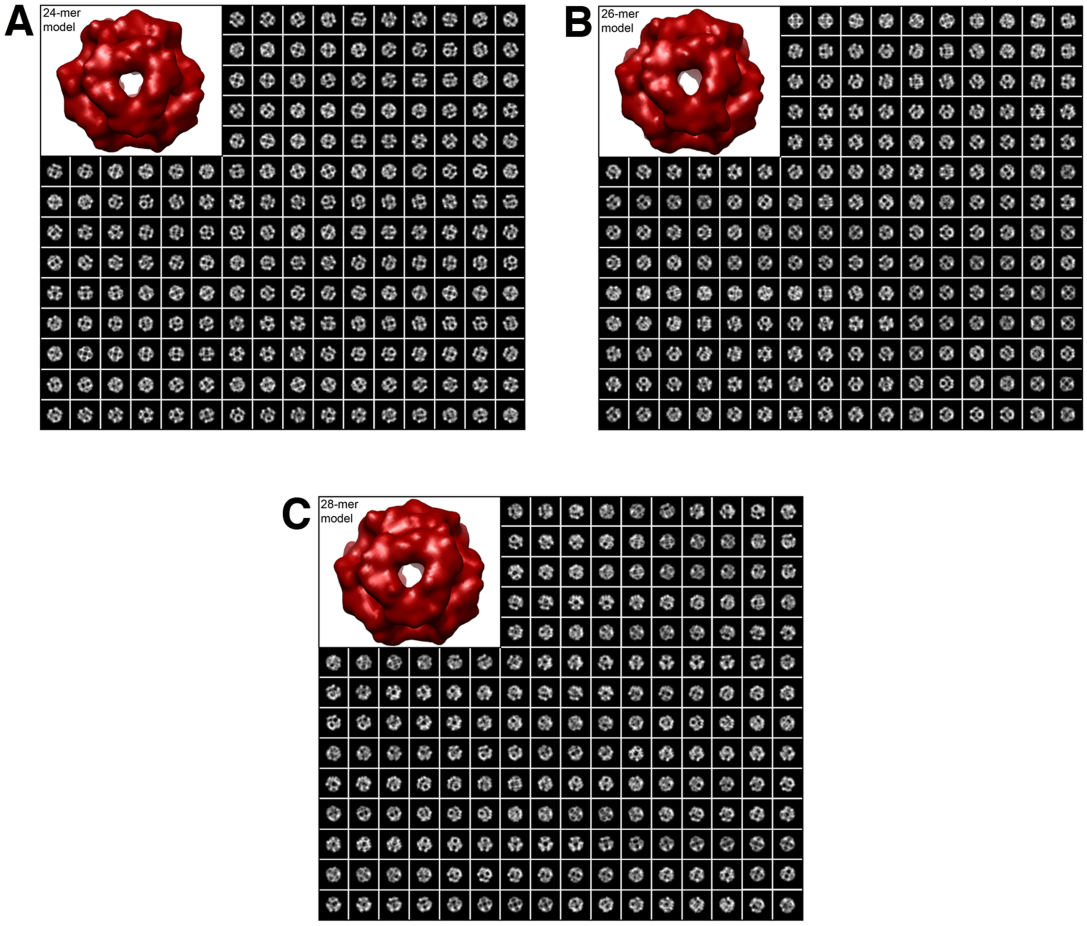

Figure 4.12 Back projections of 24-mer, 26-mer & 28-mer models. (**A**) αB 24-mer model and its calculated back projections. (**B**) αB 26-mer model and its calculated back projections. (**C**) αB 28-mer model and its calculated back projections.

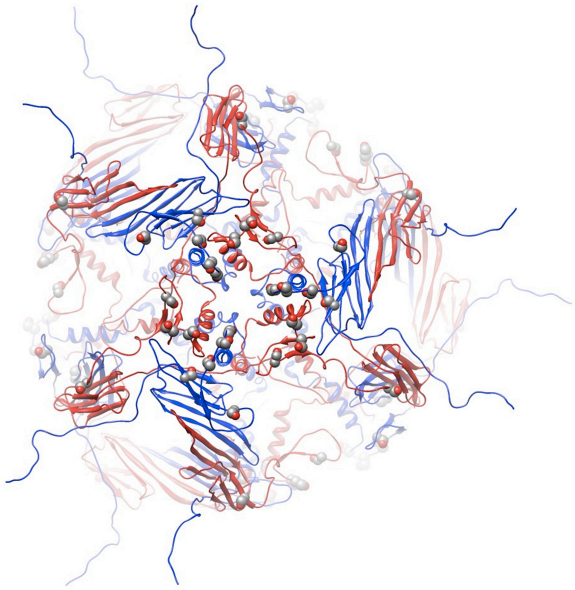

Figure 4.13 Location of phosphorylation sites in model of αB 24-mer. View onto a threefold axis of αB 24-mer showing side-chain atoms of known phosphorylation sites Ser41, Ser43, Ser45, Ser 59, and Ser76 as gray and red spheres. Not shown for clarity are phosphorylation sites Ser19, Ser21, and Ser53 that also reside in the threefold axes.

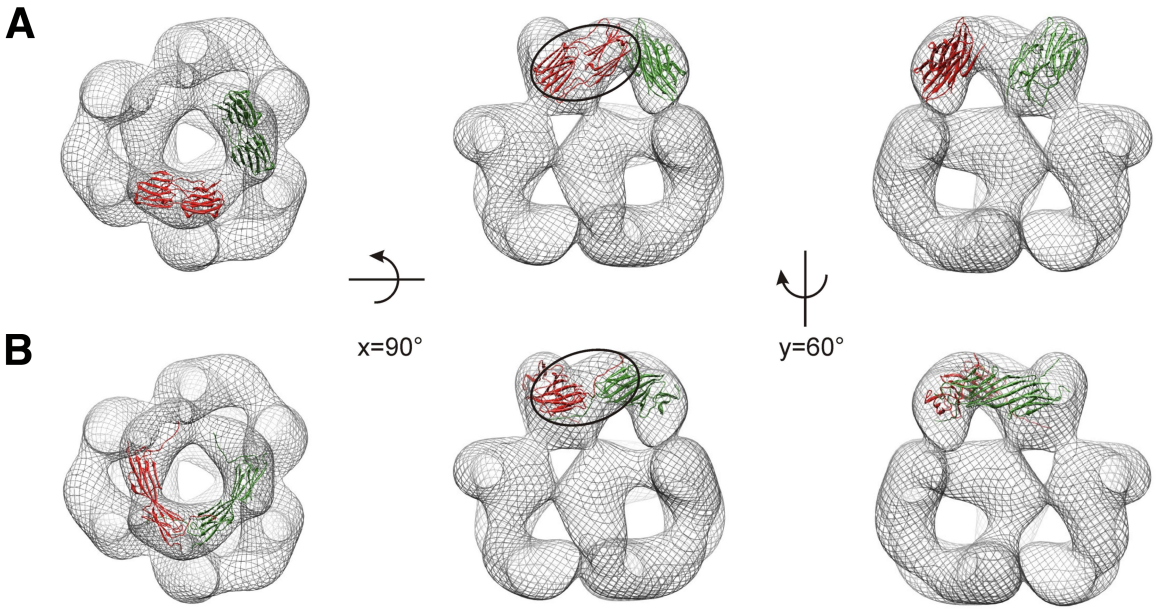

Figure 4.14 Comparison of the archaeal hsp16.5 coordinates (1shs) (A) and human αB coordinates (2klr) (B) fit into the αB density map (EMDataBank entry EMD-1776). Because of the more compact shape of the archael dimer, it fits as a whole into the density at each corner of the triangle. In contrast, two monomers of two adjacent dimers of the human αB fit back to back into the same density (Middle), indicated with an ellipse. This placement is consistent with the observations of Leganowski et al. who found an intermolecular interface on the back side of the dimer in their crystals (12).

Table 4.1 Distance restraints from solid-state NMR

αB fragment	restraints from 3D NCACX and NCOCX
residues 2-38 including α1 and α2	F14N-F14Cα-F17Cα F14N-F14Cα-P16Cδ F27N-F27Cα-L23Cβ F28N-F28Cα-L32Cβ F28N-F27C'-H31Cβ G29N-G29Cα-L32C' E30N-E30Cα-F27C' E30N-E30Cα-F28C' H31N-H31Cα-F27Cα H31N-H31Cα-F28Cα
residues 44-66 β1-loop-β2	Y48N-Y48Cα-T63Cα L49N-L49Cα-D62Cα L49N-L49Cα-T63-Cα L49N-L49Cα-F61-Cβ F61N-F61Cα-R50Cβ

Table 4.2 Similiarity Statistics

αB fragment	Homologous fragment	Pdb code	Similarity	Identity	E-value	# restraints
12-66	12-62	1vlq	47%	31%	7.35	15
2-25	25-48	2p7h	54%	38%	11.18	2
19-38	5-27	1px5	65%	43%	7.80	8

WORKS CITED IN CHAPTER 4

1. Bukau, B., Weissman, J., & Horwich, A. (2006). Molecular chaperones and protein quality control. *Cell*, **125**, 443-451.

2. Haslbeck, M., Franzmann, T., Weinfurtner, D., & Buchner, J. (2005). Some like it hot: the structure and function of small heat-shock proteins. *Nat. Struct. Mol. Biol.* **12**, 842-846.

3. Ecroyd, H. & Carver, J.A. (2009). Crystallin proteins and amyloid fibrils. *Cell. Mol. Life Sci.* **66**, 62-81.

4. Horwitz, J. (1992). Alpha-crystallin can function as a molecular chaperone. *Proc. Natl. Acad. Sci. USA*, **89**, 10449-10453.

5. Vicart, P. *et al.* (1998). A missense mutation in the alphaB-crystallin chaperone gene causes a desmin-related myopathy. *Nat. Genet.* **20**, 92-95.

6. Goldstein, L. E. *et al.* (2003). Cytosolic beta-amyloid deposition and supranuclear cataracts in lenses from people with Alzheimer's disease. *Lancet.* **361**, 1258-1265.

7. Kato, K. *et al.* (2001). Ser-59 is the major phosphorylation site in alphaB-crystallin accumulated in the brains of patients with Alexander's disease. *J. Neurochem.* **76**, 730-736.

8. Ousman, S. S. *et al.* (2007). Protective and therapeutic role for alphaB-crystallin in autoimmune demyelination. *Nature*, **448**, 474-479.

9. Augusteyn, R. C. & Koretz, J. F. (1987). A possible structure for alpha-crystallin. *FEBS Lett.* **222**, 1-5.

10. Jehle, S. *et al.* (2010). Solid-state NMR and SAXS studies provide a structural basis for the activation of alphaB-crystallin oligomers. *Nat. Struct. Mol. Biol.* **17**, 1037-1042.

11. Bagneris, C. *et al.* (2009). Crystal structures of alpha-crystallin domain dimers of alphaB-crystallin and Hsp20. *J. Mol. Biol.* **392**, 1242-1252.

12. Laganowsky, A. *et al.* (2010). Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci.* **19**, 1031-1043.

13. Peschek, J. *et al.* (2009). The eye lens chaperone alpha-crystallin forms defined globular assemblies. *Proc. Natl. Acad. Sci. USA,* **106**, 13272-13277.

14. Aquilina, J. A., Benesch, J. L., Bateman, O. A., Slingsby, C. & Robinson, C. V. (2003). Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of

oligomers in alphaB-crystallin. *Proc. Natl. Acad. Sci. USA*, **100**, 10611-10616.

15. Carver, J. A., Aquilina, J. A., Truscott, R. J. & Ralston, G. B. (1992). Identification by 1H NMR spectroscopy of flexible C-terminal extensions in bovine lens alpha-crystallin. *FEBS Lett.* **311**, 143-149.

16. Meehan, S. *et al.* (2007). Characterisation of amyloid fibril formation by small heatshock chaperone proteins human alphaA-, alphaB- and R120G alphaB-crystallins. *J. Mol. Biol.* **372**, 470-484.

17. Jehle, S. *et al.* (2009). alphaB-crystallin: a hybrid solid-state/solution-state NMR investigation reveals structural aspects of the heterogeneous oligomer. *J. Mol. Biol.* **385**, 1481-1497.

18. Higman, V. A. *et al.* (2009). Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively 13C-labelled proteins. *J. Biomol. NMR*, **44**, 245-260.

19. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213-223.

20. Kim, K. K., Kim, R. & Kim, S. H. (1998). Crystal structure of a small heat-shock protein. *Nature*, **394**, 595-599.

21. Kundu, B., Shukla, A., Chaba, R. & Guptasarma, P. (2004). The excised heat-shock domain of alphaB crystallin is a folded, proteolytically susceptible trimer with significant surface hydrophobicity and a tendency to self-aggregate upon heating. *Protein Expr. Purif.* **36**, 263-271.

22. Merck, K. B., De Haard-Hoekman, W. A., Oude Essink, B. B., Bloemendal, H. & De Jong, W. W. (1992). Expression and aggregation of recombinant alpha A-crystallin and its two domains. *Biochim. Biophys. Acta*, **1130**, 267-276.

23. van Montfort, R. L., Basha, E., Friedrich, K. L., Slingsby, C. & Vierling, E. (2001). Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat. Struct. Biol.* **8**, 1025-1030.

24. Stamler, R., Kappe, G., Boelens, W. & Slingsby, C. (2005). Wrapping the alpha-crystallin domain fold in a chaperone assembly. *J. Mol. Biol.* **353**, 68-79.

25. Augusteyn, R. C. (1998). alpha-Crystallin polymers and polymerization: the view from down under. *Int. J. Biol. Macromol.* **22**, 253-262.

26. Haley, D. A., Horwitz, J. & Stewart, P. L. (1998). The small heat-shock protein, alphaBcrystallin, has a variable quaternary structure. *J. Mol. Biol.* **277**, 27-35.

27. Haley, D. A., Bova, M. P., Huang, Q. L., McHaourab, H. S. & Stewart, P. L. (2000). Small heat-shock protein structures reveal a continuum from symmetric to variable assemblies. *J. Mol. Biol.* **298**, 261-272.

28. Mittelbach, R. & Glatter, O. (1998). Direct structure analysis of small-angle scattering data from polydisperse colloidal particles. *J. Appl. Crystallogr.* **31**, 600-608.

29. Vass, S. *et al.* (2008). Ambiguity in determining the shape of alkali alkyl sulfate micelles from small-angle scattering data. *Langmuir*, **24**, 408-417.

30. Frank, J. *et al*. (1996). SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190-199.

31. Ghosh, J. G. & Clark, J. I. (2005). Insights into the domains required for dimerization and assembly of human alphaB crystallin. *Protein Sci.* **14**, 684-695.

32. Ghosh, J. G., Shenoy, A. K., Jr. & Clark, J. I. (2006). N- and C-Terminal motifs in human alphaB crystallin play an important role in the recognition, selection, and solubilization of substrates. *Biochemistry*, **45**, 13847-13854.

33. MacCoss, M. J. *et al*. (2002). Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. USA*, **99**, 7900-7905.

34. Liu, Y. *et al*. (2006). A novel alphaB-crystallin mutation associated with autosomal dominant congenital lamellar cataract. *Invest. Ophthalmol. Vis. Sci.* **47**, 1069-1075.

35. Biswas, A. *et al*. (2007). Paradoxical effects of substitution and deletion mutation of Arg56 on the structure and chaperone function of human alphaB-crystallin. *Biochemistry*, **46**, 1117-1127.

36. Selcen, D. & Engel, A. G. (2003). Myofibrillar myopathy caused by novel dominant negative alpha B-crystallin mutations. *Ann. Neurol.* **54**, 804-810.

37. Routledge, K. E., Tartaglia, G. G., Platt, G. W., Vendruscolo, M. & Radford, S. E. (2009). Competition between intramolecular and intermolecular interactions in an amyloidforming protein. *J. Mol. Biol.* **389**, 776-786.

38. Pechmann, S., Levy, E. D., Tartaglia, G. G. & Vendruscolo, M. (2009). Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc. Natl. Acad. Sci. USA*, **106**, 10159-10164.

39. Petoukhov, M. V. & Svergun, D. I. (2007). Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr. Opin. Struct. Biol.* **17**, 562-571.

40. Pettersen, E. F. *et al*. (2004). UCSF Chimera – a visualization system for exploratory

research and analysis. *J. Comput. Chem.* **25**, 1605-1612.

41. De Vries, S. J., van Dijk, M. & Bonvin, A. M. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883-897.

Chapter V

Computational design of self-assembling protein nanomaterials with atomic level accuracy

Neil P. King[1], William Sheffler[1], Michael R. Sawaya[2], **Breanna S. Vollmar[3]**, John P. Sumida[4], Ingemar André[5], Tamir Gonen[3], Todd O. Yeates[6,7], and David Baker[1,8]*

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA.[2]Howard Hughes Medical Institute, UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA 90095, USA.[3]Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA.[4]Department of Medicinal Chemistry, University of Washington, Seattle, WA 98177, USA.[5]Department of Biochemistry and Structural Biology, Lund University, Lund, Sweden.[6]UCLA Department of Chemistry and Biochemistry, Los Angeles, CA 90095, USA.[7]UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA 90095, USA.[8]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.

This work was published in the journal of Science (2012), volume 336, pages 1171-1174.

SUMMARY

We describe a general computational method for designing proteins that self-assemble to a desired symmetric architecture. Protein building blocks are docked together symmetrically to identify complementary packing arrangements, and low-energy protein-protein interfaces are then designed between the building blocks in order to drive self-assembly. Here we use trimeric protein building blocks to design a 24-subunit, 13 nm diameter complex with octahedral symmetry and two related variants of a 12-subunit, 11 nm diameter complex with tetrahedral symmetry. The designed proteins assembled to the desired oligomeric states in solution, and crystal structures of the complexes revealed that the resulting materials closely match the design models. The method can be used to design a wide variety of self-assembling protein nanomaterials.

INTRODUCTION

Molecular self-assembly is an elegant and powerful approach to patterning matter on the atomic scale. Recent years have seen advances in the development of self-assembling biomaterials, particularly those composed of nucleic acids (1). DNA has been used to create, for example, nanoscale shapes and patterns (2), molecular containers (3), and three-dimensional macroscopic crystals (4). Methods for designing self-assembling proteins have progressed more slowly, yet the functional and physical properties of proteins make them attractive as building blocks for the development of advanced functional materials (5, 6). The sophisticated protein-based molecular machines observed in natural systems—which often require self-assembly to function as, for example, cellular motors, pumps, or scaffolds—provide a suggestion of the practical potential of designed protein materials.

In any self-assembling structure, interactions between the subunits are required to drive assembly. Previous approaches to designing self-assembling proteins have satisfied this requirement in various ways, including the use of relatively simple and well-understood coiled-coil and helical bundle interactions (7-11), engineered disulfide bonds (12, 13), chemical cross-links (14), metal-mediated interactions (15, 16), templating by non-biological materials in conjunction with computational interface design (17), or genetic fusion of multiple protein domains or fragments which naturally self-associate (18, 19). In contrast, natural protein assemblies are most often held together by many weak, noncovalent interactions which together form large, highly complementary, low energy protein-protein interfaces (20). Such interfaces spontaneously self-assemble and allow precise definition of the orientation of subunits relative to one another, which is critical for obtaining the desired material with high accuracy (18). Designing assemblies with these properties has been difficult due to the complexities of

modeling protein structures and energetics. For instance, a pioneering study used interface design by visual inspection to design new oligomeric structures, yet the experimentally determined dimeric interfaces were largely unanticipated (21). However, recent advances (22-26), including the *de novo* design of a heterodimeric protein interface with atomic level accuracy (27, 28), suggest that our ability to computationally model and design protein-protein interactions is rapidly maturing.

We describe a general computational method for designing self-assembling protein materials which consists of two steps: 1) symmetrical docking of protein building blocks in a target symmetric architecture, followed by 2) design of low energy protein-protein interfaces between the building blocks to drive self-assembly. Here we use as building blocks oligomeric proteins that share an element of symmetry with the target architecture. This reduces by one the number of new protein-protein interfaces that must be designed, since the interface within the oligomer contributes to the self-assembly of the subunits to the target material. Furthermore, the energetic contribution of each designed interaction is multiplied by the symmetry of the building block, which reduces the number of distinct new interactions required to overcome the entropic cost of self-assembly (21).

RESULTS AND DISCUSSION

Design method

We used the method to design cage-like protein nanomaterials with either tetrahedral (T) or octahedral (O) point group symmetry (Fig. 5.1). An assembly with symmetry T requires 12 copies of a protein molecule arranged in 12 symmetry-related orientations, whereas symmetry O requires 24 molecules. Both point groups can be generated from sets of three-fold rotational

symmetry axes (Fig. 5.1A), allowing the use of protein trimers with C3 symmetry as building blocks; in each case, only a single new interface between the trimeric building blocks is required for self-assembly. 271 naturally trimeric protein structures were docked symmetrically in both the T and O target architectures by aligning the three-fold axis of each building block with the three-fold axes in the target architecture and then systematically sampling the two remaining rigid body degrees of freedom, radial displacement and axial rotation, in increments of 1 Å and 1°, respectively (Fig. 5.1B,C). For each docked configuration in which no clashes between the backbone and beta carbon atoms of adjacent building blocks were present, a simple proxy for interface size and complementarity was computed to gauge the "designability" of the configuration (Fig. 5.1C,D). Around each of the 10 (O) or 20 (T) most designable configurations for each building block, a set of input structures for design was generated by sampling the radial displacement and axial rotation of the subunits more finely (0.1 Å, 0.5°). For each of these input structures, symmetric RosettaDesign calculations (29, 30) were used to design a new amino acid sequence for the protein that resulted in low-energy, symmetric protein-protein interactions between the trimeric building blocks (Fig. 5.1E,F). Designs with the lowest predicted binding energies and geometrically complementary interfaces of sufficient size were further optimized using RosettaDesign and interactive design in foldit (31). 8 T and 33 O designs derived from 15 distinct natural trimeric proteins, containing on average 9 mutations per monomer, were selected for experimental characterization (Table 5.1).

Evaluation of expressed and purified octahedral and tetrahedral designs

Genes encoding the designed proteins and the corresponding wild-type trimers were constructed and cloned into an expression vector that appended an 11-residue peptide substrate

for fluorescent modification by the *E. coli* acyl-carrier protein synthase AcpS (32). *E. coli* cells expressing the proteins were lysed, the proteins were fluorescently labeled in the clarified lysates by the addition of AcpS and the CoA-488 fluorophore, and the apparent size of each protein was visualized by subjecting the labeled lysates to PAGE under non-denaturing (native) conditions. Out of 7 T and 17 O designs that expressed solubly (Table 5.1), one designed protein of each architecture revealed a shift in apparent size relative to the corresponding wild-type trimer that suggested self-assembly to the desired material (Fig. 5.2A). Size exclusion chromatography (SEC) of the labeled lysates confirmed the change in apparent molecular weight for the two designs (Fig. 5.3). Genes encoding the octahedral design ("O3-33"; 9 mutations from the wild-type protein), the tetrahedral design ("T3-08"; 8 mutations), and the corresponding wild-type trimeric proteins were then subcloned into an expression vector that appended C-terminal $(His)_6$ tags, after which the proteins were expressed and purified by nickel affinity chromatography and SEC.

The designed protein O3-33 eluted from the SEC column as a single peak with an apparent size of about 24 subunits (Fig. 5.2B). The wild-type protein from which O3-33 was derived (PDB ID 3N79) did not assemble to a higher order structure; it eluted from the column mostly as trimers, with a small peak corresponding to a dimer of trimers (Fig. 5.2C). Analytical ultracentrifugation revealed that the designed protein sedimented as a single discrete species with a Stokes radius of 7.3 nm, in close agreement with the radius of the designed 24-subunit assembly (Fig. 5.4). A point mutation (Ala167Arg) that introduced unfavorable steric clashes at the designed interface disrupted the material, suggesting that the observed self-assembly is due to the designed interface (Fig. 5.2D). Negative stain electron microscopy (EM) of O3-33 revealed fields of monodisperse particles of the expected size (~13 nm), many of which strikingly

resembled projections of the design model along its 2-fold, 3-fold, or 4-fold symmetry axes (Fig. 5.5A). A single particle reconstruction of O3-33 obtained by EM analysis under cryogenic conditions clearly recapitulated the architecture of the design model, verifying that the protein assembles in solution as designed (Figs. 5.5C and 5.6).

We solved crystal structures of O3-33 to evaluate the accuracy of our design protocol at high resolution. Structures from two different crystal forms confirmed that the designed material adopts the target architecture and that the designed interface is responsible for driving self-assembly; the higher resolution (2.35 Å) crystal form is shown in Figure 5.5. The structure proved remarkably similar to the design model: the backbone RMSD over all 24 chains is 1.07 Å, and is lower if calculated using only the residues at the interface (0.85 Å). The high resolution of the structure allowed confident determination of the side chain configurations at the designed interface, revealing that the atomic contacts closely match those in the design model (Fig. 5.5E). The asymmetric unit of the designed interface consists of one alpha helix packing against a beta strand, a loop, and the symmetrically related helix in a neighboring building block. Several ordered water molecules were resolved at the designed interface that contribute bridging hydrogen-bonding interactions between neighboring building blocks. Truncation of designed interface residues to alanine disrupted octahedral self-assembly (Fig. 5.7). For example, the Ser156Ala mutation, which alters O3-33 by the removal of only two atoms out of 2,827 total atoms in the subunit, significantly impaired assembly. This result underscores the importance of both the detailed atomic contacts designed by our protocol and the multiplicative effect of the symmetry of the system: the Ser156Ala mutation results in the loss of 24 interface hydrogen bonds in the fully assembled material.

