

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Towards Membrane Protein Structure Determination

Permalink

<https://escholarship.org/uc/item/08r9j8ns>

Author

Savage, David

Publication Date

2007-02-22

Peer reviewed|Thesis/dissertation

Towards Membrane Protein Structure Determination

by

David F. Savage

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

copyright © 2007

by

David F. Savage

Acknowledgements

Many thanks to the staff and faculty of the UCSF Graduate Program in Biophysics. First, Julie Ransom, the den mother to all, was always there for advice and help. I also appreciated the stimulating conversations I had with Dan Minor, Dave Julius, Ken Dill, Matt Jacobson, and Volker Doetsch. Finally, to my thesis committee, Dave Agard and Robert Edwards for their help focusing down my research.

Thanks to all of the friends I met at UCSF. To Vince, Mike, Mark and my fellow classmates - I always enjoyed Friday breakfast at Raintree. Next, to Sparklemotion (and the AgroDome). If there was a biophysics world basketball championship I'm sure we would win. To all of the Stroudians, notably Larry, the Joes, Sanjay, Frank, Bill, John, Sun, and Janet for their help. Corey for his contributions in Chapters 5 and 6. To Adrian who convinced me to join the lab and Pascal who taught and put up with me for so long. To Yaneth for all the jokes and great Mexican food and Zach for everything, but mostly his organizational tips. I will miss you all so much! And to Mike, John, and Anne Marie, my family outside the lab for four years.

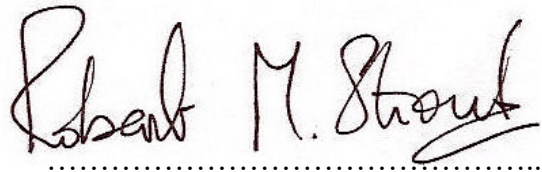
To Stroud, whose enthusiasm for life and science knows no bounds - I will carry it forever. To Sarah, the most passionate person I've ever known and my partner in life beyond San Francisco. Finally, to my parents, whose love, creativity, and intellectual curiosity is the basis for any good I've ever achieved in life.

Towards Membrane Protein Structure Determination

Abstract

Membrane protein structure determination, as well as functional characterization, has significantly lagged behind similar investigations of soluble proteins. Membrane proteins are typically expressed at low levels, must be solubilized from the membrane in detergents, and even when purified often heterogenous. Thus, membrane proteins present many unique challenges. To overcome these challenges we have selected a model membrane protein, AqpZ, of the aquaporin family of water channels. Such a model system can be used not only to investigate the structure-function properties of this class of channels, but also to elucidate fundamental conventions which can be applied to all membrane proteins. We have expressed, purified, and solved the x-ray crystal structure of AqpZ. To determine the mechanism of water selectivity a conserved region of the channel was engineered to allow conduction of other substrates and functional analysis shows the basis to be chemical rather than steric in nature. Furthermore, aquaporins were characterized with the use of mercurial compounds, and using AqpZ we have also elucidated a steric inhibition mechanism by mercurials. From these and other studies, it became apparent that protein expression was a fundamental barrier to working with membrane protein so we therefore set out to develop new expression systems and protocols targeted towards membrane proteins. We have adapted a reconstituted cell-free protein expression system that can express proteins at the milligram level and shown these proteins can be solubilized and purified to homogeneity. Finally, with an eye towards structural genomics we have introduced a series of protocols to streamline

structure determination. As a test case the structure of the monotopic membrane protein CcmG was solved using these protocols with a minimum amount of input work. Thus we have demonstrated, from both a single protein and a structural genomics level, a unified approach for moving towards membrane protein structure determination.

A handwritten signature in black ink that reads "Robert M. Stroud". The signature is written in a cursive style with a horizontal dotted line underneath it.

Robert M. Stroud, Ph.D.

Advisor and Committee Chairman

Table of Contents

CHAPTER 1: INTRODUCTION	1
CHAPTER 2: ARCHITECTURE AND SELECTIVITY IN AQUAPORINS: 2.5Å X-RAY	
STRUCTURE OF AQUAPORIN Z	8
ABSTRACT.....	9
INTRODUCTION	10
MATERIAL AND METHODS	12
RESULTS.....	14
DISCUSSION	18
REFERENCES	34
CHAPTER 3: A STRUCTURE-FUNCTION STUDY OF THE AQUAPORIN SELECTIVITY	
FILTER	39
ABSTRACT.....	40
INTRODUCTION	41
RESULTS.....	44
DISCUSSION	49
MATERIALS AND METHODS.....	52
REFERENCES	65
CHAPTER 4: STRUCTURAL BASIS OF AQUAPORIN INHIBITION BY MERCURY.....	68
ABSTRACT.....	69
INTRODUCTION	70
RESULTS.....	73
DISCUSSION	78
MATERIALS AND METHODS.....	86
REFERENCES	101

CHAPTER 5: CELL-FREE COMPLEMENTS <i>IN VIVO</i> EXPRESSION OF THE <i>E. COLI</i>	
MEMBRANE PROTEOME.....	106
ABSTRACT.....	107
INTRODUCTION	108
RESULTS.....	110
DISCUSSION	115
MATERIALS AND METHODS.....	121
REFERENCES	146
CHAPTER 6: DEVELOPMENT OF A MEMBRANE PROTEIN STRUCTURE PIPELINE.....	150
ABSTRACT.....	151
INTRODUCTION	152
RESULTS.....	155
DISCUSSION	159
MATERIALS AND METHOD	166
REFERENCES	178
APPENDIX 1	181

List of Tables

2-1.....	26
3-1.....	56
3-2.....	57
4-1.....	91
5-1.....	127
5-2.....	133
5-3.....	136
5-4.....	137
5-5.....	138
6-1.....	169

List of Figures

2-1.....	28
2-2.....	30
2-3.....	31
2-4.....	32
2-5.....	33
3-1.....	59
3-2.....	60
3-3.....	61
3-4.....	62
3-5.....	63
3-6.....	64
4-1.....	93
4-2.....	94
4-3.....	96
4-4.....	97
4-5.....	99
4-6.....	100
5-1.....	140
5-2.....	141
5-3.....	142
5-4.....	143
5-5.....	144

List of Figures Continued

5-6.....	145
6-1.....	171
6-2.....	172
6-3.....	173
6-4.....	174
6-5.....	175
6-6.....	176
6-7.....	177

Chapter 1

Introduction

My entire knowledge of the membrane protein field as a first year graduate student with a computer science background was this: they're hard. I knew nothing of the "funniness" associated with the "beast" or any more of Larry Miercke's anecdotal warnings. As usual, it's best to be naive.

Historically membrane proteins are difficult to work with. They are expressed at low levels and require the hydrophobic membrane to maintain their native state. The number of X-ray structures of membrane proteins bears this difficulty out, and less than 1% of the total number of structures are of membrane proteins. As a testament to the near total failure of expression systems, the first structure of a heterologously expressed eukaryotic membrane protein wasn't determined until 2005. This relatively recent result is significant, however. We are arguably entering the golden age of membrane protein structural biology.

I joined the Stroud lab just as our paper simulating water conduction through GlpF was published in *Science*. At the time, it seemed like membrane proteins were everywhere. New, groundbreaking structures came out elucidating fundamental properties with dramatic and atomic detail. We saw homoligomeric channels, the use of helical dipoles in the membrane as electrostatic "ray guns," and striking two-fold symmetry. Furthermore, these structures shed light on physiologic data or excitingly, for me the computer scientist, could be used as a jumping off point for computer simulations. Thus, I was hooked.

A graduate student in need of a project, I jumped into the aquaporin literature and knew that I found my protein when I read these lines from Peter Agre:

AqpZ, the aquaporin from *E. coli*, was identified as a good candidate for purification following overexpression in bacteria and may be a good substrate for structure-function studies....

For all of these reasons, we believe that AqpZ will be an exceedingly interesting and useful member of the aquaporin family.

AqpZ had all the hall marks of a membrane protein success story. It could be overexpressed, solubilized in crystallography friendly detergents, purified easily, and its function could be assayed. Furthermore, AqpZ also dovetailed nicely with the theme of our lab. The *E. coli* genome encodes only two aquaporins, the water-selective AqpZ and the glycerol channel GlpF, for which we had already solved the structure, so here I could also dissect the molecular underpinnings of substrate selectivity.

Thus, I set out to crystallize AqpZ. Agre's lines were prescient and AqpZ was indeed "useful" and "exceedingly interesting." With the help of others in the Stroud lab (and a fair amount of luck) I was able to determine the structure to 2.5 Å as described in Chapter 2. This structure revealed a channel nearly identical to GlpF except, as expected from the structure of the prototypical water channel AQP1, the 20 Å long pore was slightly more narrow. The most constricted region of the channel, termed the selectivity filter, was identical to the AQP1 structure. In water selective aquaporins this region is

smaller and contains a conserved histidine, which introduces a hydrogen bonding partner. In non-selective aquaglyceroporins this region is slightly larger and contains two conserved aromatic residues which form a hydrophobic triangle to permit passage of the amphipathic glycerol.