The designed protein T3-08 appeared by SEC to be in a slow equilibrium between two

states comprising three and approximately 12 subunits (Fig. 5.2E). The corresponding wild-type trimeric protein (PDB ID 3FTT) eluted from the column as trimer only (Fig. 5.2G). Disruption of the designed interface by a point mutation, Ala52Gln, again suggested that the designed interface is responsible for the observed self-assembly (Fig. 5.2H). A crystal structure of T3-08 revealed that the protein assembles to the desired tetrahedral architecture, but the trimeric building blocks are slightly rotated about the shared trimeric/tetrahedral three-fold rotational axes, subtly altering the atomic contacts at the designed interface relative to the design model and resulting in a backbone RMSD of 2.66 Å over all 12 subunits (Fig. 5.8).

We designed two additional variants of T3-08 (Table 5.1) in order to determine whether we could preferentially stabilize the designed configuration over the unanticipated configuration observed in the T3-08 crystal structure. One of the variants, T3-10, which contained 3 mutations relative to T3-08 intended to provide better hydrophobic packing near the tetrahedral three-fold interface (Fig. 5.9), was purified by nickel affinity chromatography and appeared by SEC to self-assemble efficiently to the tetrahedal state, yielding little detectable trimer (Fig. 5.2F). Negative stain EM images of T3-10 revealed monodisperse particles of the expected size (~11 nm), averages of which closely resembled projections of the design model along its 2-fold and 3-fold symmetry axes (Fig. 5.10A,B). A crystal structure of T3-10 verified that the original designed configuration was stabilized as intended; the backbone RMSD between the T3-10 crystal structure and the T3-08/T3-10 design models is 0.62 Å (Fig. 5.10B,C). As observed for O3-33, the atomic contacts at the designed interface, which consists of two alpha helices and two short loops, closely match those in the design model (Fig. 5.10D). This result illustrates how small alterations to the protein sequence at the designed interface may allow fine control over the structure of the resulting material.

CONCLUDING REMARKS

Our results establish a method by which self-assembling protein materials may be designed with high accuracy. The design strategy, combining symmetrical docking with interface design, is conceptually simple and generally applicable to the design of a broad range of symmetric materials. In addition to the finite, cage-like materials described here, unbounded materials in one, two, or three dimensions (i.e., fibers/helices, layers, or crystals) may be designed by choosing an appropriate target symmetric architecture. Although in the present study we used naturally occurring oligomeric proteins as building blocks, novel oligomeric building blocks could first be designed from monomers and after structural validation, used in the design of higher order assemblies with the attendant advantages of hierarchical assembly, or, with improvements in our symmetrical docking protocol, larger self-assembling systems could be designed directly from monomeric building blocks. The atomic-level accuracy of our designed materials demonstrates that using designed protein-protein interfaces to drive self-assembly results in highly ordered materials with superior rigidity and monodispersity. With further development, designed self-assembling protein materials similar to those described here could form the basis of advanced functional materials and custom-designed molecular machines with wide-ranging applications.

MATERIALS AND METHODS

Generating a set of trimeric scaffold proteins.

A set of trimeric proteins for use as input structures to our design protocol was obtained by first searching the PISA database (33) for homotrimeric proteins with an average chain length of less than 200 amino acids; this list was cross-referenced against the results of a search of the

3D Complex database (34) for homotrimeric proteins with C3 symmetry and the intersection of the two lists was then clustered at 95% sequence identity to identify a representative structure from each cluster. The final list included the 271 proteins identified by PDB ID below. Many of the proteins in the list were judged visually during the design process to not be trimeric.

Coordinates for each of the 271 proteins were downloaded from the Protein Data Bank (http://www.rcsb.org/pdb/), the trimeric axis was aligned along the vector [0,0,1], and the center of mass of the trimer translated to the origin. A single subunit of each trimer was used as input into our design protocol without further modification.

Our trimeric scaffold protein set included the following structures:

1a3f	1a8m	1aa0	1afa	1ahs	1aly	1ax8	1b4b	1baw	1ble	1c5e	1ca4
1dbf	1dg6	1di7	1dpt	1duc	1dxc	1e20	1env	1eq7	1f3g	1f7l	1f7r
1gnk	1gr3	1h7z	1h9j	1hfo	1htn	1hup	1iby	1ihc	1ij0	1ij1	1iv1
1jpx	1jq0	1js0	1k33	1kd7	1kfn	1knb	1kqa	1kr4	1l1s	1lr0	1mff
1mof	1nig	1nob	1nog	1nq3	1nza	1o51	1o8n	1o91	1ocy	1oni	1otg
1ox3	1p1l	1p4t	1p9h	1pf5	1pil	1piq	1q0p	1q3i	1q89	1qah	1qd9
1qhv	1qu9	1qzu	1r13	1rhy	1rlh	1rnj	1s55	1s9z	1t0a	1t2n	1td3
1tul	1u9d	1ufy	1ug4	1uiz	1uku	1usn	1uwn	1v3w	1vbv	1vck	1ve0
1vi4	1vmh	1w16	1wdf	1wfx	1woz	1wyy	1x25	1x90	1xhd	1xho	1ygs
1yq8	1yvs	1z8k	1zvc	1zvq	2b33	2baz	2bs5	2bsf	2bzu	2cgy	2cmp
2csl	2cwj	2d4l	2d4z	2dt4	2ebo	2ed6	2eg1	2f7w	2fb6	2fkk	2flz
2fvh	2g2c	2g2d	2gw8	2h3m	2h6l	2h8a	2hew	2hx0	2ic7	2j97	2j9c
2jbz	2nuh	2ose	2ot5	2otm	2p23	2p4v	2p6c	2p6h	2p6y	2pd2	2pmp
2pno	2pnv	2q5u	2qe3	2qg8	2qgr	2qih	2r32	2r6q	2re9	2rfr	2rhd
2tnf	2ux6	2v18	2vhe	2vky	2w6b	2w9n	2wb3	2wh7	2wl7	2wpq	2wps
2wpy	2wq1	2wst	2ww6	2ww7	2wz7	2x29	2x4j	2xbp	2xcd	2xcz	2xdh
2xmw	2yw4	2yzj	2zfc	3b6n	3b8l	3bq4	3bzq	3c19	3ce8	3ci1	3ck2
3cnc	3cp1	3d8m	3de9	3dww	3e6q	3eby	3efg	3ejc	3ejv	3elo	3emo
3esj	3exv	3exw	3f6m	3f91	3fk3	3ftt	3fwt	3fwu	3gmj	3gtz	3gud
3gwm	3h5i	3h7z	3h88	3h9n	3hhl	3hon	3hro	3hwu	3hza	3i7t	3i9f
3i9z	3ipz	3irc	3is8	3ivl	3ixc	3jsc	3jv1	3k9a	3kjj	3kwd	3kxr
3l8r	3laa	3lqw	3m06	3m1x	3mc3	3md1	3mdx	3mgw	3mh5	3mpv	3mxu
3n3f	3n79	3nbt	3ne1	3nv1	3o46	3oi9					

Symmetry definition files

A general framework for modeling arbitrary symmetric systems has been implemented in Rosetta (30). In the present study, we used this framework to dock subunits of protein trimers in configurations with octahedral and tetrahedral symmetry while maintaining the relative orientation of the subunits within the trimer. The symmetry of the target architecture (for a visual

example, see Fig. 1A in the main text) is passed to Rosetta in the form of a symmetry definition file. The symmetry definition files used in this study to model octahedra and tetrahedra constructed from trimers ("O3" and "T3" complexes) can be found in the zipped archive design_models.zip.

Symmetrical docking

An application was written within the Rosetta macromolecular modeling framework to dock oligomeric protein building blocks into higher order symmetries, searching for docked configurations that yield complementary interfaces suitable for interface design. The application, matdes_dock, takes as input a .pdb file containing a single subunit of a starting structure and a symmetry definition file. The subunits are first arranged symmetrically at the origin according to the defined symmetry, and then the full space of contacting symmetric configurations are sampled by systematically varying the translational and rotational degrees of freedom (DOFs) in the system in increments of 1 Å and 1°, respectively. Negative translations may also be sampled (by setting the boolean matdes::dock::neg_r command line option to 1) such that both orientations of the building block (inside/outside) are tested. Configurations in which backbone or beta carbon atoms from different building blocks clash (distance between backbone amide nitrogen and carbonyl oxygen atoms <= 2.6 Å; distance between all other backbone/beta carbon atom pairs <= 3.0 Å) are discarded. In each non-clashing configuration, a measure of the suitability of the configuration for design ("designability") was calculated; we found that summing the number of beta carbon contacts between building blocks, where a contact is defined as two beta carbon atoms within 10 Å, was a good measure of designability and was rapidly computable. The resulting data for each input protein, which can be saved by redirecting the output of the run to a

log file, can be visualized as a three-dimensional heatmap like that shown in Fig. 1C of the main text. The 10 (for O designs) or 20 (for T designs) docked configurations for each protein with the highest number of summed beta carbon contacts were used as inputs to the sequence design application described below.

An example of the command line options used with the matdes_dock application is shown here:

```
-database <rosetta_database_path>
-s <input_structure>
-symmetry_definition O3.sym
-matdes::pdbID 3n79
-matdes::prefix O3_
-matdes::num_subs_building_block 3
-matdes::num_subs_total 24
-matdes::dock::neg_r 0
```

Symmetrical interface design

An application was written within the Rosetta framework to identify stabilizing mutations at interfaces between oligomeric protein building blocks docked in higher order symmetries. The application, matdes_design, takes as input a .pdb file containing a single subunit of a starting structure and a symmetry definition file. In addition, settings for the rigid body DOFs (obtained from matdes_dock) are used to specify the initial docked configuration of the subunits. As part of the protocol, a grid of nearby docked configurations centered on the initial configuration are designed in independent trajectories; the width and number of samples in both dimensions of the grid (radial displacement and axial rotation) are input as command line options. In each docked configuration, amino acid residues within the independent subunit fulfilling four criteria are selected as positions which can be changed during design. First, the position must be at an inter-building block interface, defined as a position in which the beta carbon is within a user-defined cutoff distance to a beta carbon in a different building block (in this study the default 10 Å cutoff

was used). Second, the position must not be making contacts within the oligomeric building block (any atoms within 5 Å), with the exception that if a residue is making intra-building block contacts but has a high clash (fa_{rep}) score in Rosetta, it is selected for design. Third, the residue must have a nonzero solvent accessible surface area in the monomeric state. Fourth, interface proline and glycine residues are not designed, but proline side chains are allowed to repack. Residues passing all four criteria ("design positions") are then designed using RosettaDesign while all other residues in the protein remain fixed. All design calculations are performed on the independent subunit and propagated symmetrically as described (30). By default the standard Rosetta scorefunction, score12, is used during design. Once a new sequence for the protein has been designed, the side chains configurations in the design positions are minimized. Several additional metrics used to gauge the quality of the designed interface are calculated, including: 1) the number of neighboring residues within the monomer for each designed position ("average degree" in ref. 35), 2) the contacting interface surface area and shape complementarity (36), 3) the RosettaHoles packing score (37), and 4) the predicted binding energy of the interface, defined as the energy of the bound state minus the energy of the unbound state (building block alone) after repacking of the side chains at the design positions.

In this study, 55 independent design trajectories were run for each docked configuration of each protein building block. The 55 trajectories sampled the rigid body DOFs in a 2 Å by 2° grid around the starting configuration in 0.1 Å and 0.5° increments. The many thousands of design models for each symmetric architecture were filtered based on interface size, shape complementarity, and predicted binding energy. Designs that passed all three filters were analyzed in detail using the above metrics in combination with a per-residue measure of the preconfiguration of each designed side chain in the unbound state (38) and a per-residue measure

of the change in the predicted binding energy of the complex upon mutation of each designed side chain to alanine. During this stage of the design of T3-08, the side chain of arginine 27 was kept in its native configuration; its guanidinium group forms three side chain-backbone hydrogen bonds within the monomer, two of which cap an adjacent helix, that were deemed important to the stability of the monomer. The protein folding video game foldit (30) was used to interactively test mutations in certain O3 designs.

An example of the command line options used with the matdes_design application is shown here:

```
-database <rosetta_database_path>
-s <input_structure>
-symmetry_definition input/T3.sym
-ex1 -ex2
-holes::dalphaball <path_to_DAlphaBall.gcc>
-matdes::num_subs_building_block 3
-matdes::num_subs_total 12
-matdes::prefix T3_
-matdes::pdbID 3ftt
-matdes::radial_disp 32.0
-matdes::angle 110.5
-matdes::design::grid_size_angle 2
-matdes::design::grid_size_radius 2
-matdes::design::grid_nsamp_angle 5
-matdes::design::grid_nsamp_radius 11
```

Source Code

Source code is freely available to academic users through the Rosetta Commons agreement (http://www.rosettacommons.org). Revision 42921 was used to design the materials presented.

Amino acid sequences

Point mutants referred to in the main text or Fig. 5.7 contain no changes from the sequences below other than those indicated by name. The amino acid sequences of the two C-terminal tags used in this study are also presented.

>3n79-wt
MSQAIGILELTSIAKGMELGDAMLKSANVDLLVSKTICPGKFLLMLGGDIGAIQQAIETGTSQAGEMLVDSLVLANIHPS
VLPAISGLNSVDKRQAVGIVETWSVAACISAADRAVKGSNVTLVRVHMAFGIGGKCYMVVAGDVSDVNNAVTVASES
AGEKGLLVYRSVIPRPHEAMWRQMVEG

>O3-33
MSQAIGILELTSIAAGMELGDAMLKSANVDLLVSKTISPGKFLLMLGGDIGAIQQAIETGTSQAGELLVDSLVLANIHPSV
LPAISGLNSVDKRQAVGIVETWSVAACISAADRAVKGSNVTLVRVHMAFGIGGKCYMVVAGDVSDVALAVTVASSSA
GAYGLLVYASLIPRPHEAMWRQMVEG

>3ftt-wt
MTEKEKMLAEKWYDANFDQYLINERARAKDICFELNHTRPSATNKRKELIDQLFQTTTDNVSISIPFDTDYGWNVKLGK
NVYVNTNCYFMDGGQITIGDNVFIGPNCGFYTATHPLNFHHRNEGFEKAGPIHIGSNTWFGGHVAVLPGVTIGEGSVIG
AGSVVTKDIPPHSLAVGNPCKVVRKIDNDLPSETLNDETIK

>T3-08
MTEKEKMLAEKWYDANFDQTLINERLRAKVICFALNHTNPSATLKRKVLIDALFQTTTDNVSISIPFDTDYGWNVKLGK
NVYVNTNCYFMDGGQITIGDNVFIGPNCGFYTATHPLNFHHRNEGFEKAGPIHIGSNTWFGGHVAVLPGVTIGEGSVIG
AGSVVTKDIPPHSLAVGNPCKVVRKIDNDLPSETLNDETIK

>T3-10
MTEKEKMLAEKWYDANFDQTLINERLRAKVICFALNHTNPVATMMRKVLIDALFQTTTDNVSISIPFDTDYGWNVKLG
KNVYVNTNCYFMDGGQITIGDNVFIGPNCGFYTATHPLNFHHRNEGFEKAGPIHIGSNTWFGGHVAVLPGVTIGEGSVI
GAGSVVTKDIPPHSLAVGNPCKVVRKIDNDLPSETLNDETIK

>A1 tag (for fluorescent labeling)
LEGGDSLDMLEWSL

>His_6 tag
LEHHHHHH

Protein expression, fluorescent labeling, and purification

Codon-optimized genes encoding the designed and corresponding wild-type proteins were constructed from sets of purchased oligonucleotides (Integrated DNA Technologies) by recursive PCR (39) and ligated into the pET29b expression vector (Novagen) using the *NdeI* and *XhoI* restriction endonuclease sites. The genes included the A1 peptide tag (31) for fluorescent labeling with the AcpS system at either the N terminus or the C terminus, and all genes ended with a stop codon to prevent translation of the C-terminal $(His)_6$ tag. $(His)_6$-tagged versions of the proteins were constructed by subcloning the genes without the A1 tags or stop codons into pET29b using the *NdeI* and *XhoI* restriction sites. Site-directed mutagenesis was performed to generate plasmids for the expression of mutant O3-33 and T3-08 constructs and T3-10.

Expression plasmids were transformed into BL21(DE3) *E. coli* cells. Cells were grown in LB medium supplemented with 50 mg L^{-1} of kanamycin (Sigma) at 37 °C until an OD_{600} of 0.8 was reached. Protein expression was induced by addition of 0.5 mM isopropyl-thio-β-D-galactopyranoside (Sigma) and allowed to proceed for 5 h at 30 °C (O3-33 and T3-08) or 18 °C (T3-10) before cells were harvested by centrifugation.

For fluorescence analysis, cells were lysed by sonication in 25 mM TRIS pH 7.5, 300 mM NaCl, 8.5 mM $MgCl_2$, 1 mM dithiothreitol (DTT) supplemented with 1 mM phenylmethanesulfonyl fluoride, and the lysates cleared by centrifugation. Supernatants were transferred to fresh tubes and the proteins were labeled by the addition of CoA-488 (New England Biolabs) to a final concentration of 5 μM and the *E. coli* acyl carrier protein synthase AcpS (New England Biolabs) to a final concentration of 1 μM (31). The labeling reaction was allowed to proceed at 37 °C for 30 min, after which the labeled lysates were analyzed by native PAGE or fluorescence SEC. SDS-PAGE of the labeled lysates was performed to determine the extent to which the endogenous *E. coli* proteins were labeled; very little background labeling was observed. Gels were imaged fluorescently using a Fotodyne gel imaging system equipped with appropriate LEDs and filters.

For purification, cells were lysed by sonication in 50 mM TRIS pH 8.0, 250 mM NaCl, 1 mM DTT, 20 mM imidazole supplemented with 1 mM phenylmethanesulfonyl fluoride, and the lysates were cleared by centrifugation and filtered through 0.22 μM filters (Millipore). The designed, mutant, or corresponding wild-type proteins were purified from the filtered supernatants by nickel affinity chromatography on HisTrap HP columns (GE Life Sciences) and eluted using a linear gradient of imidazole (0.02–0.5 M). Fractions containing pure protein of interest were pooled, concentrated, and analyzed directly (Fig. 5.2B-H in the main text; see next

section) or further fractionated on a Superdex 200 30/100 gel filtration column (GE Life Sciences) using 25 mM TRIS pH 8.0, 150 mM NaCl, 1 mM DTT as running buffer. Gel filtration fractions containing pure protein in the desired assembly state were pooled, concentrated using centrifugal filter devices (Sartorius Stedim Biotech), and stored at room temperature.

Analytical size exclusion chromatography

All analytical SEC was performed on a Superdex 200 30/100 gel filtration column (GE Life Sciences) using 25 mM TRIS pH 8.0, 200 mM NaCl, 1 mM DTT as running buffer. Elution of fluorescently labeled proteins was observed using an Agilent 1260 Infinity Fluorescence Detector equipped with a 8 μL flow cell; the excitation and emission wavelengths were set to 506 nm 550 nm, respectively. Nickel-purified 3n79-wt, O3-33, and O3-33(A167R) were loaded for SEC analysis at a concentration of 0.5 mg mL^{-1}; nickel-purified 3ftt-wt, T3-08, T3-08(A52Q), and T3-10 were loaded at a concentration of 4 mg mL^{-1}. It was found that T3-10, when applied to the sizing column immediately after nickel purification, eluted as a mixture of two peaks corresponding to 12 and about 9 subunits; incubation at room temperature or 37 °C gradually decreased the fraction of protein in the 9mer peak and increased that in the 12mer peak over a period of days. The data shown in Figure 5.2F in the main text was collected after incubating pooled T3-10 nickel fractions at 37 °C for 48 hours. The apparent molecular weights of the designed proteins were estimated by comparison to the corresponding wild-type trimers and a set of globular protein standards.

Analytical ultracentrifugation

The sedimentation profile of O3-33 was obtained in a buffer of 50 mM potassium phosphate monobasic, 50 mM sodium phosphate dibasic, 25 mM ammonium sulfate, pH 7.5 using a Beckman Proteome XL-I instrument. Two concentrations (4.8 and 2.4 mg mL^{-1}) of O3-33 were dialyzed overnight prior to the experiment in order to ensure buffer identity between sample and reference solutions. A sedimentation velocity run was performed using two sector AUC cells with a charcoal-epon centerpiece. 440 μL of sample and reference were loaded into each cell and the cells were centrifuged at 10,000 rpm in an 8-hole rotor. Absorbance at 280 nm was used to monitor the protein. The sample was observed to have fully sedimented to the bottom of the cell in 14 hours; the experiment was allowed to continue for an additional 7 hours to ensure completion. The data was analyzed using a c(s) distribution analysis with the program SEDFIT (40). A sedimentation distribution from 0 to 200 s was determined using a resolution of 0.4 s. The program SEDNTERP was used to calculate buffer viscosity and density using the information provided.

In the c(s) distribution analysis, the experimental boundary profile was fitted, using both a Simplex and a Marquadt-Levenberg algorithm, to the equation

$$a(r,t)=c(s)\cdot\chi 1(s,D,r,t)\cdot ds,$$

where c(s) is the concentration at each value of the sedimentation coefficient in the distribution and $c_1(s,D,r,t)$ is the normalized Lamm equation solution for an individual species. Since multiple best fit solutions can result from this analysis the simplest distribution that fits the data was determined by performing a Maximum Entropy regularization of the fit using a 95% confidence level (P-value = 0.68). Integration of each peak observed in the c(s) distribution provides the loading signal of that species as well as a precise signal averaged s-value. A

molecular weight estimate can be determined by using the recovered sedimentation coefficients corrected for density and viscosity changes at 20 °C in the Svedberg equation,

$$sD=M(1-v)RT$$

Where s is the sedimentation coefficient in units of 10^{-13} sec, D is the diffusion coefficient (from which the Stokes radius is derived), M is the molar mass, v is the partial specific volume, R is the gas constant, and T is the temperature.

Negative-stain electron microscopy (EM)

Samples of purified O3-33 or T3-10 were applied to freshly glow discharged carbon coated 400 mesh copper grids (Ted Pella, Inc.) and stained with 0.075% uranyl formate as described previously (41). Samples were viewed either on a 100 kV or 120 kV transmission electron microscope (FEI, Hillsboro, OR). Images were recorded using a bottom mount Gatan slow scan charge coupled device camera at a nominal magnification of 28,000x and 70,742x at the specimen level for O3-33 and T3-10, respectively. For averaging, 9,411 unique T3-10 particles were selected using Ximdisp (42) and extracted into 100x100 pixel subframes. The stacked subframes were subjected to several rounds of MSA and MRA using IMAGIC (43) to generate class averages.

Electron cryomicroscopy data collection and single particle reconstruction

Vitrified samples of O3-33 were prepared by applying 3 μL of 0.3 mg mL^{-1} protein onto a Quantifoil holey carbon grid, blotting with filter paper, and plunging into liquid ethane using an FEI Vitrobot. Frozen samples were loaded onto a Gatan cryo-holder and inserted into a FEI Tecnai F20 operating at 200 kV equipped with a field emission gun. Images were collected at a

nominal magnification of 100,000x using an 8kx8k TVIPS F816 camera and UCSFImage, a computer assisted data collection program (courtesy Dr. Yifan Cheng, UCSF). Images were binned four times yielding a pixel size of 2.948 Å pixel^{-1} to achieve faster processing of data. Approximately 42,000 particles were selected from 565 images using Electron Micrograph Utility (cryoem.ucsf.edu). Class averages were determined using five consecutive rounds of MSA (multivariate statistical analysis) and MRA (multireference alignment) in IMAGIC (43). CTF parameters for each image were determined using CTFFIND3 (44). An initial model for refinement and three-dimensional reconstruction was generated by calculating a map filtered to 40 Å from the O3-33 design model using the pdb2mrc command in EMAN (45). Initial parameters were generated during cycles of randomized search and refinement using FREALIGN v8.08 (46) using data binned twice. After initial parameters were determined, consecutive cycles of local refinement and reconstruction with applied octahedral symmetry were carried out until there was no improvement in the alignment. Resolution of the three-dimensional model was determined using the 0.5 cutoff of the Fourier shell correlation curve, which indicated about 20 Å. The reconstructions were normalized using the proc3d command in EMAN (45). The O3-33 design model was fit into the final electron density map using UCSF Chimera (47).

Crystallization of O3-33 in space groups R32 and P4

Initial crystallization screens of the His-tagged protein were set up by the Crystallization Core Facility at UCLA. All crystals were obtained by hanging drop vapor diffusion at room temperature. The protein was in a buffer containing 25 mM TRIS pH 8.0, 150 mM sodium chloride, and 1 mM DTT at a concentration of 28 mg mL^{-1}. Crystals of various morphologies

appeared within hours under numerous conditions. The conditions for crystals of two different morphologies were optimized.

The best quality crystals obtained were hexagonal plates and belonged to space group R32 (O3-33-R32). The optimized reservoir solution was composed of 0.8 M sodium phosphate monobasic, 1.2 M potassium phosphate dibasic, and 0.1 M sodium acetate pH 4.5. The drop was composed of 0.5 μL of protein, 0.4 μL of reservoir, and 0.1 μL of 0.1M L-proline as an additive (final concentration 10 mM L-proline). The crystals reached full size (0.3 x 0.3 x 0.05 mm) in approximately 1 day and diffracted to 2.35 Å. The crystals were cryo-protected by soaking for 2 minutes in 2 M lithium sulfate. The asymmetric unit contained 8 molecules.