To determine the basis of selectivity, I serially mutated the selectivity filter residues of AqpZ to those of GlpF. AqpZ is a robust platform for mutational analysis and the structures of these mutants containing an increasingly GlpF-like selectivity filter were solved. Aquaporins can be assayed in osmotically driven proteoliposome assays, so at the same time I was spending nearly equal amounts of time developing assays for functional tests. Finally satisfied with the reproducibility of our so-called recons, I measured the water and glycerol permeabilities of the proteins and realized, that while the serial mutations decreased water conduction, they do nothing for glycerol conduction. I was thusly disappointed with “engineering” a broken channel, but with faith in the functional data, I went back and analyzed the structures. These structures reveal a channel which is tuned for selectivity beyond the selectivity filter. Indeed, the GlpF-like AqpZ mutant has a pore in size similar to that of AqpZ, caused by a rearrangement of several loops and transmembrane helices. Because of this serendipitous lack of change in pore size though, we can clearly say the barrier to glycerol conduction is steric, while the barrier to water conduction, at least in channels sized roughly the radius of water, is polarity. Thus, we elucidated the structural basis of selectivity in aquaporins.

At least for water and glycerol. Current aquaporin dogma, informed by the early discoveries and functional characterization of AQP1 and GlpF, says that water and glycerol are the two major substrates conducted by these channels. But, there is an undercurrent of literature saying otherwise and implicating molecules such as urea, ethanol, cations, anions, and various gases. One specific argument is that although a red blood cell has, on average, 200,000 copies of AQP1 it has no need for a water channel. Because of this, Walter Boron at Yale has shown with physiological conduction data that AQP1 may instead be a CO₂ channel. A few years ago the idea of a gas channel was somewhat extraordinary to me, but Sharam Khademi in our lab determined the structure of AmtB to an unprecedented 1.35 Å resolution and showed both structural and functional evidence of ammonia conduction. While at a Gordon Conference I had the pleasure of having lunch with Prof. Boron and we talked about the possibility of CO₂ conduction through both the monomer channel and the hydrophobic tetrameric axis. To dissect CO₂ conduction his lab has employed the use of inhibitory mercurial compounds and we talked about the structural implications of mercury binding. Finally, he asked me why no one had tried to solve the structure of a mercury-aquaporin complex to characterize changes in the pore. Unable to answer, except for an agreement, I found a new project.

AqpZ has been functionally verified as the bacterial homolog of AQP1, and since it can be expressed heterologously and easily purified, makes for an excellent model system. It had previously been shown that specific mutation of one of AQP1's five cysteines to serine will abrogate mercury sensitivity. The crystal structure of AQP1 has shown this

cysteine, C189, is directly in the pore and that a steric blockage was quite likely, but it did not rule out the possibility of drastic conformational change to the water coordinating carbonyls in the otherwise hydrophobic pathway. We therefore set out, as described in Chapter 4, to determine the structure of an AqpZ mutant with the known sensitive cysteine and mercury bound. This experiment did work, but we surprisingly discovered two mercury atoms bound, one which could sterically block the pore and one interstitially bound which could perturb the water-coordinating carbonyls. We hypothesized a steric blockage though and so created a new mutant to optimize binding at the pore site. Inhibition studies demonstrated that the latter mutant was indeed the most sensitive to mercury and the structure showed mercury was bound only at the pore. Thus, aquaporins are inhibited sterically by mercury.

Finally, during the course of the aquaporin research above, the Stroud lab received two major NIH grants for developing membrane protein expression system and structure determination. This led us to develop, with the help of former UCSF professor Volker Doetsch, an *E. coli*-based cell-free expression system. This system expresses milligram quantities of protein, is inexpensive, and can be set up to run in parallel, making it an excellent tool for a high throughput structural genomics approach. Cell-free would have no benefits, however, if it couldn't express well-behaved proteins, so to compare against a conventional *in vivo* system we cloned a large test set of *E. coli* membrane proteins and proceeded to express and characterize them. Furthermore, because we would also be expressing a purifying new proteins, we would also be feeding the Center for Structures of Membrane Proteins with new targets.