A second crystal form appeared as cubes or rectangular blocks and belonged to space group P4 (O3-33-P4). The optimized reservoir solution was composed of 1.0 M sodium chloride, 0.35 M lithium sulfate, and 0.1 M sodium acetate pH 4.5. The drop was composed of 1.5 μL of protein and 1.5 μL of reservoir. The crystals reached full size (0.2 x 0.2 x 0.3 mm) in approximately 3 days and diffracted to 3.15 Å. The crystals were cryo-protected by soaking for 30 seconds in 2 M lithium sulfate. The asymmetric unit contained 12 molecules. Analysis of the diffraction pattern from this crystal form indicated the presence of a translocation disorder with translocation vector of 0.500, 0.500, +/-0.4015. The effects of the disorder were diminished using the procedures described in ref. 48.

Crystallization of T3-08

Crystal screening was performed as described for O3-33. The protein was in a buffer containing 25 mM TRIS, pH 8.0, 150 mM sodium chloride, and 1 mM DTT at a concentration of 24 mg mL^{-1}. Crystals appeared as trigonal pyramids or bi-pyramids and belonged to space group

F23. The reservoir solution was composed of 2.0 M sodium chloride and 0.1 M sodium acetate pH 4.6. The drop was composed of 0.2 μL of protein, 0.2 μL of reservoir. The crystals reached full size (0.2 x 0.2 x 0.2 mm) in approximately 2 days and diffracted to 3.35 Å. The crystals were cryo-protected by soaking for 30 seconds in a solution containing 65% reservoir and 35% ethylene glycerol. The asymmetric unit contained 1 molecule.

Crystallization of T3-10

Crystal screening was performed as described for O3-33. The protein was in a buffer containing 25 mM Tris, pH 7.5, 150 mM sodium chloride, and 1 mM DTT at a concentration of 9 mg mL^{-1}. Crystals appeared as quadrilateral plates and belonged to space group C2. The reservoir solution was composed of 7.5% polyethylene glycol 2000 monomethylether and 0.1 M sodium citrate pH 6.0. The drop was composed of 3 mL of protein, 3 mL of reservoir. The crystals reached full size (0.3 x 0.3 x 0.1 mm) in approximately 4 days and diffracted to 2.25 Å. The crystals were cryo-protected by soaking for 30 seconds in a solution containing 65% reservoir and 35% glycerol. The asymmetric unit contained 6 molecules.

Data collection and structure determination

Diffraction data sets were collected from O3-33-R32, O3-33-P4, and T3-08 at the Advanced Photon Source (APS), beamline 24-ID-C, equipped with an ADSC Quantum 315 CCD detector. The data set from the T3-10 crystal was collected at the same facility, but with a Pilatus pixel detector. All data were collected at 100 K at the resolution limits given above. Cryo-protection was performed as described above. Data were collected using 1.0° oscillations. The wavelength of the X-rays used for diffraction data from O3-33-R32, O3-33-P4, T3-08, and

T3-10 were 0.9794 Å, 0.9201 Å, 0.9791 Å, and 0.9793 Å, respectively. Data reduction and scaling for O3-33-R32 and T3-10 were performed using DENZO/SCALEPACK (49). XDS/XSCALE was used to process and scale the remaining data sets (50).

The crystal structures were determined by molecular replacement using the program PHASER (51). The search model used for the O3-33 structures was based on PDB entry 3N79, PduT C38S mutant from *Salmonella enterica typhimurium* (52). The search model used for the T3-08 structure was based on PDB entry 3V4E, galactoside O-acetyltransferase from *Staphylococcus aureus*. The search model for the T3-10 structure was the refined T3-10 structure. The models were refined using REFMAC5 (53) and Buster/TNT (54) with TLS parameterization of domain disorder (55). After each refinement step, the model was visually inspected in COOT (56), using both 2Fo-Fc and Fo-Fc difference maps. The models were validated with the following structure validation tools: PROCHECK (*57*), ERRAT (*58*), and VERIFY3D (59). Greater than 98% of the residues are within the most favoured and additional allowed regions of the Ramachandran plot for all structures. The O3-33 structures have an additional 1.4% of residues in the generous and disallowed regions, corresponding to well defined areas of the electron density map. The ERRAT scores for O3-33-R32, O3-33-P4, T3-08, T3-10 are 97.8%, 97.8%, 84.2%, and 81.1% respectively, indicating that these percentages of residues fall below the 95% confidence limit of being erroneously modeled. Data collection and refinement statistics are reported in Table 2. The coordinates of the final models and the merged structure factors have been deposited in the Protein Data Bank with PDB codes 3VCD, 4DDF, 4DCL, and 4EGG. All images of protein structures were made using PyMOL (60).

ACKNOWLEDGEMENTS

We thank Jason Navarro for assistance with protein crystallization at the UCLA crystallization core facility, which is supported by DOE-BER Grant Number DE-FC03-02ER63421**.** We are grateful to Duilio Cascio and the 24-ID-C beamline staff for their assistance in data collection. This work is based upon research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines, which are supported by award RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under contract no. DE-AC02-06CH11357. AUC was performed in the Bioanalytical Pharmacy Core at the University of Washington, which is supported by the Washington State Life Sciences Discovery Fund and the Center for the Intracellular Delivery of Biologics. We are grateful to Matt Iadanza for EM analysis of T3-10. We would like to thank Yifan Cheng and Xueming Li (UCSF) for giving us access to their electron cryomicroscope for data collection, for helpful discussions, and for sharing scripts with us. We also thank Nikolas Grigorieff (Brandeis) for helpful discussions. The Gonen laboratory is supported by HHMI. Work by NPK, WS, and DB was supported by DOE, HHMI, and the International Aids Vaccine Initiative. Coordinates and structure factors were deposited in the Protein Data Bank with the accession codes 3VCD (O3-33, R32 crystal form), 4DDF (O3-33, P4 crystal form), 4DCL (T3-08), and 4EGG (T3-10).

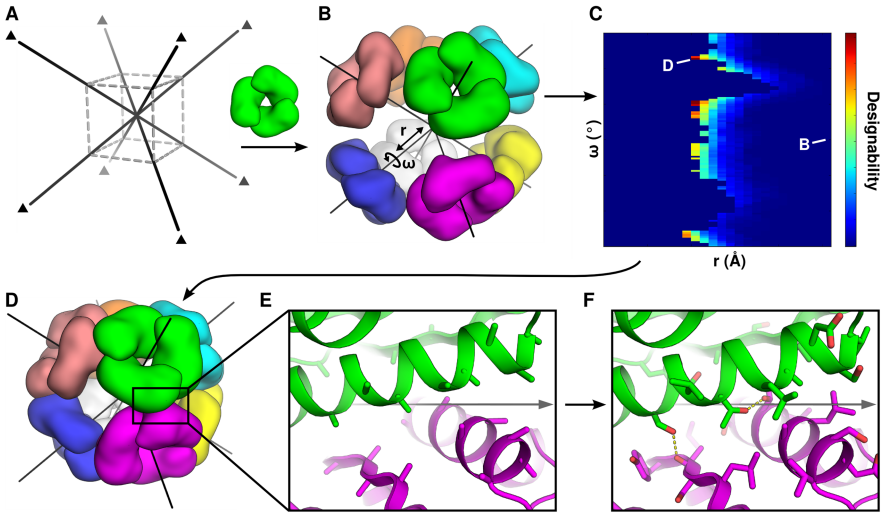

Figure 5.1 General approach to designing self-assembling protein nanomaterials. (**A**) First, a target symmetric architecture is chosen. Octahedral point group symmetry is used in this example; the three-fold rotational axes are marked here by triangles and shown as black lines throughout. The dashed cube is shown to orient the viewer. A symmetric oligomer which shares an element of symmetry with the target architecture, here a C3 symmetric trimer (green), is selected as a building block. (**B**) Multiple copies of the building block are symmetrically arranged in the target architecture by aligning their shared symmetry axes. The pre-existing organization of the oligomeric building block fixes several (in this case four) rigid body degrees of freedom (DOFs). The two remaining DOFs, radial displacement (r) and axial rotation (ω), are indicated. (**C**) Symmetrical docking is performed by systematically varying the two DOFs (moves are applied symmetrically to all subunits) and computing the suitability of each configuration for interface design (red: more suitable; blue: less suitable). Points corresponding to the docked configurations in panels (**B**), in which the building blocks are not in contact, and (**D**), a highly complementary interface, are indicated. (**E**) Closer view of the interface in (**D**). The interface lies on an octahedral two-fold symmetry axis shown as a grey line. In all steps before interface design, only backbone (shown in cartoon) and carbon beta (shown in sticks) atoms are considered. (**F**) Sequence design calculations are used to create low-energy protein-protein interfaces that drive self-assembly of the desired material. Designed hydrogen bonds across the interface are indicated by dashed lines.

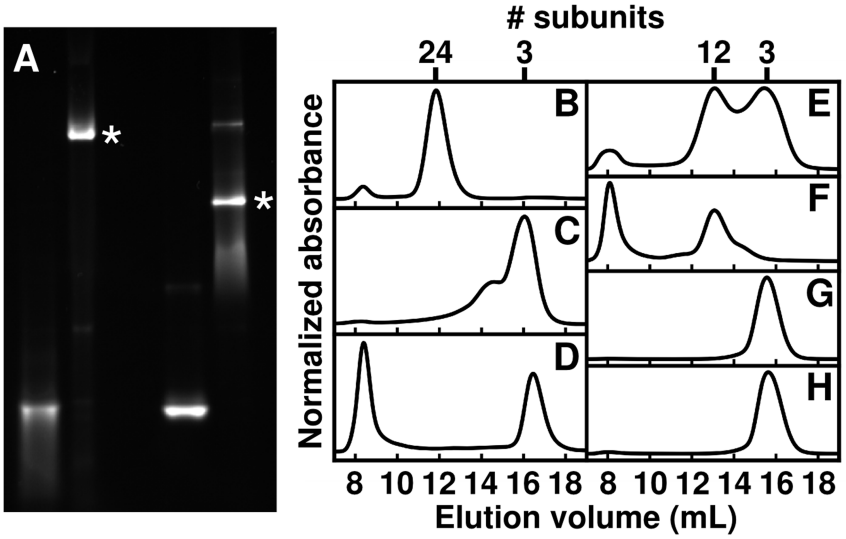

Figure 5.2 Experimental characterization of O3-33, T3-08, and T3-10. (**A**) Native PAGE of fluorescently labeled (from left) 3n79-wt, O3-33, 3ftt-wt, and T3-08 in lysates. Bands corresponding to the designed octahedral (O3-33) and tetrahedral (T3-08) assemblies are indicated with asterisks. SEC chromatgrams of nickel-purified (**B**) O3-33, (**C**) 3n79-wt, (**D**) O3-33(Ala167Arg), (**E**) T3-08, (**F**) T3-10, (**G**) 3ftt-wt, and (**H**) T3-08(Ala52Gln) collectively demonstrate that the assembly of the designed proteins is a result of the designed interfaces.

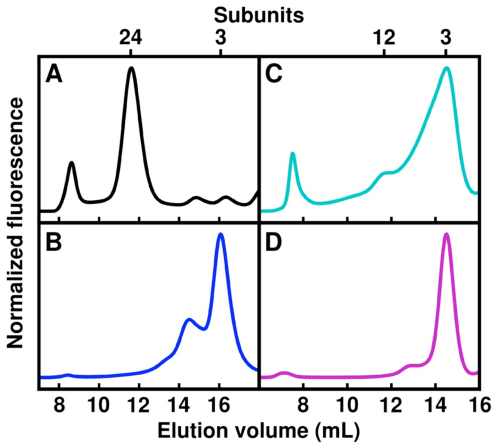

Figure 5.3 Fluorescent SEC chromatograms of (A) O3-33, (B) 3n79-wt, (C) T3-08, and (D) 3ftt-wt in labeled cell lysates. The chromatograms of the designed proteins reveal peaks at earlier elution volumes than the corresponding wild-type trimers, mirroring the result observed by native PAGE in Fig. 1A of the main text. The elution volumes for the chromatograms in (**C**) and (**D**) appear shifted to the left relative to those in (**A**) and (**B**) because the column was connected in a different manner such that dead volume before and after the column was minimized; each pair of chromatograms is internally consistent. The data shown in (**A**) and (**B**) is used again in Fig. 5.7 for comparison to several point mutants of O3-33.

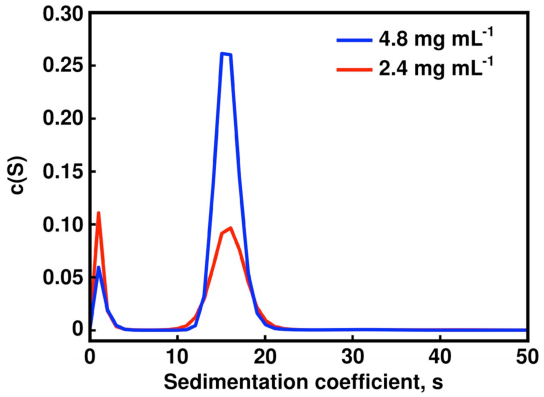

Figure 5.4 Analytical ultracentrifugation analysis of O3-33. c(s) distribution of O3-33 at two concentrations determined using SEDFIT at a resolution of 1.0 s. The distribution for each sample shows a single peak corresponding to an average s of 17.2, yielding a Stokes radius of 7.3 nm for the designed material.

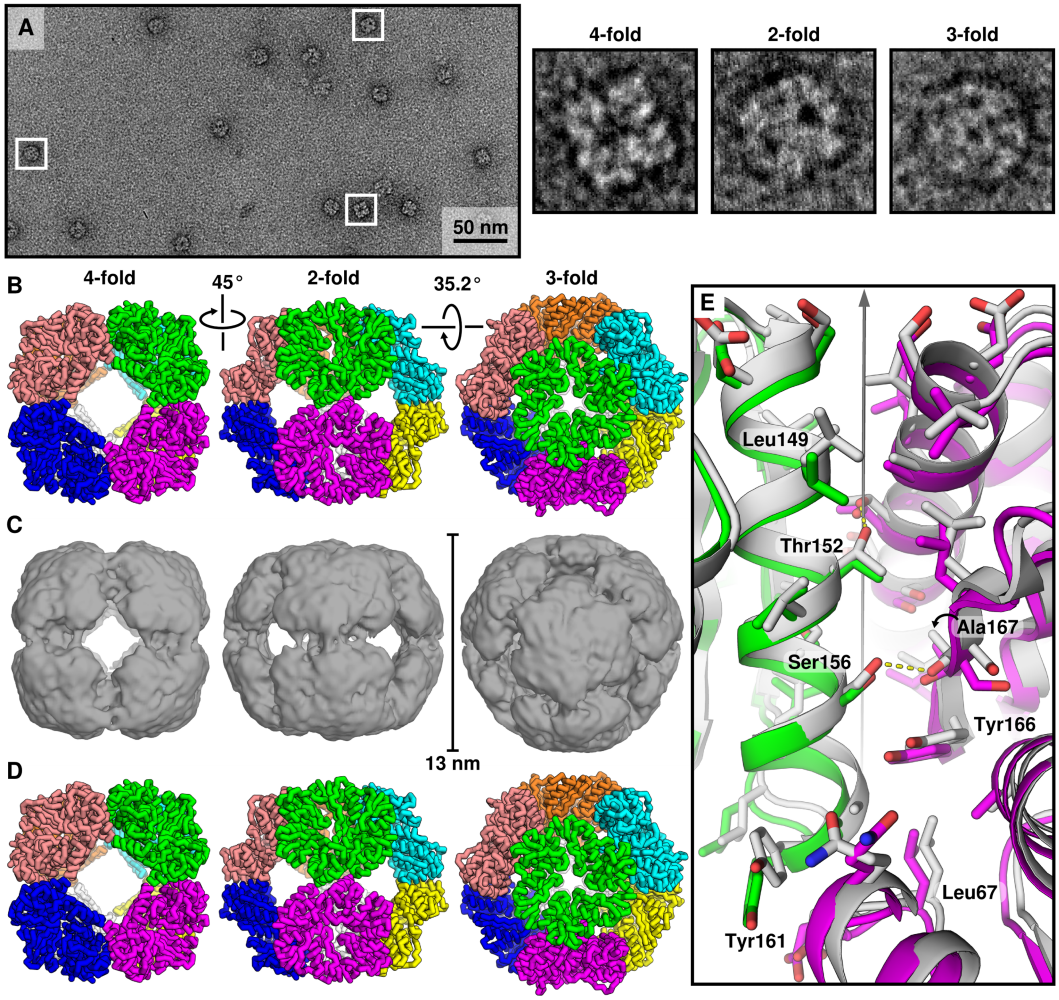

Figure 5.5 Structural characterization of O3-33. (**A**) A representative negative stain electron micrograph of O3-33. Selected particles (boxed in white) that resemble views of the design model along its 4-fold, 2-fold, and 3-fold rotational axes, shown in (**B**), are enlarged at right. (**B**) The O3-33 design model, depicted in ribbon format. Each trimeric building block is shown in a different color. (**C**) The density map from a 20 Å resolution cryo-EM reconstruction of O3-33 clearly recapitulates the architecture of the design model. (**D**) The crystal structure of O3-33 (R32 crystal form). Images in (**B**) to (**D**) are shown to scale along the three types of symmetry axes present in point group O. (**E**) The designed interface in O3-33, highlighting the close agreement between the crystal structure (green and magenta) and the design model (white). Oxygen atoms are red; nitrogens, blue. Hydrogen bonds between the building blocks are shown as yellow dashes, and an octahedral 2-fold rotational axis that passes through the interface is shown as a gray line. Residues in which substitution disrupted self-assembly (see Fig. 5.7) are labeled.

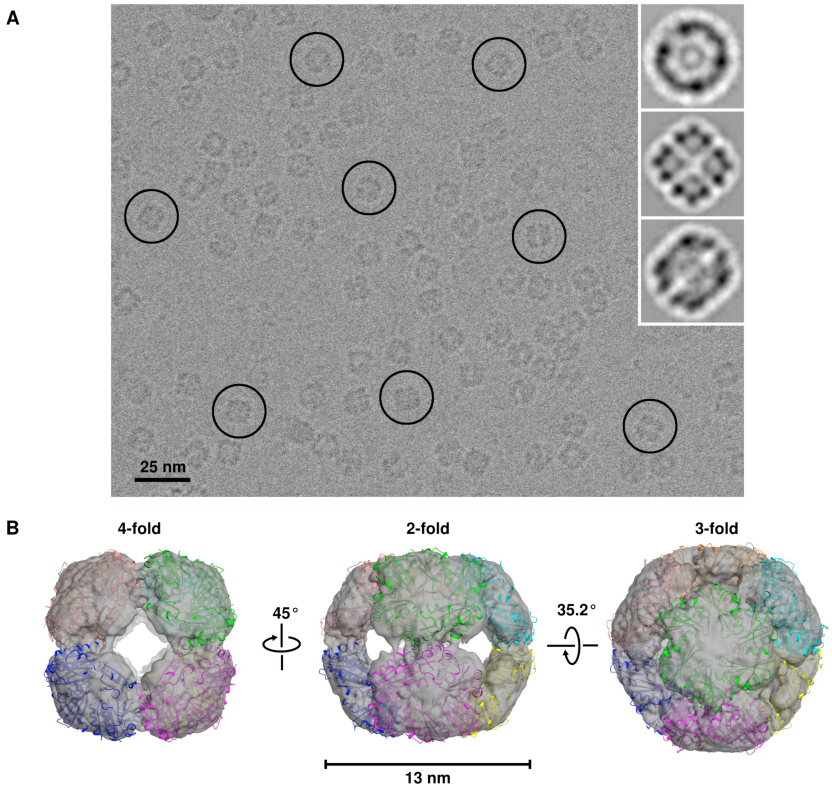

Figure 5.6 Electron cryomicroscopy of O3-33. (**A**) Representative electron micrograph of O3-33 frozen in vitrified ice. Inset, three classes of projection averages showing views of the three-fold axis (top), four-fold axis (middle) and two-fold axis (bottom) calculated using the MSA and MRA programs in IMAGIC (*41*). (**B**) The density map from the single particle reconstruction is shown as a gray surface as in Fig. 5.5C in the main text, with the design model, represented in cartoon, fit into the map. Each trimeric building block is shown in a different color.

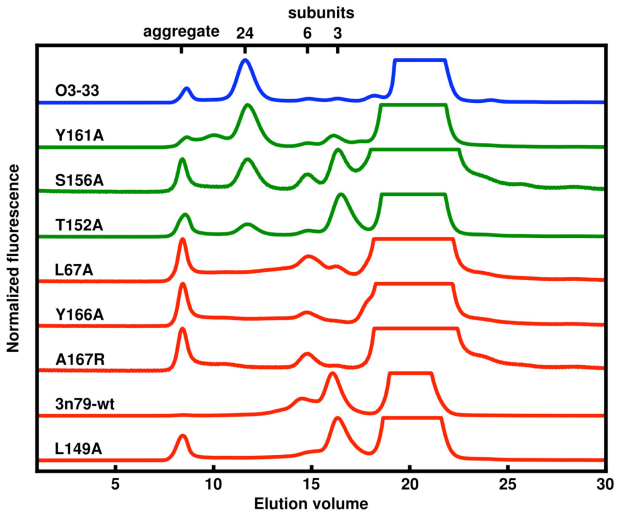

Figure 5.7 Fluorescent SEC chromatograms of 3n79-wt, O3-33, and various interface point mutants in lysates. Each trace is labeled with the protein name (3n79-wt, O3-33) or the mutation; all mutants are derived from O3-33. The mutations are classified as disrupting (green traces) or severely disrupting (red traces) depending on the extent to which they disrupt self-assembly of the designed 24-subunit complex. The plateau-like peak around 20 mL is unbound fluorophore from the labeling reaction saturating the fluorescence detector.

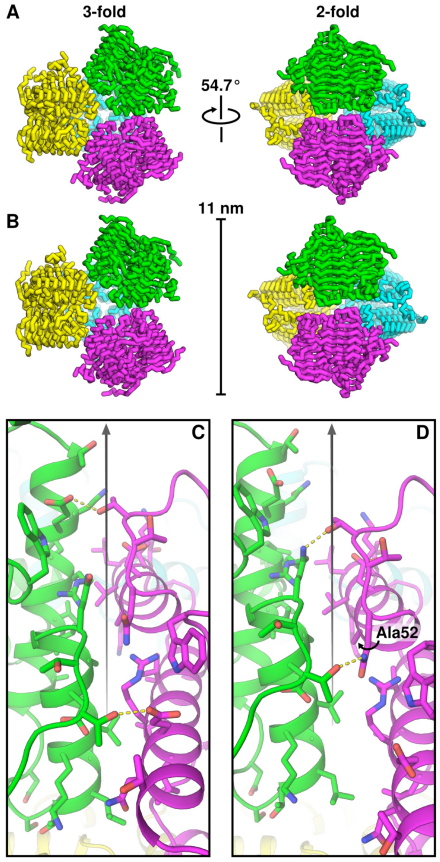

Figure 5.8 Structural characterization of T3-08. Images in (**A**) and (**B**) are shown to scale. (**A**) The T3-08 design model, depicted along its two types of symmetry axes. Each trimeric building block is shown in a different color. (**B**) The crystal structure of T3-08. (**C**) The designed interface in the T3-08 design model, depicted as in Fig. 5E of the main text. (**D**) The interface in the T3-08 crystal structure. Each trimeric building block is slightly rotated about its three-fold axis relative to the design model, which subtly alters the atomic interactions at the designed interface. The rotation of the building blocks is symmetric and the tetrahedral symmetry of the material is therefore maintained. Alanine 52, which disrupts assembly upon reversion to the wild-type glutamine, is labeled. Images in (**C**) and (**D**) are shown from the same viewpoint in separate panels for clarity.

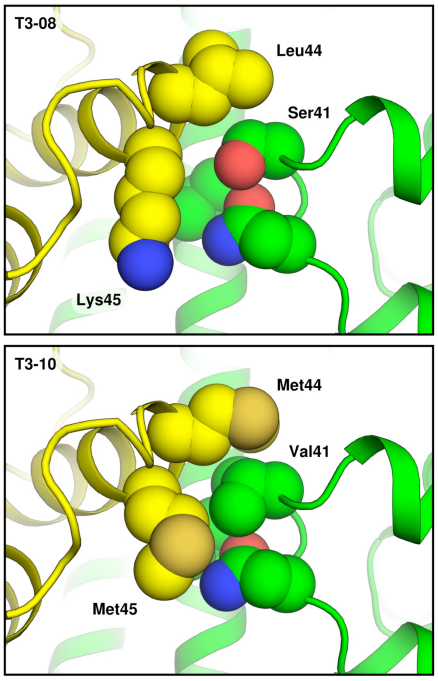

Figure 5.9 Mutations in T3-10 that stabilize the original designed configuration. The design models of T3-08 (top) and T3-10 (bottom) are shown, focusing on the labeled mutations which result in greater hydrophobic packing across the designed interface.