As described in Chapter 5, the cell free system thus proved to be quite useful for expressing membrane proteins, but the most important result, I believe, is the pipeline approach we have outlined in Chapter 6. This pipeline was created by incorporating the large body of empirical knowledge from the Stroud and other membrane protein labs. I must say too, I am extremely biased by my early work on AqpZ. I learned from AqpZ that beauty can be found in the structure and function of nearly any protein, but first, you must have protein! Thus, at least for the near future, in studying membrane proteins it is a good idea to focus on those that express and are well-behaved. Therefore, the main thrust of the pipeline is to determine a large number of new structures with the least amount of effort. Basically, cast a wide net to quickly identify the lowest hanging fruit. Essentially, it contains no new information - instead, its power comes from simplicity. A typically pipeline has unneeded dimensionality, such as screening multiple constructs in a multitude of detergents. We have removed this dimensionality which has created ambiguity and plagued other efforts and focused only on a few key steps. I believe this pipeline would be incredibly successful if implemented. At least while the low hanging fruit are still there, of course. Thankfully, 30% of all proteins are membrane proteins!

Chapter 2

Architecture and Selectivity in Aquaporins: 2.5Å X-ray Structure of Aquaporin Z

Research completed in collaboration with

Pascal F Egea, Yaneth Robles Colmenares, Joseph D. O'Connell III, and Robert M.

Stroud

This chapter originally appeared in PloS Biology; 2003; 1(3);334-340.

Abstract

Aquaporins are a family of water and small molecule channels found in organisms ranging from bacteria to animals. One of these channels, the *E. coli* aquaporin Z (AqpZ), has been shown to selectively conduct only water at high rates.

We have expressed, purified, crystallized, and solved the X-ray structure of AqpZ. The 2.5Å resolution structure of AqpZ suggests aquaporin selectivity results from both a steric mechanism due to pore size and specific amino acid substitutions that regulate the preference for a hydrophobic or hydrophilic substrate.

This structure provides direct evidence on the molecular mechanisms of specificity between water and glycerol in this family of channels from a single species. It is the first atomic resolution structure of a recombinant aquaporin and so provides a platform for combined genetic, mutational, functional and structural determinations of the mechanisms of aquaporins, and more generally, the assembly of multimeric membrane proteins.

Introduction

The self and non-self properties of cells are defined by components within the cellular membrane, and as a physical barrier, the membrane separates internal contents of the cell from the extracellular milieu. By necessity, nature has devised mechanisms for the highly selective passive and active transport of materials across the membrane into the cell, and out of the cell so that cellular viability is maintained (Alberts et al. 2002).

The aquaporin (AQP) family, composed of transmembrane water-conducting channels (aquaporins) and glycerol (and water) conducting channels (aquaglyceroporins), is a major group of selective transporters (Heller et al. 1980, Preston et al. 1992, Park and Saier 1996). The diversity of the aquaporin family is embodied by the human proteome, where at least ten different aquaporins are expressed in tissues and cells such as brain, kidneys, and erythrocytes. AQPs play a fundamental role in osmoregulation, and mutations are responsible for human diseases ranging from diabetes insipidus to congenital cataract formation (Borgnia et al. 1999).

The structural architecture of aquaporins was first determined by electron microscopy (Walz et al. 1997, Murata et al. 2000). Later, high-resolution X-ray structures of the recombinant *E. coli* aquaglyceroporin glycerol facilitator (GlpF) (Fu et al. 2000), and bovine aquaporin 1 (AQP1) obtained from red blood cells elucidated the mechanisms by which aquaporins preserve the electrochemical membrane potential and selectively conduct water and linear polyalcohols (Fu et al. 2000, Sui et al. 2001). Until now, there

was no structure of a recombinantly expressed water-selective aquaporin, which would allow the systematic analysis of amino acid substitution by mutagenesis, structure and function.

Besides GlpF, the *E. coli* genome contains a second aquaporin, Aquaporin Z (AqpZ). AqpZ is a highly efficient water channel and conducts water at rates six times that of GlpF (Calamita et al. 1995, Borgnia et al. 1999). AqpZ has been used to probe substrate selectivity and shown promise as a structural target, providing one of the early electron microscopy studies of an aquaporin. (Calamita et al 1998, Ringler et al. 1999, Borgnia et al. 2001). The pair of AqpZ and GlpF exist in the same organism, implying similar lipid, chemical, and osmotic environments, and thus present a unique opportunity to study aquaporin structure and function in a genetically and biochemically tractable system. In anticipation of the mutagenic probes of function and structure, we report the X-ray structure of wild-type AqpZ to 2.5Å resolution.