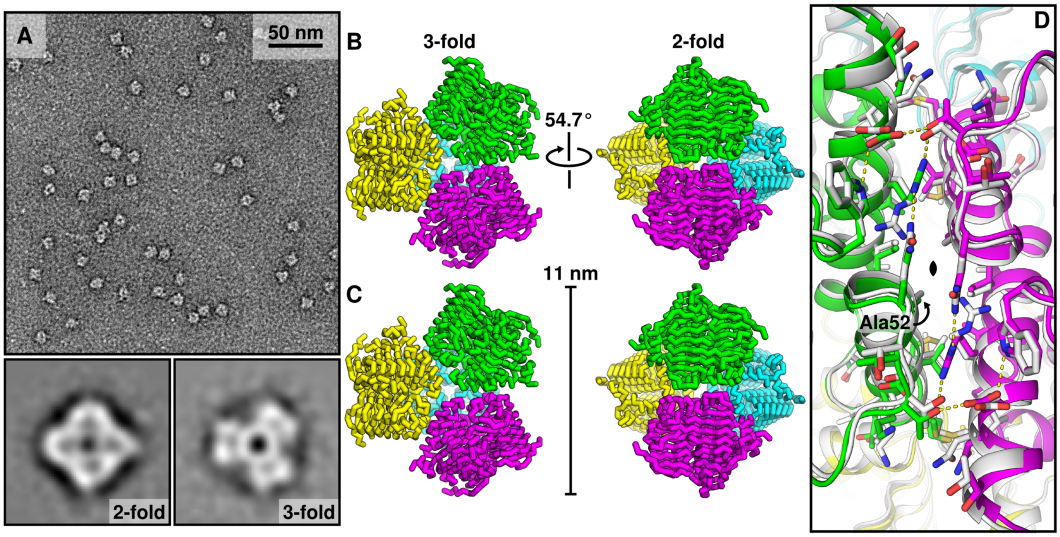

Figure 5.10 Structural characterization of T3-10. (**A**) A representative negative stain electron micrograph of T3-10. At bottom, averages of the particles resemble views of the design model along its 2-fold and 3-fold rotational axes, shown in (B). (**B**) Backbone representation T3-08/T3-10 design model, depicted as in Fig. 5B. (**C**) The T3-10 crystal structure. Images in (**B**) and (**C**) are shown to scale along the two types of symmetry axes present in point group T. (**D**) The designed interface in T3-10, revealing the close agreement of the crystal structure (green and magenta) to the design model (white). A network of polar interactions observed in the crystal structure at the designed interface is indicated by yellow dashes. The interface is viewed along an indicated tetrahedral 2-fold rotational axis. Alanine 52, which when mutated to glutamine in T3-08 disrupts assembly of the designed material, is labeled.

Table 5.1 Mutations and experimental screening results for all designed proteins

Design name	Scaff-old PDB	(#) Mutations*	Soluble expr-ession	Assembly
O3-01	1h9j	(8) K17Y,D26Y,K34Y,A36V,Q48F,A50W,A51V,K53L	no	N/A
O3-02	1h9j	(8) K17Y,D26Y,K34A,A36V,Q48F,A50W,A51V,K53L	N/A†	N/A
O3-03	1nza	(8) Y41F,W43A,E46V,E49A,D50V,K72W,A73L,Y77L	yes	no
O3-04	1nza	(8) Y41F,W43A,E46V,E49A,D50V,K72A,A73L,Y77L	yes	no
O3-05	1nza	(10) E13D,Y41F,W43A,Q44N,E46V,E49A,D50V,K72A,A73L,Y77L	yes	no
O3-06	1qd9	(12) I14L,E40A,M41L,N43V,D45Y,K47V,D80R,Q83L,E86L,E89L,K112F,D113V	no	N/A
O3-07	1ve0	(6) R15V,D72V,I79R,D81L,D120A,R123E	yes	no
O3-08	1ve0	(6) R15V,D72A,I79R,D81L,D120A,R123E	yes	no
O3-09	1x25	(12) V13F,R37A,E40A,I41L,V42I,K43I,D45V,K47A,N85D,D86V,A92Y,D113V	yes	no
O3-10	1x25	(12) V13F,R37A,E40A,I41L,V42I,K43S,D45V,K47A,N85D,D86V,A92Y,D113V	yes	no
O3-11	2flz	(11) R11F,D61Y,D87N,V91W,E94F,D96A,R97V,K98V,H99A,E136L,F139A	no	N/A
O3-12	2flz	(12) R11F,D61Y,D87N,V91W,E94F,D96A,R97V,K98V,H99A,E136L,F139A,T142A	no	N/A
O3-13	2flz	(8) D76M,Q80S,Q83L,D87F,V91Q,H124Y,E127W,N131A	no	N/A
O3-14	2flz	(7) D76M,Q80S,Q83L,D87A,V91Q,H124Y,N131A	yes	no
O3-15	2flz	(8) D76M,Q80A,Q83L,D87F,V91Q,H124Y,E127W,N131A	yes	no
O3-16	2fvh	(8) R51A,A54L,E55A,R66A,D67N,E71M,E75Y,D88L	no	N/A
O3-17	2fvh	(9) R29S,R51A,A54L,E55A,R66A,D67S,E71M,E75Y,D88L	no	N/A
O3-18	2p6c	(8) K14Y,K15I,R16A,E18A,Y73V,D76L,H80Y,D123Y	yes	no
O3-19	2p6c	(7) K14Y,K15A,R16A,Y73V,D76L,H80Y,D123Y	yes	no
O3-20	2p6c	(8) K14Y,K15I,R16A,E18A,Y73V,D76L,H80Y,D123Y	yes	no
O3-21	3jv1	(9) E74Y,E140I,D180A,D184F,T187Y,S191Y,D197Y,T198A,N201A	yes	no
O3-22	3jv1	(10) E54D,E74Y,E140I,D180A,D184F,T187Y,S191Y,D197Y,T198A,N201A	yes	no
O3-23	3jv1	(13) R51I,E54D,E74Y,E140I,D180A,H181A,D184F,S185A,T187Y,S191Y,D197Y,T198A,N201A	yes	no
O3-24	3jv1	(9) E54I,E74Y,D138A,E140V,D180R,H181L,D184F,T187F,D197Y	no	N/A

Table 5.1 continued

O3-25	3jv1	(11) E54I,E74Y,D138A,E140V,D180R,H181L,D184F,T187F,D197Y,T198A,N201A	no	N/A
O3-26	3m1x	(9) D43A,K47V,T48W,E50M,E51A,K85A,V89A,E96L,K115A	no	N/A
O3-27	3m1x	(8) K47V,T48W,E50M,E51A,K85A,V89A,E96L,K115Y	no	N/A
O3-28	3m1x	(10) D43A,K47V,T48W,E50M,E51A,K54R,K85A,V89A,E96L,K115A	no	N/A
O3-29	3n3f	(7) A5L,D11V,K15M,H17I,L18Y,I20Y,D29A	no	N/A
O3-30	3n3f	(10) A5L,S7Q,D11V,K15M,H17I,L18Y,I20Y,E21S,R28V,D29A	no	N/A
O3-31	3n79	(10) K15A,E66A,M67L,N148A,N149L,E156M,E160A,K161Y,R167A,V169L	yes	no
O3-32	3n79	(10) K15A,E66D,M67L,N148A,N149L,E156M,E160A,K161Y,R167M,V169L	no	N/A
O3-33	3n79	(9) K15A,M67L,N148A,N149L,E156S,E160A,K161Y,R167A,V169L	yes	yes
T3-01	1qah	(5) Q44L,E52N,N86A,T90L,K116S	yes	no
T3-02	1qd9	(8) N43A,K47N,E48I,E82A,Q83S,A85N,R109S,D113A	yes	no
T3-03	1x25	(9) I41L,K43S,K47S,N82A,M83A,S89N,R109A,K112N,D113A	yes	no
T3-04	1x25	(9) I41L,K43S,K47S,N82A,M83A,S89N,R109A,K112L,D113A	yes	no
T3-05	2cwj	(14) V8N,E37A,E38A,F41T,K42A,E43S,D75T,I76A,S77A,R78S,S80N,E81L,A104S,L107N	no	N/A
T3-06	2flz	(6) D76I,Q110N,R117V,Q121S,E127A,N131A	yes	no
T3-07	2flz	(7) D76I,Q110N,R117V,Q121S,E127A,N131A,E136D	yes	no
T3-08	3ftt	(8) Y20T,A26L,D30V,E34A,R39N,N44L,E48V,Q52A	yes	yes
T3-09‡	3ftt	(9) Y20T,A26L,D30V,E34A,R39N,N44M,K45L,E48V,Q52A	yes	no
T3-10‡	3ftt	(10) Y20T,A26L,D30V,E34A,R39N,S41V,N44M,K45M,E48V,Q52A	yes	yes

* The residue numbers and original amino acid identities listed correspond to the PDB file from which the design was derived. All mutations are at the designed interfaces.
† This design was never tested experimentally.
‡ These designs were based off of the T3-08 design model after the T3-08 crystal structure was obtained.

Table 5.2 Data collection and refinement statistics

	O3-33-R32	O3-33-P4	T3-08	T3-10
Data collection				
Space group	R32	P4	F23	C2
Cell dimensions				
a, *b*, *c* (Å)	137.9, 137.9, 560.6	169.5, 169.5, 119.9	153.7, 153.7, 153.7	183.8, 90.8, 207.7
a, *b*, *g* (°)	90.0, 90.0, 120.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 125.3, 90.0
protomers/asu	8	12	1	6
Resolution (Å)	2.35	3.15	3.35	2.25
R_{merge}	0.066 (0.509)*	0.101 (0.576)	0.056 (0.531)	0.106 (0.553)
I/sI	26.6 (2.4)	8.7 (1.5)	24.1 (4.4)	15.0 (3.0)
Completeness (%)	0.95 (0.89)	0.98 (0.92)	0.966 (0.988)	0.968 (0.934)
Redundancy	6.9 (3.7)	3.3 (1.9)	6.7 (6.1)	6.0 (2.4)
Refinement				
Resolution (Å)	2.35	3.15	3.35	2.25
No. reflections	81,684	52,971	4,298	67,098
R_{work}/R_{free}	0.180/0.213	0.183/0.221	0.231/0.240	0.208/0.239
No. atoms				
Protein	10,496	15,732	1,457	8,748
Ligand/ion	506	78	0	48
Water	375	6	0	297
B-factors (Å^2)				
Protein	70.1	76.1	129.2	66.0
Ligand/ion	78.1	114.5	-	66.0
Water	70.1	20.4	-	57.3
R.m.s deviations				
Bond lengths (Å)	0.010	0.010	0.008	0.010
Bond angles (º)	1.2	1.3	1.1	1.1

*Highest resolution shell is shown in parentheses.

WORKS CITED IN CHAPTER 5

1. Seeman, N. C. (2010). Nanomaterials based on DNA. *Annu. Rev. Biochem.* **79**, 65-87.

2. Rothemund, P. W. (2006). Folding DNA to create nanoscale shapes and patterns. *Nature*, **440**, 297-302.

3. Andersen, E. S. *et al*. (2009). Self-assembly of a nanoscale DNA box with a controllable lid. *Nature*, **459**, 73-76.

4. Zheng, J. *et al*. (2009). From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature*, **461**, 74-77.

5. Zhang, S. (2003). Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* **21**, 1171-1178.

6. Douglas, T. & Young, M. (2006). Viruses: making friends with old foes. *Science*, **312**, 873-875.

7. Lovejoy, B. *et al*. (1993). Crystal structure of a synthetic triple-stranded alpha-helical bundle. *Science*, **259**, 1288-1293.

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

9. Gribbon, C. *et al*. (2008). MagicWand: a single, designed peptide that assembles to stable, ordered alpha-helical fibers. *Biochemistry*, **47**, 10365-10371.

10. Zaccai, N. R. *et al*. (2011). A de novo peptide hexamer with a mutable channel. *Nat. Chem. Biol.* **7**, 935-941.

11. Koder, R.L. *et al*. (2009). Design and engineering of an O(2) transport protein. *Nature*, **458**, 305-309.

12. Ballister, E. R., Lai, A. H., Zuckermann, R. N., Cheng, Y. & Mougous, J. D. (2008). In vitro self-assembly of tailorable nanotubes from a simple protein building block. *Proc. Natl. Acad. Sci. USA*, **105**, 3733-3738.

13. Usui, K. *et al*. (2009). Nanoscale elongating control of the self-assembled protein filament with the cysteine-introduced building blocks. *Protein Sci.* **18**, 960-969.

14. Ringler, P. & Schulz, G. E. (2003). Self-assembly of proteins into designed networks. *Science*, **302**, 106-109.

15. Salgado, E. N., Faraone-Mennella, J. & Tezcan, F.A. (2007). Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am.*

Chem. Soc. **129**, 13374-13375.

16. Salgado, E. N., Radford, R. J. & Tezcan, F. A. (2010). Metal-directed protein self-assembly. *Acc. Chem. Res.* **43**, 661-672.

17. Grigoryan, G. *et al*. (2011). Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science*, **332**, 1071-1076.

18. Padilla, J. E., Colovos, C. & Yeates, T.O. (2001). Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. USA*, **98**, 2217-2221.

19. Sinclair, J. C., Davies, K. M., Vénien-Bryan, C. & Noble, M. E. (2011). Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* **6**, 558-562.

20. Janin, J., Bahadur, R. P. & Chakrabarti, P. (2008). Protein-protein interaction and quaternary structure. *Q. Rev. Biophys.* **41**, 133-180.

21. Grueninger, D. *et al*. (2008). Designed protein-protein association. *Science*, **319**, 206-209.

22. Kortemme, T. *et al*. (2004). Computational redesign of protein-protein interaction specificity. *Nat. Struct. Mol. Biol.* **11**, 371-379.

23. Huang, P. S., Love, J. J. & Mayo, S. L. (2007). A de novo designed protein protein interface. *Protein Sci.* **16**, 2770-2774.

24. Jha, R. K. *et al*. (2010). Computational design of a PAK1 binding protein. *J. Mol. Biol.* **400**, 257-270.

25. Karanicolas, J. *et al*. (2011). A de novo protein binding pair by computational design and directed evolution. *Mol. Cell,* **42**, 250-260.

26. Stranges, P. B., Machius, M., Miley, M. J., Tripathy, A. & Kuhlman, B. (2011). Computational design of a symmetric homodimer using β-strand assembly. *Proc. Natl. Acad. Sci. USA*, **108**, 20562-20567.

27. Fleishman, S.J. *et al*. (2011). Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*, **332**, 816-821.

28. Fleishman, S. J. *et al*. (2011). Hotspot-centric de novo design of protein binders. *J. Mol. Biol.* **413**, 1047-1062.

29. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA,* **97**, 10383-10388.

30. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & André, I. (2011). Modeling symmetric macromolecular structures in Rosetta3. *PLoS ONE*, **6**, e20450.

31. Cooper, S. *et al*. (2010). Predicting protein structures with a multiplayer online game. *Nature*, **466**, 756-760.

32. Zhou, Z. *et al*. (2007). Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem. Biol.* **2**, 337-346.

33. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774-797.

34. Levy, E. D., Pereira-Leal, J. B., Chothia, C. & Teichmann, S. A. (2006). 3D complex: a structural classification of protein complexes. *PLoS Comput. Biol.* **2**, e155.

35. Fleishman, S. J. *et al*. (2011). Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *J. Mol. Biol.* **414**, 289-302.

36. Lawrence, M. C. & Colman, P. M. (1993) Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946-950.

37. Sheffler, W. & Baker, D. (2010). RosettaHoles2: a volumetric packing measure for protein structure refinement and validation. *Protein Sci.* **19**, 1991-1995.

38. Fleishman, S. J., Khare, S. D., Koga, N. & Baker, D. (2011). Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci.* **20**, 753-757.

39. Prodromou, C. & Pearl, L. H. (1992). Recursive PCR: a novel technique for total gene synthesis. *Protein Eng.* **5**, 827-829.

40. Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606-1619.

41. Ohi, M., Li, Y., Cheng, Y. & Walz, T. (2004). Negative staining and image classification - powerful tools in modern electron microscopy. *Biol Proced Online*, **6**, 23-34.

42. Smith, J. (1999). XIMDISP—A visualization tool to aid structure determination from electron microscope images. *J. Struct. Biol.* **125**, 223-228.

43. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. (1996) A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17-24.

44. Mindell, J. A. & Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334-347.

45. Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82-97.

46. Grigorieff, N. (2007). FREALIGN: high-resolution refinement of single particle structures. *J. Struct. Biol.* **157**, 117-125.

47. Pettersen, E. F. *et al*. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-12.

48. Tsai, Y., Sawaya, M. R. & Yeates, T. O. (2009). Analysis of lattice-translocation disorder in the layered hexagonal structure of carboxysome shell protein CsoS1C. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 980-988.

49. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology Macromolecular Crystallography part A* (Carter Jr., C.W & Sweet, R. M., eds.), **276**, pp. 307-326, Academic Press, New York.

50. Kabsch, W. (2010). XDS. *Acta Cryst. D*, **66**, 125-132.

51. McCoy, A. J. *et al*. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674.

52. Crowley, C. S. *et al*. (2010). Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. *J. Biol. Chem.* **285**, 37838-37846.

53. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240-255.

54. Blanc, E. *et al*. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2210-2221.

55. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Meth. Enzymol.* **374**, 300-321.

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

57. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

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

59. Lüthy, R., Bowie, J. U. & Eisenberg, D. (1992). Assessment of protein models with

three-dimensional profiles. *Nature*, **356**, 83-5.

60. The PyMOL Molecular Graphics System, Version 1.4, Schrödinger, LLC (http://www.pymol.org).

Chapter VI

The hepatitis C virus IRES manipulates the 80S ribosome

Megan E. Filbin[2], **Breanna S. Vollmar[3]**, Dan Shi[3], Tamir Gonen[3] & Jeffrey S. Kieft[1,2]

[1]Howard Hughes Medical Institute,
University of Colorado Denver School of Medicine, Aurora, CO, 80045, USA
[2]Department of Biochemistry and Molecular Genetics,
University of Colorado Denver School of Medicine, Aurora, CO, 80045, USA
[3]Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA

This work was published in the journal of Nature Structure Molecular Biology (2013), volume 20, pages 150-158.

SUMMARY

The internal ribosome entry site (IRES) of the hepatitis C virus (HCV) drives noncanonical initiation of protein synthesis necessary for viral replication. Functional studies of the HCV IRES have focused on 80S ribosome formation but have not explored its role after the 80S ribosome is poised at the start codon. Here, we report that mutations of an IRES domain that docks in the 40S subunit's decoding groove cause only a local perturbation in IRES structure and result in conformational changes in the IRES–rabbit 40S subunit complex. Functionally, the mutations decrease IRES activity by inhibiting the first ribosomal translocation event, and modeling results suggest that this effect occurs through an interaction with a single ribosomal protein. The ability of the HCV IRES to manipulate the ribosome provides insight into how the ribosome's structure and function can be altered by bound RNAs, including those derived from cellular invaders.

INTRODUCTION

HCV is the most common blood-borne virus in the United States and is estimated to infect ~3% of the world's population. HCV's genome is a single-stranded, positive-sense RNA molecule; thus, after release of the genome into the cytoplasm, the first step of viral replication is translation of a single open reading frame to yield the viral proteins. The HCV genomic RNA is therefore similar to cellular mRNA, serving as the template for protein synthesis by translation machinery. However, unlike cellular mRNAs, which originate in the nucleus and are capped on their 5′ ends by 7-methylguanosine and polyadenylated on their 3′ ends (both important translation initiation signals), HCV genomic RNA is delivered directly to the cytoplasm lacking both a cap and a poly(A) tail (Fig. 6.1A**)**. As a result, translation of HCV RNA is initiated by a mechanism that differs substantially from the one used by the cell to translate its own mRNA. Specifically, HCV uses an IRES RNA at the 5′ end of the viral genome (1) to hijack the cellular translation machinery. The IRES RNA sequence is highly conserved among HCV isolates and genotypes (2,3). This conservation underscores the importance of the HCV IRES RNA to the viral replication cycle and reflects the specificity of the interactions between the IRES RNA and the translation machinery.

Mechanistic studies of the HCV IRES have focused on how it assembles an 80S ribosome at the AUG start codon, and they have revealed a mechanism different from canonical translation initiation (4). In the canonical pathway, the 5′ cap is recognized by the eukaryotic initia- tion factor 4F (eIF4F) complex, followed by binding of the 43S particle (which includes the 40S subunit; the ternary complex formed by eIF2, GTP and initiator methionyl-tRNA (eIF2–GTP–Met-tRNAiMet); eIF3 and other factors). The subunit then scans the mRNA to locate the start codon, at which time factor release and 60S subunit association yield an 80S ribosome (5).

In contrast, the HCV IRES first directly binds the 40S subunit (6–9), then eIF3 and the ternary complex bind (10–12) (Fig. 6.1B). Subsequent GTP hydrolysis, eIF release and binding of a 60S subunit yield an 80S ribosome placed directly at the start codon (13,14). In addition, the HCV IRES can, under conditions of cellular stress, use eIF2-independent pathways to generate 80S ribosomes (15,16).

The function of HCV IRES is conferred by its structure. The IRES adopts an extended global architecture (17), within which specific RNA structural domains drive different steps of 80S ribosome formation (18) (Fig. 6.1C). Domain III binds the 40S subunit and eIF3 (6,7,19); domain IV provides the AUG initiation codon for interaction with the ternary complex (11,12,20); the pseudoknot is important for placement of the AUG in the 40S subunit decoding groove (21); and domain II (dII) is involved in eIF2-catalyzed GTP hydrolysis (14), removal of eIF3j (13), 60S subunit joining (11) and the configuration of RNA in the decoding groove (22). These findings, combined with the observation that HCV IRES binding alters the conformation of the 40S subunit (23), indicate that the IRES RNA is an active manipulator of the translation machinery, not just a binding site for the ribosome and factors.

Among the aforementioned IRES domains, dII is particularly notable: it induces changes in the conformation of the 40S subunit (23), and it docks within the ribosome's decoding groove to interact with ribosomal protein rpS5 (24,25), which is known to contact E-site tRNA (26–28). We therefore set out to study the role of dII sub- domain dIIb, the part of the IRES that penetrates deep into the ribosome's decoding groove. We have discovered that dIIb muta- tions alter the conformation of the IRES–40S subunit complex and inhibit the first round of ribosomal translocation. The modeled position of this domain (adjacent to rpS5) suggests this effect may be due to altered contact with rpS5. This is, to our knowledge, the first evidence that a step after

80S ribosome assembly is directly influenced by the HCV IRES, and it may have implications for the understanding of translation initiation in general.

RESULTS

IRES dIIb affects the rate of protein synthesis

We have previously reported that mutation of the dIIb apical loop (Fig. 6.1D) changes the configuration of the HCV RNA in the decod- ing groove of 40S subunit–HCV IRES RNA complexes and reduces IRES-driven protein synthesis (22). This is consistent with other stud- ies showing that dIIb is important for HCV IRES–mediated translation initiation (11,29,30). To monitor protein production as a function of time, we measured the ability of three IRES RNAs in which dIIb was mutated to translate a downstream luciferase (LUC) reporter sequence (uncapped and not polyadenylated) in rabbit reticulocyte lysate (RRL) over a 90-min time course (Fig. 6.1D,E). All three dIIb mutants produced LUC at a rate that was slower than that of wild-type (WT) IRES RNA but faster than that of an IRES mutant lacking dII (ΔdII). We observed differences in total LUC production after only 15 min (Fig. 6.1F), and the mutants continued to make protein at a roughly constant rate but more slowly than WT. These results indi- cate that the activity of the IRES with dIIb mutations is inhibited but not abolished. Assuming identical rates of elongation and ribosome termination on each RNA sample, we conclude that the dIIb muta- tions slow, but do not halt, translation initiation.

IRES–40S subunit complex conformation is influenced by dIIb

Cryo-EM reconstructions of HCV IRES–40S subunit complexes show that binding of full-length WT HCV IRES induces structural changes in the subunit, but the ΔdII mutant does

not cause this conformational change (23). This raised the question of what the conformation of the 40S subunit would be when bound to the dIIb mutant RNAs generated for this study. To answer this, we visualized IRES RNAs with the apical GCC (nucleotides 82–84) in dIIb deleted (ΔGCC) in complex with the 40S subunit by EM, using both negative staining (to assess sample purity, homogeneity and concentration) and cryo-EM (to generate a three-dimensional (3D) structure) (Fig. 6.2A). We obtained the reconstruction of the complex at a resolution comparable to previous HCV IRES–40S subunit reconstructions (17–20 Å) (23). Comparison of the ΔGCC IRES–40S subunit and WT IRES–40S subunit structures revealed notable differences in the position and orientation of IRES domains and in the conformation of the 40S subunit (Fig. 6.2B). Specifically, in the mutant, dII did not loop away from the subunit's surface to contact the side of the head and enter the decoding groove; rather, it lay across the platform. This change in the position of dII was accompanied by a rotation of the overall IRES orientation relative to the body of the 40S subunit. This cryo-EM structure does not eliminate the possibility that the IRES location and conformation is an average of several similar structures. In other words, the GCC deletion might cause the conformation of the IRES and its position on the 40S subunit to be more dynamic, which could explain why structural features visible in the WT IRES are not seen in the ΔGCC IRES and why the IRES appears to be rotated relative to the subunit. However, this possibility does not affect the conclusion that this relatively small deletion mutation at the tip of dIIb alters the IRES–40S subunit interaction, which is consistent with our observation that the ΔGCC mutation lowers, but does not eliminate, IRES activity. When the 40S subunit was bound to the ΔGCC mutant IRES, some features of its conformation were similar to the WT IRES–bound state, but some were markedly different. In both states, the latch formed between ribosomal RNA (rRNA) helices 18 and 34 was closed (this latch is open in apo-40S) (23,31).

However, the entry site (where mRNA enters the decoding groove) was much more open in the ΔGCC mutant IRES complex; this feature was clearly visible in class averages assembled from individual particle images (Fig. 6.2A). This result is consistent with a previous study showing that mutation of dIIb alters the configuration of the mRNA in the decoding groove (22). In summary, the dIIb mutant IRES binds the 40S subunit using the same side of the subunit as the WT IRES. But the structures are considerably different; thus, deletion of the apical nucleotides of dIIb substantially affects the global conformation of the complex. These changes are different from those caused by ΔdII23, suggesting that mutation of dIIb perturbs translation initiation differently from the ΔdII mutation.

dIIb influences a step after 80S ribosome formation

The structure of the mutant IRES–40S complex and the translation initiation efficiencies of all three dIIb mutants raises the question of which step in initiation is affected by these mutations. Removal of domain II or replacement of the entire dIIb apical loop with an ultrastable UUCG tetraloop has been reported to inhibit 80S ribosome formation (11,12,14), which suggested that our targeted dIIb mutations would do the same. To test this, we assayed ribosome assembly in RRL, separating the resultant ribosomal complexes by ultracentrifugation in a sucrose density gradient. Consistent with previous reports (11), we observed robust 80S formation with WT IRES and minimal 80S formation in the ΔdII mutant after 15 min (Fig. 6.3A). However, contrary to our expectation, all of the targeted dIIb mutants were as efficient as WT IRES RNA at forming IRES–40S complexes, IRES–48S* complexes (asterisk denotes complex formed by a noncanonical pathway) and IRES–80S ribosomes (Fig. 6.3A). This result indicated that these mutations do not inhibit the formation of 80S ribosomes on the HCV IRES

RNA. Furthermore, we observed no difference in the rate of 80S formation between dIIb mutant and WT IRES during a 10-min time course (Fig. 6.3B). These data, combined with the translation initiation data (Fig. 6.1E), confirm that mutation of dIIb does not inhibit 80S ribosome formation on the IRES. Thus, we conclude that the altered position of the ΔGCC mutant IRES RNA on the 40S subunit and the different conformation of the 40S subunit (relative to the WT IRES–40S complex) can support 80S ribosome formation, but some subsequent event is slowed by mutation of dIIb.

Modeled local interaction between dIIb and rpS5

To determine which interactions of dIIb with the ribosome are disrupted by mutation, we modeled the placement of dII on the ribosome by docking a 40S subunit structure (32) and the structure of dII (33) into the cryo-EM density of the complete WT HCV IRES–40S subunit complex (22,23) (Fig. 6.3C). Our placement of dII is consistent with a previously published model based on an IRES–80S complex (24), with additional detail provided by the crystal structure of a 40S subunit. We compared our model with the crystal structure of a bacterial ribosome bound to tRNAs to determine how similar the contacts with the 40S subunit are between our modeled dII and a bound E-site tRNA (Fig. 6.3D). We observed only one instance of overlap, at the anticodon loop of the tRNA and the modeled position of the HCV IRES dIIb, which are both positioned directly against the β-hairpin structure of rpS5 (or its bacterial homolog rpS7) (Fig. 6.3C). Specifically, we observed that the mutated nucleotides in dIIb are directly adjacent to the rpS5 β-hairpin (Fig. 6.3E). The modeled position of dIIb is consistent with observed cross-linking of dII to rpS5 (25), with the previous IRES–80S ribosome model (24) and with the role of the β-hairpin in contacting tRNA (26–28). The putative location of our dIIb mutations

adjacent to the β-hairpin of rpS5 suggests that they disrupt a specific contact between dIIb and rpS5. Previous chemical probing of these mutant RNAs showed no global change in IRES secondary structure, but it remains possible that the mutations change the overall structure of dII. To assess this, we characterized the structures of the dIIb mutant RNAs using NMR. We generated WT and mutant samples comprising nucleotides 76–100 of the HCV IRES (Fig. 6.4A), which contain dIIb and are identical to RNAs used to solve the structure of dII (33). Comparison of the one-dimensional 1H spectra obtained in water showed very little difference in the chemical shifts or relative intensity of the peaks obtained for WT and mutant RNAs (Fig. 6.4B). We observed the largest chemical shift in the imino protons of G87 and G88, which are adjacent to the apical loop of dIIb. The visibility of these imino proton peaks in all spectra indicates that base-pair formation is unaltered by dIIb mutation.

The spectrum of mutant ΔapexC (dIIb mutant with only apical C83 deleted) was most similar to that of the WT, suggesting that the ΔapexC mutation elicits the post–80S ribosome functional effect with the smallest change in loop structure. On the basis of our prediction that the single deleted nucleotide (C83) would contact the β-hairpin of rpS5 (Fig. 6.3E), we selected this mutant for additional NMR experiments. When we overlaid the WT and ΔapexC spectra obtained by 2D 1H-NOESY in water, the portions that contained the imino-imino cross-peaks overlapped almost perfectly, with the largest shift occurring at the G87 imino (Fig. 6.4C). Likewise, we observed only small chemical-shift changes in spectra showing the cross-peaks between the imino and other protons (Fig. 6.4D). For example, cross-peaks between the G87 imino and C79 amino protons were shifted but intense (Fig. 6.4D), again confirming that this base pair at the base of dIIb still forms in the mutant RNA (Fig. 6.4E). Changes to the spectra were limited to nucleotides adjacent to the deletion, showing the localization of structural

changes to the dIIb apical loop. This result is consistent with published chemical-probing data, based on selective 2′-OH acylation analyzed by primer extension, for the ΔapexC and other mutant RNAs, in that changes in the chemical probing pattern were limited to the apical loop in the IRES–40S subunit complex (22) (Fig. 6.4F). Taken together, our data show that C83 deletion—and, we would predict, the other dIIb mutations shown in Figure 1—induces a local structural perturbation that, we propose, disrupts dIIb interaction with the β-hairpin of rpS5 and is accompanied by a global change in the structure of the IRES–40S complex and the inhibition of a step after 80S ribosome formation.

AUG start codon placement is unaffected by dIIb mutations

To determine which step after 80S ribosome formation is affected by dIIb mutation, we first explored the possibility that the mutations cause incorrect placement of the AUG start codon. To test this, we used primer extension inhibition (toeprinting) analysis (10,34–36) on WT and dIIb mutant RNA in the unbound and 80S ribosome–bound states (the latter is formed in RRL with cycloheximide) (Fig. 6.5A, compare lanes 5, 7, 9 and 11 with 6, 8, 10 and 12). As expected, we observed no toeprint for unbound RNAs; however, when bound to the 80S ribosome, all mutants produced similar toeprints at nucleotide positions +15 and +16 downstream of the AUG initiation codon (where A is in position +1) (Fig. 6.5A,B). This indicates that the AUG is properly positioned in all of these complexes and that the mutations probably affect a step after 80S ribosome assembly on the IRES.

dIIb mutants initiate in the correct reading frame

Mutations to the β-hairpin of rpS7 have been shown to increase frameshift rate (37). We

therefore considered the possibility that the dIIb mutants, by disrupting contact with rpS5, could cause frameshift during or after the first translocation event and, thus, an apparent decrease in translation. To test this, we conducted translation assays using reporters to which one or two nucleotides were added after the IRES start codon but before the start codon of the LUC open reading frame. If the ribosome sometimes slips out of frame on the dIIb mutant RNAs, the addition of these nucleotides would rescue a 1- or 2-nucleotide frameshift (Fig. 6.5C). We did not observe a partial rescue (Fig. 6.5D); in fact, introduction of these nucleotides decreased translation efficiency even more than the dIIb mutations. Hence, the decreased translation initiation efficiency observed for the dIIb mutants is not caused by frameshifting.

dIIb mutation does not affect peptide bond formation

Because slowed initial peptide bond formation after 80S ribosome assembly would decrease the overall rate of protein production and explain a decrease in protein synthesis (Fig. 6.1E), we hypothesized that dIIb mutations affect peptide bond formation. We tested the ability of 80S ribosomes formed on dIIb mutants to catalyze peptide bond formation using the aminoglycoside puromycin to accept an amide linkage from the amino acid on the P-site tRNA when the A site of the ribosome is vacant (10,38) (Fig. 6.6A,B). We generated 80S ribosome complexes in RRL supplemented with [^{35}S]Met using WT and dIIb mutant IRES RNAs truncated to end after the AUG start codon. The resulting 80S ribosome–IRES complexes contained [^{35}S]Met-tRNAiMet in the P site and a vacant A site. The amount of [^{35}S]Met transferred to added puromycin indicates the ability of the 80S ribosomes to catalyze peptide bond formation. We found that the dIIb mutant IRES–80S complexes formed peptide bonds as competently as complexes formed on WT IRES (Fig. 6.6C). In fact, two of the mutants showed a

reproducible increase in the production of the puromycin-Met product compared to WT, which is notable given that the location of dIIb and its putative interaction with rpS5 is ~100 Å from the peptidyl transferase center of the large subunit (Fig. 6.6D). This result suggests that the dIIb mutations do not decrease peptide bond formation but have effects that are felt in distal parts of the ribosome.

The first translocation event is promoted by dIIb

We hypothesized that mutation of dIIb allows 80S ribosome formation and peptide bond formation but inhibits the first round of translocation, in which the P site–bound initiator tRNA and start codon AUG move to the E site. To test this hypothesis, we used toeprinting analysis to directly detect ribosome movement on the WT and mutant IRES RNAs. We conducted these assays in RRL and in the absence and presence of the antibiotic hygromycin B, which inhibits translocation (39). In the absence of hygromycin B, all IRES RNAs produced a toeprint consistent with start codon placement in the P site of 80S ribosomes (Fig. 6.6E, lanes 5, 7, 9 and 11). We did not see stops caused by elongating ribosomes, probably because these ribosomes move to a position that prevents binding of primers or reverse transcriptase. We observed slightly higher toeprint intensity for the dIIb mutants than for WT IRES RNA, indicating that more 80S ribosomes were paused at this initial assembly location (Fig. 6.6F). When present, hygromycin B captured ribosomes that had undergone one or two rounds of translocation, as indicated by the presence of toeprint stops at positions +20, +21 and +22 (Fig. 6.6E, lane 6). In contrast, we did not observe these strong downstream toeprint stops on the mutant RNAs (Fig. 6.6E, lanes 8, 10 and 12). Again, we quantified multiple replicates to correct for loading differences and found that 80S ribosomes formed on dIIb mutants were slow to move from their

initial positions (Fig. 6.6G). Specifically, 33.3% of the WT RNA was left in an untranslocated state, whereas 54.9%, 58.0% and 62.0% of ΔGCC, ΔapexC and GCC-AUU (dIIb mutant with GCC at nucleotides 82–84 replaced with AUU) did not translocate, respectively. Thus, mutations in dIIb decreased translocation by ~50%, which is consistent with the activity measurements presented in Figure 6.1. We conclude, therefore, that mutation of HCV IRES dIIb results in 80S ribosomes that have a reduced ability to undergo the first round of translocation. This suggests that an intact dIIb is needed for the first translocation step, at which initiation transitions to elongation.

DISCUSSION

We have found that mutations in the apical loop of dII of the HCV IRES inhibit the first translocation event after the formation of the 80S ribosome. HCV IRES–driven 80S ribosome formation and the first step of translocation can therefore be decoupled, indicating that dII has a function that is distinct from previously defined roles of HCV IRES domains. The involvement of dIIb in the first round of translocation could be explained by two broad mechanisms. First, the position of dII in the E site requires it to move to make way for the P-site tRNA to translocate, as evidenced in cryo-EM reconstructions of the HCV IRES bound to a 40S subunit and 80S ribosome (23,24). On the basis of cryo-EM reconstruction of the HCV IRES–80S ribosome, it has been suggested that contacts between a different portion of dII and the L1 stalk of the large subunit could facilitate dII displacement from the E site, but this IRES-ribosome contact and its potential function have not been investigated (24). Also, a recent crystal structure of HCV IRES domain IIa bound to an inhibitor suggests that conformational changes in domain IIa or other parts of dII not studied here may effect movement of the domain from the E site (40), but this has

not been demonstrated functionally. In dIIb, mutations could potentially inhibit dII displacement, and this would slow translocation by sterically hindering the movement of tRNA. This 'failure-to-move' phenotype, then, could be ascribed to the loss of a specific IRES-rpS5 interaction necessary for dII ejection from the E site, although an analogous role for rpS5 in tRNA ejection has not been reported. This idea does not eliminate the possibility that the L1 stalk and dII conformational changes also help move dII. The second potential explanation is that dIIb actively promotes an event within the ribosome that is important for the first round of translocation—again, probably through a specific interaction of dIIb with rpS5 and subsequent conformational changes in the IRES–ribosome complex. Mutation of dIIb, then, might either interfere with this event or render the IRES unable to actively promote it. These broad mechanisms are consistent with our structural and functional data, and they are not mutually exclusive.

Given that the HCV IRES dIIb is positioned to interact directly with the β-hairpin of rpS5, examining known functions of rpS5 (and bacterial ortholog rpS7) could give insight into how dIIb influences translocation. During elongation, rpS5 has a role in maintaining the reading frame and in overall fidelity (37,41,42). In bacteria, truncation of the β-hairpin of rpS7 results in destabilization of the E-site tRNA and an increased number of frameshift and reverse translocation events (37). However, our data show no evidence of frameshift induced by the dIIb mutants. Although not explicitly shown to be dependent on rpS5 or rpS7, the fidelity of aminoacyl-tRNA selection in the A site (43) was recently shown to be influenced by the presence of tRNA in the E site during elongation, but other studies find little or no evidence for this (44–46). It seems unlikely, then, that the presence of dIIb directly influences entry of A-site tRNA. Overall, we find no known role for rpS5 that readily explains the observed effects of

HCV IRES dIIb mutation, consistent with the notion that the IRES is co-opting this ribosomal feature to manipulate the complex in a noncanonical way or that rpS5 has a yet-undiscovered role in translation initiation.

Mutation of dIIb disrupts a putative interaction with rpS5, changes the structure of the IRES–40S subunit complex and inhibits the first translocation step. These observations suggest that a specific rpS5-dIIb interaction induces allosteric changes that propagate throughout the ribosome. Indeed, there is evidence for a network of interactions within the ribosome that could propagate allostery, and also for similar conformational changes induced by bound initiation factors eIF1 and eIF1A. Specifically, rpS5 interacts with rRNA helices 29, 30 and 42 (47,48), and these helices interact with eIF1A (49). Another network links rpS5 to the eIF1 binding site through rpS14 and rRNA helix 23 (41,50,51). This is notable because binding of eIF1 and eIF1A induces a conformational change in the 40S subunit that strongly resembles the change induced by the WT HCV IRES (23,31), and eIF1A is known to act with eIF5B after 80S formation to commit the ribosome to elongation (52,53). This last point raises the possibility that the HCV IRES dIIb may induce the same effect as eIF1, eIF1A and eIF5B to promote a late step during initiation. This (speculative) notion is appealing because a minimal reconstitution of HCV IRES–driven translation initiation does not require eIF1 or eIF1A (10), so dIIb could substitute for these absent factors. Although not a part of this study, higher-resolution structures of HCV IRES mutants in complex with 80S ribosomes and chemical probing of the rRNA in these complexes before and after translocation could provide insight into the putative allosteric changes associated with this translocation-slowing phenotype.

We would like to propose the following model to explain the role of HCV IRES dIIb in events that occur within the IRES–ribosome complex before and during the first translocation

step (Fig. 6.7): first, the IRES assembles an 80S ribosome such that the ribosome is poised at the start codon with an initiator tRNA in the P site. Within this ribosome, dIIb contacts the β-hairpin of rpS5, thereby stabilizing the ribosome in a conformation that is conducive to translocation. Delivery of aminoacyl-tRNA to the ribosome by eukaryotic elongation factor 1A (eEF1A) and subsequent peptide bond formation are then followed by rapid and efficient eEF2-catalyzed translocation. In the dIIb mutants studied here, the mutation induces a local change in the apical loop structure that perturbs the interaction with rpS5, affecting 40S subunit conformation. The dIIb mutant IRES–40S complex is still capable of forming an 80S ribosome, but one whose conformational equilibrium is shifted such that its ability to translocate is inhibited. Although aminoacyl-tRNA may still be delivered to the A site and a peptide bond formed, the mutant-bound ribosome stalls at the start site. However, the ribosome samples conformations, so these ribosomes are not permanently stalled; rather, they occasionally sample a productive state in which they are able to translocate. In summary, our data support a model in which dIIb selects a productive state from the conformational ensemble, and ribosomes bound to IRESs with mutated dIIb spend more time in an unproductive state and transition to elongation less efficiently.

Our data open another door to understanding the intricacies of translation initiation and ribosome function. The ribosome is fundamentally a Brownian machine that samples many conformations—protein factors and tRNA binding shift the conformational equilibrium, providing efficiency and directionality. Thus, the ribosome is programmed to be manipulated by its binding partners. This inherent characteristic of the ribosome is crucial for canonical translation processes and allows subtle and robust regulation of ribosome function. Our results reveal that these principles are exploited by a single loop of the HCV IRES, supporting the view of the HCV IRES as a dynamic manipulator of the translation machinery and lending insight into how

the translation machinery works in cap-dependent and cap-independent pathways.

MATERIALS AND METHODS

Plasmid construction and cloning

We constructed pUC19-based plasmids containing the HCV genotype Ib wild-type (WT; nucleotides 40–372) and ΔdII mutant (nucleotides 119–372) sequences flanked by a 5′ hammerhead and 3′ hepatitis delta ribozyme as previously described (6,17,22). Plasmids with the WT or mutant HCV IRES between two luciferase-encoding genes were made by PCR amplification of the desired sequence and ligation into the EcoRI and NcoI sites of plasmid pRL57 (gift of A. Willis). The plasmid used to generate RNA for toeprinting analysis (containing 85 additional 3′ nucleotides on the wild-type HCV genotype 1a as well as the primer binding site) was a kind gift of P. Lukavsky (14). We generated the genotype 1b ΔdII mutant used in toeprinting by PCR amplification of the desired sequence and ligation into the HindIII and XbaI sites of this plasmid. All mutants were made using the QuikChange mutagenesis kit (Stratagene).

RNA preparation

We made RNAs for assembly assays and puromycin experiments using DNA generated by PCR using M13 (-41) forward and reverse primers (5′-GGTTTTCCCAGTCACGAC-3′ and 5′-GGAAACAGCTATGACCATG-3′, respectively) and the relevant plasmid template. The PCR products were used for *in vitro* transcription reactions as described (58). We purified and concentrated RNA as described (22). Monocistronic *Photinus* luciferase RNAs were made from PCR templates using forward T7-HCV and reverse photinus primers (5′-TAATACGACTCACTATAGGGCTCCCCTGTGAGGAACTACTGTCTT-3′ and 5′-

TTACACGGCGATCTTTCCGCCCTTCTT-3′, respectively) using the T7 MegaScript kit (Ambion). RNAs were DNase treated, then purified with TRI Reagent (Sigma) and chloroform followed by 100% isopropanol 75% ethanol precipitations, respectively. We made RNAs for toeprinting from EcoRI-linearized plasmids in the same manner as the luciferase RNAs.

Radiolabeling RNA and primers

For assembly assays, we 5′ radiolabeled RNA as described (22) and diluted it to approximately 1,000 c.p.m. μl^{-1}. DNA primers were 5′ radiolabeled in a reaction containing 800 pmol primer in the same conditions as the RNA, mixed with 20 µl 9 M urea loading buffer, loaded directly onto a 10% urea denaturing gel, purified and diluted to approximately 25,000 c.p.m. μl^{-1}.

Ribosome assembly assays

We diluted 5′-radiolabeled HCV IRES RNAs to ~1,000 c.p.m. µl−1, heated to 85 °C for 30 s, then cooled on the desktop. 1 µl of this RNA was then added to a mixture containing 30 µl rabbit reticulocyte lysate (RRL), 0.5 µl amino acid mixture minus leucine, 0.5 µl amino acid mixture minus methionine (all provided in RRL translation kit, micrococcal-nuclease treated; Promega) and 18 µl RNase-free water. Reactions then were incubated at 30 °C for the desired time, halted by the addition of ribosome association dilution buffer (50 mM Tris (pH 7.5), 50 mM NaCl, 5 mM MgCl2 and 1 mM DTT) and placed on ice. The reactions were analyzed with 10–35% sucrose gradients in ribosome association dilution buffer by ultracentrifugation in a SW41 Ti swinging bucket rotor at 36,000 r.p.m. (~222,000*g*) for 3 h. We fractionated the gradients into ~0.5 ml fractions using a BIOCOMP Gradient Station and Gilson FC203B fraction

collector. Two hundred microliters of each fraction were blotted onto membranes, air dried and analyzed using a PhosphorImager. We quantified the spots using ImageQuant software and reported each as a fraction of the total radiation. Note that 40S and 48S* were indistinguishable in a 10–35% sucrose gradient.

Luciferase assays

We conducted translation assays as described (22) with the following exceptions: for the time-point experiment, reactions were brought up to 125 µl volume so that 25 µl could be removed at each time point, halted with 200 µl cold (1). Passive Lysis Buffer (Promega) and placed at –80 °C to ensure no further activity.

Toeprinting assays

We completed toeprinting assays essentially as described (22). For 48S-bound IRES, 10.75 µl RRL was mixed with 0.5 µl RNasin Plus and 0.5 mM guanylyl imidodiphosphate and incubated at 30 °C for 5 min, after which 0.5 µg toeprint RNA was added to a final volume of 15 µl. For 80S-bound IRES, 10.8 µl RRL was mixed with 0.5 µl RNasin Plus (Promega) alone or with 3 mg ml^{-1} cycloheximide or 2 mg ml^{-1} hygromycin b and incubated at 30 °C for 5 min. This incubation was followed by addition of 0.5 µg toeprint RNA to a final volume of 15 µl. We made the ladder used for analysis with wild-type toeprint RNA reverse transcribed with SuperScript III Reverse Transcriptase (Life Technologies) with annealing and extension temperatures at 45 °C.

Puromycin assays

We first biotinylated 3′-truncated IRES RNAs (nucleotides 40–344, which stop after the AUG codon) using the 5′ EndTag Nucleic Acid Labeling System (Vector Laboratories). Briefly, 65 μg of RNA was phosphorylated with ATPγS using T4 polynucleotide kinase in a 20-μl reaction, at 37 °C for 1 h. Biotinylation then was carried out upon the addition of ~385 μg biotin (long arm) maleimide for 1 h at 65 °C. Reactions then were extracted with equal volume phenol-chloroform-isoamyl alcohol, pH 6.7 (25:24:1) (Fisher), precipitated with 1/10 volume 3 M acetic acid (pH 5.2) and three volumes 100% cold ethanol overnight. RNA was pelleted and washed with 70% ethanol, dried and resuspended in 10 μl RNase-free water. Concentration was determined by absorbance at 260 nm. To conduct the assay, we mixed 30 μl RRL with 16.5 μl RNase-free water and 1 μl L-[^{35}S]methionine (>1,000 Ci mmol−1) and incubated the mixture at 30 °C for 15 min to allow aminoacylation of initiator methionine tRNA. Biotinylated RNA (2.25 μg) was added to the reaction and incubated at 30 °C for 25 min for 80S ribosome formation. Reactions were mixed with one tube (0.6 ml) streptavidin paramagnetic beads (MagneSphere, Promega), prewashed three times with 0.5. saline sodium citrate (SSC) buffer (7.5 mM trisodium citrate dehydrate (pH 7.2), 75 mM NaCl) and once with 300 μl ribosome association dilution buffer), resuspended in 50 μl ribosome association dilution buffer and incubated at 30 °C for 10 min. Complexes were then washed six times with 500 μl ribosome association dilution and resuspended such that reactions were split into duplicates with and without 1 mM puromycin. The assay was then carried out at 35 °C for 60 min. Puromycin was extracted with 500 μl 200 mM potassium phosphate buffer (pH 8) and 500 μl ethyl acetate with continuous shaking for 10 min at 35 °C. The upper ethyl acetate layer was removed and mixed with 7 ml ScintiSafe (Fisher) liquid scintillation fluid, and counts per minute were averaged between two 10-min

count times. This method of immobilizing the IRES-80S ribosomes greatly reduced background levels of puromycin-[^{35}S]methionine formation compared to results for 80S ribosomes purified by ultracentrifugation in sucrose gradients.

Nuclear magnetic resonance

We collected NMR spectra at 25 °C on a Varian 900 MHz spectrometer using Standard Varian Biopack pulse sequences for all experiments. This included both one-dimensional spectra as well as all homonuclear two-dimensional spectra employing a 3919 WATERGATE for water suppression. NOESY (Biopack pulse sequence, WBNOESY) spectrum was collected with 256 indirect points. Two-dimensional data were processed using a Gaussian weighting function in the direct dimension and a sine-bell weighting function in the indirect dimension.