Material and Methods

Expression and Purification. AqpZ was cloned by PCR from isolated *E. coli* genomic DNA into the pET 28 expression vector with kanamycin selection and N-terminal 6xHis affinity tag (Novagen). The *E. coli* strain C43 (Miroux and Walker 1996) was transformed, grown to .6-1 OD at 600nm in LB with 20mg/L kanamycin, and induced with 1mM isopropyl β -D-thiogalactoside. Cells were harvested and lysed by sonication in 20mM Tris pH 7.4, 100mM NaCl, .5 mM phenylmethylsulfonyl fluoride, and 5mM 2-mercaptoethanol. Cellular debris were pelleted at 10,000 x g for 45 min and discarded. Membranes were recovered from supernatant by 100,000 x g centrifugation for 90min. AqpZ was solubilized from membranes by agitation in 20mM Tris pH 7.4, 100mM NaCl, 5mM 2-mercaptoethanol, 10% glycerol, and 270 mM n-Octyl- β -D-glucopyranoside (OG) (Anatrace) for 12-16 hours. Solubilized protein was bound in batch to Ni-NTA resin (Qiagen), washed, and eluted with 20mM Tris pH 7.4, 100mM NaCl, 5mM β -mercaptoethanol, 10% glycerol, 40 mM n-Octyl- β -D-glucopyranoside, and 250 mM imidazole. Imidazole was removed using a Biorad Econo-Pac DG10 desalting column and the histidine tag was removed following the protocol of Borgnia et al (1999). The final purification step was performed on a Pharmacia Superose 200 column.

Crystallization. Purified AqpZ was concentrated to ~20mg/mL and crystallized in 28% polyethylene glycol monomethyl ether 2000, 100mM sodium cacodylate pH 6.5, 200mM MgCl₂, 4% 2-propanol in hanging drop plates (Nextal Biotechnologies) by vapor

diffusion at room temperature. Crystals grew to 300 μ m x 300 μ m x 150 μ m, in several days. Crystals were flash frozen in a 90K nitrogen gas stream.

Data Collection and Model Building. Diffraction intensities were collected at the Advanced Light Source Beamline 8.3.1 on a Quantum CCD detector. Data were processed with DENZO/SCALEPACK (Otwinowski and Minor 1997). The structure was solved by molecular replacement using the AQP1 (Protein Data Bank code 1J4N) structure as a search model. The model was refined with CNS and built using Moloc (Muller et al. 1988, Brünger 1996).

Results

Structure of AqpZ. Three-dimensional crystals of AqpZ were grown and diffraction data to better than 2.5Å were collected under cryoconditions (space group P4, $a = 93.6\text{Å}$ $c = 80.4\text{Å}$, with two protomers per asymmetric unit, henceforth called protomer A and B). The structure was solved by molecular replacement and refined to R_{cryst} of 23.3% and R_{free} of 26.9% using reflections to 2.5Å (Table 1). There are two tetramers in the unit cell composed of 4 copies of protomers A as a tetramer, and 4 of B respectively. The positioning of these tetramers in the unit cell shows quasi-I4 symmetry in which the body-centered tetramer is slightly rotated around the four-fold axis. Protomers A and B are involved in different crystal packing interactions; their root mean square difference (r.m.s.d) is .44Å. Pore electron density was stronger for protomer A. Non-crystallographic symmetry did not prove useful in model building, presumably because of the differences between protomer A, and B.

The protomer structure of AqpZ displays the canonical aquaporin fold of six transmembrane helices and two half membrane-spanning helices (M1-M8) in a right-handed helical bundle. As shown previously, the protomer oligomerizes to form a homotetramer (Fig. 1A). The amino terminus begins on the cytoplasmic side. M1 crosses the membrane and loops to M2, which recrosses the membrane. M2 is followed by a loop from residues 54 to 62 that contains the four carbonyls that project into the pore near the cytoplasmic side. Following this loop is helix M3, which contains the signature Asparagine-Proline-Alanine (NPA) motif and is oriented in such a way as to point its

positive dipole towards the central water position in the channel. The first domain of AqpZ ends with M4 ending on the periplasmic side. This is followed by a loop from residue 103 to 131 that descends into the periplasmic vestibule and leads into the carboxy terminal segment. M5-M8 reiterates the amino terminus topology, except now beginning on the periplasmic and ending on the cytoplasmic side. This pseudo two-fold symmetry creates a general architecture in which the main chain carbonyls establish water-binding sites along the channel and side chain variation determines channel size and chemistry (Fig. 1B).