Negative-stain electron microscopy

We prepared the ΔGCC mutant IRES RNA as described above and the 40S ribosomal subunits from RRL as previously described (6). We assembled IRES–40S subunit complex as previously described (23). We applied the complex to freshly glow-discharged carbon coated 400 mesh copper grids and stained with 0.75% uranyl formate as described (59). Samples were viewed on a 120-kV transmission electron microscope (FEI, Hillsboro, OR). Images were recorded at a nominal magnification of x40,000 using a bottom mount 4k x 4k Gatan slow-scan charge coupled device (CCD) camera.

Electron cryomicroscopy

Purified ΔGCC mutant IRES RNA in RNase-free water was heated to 70 °C for 2 min

then cooled to room temperature. Buffer solution was added to a final concentration of 20 mM Tris-HCl (pH 7.4), 100 mM acetic acid, 2.5 mM MgCl2, 1 mM DTT and 40 mM KCl. Purified 40S subunits were added at a 1:1 molar ratio with the IRES RNA to a final concentration of 500 nM complex. Complex was stored on ice until diluted (generally to 100 nM) and used in microscopy.

We prepared vitrified samples of the ΔGCC mutant IRES-40S subunit complexes at 100 nM using an FEI Vitrobot. Briefly, 3.5 µl was applied to a Quantifoil Holey Carbon grid (Vitrobot chamber was at 4 °C and 100% humidity). After a 20-s pause, the grid was blotted with filter paper (force = 0, blot time = 2 s) and plunged into liquid ethane. Frozen samples were loaded onto a Gatan cryo-holder and inserted into an FEI Tecnai F20 operating at 200 kV equipped with a field emission gun. Images were collected at a nominal magnification of x62,000 using a 4k x 4k Tietz CMOS detector. Images were binned two times yielding a pixel size of 2.66 Å per pixel. Approximately 29,000 particles were selected from 1,790 images using Electron Micrograph Utility (http://cryoem.ucsf.edu/). Class averages were determined using five consecutive rounds of MSA (multivariate statistical analysis) and MRA (multireference alignment) in IMAGIC (54). Contrast transfer function parameters for each image were determined using CTFFIND3 (60). An initial model for refinement and three-dimensional reconstruction was generated by filtering a previously published apo-40S reconstruction (EMD-1346) model to 40 Å (31). Initial parameters were generated during cycles of randomized search and refinement using FREALIGN v. 8.08 (61). After initial parameters were determined, consecutive cycles of local refinement and reconstruction were carried out until there was no apparent improvement in the alignment. Resolution of the three-dimensional model was calculated with the program RMEASURE (62) and determined to be ~17.5 Å. The density was

normalized using MAPMAN (63) and filtered to 20 Å using BFACTOR. Difference maps presented in the figures were calculated using MAPMAN (63). We calculated the density corresponding to the WT IRES from a difference map between the WT IRES–40S and apo-40S (23), and the mutant IRES was calculated from a difference map between the ΔGCC IRES–40S reconstruction and apo-40S (23). Reconstructions and difference maps were assembled as displayed in Figure 2B using UCSF Chimera (64).

Accession Codes

The cryo-EM map of the mutant HCV IRES–rabbit 40S subunit complex has been deposited in the Electron Microscopy Data Bank with accession number 5527.

ACKNOWLEDGEMENTS

The authors thank the members of J.S.K.'s lab and R. Davis, D. Bentley, D. Barton and T. Evans for useful suggestions and discussions, and M. Ruehle, T. Blumenthal, T. Cech and M. Johnston for critical reading of this manuscript. We also thank C. Spahn (Institut für Medizinische Physik und Biophysik, Charite–Universitätsmedizin Berlin) for data files and advice with structural modeling; P. Lukavsky (Central European Institute of Technology, Masaryk University) for NMR resonance assignments and the pUC18 plasmid for toeprinting experiments; A.Willis (Medical Research Council Toxicology Unit) for the pRL plasmid for LUC experiments and G. Armstrong and E. Eisenmesser for assistance in NMR data collection and processing. This work was supported by US National Institutes of Health grant GM081346 to J.S.K. M.E.F. was supported as an American Heart Association predoctoral fellow (grant no. 0815655G). J.S.K. is an Early Career Scientist of the Howard Hughes Medical Institute. T.G.'s

laboratory is supported by the Howard Hughes Medical Institute.

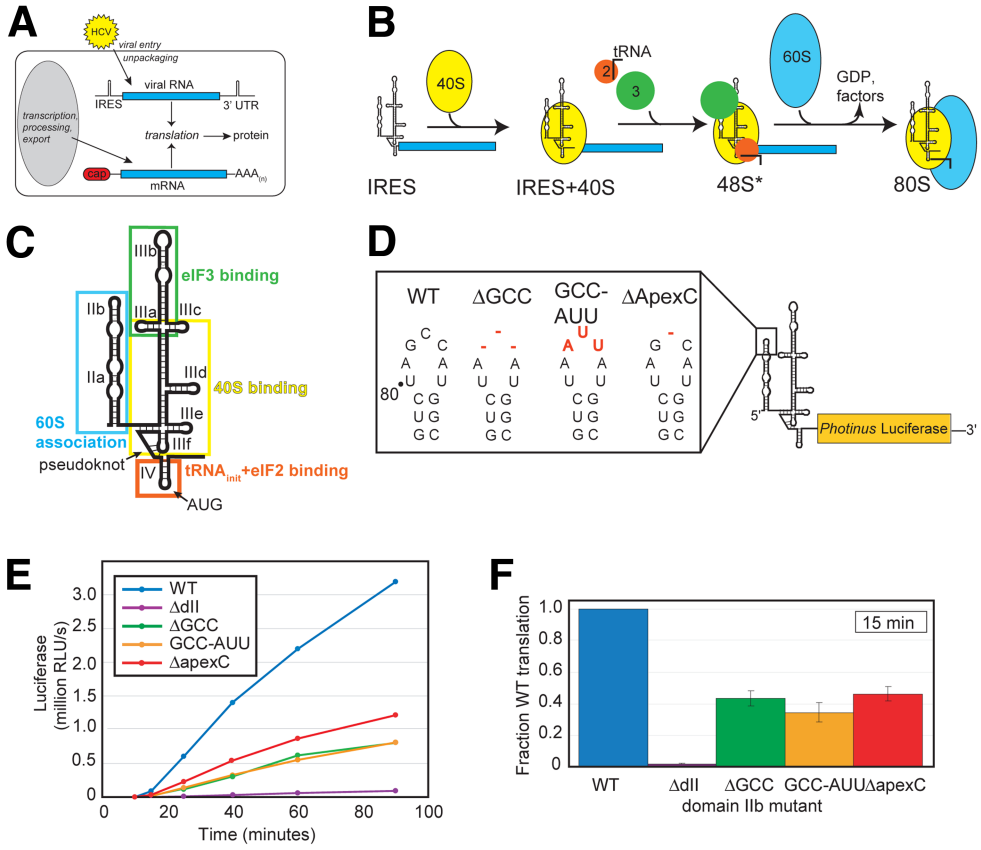

Figure 6.1 ***In vitro*** **translation analysis of dIIb mutations**. (**A**) HCV viral RNA and cellular mRNA differ in their origin and features. HCV viral RNA is delivered directly to the cytoplasm lacking a 7-methylguanosine cap and poly(A) tail, whereas cellular mRNA is produced and processed in the nucleus before exportation to the cytoplasm with a cap and tail. However, both are translated by the same cellular machinery, mandating different mechanisms of initiation. (**B**) Simplified diagram of HCV IRES 80S ribosome assembly mechanism. The IRES binds the 40S subunit (yellow), then eIF3 (green) and the ternary complex (eIF2–GTP– Met-tRNAiMet; orange), and finally, after GTP hydrolysis and eIF release, the 60S subunit (blue) joins to form an 80S ribosome. The noncanonical 48S* complex requires fewer initiation factors compared to canonical 48S complexes. (**C**) Cartoon representation of the secondary structure of the HCV IRES. The location of the start AUG is shown. Boxed areas indicate the parts of the IRES involved in different steps of the mechanism shown in B. Ternary complex (eIF2–GTP– Met-tRNAiMet) indicated by orange. (**D**) Schematic of mutations (red) made to dIIb in the context of the uncapped and unpolyadenylated LUC reporter. WT RNA is shown at left. (**E**) Time course of a translation assay from 0 to 90 min as measured by produced LUC relative light units (RLUs). (**F**) Fifteen-minute translation assay with RLUs calculated as a fraction of the WT IRES. Error bars represent s.e.m. for three biological replicates performed in triplicate.

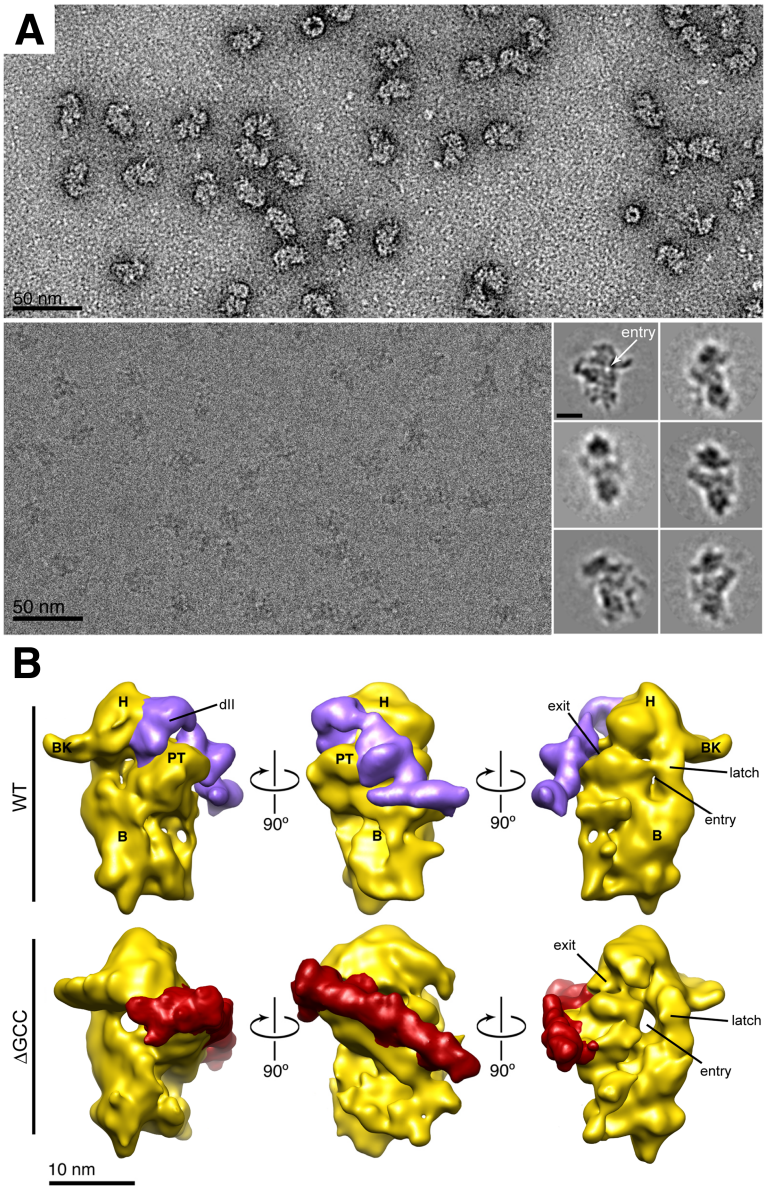

Figure 6.2 EM of HCV IRES mutant ΔGCC in complex with the 40S subunit. **(A)** Representative electron micrographs of ΔGCC IRES–40S complexes, negatively stained (top) and embedded in vitrified ice (bottom). Inset, six classes of projection averages of particles in vitrified ice showing different views of the complex calculated using MSA and MRA programs in IMAGIC54. The readily observed open entry tunnel is indicated (arrow). Scale bar for inset, 10 nm. (**B**) Structures of WT (top, purple)23 and ΔGCC (bottom, red) HCV IRES bound to the 40S ribosomal subunit (yellow). Solvent-accessible (left), exit-channel (middle) and solvent-inaccessible (right) sides are shown. Key structural features are labeled (H, head; PT, platform; BK, beak; B, body). The entry and exit sites for mRNA as well as the location of dII are indicated.

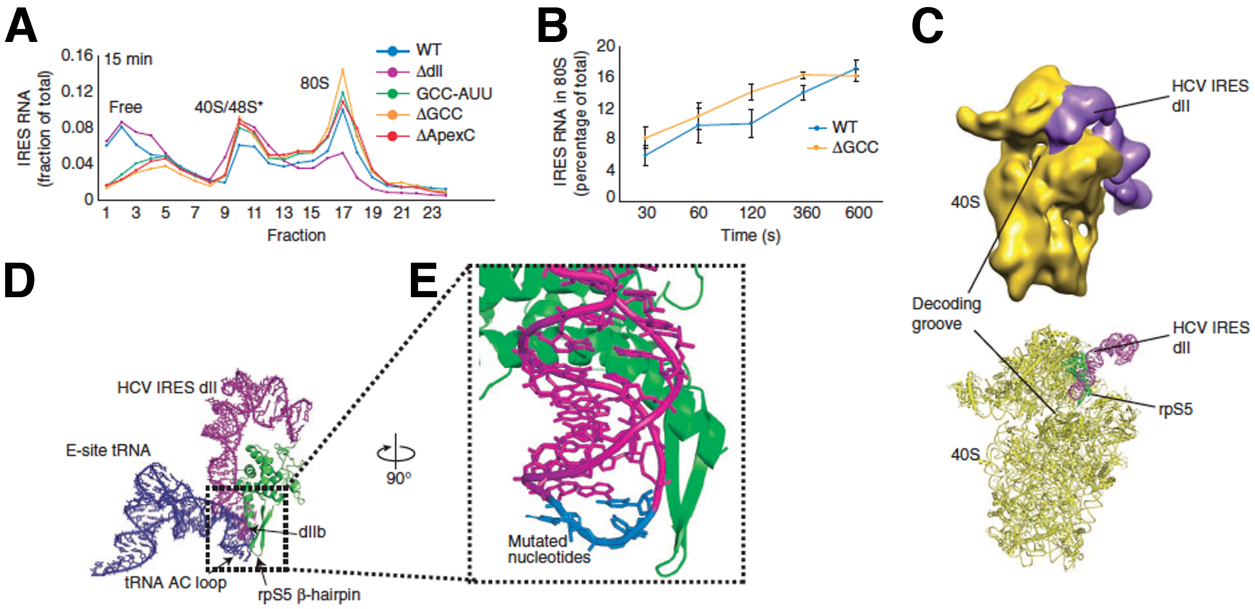

Figure 6.3 WT and mutant IRES ribosome assembly assays and position of domain IIb. (**A**) Graph of radiolabeled IRES RNA migration through a sucrose gradient after 15 min incubation in RRL followed by ultracentrifugation. Free, IRES RNA not bound by any translation components. (**B**) Amount of 80S complex formed at time points from 0.5 min to 10 min. Error bars represent s.e.m. from three independent experiments. (**C**) Top, cryo-EM reconstruction of the full-length HCV IRES RNA (purple) bound to mammalian 40S subunit (yellow) (23). Bottom, crystal structure of 40S subunit from *Tetrahymena thermophila*32 (yellow) and the NMR structure of HCV IRES dII (33) (purple) placed into the cryo-EM reconstruction (not shown). RpS5 is shown in green. (**D**) Comparison of the orientation of E-site bound tRNA (blue) and HCV IRES dII (purple) within the decoding groove. The position of dII was determined on the basis of the model shown in D **(**22,24), and the E-site tRNA position, from a crystal structure of the *T. thermophilus* 70S ribosome (55). rpS5 (rpS7 in bacteria) is shown in green, and its β-hairpin and the tRNA anticodon (AC) loop are indicated. (**E**) Close-up view of the position of dIIb (purple) near the β-hairpin of rpS5 (green). The nucleotides that were mutated in this study are shown in blue.

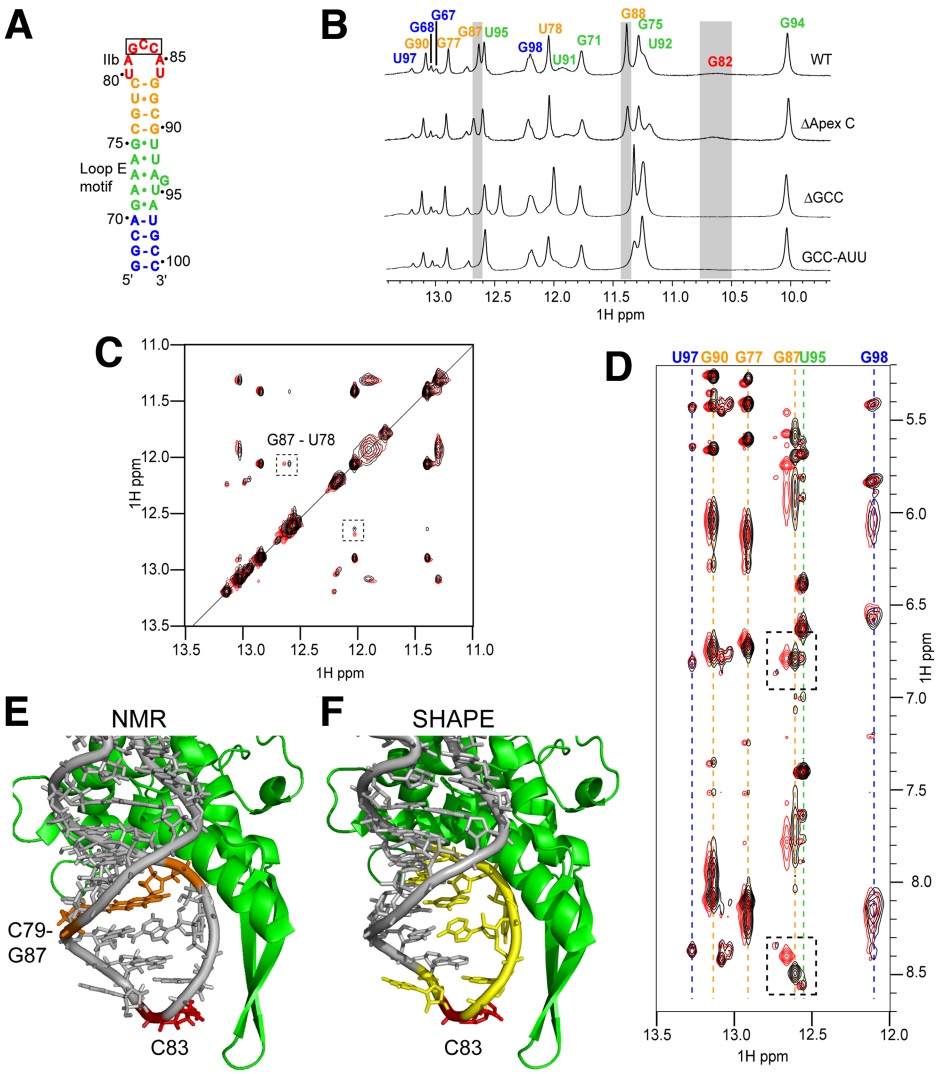

Figure 6.4 Characterization of the structural changes induced by dIIb mutation. (**A**) Secondary structure of the RNA sequence (previously solved)33 used to characterize the structural changes induced by mutation of dIIb. Elements are color-coded to match B–F. (**B**) One-dimensional 1H-NMR spectra (in water) of the WT and dIIb mutant RNAs. This part of the spectrum contains resonances from the base imino protons. Assignments for WT are labeled. The gray boxes indicate the most shifted resonances in all three mutant spectra. (**C**) Overlaid 2D 1H-NOESY NMR spectra (in water) for WT (black) and ΔapexC (red). The portion containing the cross-peaks between imino protons is shown, and the cross-peak for G87 and U78 imino protons is indicated. (**D**) Overlaid spectra and color scheme as in **c**, showing the cross-peaks between imino and other protons. Dashed boxes indicate locations of the cross-peaks between the G87 imino and the C79 amino protons. Imino proton resonances are indicated. (**E**) Close-up view of the tip of dIIb against the β-hairpin of rpS5 (green). The C79-G87 base pair (orange) and the location of the single base deletion in ΔapexC (C83, red) are indicated. (**F**) View as in E, with the locations of previously reported increases in chemical modification in the ΔapexC mutant22 shown (yellow). SHAPE, selective 2-OH acylation analyzed by primer extension.

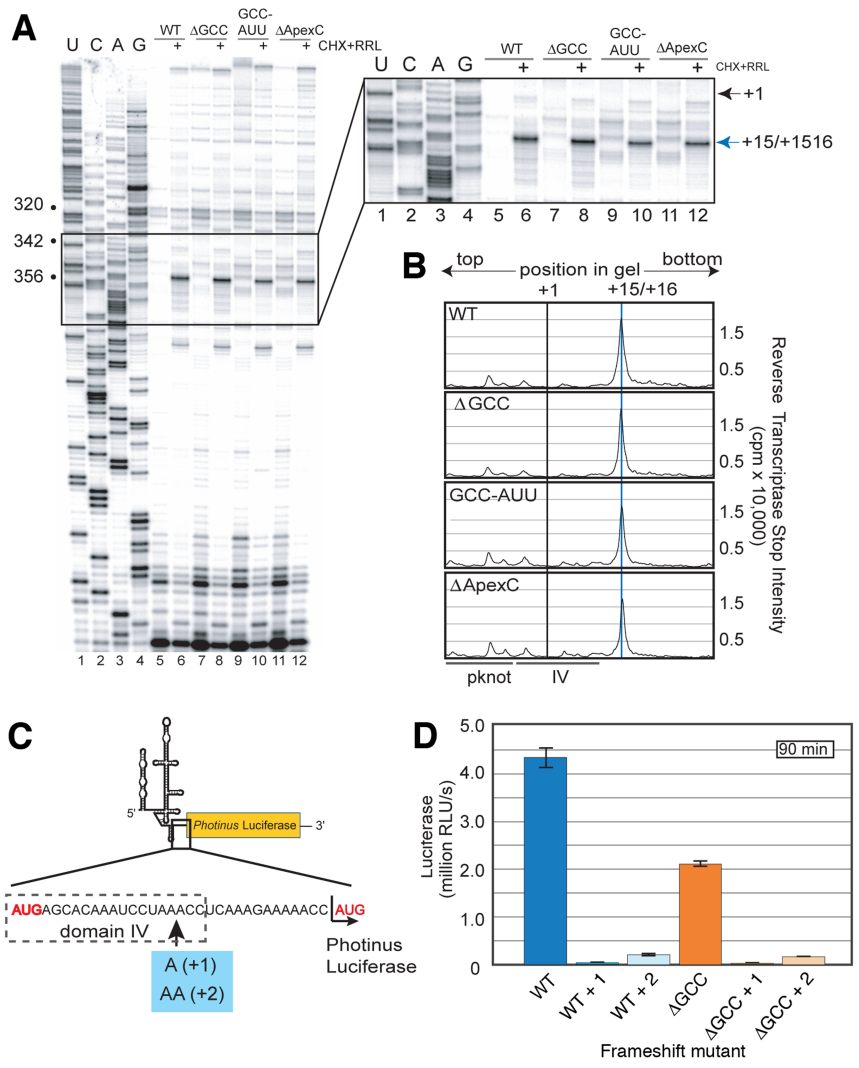

Figure 6.5 Biochemical analysis of AUG docking and potential frameshift. (**A**) Denaturing sequencing gel of the reverse transcription and toeprinting of WT and mutant IRES RNAs; the relevant part of the gel is magnified at right. Lanes 1–4, dideoxy sequencing reaction; lanes 5, 7, 9 and 11, free IRES; lanes 6, 8, 10 and 12, IRES–80S complexes (formed by incubation in RRL with cycloheximide (CHX)). Nucleotide numbers (bullets; left), the A of the AUG (black arrow; +1) and the toeprint (blue arrow; +15 and +16) are indicated. (**B**) Graph of quantified, normalized and background-corrected IRES–80S toeprints as in A. Positions +1 (black) and +15 and +16 (blue) are indicated. pknot, pseudoknot; dIV, domain IV. (**C**) Cartoon representation of the uncapped, unpolyadenylated monocistronic LUC reporter; the RNA region between the viral AUG and the LUC AUG (both shown in red) is magnified. One or two adenosines (blue box) were added for frameshift analysis. (**D**) LUC activity detected during a 90-min translation assay of WT and ΔGCC reporters, unmodified or with one (+1) or two (+2) additional adenosine residues. RLUs, relative light units. Error bars represent s.e.m. for three biological replicates performed in triplicate.

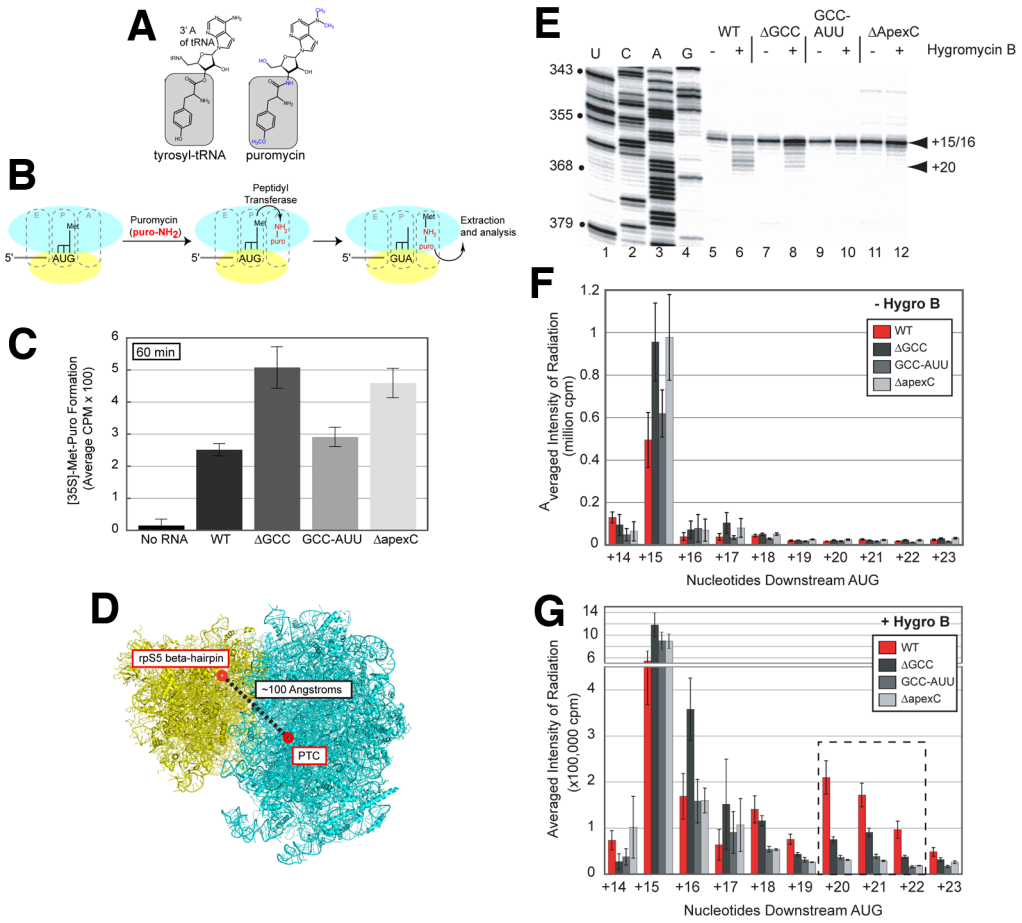

Figure 6.6 Puromycin and toeprinting assays with antibiotic. (**A**) Structures of tyrosyl-tRNA (left) and puromycin (right); amino acid residue and puromycin mimic boxed in gray; differences are shown in blue. (**B**) Cartoon representation of the puromycin assay (the 40S subunit is shown in yellow, the 60S in blue). From left to right: IRES–40S formation, then Met-tRNAiMet and puromycin binding in the 60S subunit P and A sites, respectively. Next, peptidyl transferase activity results in methionine bound to puromycin via a noncanonical amide linkage (Met-NH2-puro), then extraction. (**C**) Quantified and background-corrected graph of Met-NH2-puro formation after 60 min on WT and mutant IRES RNAs. Error bars represent s.e.m. for three independent duplicate experiments. (**D**) Yeast 80S ribosome crystal structure56 (the 40S subunit is shown in yellow, the 60S in cyan). Approximate locations and distance between the IRES dIIb-rpS5 interaction and the peptidyl transferase center (PTC) are shown. (**E**) Toeprint gel; lanes 1–4, dideoxy sequencing reactions; lanes 5, 7, 9 and 11, free IRES–80S complexes in RRL without any antibiotic; lanes 6, 8, 10 and 12, initiating and elongating IRES–80S complexes formed in RRL with hygromycin B. Initiating complexes (black arrowhead; +15 and +16), elongating complexes (blue arrowhead; +20) and nucleotide numbers (bulleted) are indicated. (**F,G**) Graph of quantified, normalized and background-corrected toeprints of WT IRES (red) and dIIb mutants (gray) without (**F**) and with (**G**) hygromycin B. Error bars represent s.d. for three independent experiments.

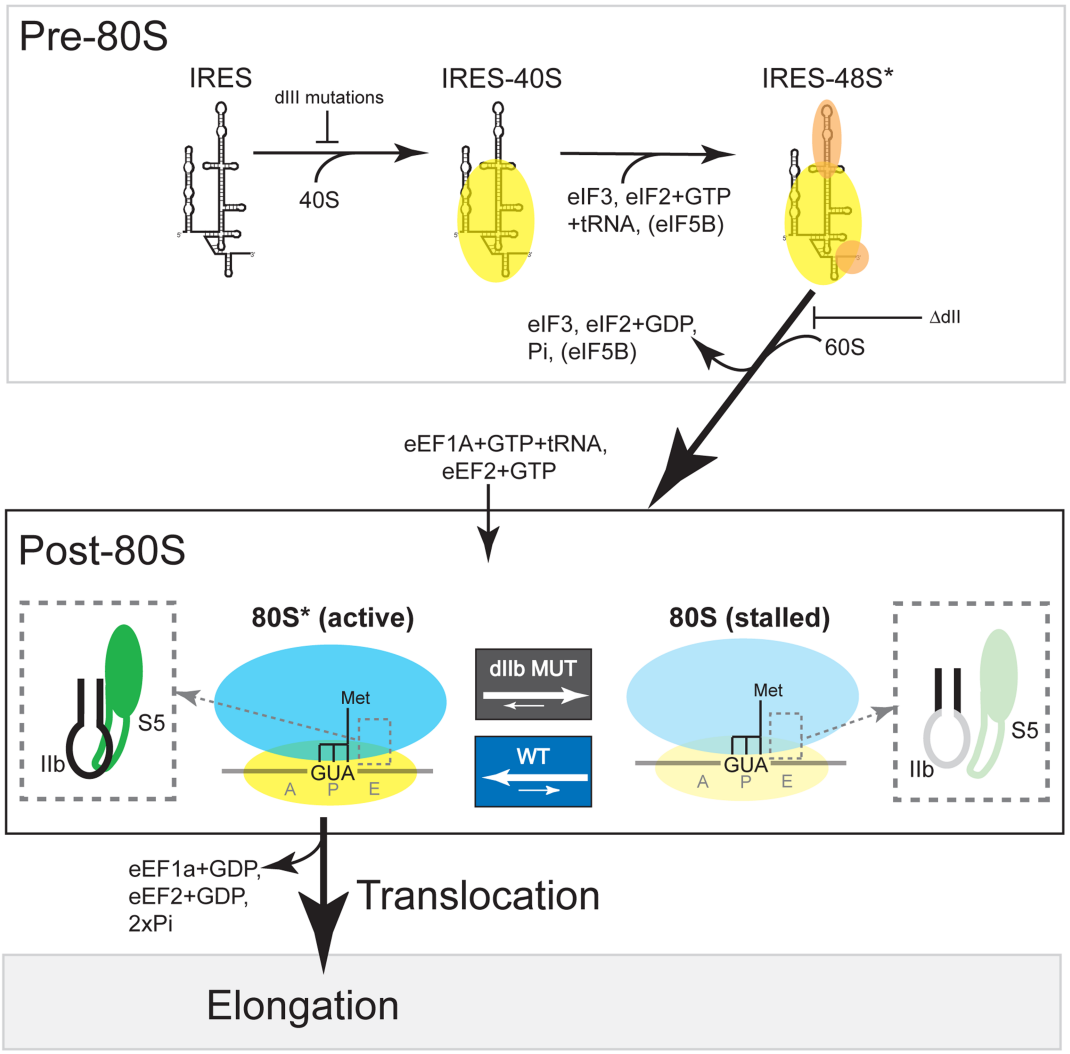

Figure 6.7 Model of the role of dIIb in HCV IRES translation initiation. **Top**, simplified pathway for HCV IRES–driven initiation of translation up to 80S ribosome formation. This process includes 40S subunit binding, ternary complex (eIF2– Met-tRNAiMet–GTP) binding and eIF3 binding, then subunit joining and factor release. **Bottom**, the post–80S assembly events that we hypothesize occur within the newly formed 80S ribosome. dIIb (black) interacts with the β-hairpin of rpS5 (green; dashed box at left), and this favors a conformation that is able to accept the delivery of a tRNA into the A site (catalyzed by eEF1A) and the subsequent translocation (catalyzed by eEF2). We term this fully competent conformation 80S*. When dIIb is mutated (right), the local structure of the dIIb apical loop shifts toward an inactive conformation, and the productive interaction with rpS5 is lost (dashed box at right). This favors an 80S ribosome state that stalls before translocation. In our model, the active (80S*) and inactive (80S) states are in dynamic equilibrium, and the presence of an intact dIIb shifts the equilibrium toward 80S*, thus promoting progression to elongation.

WORKS CITED IN CHAPTER 6

1. Tsukiyama-Kohara, K., Iizuka, N., Kohara, M. & Nomoto, A. (1992). Internal ribosome entry site within hepatitis C virus RNA. *J. Virol.* **66**, 1476–1483.

2. Bukh, J., Purcell, R.H. & Miller, R.H. (1992). Sequence analysis of the 5′ noncoding region of hepatitis C virus. *Proc. Natl. Acad. Sci. USA*, **89**, 4942–4946.

3. Simmonds, P. *et al*. (1993). Sequence variability in the 5′ non-coding region of hepatitis C virus: identification of a new virus type and restrictions on sequence diversity. *J. Gen. Virol*. **74**, 661–668.

4. Fraser, C.S. & Doudna, J.A. (2007). Structural and mechanistic insights into hepatitis C viral translation initiation. *Nat. Rev. Microbiol*. **5**, 29–38.

5. Jackson, R.J., Hellen, C.U. & Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 113–127.

6. Kieft, J.S., Zhou, K., Jubin, R. & Doudna, J.A. (2001). Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA*, **7**, 194–206.

7. Kolupaeva, V.G., Pestova, T.V. & Hellen, C.U. (2000). An enzymatic footprinting analysis of the interaction of 40S ribosomal subunits with the internal ribosomal entry site of hepatitis C virus. *J. Virol*. **74**, 6242–6250.

8. Lytle, J.R., Wu, L. & Robertson, H.D. (2001). The ribosome binding site of hepatitis C virus mRNA. *J. Virol*. **75**, 7629–7636.

9. Lytle, J.R., Wu, L. & Robertson, H.D. (2002). Domains on the hepatitis C virus internal ribosome entry site for 40s subunit binding. *RNA*, **8**, 1045–1055.

10. Pestova, T.V., Shatsky, I.N., Fletcher, S.P., Jackson, R.J. & Hellen, C.U. (1998). A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev*. **12**, 67–83.
11. Otto, G.A. & Puglisi, J.D. (2004). The pathway of HCV IRES-mediated translation initiation. *Cell,* **119**, 369–380.

12. Ji, H., Fraser, C.S., Yu, Y., Leary, J. & Doudna, J.A. (2004). Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. *Proc. Natl. Acad. Sci. USA,* **101**, 16990–16995.

13. Fraser, C.S., Hershey, J.W. & Doudna, J.A. (2009). The pathway of hepatitis C virus mRNA recruitment to the human ribosome. *Nat. Struct. Mol. Biol.* **16**, 397–404.

14. Locker, N., Easton, L.E. & Lukavsky, P.J. (2007). HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly. *EMBO J.* **26**, 795–805.

15. Terenin, I.M., Dmitriev, S.E., Andreev, D.E. & Shatsky, I.N. (2008). Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. *Nat. Struct. Mol. Biol.* **15**, 836–841.

16. Kim, J.H., Park, S.M., Park, J.H., Keum, S.J. & Jang, S.K. (2011). eIF2A mediates translation of hepatitis C viral mRNA under stress conditions. *EMBO J.* **30**, 2454–2464.

17. Kieft, J.S. *et al.* (1999). The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold. *J. Mol. Biol.* **292**, 513–529.

18. Lukavsky, P.J. (2009). Structure and function of HCV IRES domains. *Virus Res.* **139**, 166–171.

19. Sizova, D.V., Kolupaeva, V.G., Pestova, T.V., Shatsky, I.N. & Hellen, C.U. (1998). Specific interaction of eukaryotic translation initiation factor 3 with the 5′ nontranslated regions of hepatitis C virus and classical swine fever virus RNAs. *J. Virol.* **72**, 4775–4782.

20. Honda, M., Brown, E.A. & Lemon, S.M. (1996). Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA,* **2**, 955–968.

21. Berry, K.E., Waghray, S. & Doudna, J.A. (2010). The HCV IRES pseudoknot positions the initiation codon on the 40S ribosomal subunit. *RNA,* **16**, 1559–1569.

22. Filbin, M.E. & Kieft, J.S. (2011). HCV IRES domain IIb affects the configuration of coding RNA in the 40S subunit's decoding groove. *RNA,* **17**, 1258–1273.

23. Spahn, C.M. *et al.* (2001). Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science,* **291**, 1959–1962.

24. Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J.D. & Stark, H. (2005). Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure,* **13**, 1695–1706.

25. Fukushi, S. *et al.* (2001). Ribosomal protein S5 interacts with the internal ribosomal entry site of hepatitis C virus. *J. Biol. Chem.* **276**, 20824–20826.

26. Wower, J., Scheffer, P., Sylvers, L.A., Wintermeyer, W. & Zimmermann, R.A. (1993). Topography of the E site on the *Escherichia coli* ribosome. *EMBO J.* **12**, 617–623.

27. Yusupov, M.M. *et al*. (2001). Crystal structure of the ribosome at 5.5-Å resolution. *Science*, **292**, 883–896.

28. Döring, T., Mitchell, P., Osswald, M., Bochkariov, D. & Brimacombe, R. (1994). The decoding region of 16S RNA; a cross-linking study of the ribosomal A, P and E sites using
tRNA derivatized at position 32 in the anticodon loop. *EMBO J*. **13**, 2677–2685.

29. Odreman-Macchioli, F., Baralle, F.E. & Buratti, E. (2001). Mutational analysis of the different bulge regions of hepatitis C virus domain II and their influence on internal ribosome entry site translational ability. *J. Biol. Chem*. **276**, 41648–41655.

30. Kalliampakou, K.I., Psaridi-Linardaki, L. & Mavromara, P. (2002). Mutational analysis of the apical region of domain II of the HCV IRES. *FEBS Lett*. **511**, 79–84.

31. Passmore, L.A. *et al*. (2007). The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol. Cell*, **26**, 41–50.

32. Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A. & Ban, N. (2011). Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science,* **331**, 730–736.

33. Lukavsky, P.J., Kim, I., Otto, G.A. & Puglisi, J.D. (2003). Structure of HCV IRES domain II determined by NMR. *Nat. Struct. Biol*. **10**, 1033–1038.

34. Pestova, T.V., Hellen, C.U. & Shatsky, I.N. (1996). Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol. Cell Biol*. **16,** 6859–6869.

35. Wilson, J.E., Pestova, T.V., Hellen, C.U. & Sarnow, P. (2000). Initiation of protein synthesis from the A site of the ribosome. *Cell*, **102**, 511–520.

36. Hartz, D., McPheeters, D.S., Traut, R. & Gold, L. (1988). Extension inhibition analysis of translation initiation complexes. *Methods Enzymol*. **164**, 419–425.

37. Devaraj, A., Shoji, S., Holbrook, E.D. & Fredrick, K. (2009). A role for the 30S subunit E site in maintenance of the translational reading frame. *RNA*, **15**, 255–265.

38. Monro, R.E. & Marcker, K.A. (1967). Ribosome-catalysed reaction of puromycin with a formylmethionine-containing oligonucleotide. *J. Mol. Biol*. **25**, 347–350.

39. Peske, F., Savelsbergh, A., Katunin, V.I., Rodnina, M.V. & Wintermeyer, W. (2004). Conformational changes of the small ribosomal subunit during elongation factor G-dependent tRNA-mRNA translocation. *J. Mol. Biol*. **343**, 1183–1194.

40. Dibrov, S.M. *et al*. (2012). Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches. *Proc. Natl. Acad. Sci. USA*, **109**, 5223–5228.

41. Robert, F. & Brakier-Gingras, L. (2003). A functional interaction between ribosomal proteins S7 and S11 within the bacterial ribosome. *J. Biol. Chem*. **278**, 44913–44920.

42. Galkin, O. *et al*. (2007). Roles of the negatively charged N-terminal extension of Saccharomyces cerevisiae ribosomal protein S5 revealed by characterization of a yeast strain containing human ribosomal protein S5. *RNA*, **13**, 2116–2128.

43. Geigenmüller, U. & Nierhaus, K.H. (1990). Significance of the third tRNA binding site, the E site, on E. coli ribosomes for the accuracy of translation: an occupied E site prevents the binding of non-cognate aminoacyl-tRNA to the A site. *EMBO J*. **9**, 4527–4533.

44. Petropoulos, A.D. & Green, R. (2012). Further in vitro exploration fails to support the allosteric three-site model. *J. Biol. Chem*. **287**, 11642–11648.

45. Uemura, S. *et al*. (2010). Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature*, **464**, 1012–1017.

46. Chen, C. *et al*. (2011). Allosteric vs. spontaneous exit-site (E-site) tRNA dissociation early in protein synthesis. *Proc. Natl. Acad. Sci. USA*, **108**, 16980–16985.

47. Malygin, A.A., Yanshina, D.D. & Karpova, G.G. (2009). Interactions of human ribosomal proteins S16 and S5 with an 18S rRNA fragment containing their binding sites. *Biochimie*, **91**, 1180–1186.

48. Ian'shina, D.D., Malygin, A.A. & Karpova, G.G. (2006). Binding of human ribosomal protein S5 with the 18S rRNA fragment 1203–1236/1521–1698 [in Russian]. *Mol. Biol.* (Mosk.), **40**, 460–467.

49. Yu, Y. *et al*. (2009). Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. *Nucleic Acids Res*. **37**, 5167–5182.

50. Antúnez de Mayolo, P. & Woolford, J.L. Jr. (2003). Interactions of yeast ribosomal protein rpS14 with RNA. *J. Mol. Biol*. **333**, 697–709.

51. Lomakin, I.B., Kolupaeva, V.G., Marintchev, A., Wagner, G. & Pestova, T.V. (2003). Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes Dev*. **17**, 2786–2797.

52. Acker, M.G. *et al*. (2009). Kinetic analysis of late steps of eukaryotic translation initiation. *J. Mol. Biol.* **385**, 491–506.

53. Fringer, J.M., Acker, M.G., Fekete, C.A., Lorsch, J.R. & Dever, T.E. (2007). Coupled release of eukaryotic translation initiation factors 5B and 1A from 80S ribosomes

following subunit joining. *Mol. Cell. Biol*. **27**, 2384–2397.

54. van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R. & Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J. Struct. Biol*. **116**, 17–24.

55. Selmer, M. *et al*. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.

56. Ben-Shem, A. *et al*. (2011). The structure of the eukaryotic ribosome at 3.0-Å resolution. *Science*, **334**, 1524–1529.

57. Stoneley, M., Paulin, F.E., Le Quesne, J.P., Chappell, S.A. & Willis, A.E. (1998). C-Myc 5′ untranslated region contains an internal ribosome entry segment. *Oncogene*, **16**, 423–428.

58. Keel, A.Y., Easton, L.E., Lukavsky, P.J. & Kieft, J.S. (2009). Large-scale native preparation of in vitro transcribed RNA. *Methods Enzymol*. **469**, 3–25.

59. Ohi, M., Li, Y., Cheng, Y. & Walz, T. (2004). Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol. Proced. Online* **6**, 23–34.

60. Mindell, J.A. & Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol*. **142**, 334–347.

61. Grigorieff, N. (2007). FREALIGN: high-resolution refinement of single particle structures. *J. Struct. Biol*. **157**, 117–125.

62. Sousa, D. & Grigorieff, N. (2007). Ab initio resolution measurement for single particle structures. *J. Struct. Biol*. **157**, 201–210.

63. Kleywegt, G.J. & Jones, T.A. (1996). xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr. D Biol. Crystallogr*. **52**, 826–828.

64. Pettersen, E.F. *et al*. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem*. **25**, 1605–1612.

Chapter VII

Concluding remarks

HUMAN AQUAPORIN 3

Aquaporin 3 is a human aquaglyceroporin permeable to water, glycerol, urea and small solutes. In the skin, AQP3 has been shown to have roles in the process of wound healing specifically affecting cell migration and cell proliferation (1). Currently only two aquaglyceroprins structures have been determined, GlpF from E. coli (2) and PfAQP from the parasite (3), plasmodium falciparum, revealing a wider constriction site and a amphiphilic pore to accommodate the hydrophobic backbone of glycerol. The structure of a human aquaglyceroporin still remains to be determined. In work presented here, I have been able to heterologously express and purify human AQP3 yielding 1 mg/liter of *Pichia pastoris* culture. Purified human AQP3 was characterized using negative-stain TEM showing the observed square shaped protein particle consistent with previously published images of aquaporin tetramers (4,5). Experiments to produce two-dimensional crystals for structure determination using electron crystallography were unsuccessful due to aggregation of the protein during reconstitution experiments.

Current literature of hAQP3 has focused on understanding the function in different biological processes. There is growing literature that hAQP3 is involved in targeting epidermal growth factor to the membrane (6) which is necessary for cell migration in wound healing. An interesting feature of hAQP3 is that both water and glycerol permeation through the channel is regulated by pH (7). Other hAQP3 regulatory mechanisms identified include sensitivity to cationic metals (7,8). In the hAQP3 literature there is currently a lack of research focused on how

hAQP3 maybe involved in different physiological processes and what types of regulatory mechanism are involved. Using electron crystallography to study the structure of hAQP3 would provide a model to start testing hypotheses regarding the regulation of water permeation through hAQP3 and regions that maybe responsible for targeting epidermal growth factor to the basolateral membrane of cells.

The efforts presented here to produce two-dimensional crystals, let alone attempts to reconstitute human AQP3 with lipids, represent difficulties faced by the field of electron crystallography and the structural biology of human membrane proteins. Two-dimensional crystallization has been the barrier for studying the structure of membrane proteins using electron crystallography. We currently do not understand the formation of two-dimensional crystals but have identified several factors that are important for two-dimensional crystallization such as lipid to protein ratio, detergent, lipid, stability of purified protein and dialysis conditions. During two-dimensional crystal formation, there is a multitude of interactions taking place that govern reconstitution such as protein-lipid, protein-detergent and lipid-detergent that make the process extremely complex. For each individual membrane protein, the specific interactions will be different depending on the affinity of the protein for the detergent versus the lipid. Therefore conditions that work for one membrane protein maybe drastically different than for another membrane protein even for proteins of the same family. From studies of human AQP3, I have learned that understanding the biochemistry of the purified membrane protein is essential before attempting to screen two-dimensional crystallization conditions.

E. COLI GALACTOSE PERMEASE

Glucose is taken up by cells lining the intestine, transported to the blood stream and shuttled to tissues throughout the body. This process is mediated by a family of sugar transporters that transport glucose across the hydrophobic cell membrane. GLUT1, a facilitated glucose transporter highly abundant in erythrocytes, has been studied for decades, but the mechanism by which glucose is transported has not been elucidated. Previously published work from our lab showed that the E. coli galactose permease (GalP) forms trimers when embedded in the membrane observed in a projection map calculated from negative-stain TEM images (9). Here I presented optimization of two-dimensional crystals of GalP by exchanging GalP into the detergent LDAO followed by reconstitution with DMPC. Two-dimensional crystals measuring up to 10 μm in diameter with crystalline patches measuring 400 nm were produced and used to calculate a projection map from cryo-EM images at 18-Å resolution revealing triangular arrays of trimers. The lack of significant improvement in resolution of the cryo-EM projection map of GalP over the map obtained from negative-stain TEM images maybe a result of poor embedding conditions used for the crystals to collect cryo-EM images that disrupt coherence. Future work will focus on improving embedding conditions for cryo-EM data collection and improving coherence of the crystals so that diffraction data can be collected to determine the structure of GalP embedded in a lipid bilayer.

Several structures of *E. coli* sugar transporters have been determined, including lactose permease (10), fucose permease (11) and most recently xylose permease (12), that have revealed evidence for the alternating access model for substrate transport. A remaining question regarding the sugar transporter structures determined is whether the structures determined of the membrane proteins stabilized in detergent micelle reflect conformations that exist when the transporters are

embedded in a lipid bilayer. Crystals of membrane proteins grown for structure determination using X-ray crystallography involve the formation of the lowest energy state of the protein within a detergent micelle facilitated by protein-protein contacts within the crystal. The most direct way to determine the accuracy of the observed conformations is by comparison of structures determined using X-ray and electron crystallography.

The only MFS protein studied by electron crystallography that has yielded a three-dimensional structure is the bacterial oxalate transporter at 6.5 Å resolution (13). The structure of OxlT revealed that the protein adopts an occluded conformation; the substrate bound state, in the membrane. Details regarding the specific residues involved in oxalate transport remain to be determined because of the low resolution. Unfortunately there is currently not a structure of the oxalate transporter determined using X-ray crystallography so we cannot compare results from the two different methods. However, comparison of the conformation of the oxalate transporter (13) with the known *E. coli* sugar transporters (10,11,12) reveals evidence that supports the alternating access model that has been proposed. This suggests that the conformations adopted by the sugar transporters in three-dimensional crystals provide snap shots of possible conformations that the sugar transporters may adopt in the membrane. To provide insight into the driving forces of the conformations and the true assessment of the validity of the possible conformations, the structure of a sugar transporter needs to be observed in the membrane in the unbound, bound and inhibitor states, which can only occur by using electron crystallography.