The channel. The aquaporin channel is a long ($\sim 28\text{\AA}$) and narrow ($< 4\text{\AA}$ diameter) pore that widens out to periplasmic and cytoplasmic vestibules. The channel is formed by the packing of helices M1-M3 and M5-M7 and is amphipathic, establishing a single-file water conduction pathway. The hydrophilic nature of the channel results from the four adjacent carbonyls of G59(GlpF number 64), G60(65), H61(66), and F62(67) from the N-terminal domain and the quasi-twofold related N182(199), T183(200), S184(201), and V185(202) from the C-terminal domain. The hydrophobic nature of the channel results from an abundance of valines, phenylalanines, and isoleucines within the channel.

The channel contains two highly conserved regions, the selectivity filter and the NPA region. Located $\sim 7\text{\AA}$ inside the periplasmic vestibule, the selectivity filter is the narrowest point (diameter of $\sim 1.5\text{\AA}$) in the entire channel. It is formed by the side chains of F43(48), H174(191), R189(206) and the carbonyl of T183(200). The trio of H174(191), T183(200), and R189(206) create a hydrophilic triangle opposite the hydrophobic F200.

The NPA sequences from each M1-M4, and M5-M8 domain form a constrained and interlocked junction around the quasi two-fold axis, based on asparagine, proline, and alanine from the N terminal ends of M3, and M7. The alanine side chain and the proline ring make a head-to-tail, twinned, largely hydrophobically driven, contact with the proline and alanine of the other domain. Each asparagine side chain is oriented by two almost ideal hydrogen bonds. For N63(68), these bonds are one from OD1 to the NH of A65(70) and one from NH2 to the carbonyl of V185(202). Similar interactions occur at N186(203). This highly constrains and orients both asparagine side chains to project their ND2 groups strictly into the pore, which are hydrogen bond donors to the central water molecule.

Five waters are unambiguously located in the channel. The water is arranged in single file, hydrogen bonding as donors to the projecting carbonyls from AqpZ, and as donors to neighboring waters. From periplasmic side to cytoplasmic side, there are waters located adjacent to the carbonyls of T183(200) (OAO distance of 3.0Å), S184(201) (3.2Å), H61(66) (3.0Å), and G60(65) (3.4Å). The waters are at appropriate (<3.2Å) distances from each other for hydrogen bonding. No electron density was observed adjacent to the carbonyls of G59(64) or V185(202).

In protomer B, four n-Octyl-β-D-glucopyranoside (OG) molecules are positioned at the potential location of the periplasmic membrane leaflet (Fig. 1A, 1C). The detergent head groups pack against the aromatic residues F196(224), W200(228), and W206(234) near

helix M8 and the lipid tails run towards the centerline of AqpZ. Their conformation suggests a belt-like micelle surrounding the full tetramer.

Three 2-propanol molecules are located in the cytoplasmic and periplasmic vestibules, just outside the channel (Fig. 1B). The propyl groups are packed against hydrophobic side chains, while the hydroxyl groups participate in hydrogen bonding with vestibule waters.

Discussion

Selectivity of Aquaporins. The *E. coli* genome encodes two aquaporins, GlpF and AqpZ. These two channels represent the functional diversity of the aquaporin family and as both are *E. coli* transmembrane proteins, exist in the same lipid, chemical, and osmotic environments. Both channels preserve the electrochemical gradient and display selective transport, yet have a different biological function. A comparison of the two structures delineates the nature of aquaporin selectivity, uncluttered by species differentiation.

In vitro and *in vivo* functional experiments demonstrate AqpZ's preference for water transport and GlpF's preference for glycerol (Maurel et al. 1994, Borgnia and Agre 2000). To first approximation, this preference is due to channel size, as depicted in Figure 2 for the known structures. This calculation shows that aquaporins have a smaller pore size than aquaglyceroporins, and that the selectivity filter is the narrowest point in the channel for all three proteins. In AqpZ this selectivity filter is formed by the side chains of F43(48), H174(191), R189(206), and the carbonyl of T183(200) (Fig. 2A). The presence of a bound water molecule (distances: 2.7Å to NE2 H174, 3.0Å to O T183, 2.6Å NH2 R189), confirms the selectivity filter's preference for a small hydrophilic substrate. The Aqp1 selectivity filter is nearly identical, with a cysteine substituted for threonine, which is also the basis of inhibition by mercury (Preston et al. 1992). In sharp contrast, GlpF contains the typical aquaglyceroporin substitutions of F43W, H199G, and T200F; the GlpF wild-type structure contains both a water and glycerol molecule bound at the selectivity filter.

The GlpF selectivity filter, larger and more hydrophobic than in AqpZ, is reminiscent of the maltoporin “greasy slide” sugar-binding sites (Dutzler et al. 1996, Van Gelder et al. 2002). Maltoporins, a family of bacterial outer membrane transporters, facilitate the translocation of maltooligosaccharides using a “greasy slide” hydrophobic path of seven aromatic residues along the central pore. Such a path, with a preference for non-polar groups, can increase the effective concentration of ligand near the channel and thereby increase the probability of a transport event. The periplasmic vestibule of GlpF also has a hydrophobic patch of residues leading into the selectivity filter, while in AqpZ the polar side chain of N182 chemically and structurally caps off the already hydrophilic selectivity filter.

Three molecules of 2-propanol, present at 4% in the crystallization solution are located just outside the channel in the vestibule regions (Fig1B). They pack against hydrophobic side chains, forming favorable VDW contacts with A27, F36, V39, F43, A62, T153, I178, and hydrogen bonding with nearby water. Based on its vestibular location, 2-propanol is seemingly too big for transport, though this idea has not been tested experimentally. Despite its presence at high concentration, one substrate not seen in the AqpZ structure is glycerol. This contrasts with the structure of GlpF where ordered glycerol was located at three sites in the channel, including the NPA motif. This absence confirms previous functional data and suggests a steric mechanism of selectivity.

There are five well oriented waters in the AqpZ channel, forming a chain of water nearly the length of the channel. Four waters are hydrogen bond donors to the carbonyls of G60(65), H61(66), T183(200), and S184(201). The fifth, central water molecule, is a hydrogen bond acceptor from the ND2 groups of the NPA motif asparagines (Fig. 1B, 4). Normally, a single-file column of water such as we observe should conduct protons. In 1806, Grotthuss, based on electrolysis experiments, postulated that polar water molecules could align themselves in long chains from cathode to anode, in essence forming a wire (see supplemental materials). Bernal and Fowler, using quantum mechanics to explain Grotthuss's qualitative hypothesis, postulated that protons could easily jump between neighboring waters, thereby making protons highly mobile in solution (Bernal and Fowler 1933). It is therefore remarkable that aquaporins, which inherently contain a chain of water, preserve the electrochemical gradient. A possible mechanism for this disparity is disruption of the proton jumping mechanism at the NPA region.

The NPA asparagine ND2 groups act as hydrogen bond donors to the central water, locking it in a conformation such that it can only donate hydrogen bonds to nearby single-file waters. Therefore, while the central water can readily donate a proton it can never accept one. This prevents adjacent water from performing the reorientation necessary to conduct protons, and the proton-conducting 'wire' is broken. This effect, termed global orientational tuning (Tajkhorshid et al. 2002), is also aided by the positive dipoles of M3 and M7 (Murata et al. 2000), which are aimed directly at the central water in a manner reminiscent of potassium channels (Doyle et al. 1998, Nollert and Tajkhorshid et al. 2002).

For water flux to occur, the central water must quickly be replaced by another, and the uniquely oriented carbonyls proximal to the NPA motif, those of F62 and V165, may reorient the NPA region water as it moves away from the center (Fig. 4). While the carbonyls of G59(64), G60(65), H61(66), N182(199), T183(200), and S184(201) are nearly orthogonal to the channel axis, those of F62(67) and V185(202) run parallel to the axis. As these carbonyls are proximal to the asparagines of the NPA motif, their unique conformation may be necessary to allow water to reorient as it passes the NPA region at the quasi-twofold axis. This hypothesis is supported by molecular dynamics simulations (unpublished data).

Besides breaking the proton wire, the NPA region is likely to play a role in selectivity. Sequence analysis has shown that positions 15(21) and 145(159), residues from helix M1 and M5 respectively, are correlated; aquaglyceroporins typically contain leucine at both positions, while aquaporins have an aromatic at one of the two (Heymann and Engel 2000) (Fig. 3). In AqpZ, the side chains of L15(21) and F145(159) project into the pore and narrow it to a diameter of 3Å; the GlpF diameter is 4Å (Fig. 2B). In AqpZ, the central water located opposite the NH₂ of N63(68) and N186(203) is better resolved and has a shorter hydrogen bond distance (~2.8Å vs. ~3.5) than the corresponding water in the GlpF structure without glycerol in the crystallization buffer. Furthermore, with glycerol in the buffer GlpF readily crystallizes with glycerol bound at this position. Thus, this secondary constriction emphasizes the preference for a small hydrophilic substrate in aquaporins.