αB-CRYSTALLIN

Small heat shock proteins function to protect the cell during times of stress. α-crystallins are a family of ATP-independent small heat shock proteins that are responsible for maintaining

the transparency of the lens (14,15,16). Of the two types of α-crystallins, αB is more widely expressed most notably in the brain and heart, where mutations in αB have been associated with the development of Alzheimer's and cardiac myopathies (17,18). Investigations of the structure of αB crystallin have been difficult due to the heterogeneous nature of multimers formed (19). A recent study has finally provided details regarding the structure of the stable dimer formed by αB crystallin within the multimers using solid state NMR (20). The solid state NMR structure revealed a highly curved dimer formed by the α crystallin domains, the basic building block for assembly of the heterogeneous multimers, but the flexible N-terminus was not able to be resolved. Remaining questions are what is the structure of the N-terminus and how do the multimers assemble that allow for rapid exchange of subunits and the formation of a heterogeneous mixture of multimers. In work that published in PNAS (21) and presented here, we have determined a model for the N-terminus of αB crystallin based on NMR restraints and homology modeling, as well as a pseudo atomic model of full length αB crystallin that provides a mechanism for multimerization.

In collaboration with Rachel Klevit and Hartmut Oschkinat, we have combined structural restraints from solution and solid state NMR, SAXS and negative-stain TEM to develop a model for accommodating up to 8 additional subunits of αB into a 24-mer tetrahedral assembly based on a previous reconstruction of αB from negative-stain TEM images (21). Our model suggests that there are four regions providing interactions that facilitate the assembly of αB multimers. The smallest building block of the multimer consists of β6+7 strands from each α-crystallin domain that interact forming a homodimer. A triangular array of dimers is formed by the binding of the conserved IXI motif of one dimer to a hydrophobic pocket formed by β4 and β8 strands of a neighboring dimer. N-terminal helices interact at the three-fold axis contributing to the

formation of higher order multimers by association of the pseudo hexameric units. Variable interactions of β2 strands in the N-terminus allow incorporation of dimers into the six openings on the edge of each two-fold axis. For the first time, a pseudo atomic model for full length αB was proposed that agrees with previously published biochemical data and provides a model to analyze known disease causing mutations.

In addition to providing a model for multimerization, we addressed a previously published paper that provided controversial findings that αB formed homogeneous tetrahedral 24mer subunits. The previous reconstruction of αB (22) used for building our model claimed that αB formed tetrahedral 24mer multimers contradicting the established heterogeneous distribution known for αB. Projection analysis comparing multimers of 24, 26 and 28 subunits showed that using negative-stain TEM different multimers were not distinguishable at the resolution limit of negative stain (21). Using our model for multimerization it is reasonable to believe that the previous reconstruction of αB probably contained different multimers but application of symmetry during reconstruction and the limit of negative-stain TEM differences could not be resolved.

After publication of our results in PNAS, Buchner's group followed with publication of a different pseudo atomic model of full length αB crystallin based on a subnanometer resolution single particle cryo-EM reconstruction and a model for multimerization (23) that was very similar to our work. The model for the 24mer published by Braun *et al*. (23) shows the dimer in a more tightly packed triangular array with the N-termini composed of three α-helices that pack at the edge of the three-fold axes based on structure prediction and density in the cryo-EM reconstruction. In comparison, the N-termini of our model are composed of two α-helices followed by two β-strands based on NMR restraints and homology modeling (21). The

secondary structure elements proposed by Braun *et al.* are largely based on the cryo-EM reconstruction and weakly supported by structure prediction. Braun *et al*. claim that in the cryo-EM reconstruction 3 domes of density are observed at the edge of the three fold axes which were identified as α-helices, which at the reported resolution of 9.4 Å might be visible (23). The calculation of subnanometer resolution is based on the 0.5 FSC cutoff, but FSC curves are not included in supplementary materials. Resolution of single particle cryo-EM reconstructions is a debated topic, the FSC criterion is highly susceptible to systematic error alignments, and thus details present in the map should be used to judge the reliability of FSC. Upon closer examination of the reconstructed density, the three "domes" are nothing more than subtle protrusions indicating that the reconstruction is not 9.4-Å resolution and that assignment of the protrusions to α-helices is mere speculation.

Braun *et al*. propose that αB crystallin forms stable pseudo hexameric units in solution and that equilibrium exists between different oligomeric states with the most prevalent being the 24-mer assembly (23). Several reconstruction of various oligomers ranging from 12 subunits to 48 subunits are presented showing multimers of αB crystallin typically are even numbered representing the addition of dimers. This was seen in the reconstruction of the 18-mer where an additional density was seen at the edge of two three-fold axes which supports our model that suggested a 24-mer is the most stable assembly that multimers can be made by the addition of dimers. Despite the different pseudo atomic model for full length αB crystallin proposed by Braun *et al*., cryo-EM data support our model for assembly of multimerization. The model we present provides a specific hypothesis for the particular interactions contributing to the formation of higher order multimers.

Each model for full length αB crystallin that is presented by our work (21) and Braun *et al.* (23) remains to be tested and verified by biochemistry studies such as FRET, to determine the proximity of residues in the assembled multimers. Extensive mutations studies such as cysteine scanning mutagenesis could reveal residues critical for forming the different interactions contributing to multimerization. Questions still remain as to the regulation of the multimerization and how this is involved in the function of αB crystallin to bind proteins then the release of proteins to be refolded. There are several identified phosphorylation sites that are not equally phosphorylated in isolated populations of αB crystallin (24), the role of phosphorylation in governing the oligomeric state or function is not currently understood. Both models (21,23) proposed have phosphorylation sites concentrated at the edge of the three fold axes, where the N-termini are located participating in interactions forming higher order multimeric assemblies. Investigation of the structure of the multimeric heterogeneous αB crystallin only provides a model, to understand how αB crystallin functions as a molecular chaperone, structural studies must be combined with study of disease causing mutations, regulation studies and in vitro functional assays.

DESIGNED SELF-ASSEMBLING PROTEIN NANOMATERIALS

The use of biological molecules as building blocks to design nanoscale materials has largely focused on DNA, but the use of proteins, as a design scaffold has been slower to develop. Nucleic acids have a distinct advantage over proteins in that the interactions of nucleic acids are well defined forming discrete hydrogen bonds with fewer numbers of different bases compared to the number and variability of amino acids. Using nucleic acids as a building block several different assemblies have been designed such as molecular containers (25), distinct patterns &

shapes (26) as well as macroscopic crystals (27). The vast array of protein functions seen in the cell, hold the promise that using proteins as a building block can increase the functionality of designed nanomaterials. Recent advances in protein design have enabled the possibility of designing protein-protein interactions. In David Baker's lab they have come up with a general method to design self-assembling oligomers with a defined symmetry using RosettaDesign (28,29**)** and foldit (30). The computational method consists of two steps: 1) symmetrical building blocks are docked into a target symmetrical architecture and 2) low-energy protein-protein interfaces are designed to drive self-assembly into the desired architecture. Using the method, two designs produced self-assembling protein oligomers exhibiting octahedral and tetrahedral symmetry. Constructs were verified in solution using single particle cryo-EM and specific residue contacts observed using X-ray crystallography. The remarkable accuracy of the method is indicated by the less than 1 Å RMSD of the crystal structure compared to the design. This work presented here was published in Science (31).

We collaborated with the Baker lab to initially screen designs via negative-stain TEM and later to reconstruct a three-dimensional model of the octahedral design at 15-Å resolution using single particle cryo-EM (31). The overall shape and size of each of the designs was easily observed using projection averages calculated from negative-stain TEM images. In the case of the octahedral assembly, individual particles exhibited features consistent with the design. The 15-Å resolution reconstruction of the O3-33 design from cryo-EM images revealed that in solution the oligomer exhibits the expected 432 symmetry of the octahedral point group. Unfortunately after selecting over 40,000 particles from the cryo-EM data and multiple rounds of alignment and refinement the reconstruction did not achieve sub-nanometer resolution. The main reason is thought to be the insufficient alignment of the particles and thus signal from high-

resolution features is lost. Many factors contribute to the inability of the particles to be aligned such as drift and sample freezing conditions. Despite these issues, the cryo-EM reconstruction was able to demonstrate that in solution the designed octahedral oligomers do in fact form the predicted octahedral symmetry.

Previous work to design discrete protein assemblies had used a variety of methods including cross-linking (32), engineered disulfide bonds (33,34), metal mediated interactions (35,36) and coiled coil, helical bundle interactions (37,38,39,40,41). The method published by King et al. in Science provides advancement in the field to use weak protein-protein interactions to drive self-assembly of symmetrical protein nanomaterials. Future work will be focused on engineering functionality into the protein assemblies as has been accomplished with nucleic acids, such as the DNA box that opens in response to pH (25).

MUTANT HCV IRES – 40S COMPLEX

Hepatitis C virus (HCV) infects approximately 3 million Americans each year and is the most common chronic blood borne disease in the United States. HCV is a single stranded RNA virus that replicates by hijacking the host protein production machinery using a single region of its RNA genome referred to as an internal ribosome entry site (IRES). The IRES is small region of the 5' untranslated region of the HCV RNA genome that is able to bind to the 40S ribosomal subunit and using only two host initiation factors is able to assemble the 80S ribosome and initiate translation of viral proteins (42). The HCV RNA is able to manipulate the ribosome using a cap-independent translation initiation pathway compared to native translation initiation that requires a capped 5' end and several eukaryotic initiation factors. Previous studies using cryo-electron microscopy have shown that the HCV IRES has been shown to bind to the solvent

accessible side of the 40S ribosome reaching into the E site of the decoding groove and induces a conformational change of the 40S ribosome (43). To understand the specific role of domain II of the HCV IRES, mutations were carried out at the apical end of the loop of the HCV IRES, domain IIb (dIIb), that reaches into the E-site, translation assays were conducted and single particle cryo-EM was used. We have shown that removal of only 3 nucleotides of dIIb alters the conformation of the 40S upon binding, resulting in significantly reduced rate of translation by affecting the first translocation step (44). This is the first study to show that interactions of the IRES with 40S are propagated to affects post-80S assembly affecting the transition from initiation to elongation. This work was published in the journal of Nature Structural & Molecular Biology (44).

Our lab carried out the single particle cryo-EM reconstruction of the mutant ΔGCC HCV IRES bound to the 40S ribosomal subunit (44). The three-dimensional model of the complex was reconstructed to 17-Å resolution from approximately 4328 individual particle images. The resolution of the reconstruction was limited by several factors including flexibility of the ribosome, association of the complexes in the vitrified ice and the limited orientation of the complexes. The limited orientation of the IRES-40S complexes in the vitrified ice reduced the amount of different views of the particle that were imaged. Most of the views obtained were of the profiles of the complex the solvent accessible side and the 60S-binding interface. For a symmetrical macromolecular complex limited views are less detrimental due to redundancy but with the asymmetric ribosome there is not the redundancy so as many different projections are required. Another factor that contributed to the low resolution was the low signal to noise from the lack of having many images of particular views that resulted in the inability to sufficiently align the images to achieve high resolution. Confidence in the resolution of the reconstruction

was provided by comparing features present in the density map to a map calculated from the *Tetrahymena thermophila* 40S crystal structure (45) filtered to 15 Å and 20-Å resolution.

At 17-Å resolution, differences in the conformation of the IRES bound to the 40S as well as differences in the conformation of the 40S ribosomal subunit itself could be observed. Three features were clearly observed in the reconstruction: 1) 40S mRNA exit channel, 2) the latch is closed in the 40S, and 3) ΔGCC HCV IRES binds along the platform. These observations are also present in projection averages of individual particles that were classified according to orientation in the cryo-EM images. Several projection class averages show a distinct hole or absence of density below the beak region of the 40S where as in the reconstruction of the wtIRES-40S complex this was closed. The orientation of the ΔGCC IRES along the platform domain is seen in several different projection class averages compared to the WT IRES – 40S complex where the IRES associates closely with the head, positioning dIIb in the E site. The identification of features present in the projection averages, which are a reflection of the populations present in the data and therefore are bias free, provides a measure of confidence in the final reconstruction of the ΔGCC IRES – 40S complex and the conclusions drawn.

Our paper proposes a model that the ribosome samples different conformations and that the binding of eukaryotic initiation factors stabilizes conformations that result in translation initiation. The mutant alters the equilibrium sampling less favorable conformations leading a decreased rate of translocation, ribosome stalls but then eventually conformation sampled is correct. A potential interaction is proposed based on comparing the crystal structure of the 40S ribosome (45) with a tRNA bound in the E-site with the WT IRES-40S complex reconstruction suggesting that dIIb of the IRES interacts with the beta hairpin of rpS5 of the 40S facilitating the first translocation event. From our paper we note that it is interesting that the potential rpS5-

IRES interaction is approximately > 100 Å from the peptidyl transferase center, indicating an allosteric effect propagated from the 40S to the PTC of the 60S ribosomal subunit. Future work to understand the allosteric effect is to investigate conformational changes in the 60S ribosomal subunit upon binding of mutant HCV IRES – 40S complexes using single particle cryo-EM as well as chemical probing after 80S formation. Our results have broader implications that understanding how the binding of HCV IRES can modulate the eukaryotic ribosome but also for understanding the complexity of interactions that make the ribosome a robust macromolecular machine.

FUTURE OF BIOLOGICAL TEM

From the studies presented here it is clear that the significant advantage of using TEM to study biology is by combining the results with other methods, such as NMR, X-ray crystallography and biochemical assays. The combination of different methods of TEM not only helps to overcome some of the hurdles in the field of TEM but also to form a complete story regarding the mechanisms used by macromolecular machines. With single particle cryo-EM and electron crystallography the achievement of high resolution has yet to be achieved consistently but low-resolution information is still valuable for answering many biological questions. From the assortment of studies discussed we have seen how results from negative-stain TEM, single particle cryo-EM and electron crystallography have been used. Projection class averages calculated from negative-stain TEM of αB crystallin were used to study the heterogeneity of multimerization and assess models developed from NMR experimental restraints. A single particle cryo-EM reconstruction provided conformation of the assembly of a designed octahedral assembly in solution to complement an atomic resolution structure determined by X-ray

crystallography. Large conformational changes were observed upon binding of a mutant IRES to the 40S ribosomal subunit from a single particle cryo-EM reconstruction of the GCC IRES-40S complex and combining with chemical probing assays, a model was proposed to explain the ability of a single stranded RNA to manipulate the ribosome. Electron crystallography study of GalP has shown that the sugar transporter forms trimers in the membrane with lipids facilitating the crystal contacts between proteins. To improve the resolution and applicability of TEM for studying structural biology, advancements must be made to increase the versatility of samples for data collection and data processing.

Single particle cryo-EM is an advantageous technique because protein crystals are not required, small amounts of protein are sufficient to collect data and large macromolecular complexes can be studied that are traditionally difficult to study by other methods. Currently, only large symmetrical molecular complexes have achieved atomic level resolution. The major issue underlying the ability to achieve high resolution and reconstruct small protein complexes is alignment of individual particles, the determination of the correct orientation of each of the particles in the vitrified ice, during the refinement and reconstruction phase of single particle cryo-EM. Developments in single particle cryo-EM are focused on improving the ability to reach high resolution with asymmetrical complexes and reconstruct smaller complexes. Recently a single particle cryo-EM reconstruction of the 80S ribosome reached near atomic resolution from only 30,000 individual particles using a direct electron detector and capturing movies to correct for beam induced movement (46). Another development is currently focused on studying smaller complexes using single particle cryo-EM by the use of antibody fragments bound to the protein of interest to help alignment that resulted in the reconstruction of a 65 kDa protein to 13 Å

resolution (47). With these advancements in single particle cryo-EM it has expanded the possibilities of macromolecular complexes that can be studied.

The major hurdle for electron crystallography that has been repeated throughout this body of work is the ability to form two-dimensional crystals of membrane proteins. Currently, we do not understand the complexity of interactions occurring during reconstitution and the formation of two-dimensional crystals. The simple fact is that we may never understand two-dimensional crystallization or be able to determine a "magic" condition in which all membrane proteins will form ordered arrays in a membrane. The electron crystallography field may need to look to X-ray crystallography to understand how to improve the technique for high-throughput membrane protein structure determination. In X-ray crystallography, crystals are formed by randomly screening thousands of conditions to find initial conditions conducive for the formation of three-dimensional crystals. Work is under way to develop high throughput screening methods for two-dimensional crystallization using 96-well plates for dialysis (48) as well as automation of screening conditions by negative-stain TEM (49). Large two-dimensional screens for assaying reconstitution are currently not commercially available. Most labs have a defined unique protocol for screening based on conditions that worked for a specific protein but the applicability to a wide range of proteins is questionable as is seen by an overview of the literature of published two-dimensional crystals. Electron crystallography has received the reputation for being notoriously difficult due to the inability to form two-dimensional crystals, until proven advancements in screening reconstitution conditions occur the field will not evolve to its potential.

Single particle cryo-EM and electron crystallography have evolved to be complementary techniques for other structural biology methods such as NMR and X-ray crystallography. Each

technique has the ability to achieve atomic resolution, but must overcome hurdles related to the particular sample, such as the ability to form well ordered of two-dimensional crystals or size and symmetry stipulations for single particle. Developments are currently being made to overcome these hurdles in order to expand the versatility of TEM to investigate macromolecular complexes that are vital players in all biological processes important for life.

WORKS CITED IN CHAPTER 7

1. Hara-Chikuma, M. & Verkman, A.S. (2008). Aquaporin-3 facilitates epidermal cell migration and proliferation during wound healing. *J. Mol. Med.* **86**, 221-31.

2. Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J. & Stroud, R.M. (2000). Structure of a glycerol-conducting channel and the basis for its selectivity. *Science*, **290**, 481-486.

3. Newby, Z.E.R, O'Connell, J., Robles-Colmenares, Y., Khademi, S. Miercke, L.J. & Stroud, R.M. (2008). Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite *Plasmodium falciparum. Nat. Struct. Mol. Biol.*, **15**, 619-625.

4. Viadiu, H., Gonen, T. & Walz, T. (2007). Projection of aquaporin-9 at 7Å resolution. *J. Mol. Biol.* 367: 80-88.

5. Walz, T., Smith, B.L., Agre, P. & Engel, A. (1994). The three-dimensional structure of human erythrocyte aquaporin CHIP. *The EMBO J.* 13: 2985-2993.

6. Cao, C., Sun, Y., Healey, S., Bi, Z., Hu, G., Wan, S., Kouttab, N., Chu, W. & Wan, Y. (2006). EGFR-mediated expression of aquaporin-3 is involved in human skin fibroblast migration. *Biochem. J.* **400**, 225-234.

7. Zelenina, M., Bondar, A.A., Zelenin, S. & Aperia, A. (2003). Nickel and extracellular acidification inhibit the water permeability of human aquaporin-3 in lung epithelial cells. *J. Biol. Chem*. **278**, 30037-30042.

8. Zelenina, M., Tritto, S., Bondar, A.A., Zelenin, S. & Aperia, A. (2004). Copper inhibits the water and glycerol permeability of aquaporin-3. *J. Biol. Chem*. **279**, 51939-51943.

9. Zheng, H., Taraska, J., Merz, AJ. & Gonen, T. (2010). The prototypical H^+/galactose symporter GalP assembles into functional trimers. *J. Mol. Biol.* **396**, 593-601.

10. Abramson, J. *et al.* (2003). Structure and mechanism of the lactose permease of *Escherichia coli*. *Science*, **301**, 610-615.

11. Dang, S. *et al.* (2010). Structure of a fucose transporter in an outward-open conformation. *Nature*, **467**, 734-738.

12. Sun, L. *et al.* (2012). Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature*, **490**, 361-366.

13. Hirai, T., Heymann, J.A.W., Shi, D., Sarker, R., Maloney, P.C. & Subramaniam, S. (2002). Three-dimensional structure of a bacterial oxalate transporter. *Nat. Struct. Biol.* **9**, 597-600.

14. Haslbeck, M., Franzmann, T., Weinfurtner, D., & Buchner, J. (2005). Some like it hot: the structure and function of small heat-shock proteins. *Nat. Struct. Mol. Biol.*, **12**, 842-846.

15. Ecroyd, H. & Carver, J.A. (2009). Crystallin proteins and amyloid fibrils. *Cell. Mol. Life Sci.*, **66**, 62-81.

16. Horwitz, J. (1992). Alpha-crystallin can function as a molecular chaperone. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 10449-10453.

17. Vicart, P. *et al.* (1998). A missense mutation in the alphaB-crystallin chaperone gene causes a desmin-related myopathy. *Nat. Genet.*, **20**, 92-95.

18. Goldstein, L. E. *et al.* (2003). Cytosolic beta-amyloid deposition and supranuclear cataracts in lenses from people with Alzheimer's disease. *Lancet.*, **361**, 1258-1265.

19. Aquilina, J. A., Benesch, J. L., Bateman, O. A., Slingsby, C. & Robinson, C. V. (2003). Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of oligomers in alphaB-crystallin. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 10611-10616.

20. Jehle, S. *et al.* (2010). Solid-state NMR and SAXS studies provide a structural basis for the activation of alphaB-crystallin oligomers. *Nat. Struct. Mol. Biol.*, **17**, 1037-1042.

21. Jehle, S. *et al.* (2011). N-terminal domain of αB-crystallin provides a conformational switch for multimerization and structural heterogeneity. *Proc. Natl. Acad. Sci. USA*, **108**, 6409-6414.

22. Peschek, J. *et al.* (2009). The eye lens chaperone α-crystallin forms defined globular assemblies. *Proc. Natl. Acad. Sci. USA*, **100**, 13272-13277.

23. Braun, N. *et al.* (2011). Multiple molecular architectures of the eye lens chaperone αB-crystallin elucidated by a triple hybrid approach. *Proc. Natl. Acad. Sci. USA*, **108**, 20491-20496.

24. MacCoss, M. J. *et al*. (2002). Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 7900-7905.

25. Andersen, E. S. *et al*. (2009). Self-assembly of a nanoscale DNA box with a controllable lid. *Nature*, **459**, 73-76.

26. Rothemund, P. W. (2006). Folding DNA to create nanoscale shapes and patterns. *Nature*, **440**, 297-302.

27. Zheng, J. *et al*. (2009). From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature*, **461**, 74-77.

28. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA,* **97**, 10383-10388.

29. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & André, I. (2011). Modeling symmetric macromolecular structures in Rosetta3. *PLoS ONE*, **6**, e20450.

30. Cooper, S. *et al*. (2010). Predicting protein structures with a multiplayer online game. *Nature*, **466**, 756-760.

31. King, N. *et al.* (2012). Computational design of self-assembling protein nanomaterials with atomic leel accuracy. *Science*, **336**, 1171-1174.

32. Ringler, P. & Schulz, G. E. (2003). Self-assembly of proteins into designed networks. *Science*, **302**, 106-109.

33. Ballister, E. R., Lai, A. H., Zuckermann, R. N., Cheng, Y. & Mougous, J. D. (2008). In vitro self-assembly of tailorable nanotubes from a simple protein building block. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 3733-3738.

34. Usui, K. *et al*. (2009). Nanoscale elongating control of the self-assembled protein filament with the cysteine-introduced building blocks. *Protein Sci.*, **18**, 960-969.

35. Salgado, E. N., Faraone-Mennella, J. & Tezcan, F.A. (2007). Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. *J. Am. Chem. Soc.*, **129**, 13374-13375.

36. Salgado, E. N., Radford, R. J. & Tezcan, F. A. (2010). Metal-directed protein self-assembly. *Acc. Chem. Res.*, **43**, 661-672.

37. Lovejoy, B. *et al*. (1993). Crystal structure of a synthetic triple-stranded alpha-helical bundle. *Science*, **259**, 1288-1293.

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

39. Gribbon, C. *et al*. (2008). MagicWand: a single, designed peptide that assembles to stable, ordered alpha-helical fibers. *Biochemistry*, **47**, 10365-10371.

40. Zaccai, N. R. *et al*. (2011). A de novo peptide hexamer with a mutable channel. *Nat. Chem. Biol.*, **7**, 935-941.

41. Koder, R.L. *et al*. (2009). Design and engineering of an O(2) transport protein. *Nature*, **458**, 305-309.

42. Fraser, C.S. & Doudna, J.A. (2007). Structural and mechanistic insights into hepatitis C viral translation initiation. *Nat. Rev. Microbiol.* **5**, 29-38.

43. Spahn, C.M., Kieft, J.S., Grassucci, R.A., Penczek, P.A., Zhou, K., Doudna, J.A. & Frank, J. (2001). Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science*, **291**, 1959-1962.

44. Filbin, M.E., Vollmar, B.S., Shi, D., Gonen, T. & Kieft, J.S. (2013). HCV IRES manipulates the ribosome to promote the switch from translation initiation to elongation. *Nat. Struct. Mol. Biol.* **20**, 150-158.

45. Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A., and Ban, N. (2011). Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science*, **331**, 730-736.

46. Bai, X., Fernandez, I.S., McMullan, G. & Scheres, S.H.W. (2013). Ribosome structure to near-atomic resolution from thrity thousand cryo-EM particles. *eLife*, **2**, e00461.

47. Wu, S. *et al.* (2012). Fabs enable single particle cryoEM studies of small proteins. *Structure*, **20**, 582-592.

48. Vink, M., Derr, K., Love, J., Stokes, D.L. & Ubarretxena-Belandia, I. (2007). A high-throughput strategy to screen 2D crystallization trials of membrane proteins. *J. Struct. Biol.* **160**, 295-304.

49. Coudray, N. *et al.* (2011). Automated screening of 2D crystallization trials using transmission electron microscopy: a high-throughput tool-chain for sample preparation and microscopic analysis. *J. Struct. Biol.* **173**, 365-374.