Continuing along the pore axis towards the cytoplasm, aquaporins display a narrower channel. This pore difference does not result from helix rearrangement, as the main chain r.m.s.d for AqpZ and GlpF is 1.6Å and that of AqpZ and AQP1 is 1.2Å, but comes from side chain variation. Strict conservation of helical tertiary structure suggests one can effectively apply methods such as homology modeling to predict side chain conformation and function in the pore region (Marti-Renom 2000).

C α —H...O Bonds in Aquaporins. The strict tertiary conservation of aquaporins underscores the unique features of helix packing in membrane proteins. A survey of alpha-helical membrane proteins has revealed that transmembrane helices are often packed at distances close enough for C α —H...O hydrogen bond formation (<3.5Å) (Wahl and Sundaralingam 1997, Senes et al. 2001). This type of interaction is facilitated by glycine, an amino acid that is overrepresented in transmembrane segments, because it allows short interhelix distances (Sennes et al. 2000). In an analysis of AqpZ, we identified 15 potential bonds. With an estimated energy of 2.5-3.0 kcal/mol per bond (*in vacuo*), this is a partial explanation for the stability of AqpZ in denaturing conditions. (Borgnia et al. 1999, Scheiner et al. 2001). These bonds are also likely to play a role in the dynamics of other alpha-helical membrane proteins.

The structure of lactose permease was recently solved using a thermostable cysteine to glycine mutant (Abramson et al. 2003). The site-directed mutation occurs at the interface between two transmembrane helices and appears to lock the protein in a

conformation such that it can tightly bind substrate, but not translocate it. Engineering ultra-stable glycine mutants may prove useful in structural studies of other, less robust, membrane proteins.

Detergents in Aquaporins. Four OG detergent molecules were located bound to the periplasmic surface of each AqpZ molecule in a belt-like fashion (Fig1C). The detergents are situated near the C-terminus of M7 and N-terminus of M8, capping off the helices and forming hydrogen bonds with the carbonyls of the loop between helix M7 and helix M8. The hydrophobic sugar rings and tails pack against the nearby residues, many of which are aromatic.

The GlpF and Aqp1 structures each contained three detergent molecules at virtually identical location on the outside surface, presumably due to the abundance of aromatic side chains in all three proteins at this location. This abundance is present in all aquaporins, and may be important for lipid interaction. In AqpZ there are also both acidic and basic residues interacting with the OG head group, suggesting the native lipid may be a zwitterion like phosphatidyl ethanolamine, the most common *E. coli* lipid (Neidhardt et al. 1996). The importance of native lipids has been demonstrated in the folding and function of ion channels and may be important in designing future aquaporin functional assays (Valiyaveetil et al. 2002).

Tetramer Axis. While it is clear the monomer is the functional unit, the existence of aquaporin tetramers in nature reinforces the importance of oligomerization.

Fundamentally, tetramerization is driven by the energetically favorable assembly of four protomers. The protomer-protomer interface is large, tightly packed, and formed by helices M1 and M2 of one protomer and the quasi-twofold related M5 and M6 of the neighboring protomer (Fig 5A). This interface is therefore repeated four times. In AqpZ the interface is $3,340\text{\AA}^2$, in large part due to the presence of 11 aromatic residues. The GlpF and Aqp1 interfaces are $3,060\text{\AA}^2$ and $3,180\text{\AA}^2$ respectively, with five and three aromatic residues respectively. Strikingly, the interface surface area correlates positively with biochemical stability; GlpF tends to aggregate in solution, whereas AqpZ is a stable tetramer in even mild denaturing conditions. In this protomer-protomer interface, helices M2 and M6 form the tetramer pore.

There is remarkable tertiary and quaternary structure conservation and the tetramer pore remains nearly constant in shape between the known structures. All structures appear open towards the periplasm but closed towards the cytoplasm, a necessity in preserving the electrochemical gradient, and contain a large (7-10 \AA diameter) central cavity. Except for a glutamate near the periplasmic opening in Aqp1 and GlpF, the residues lining the pore are hydrophobic, suggesting a very large energetic barrier of translocation for any polar substrate. Notably, all three X-ray structures contained electron density along the pore, signifying the presence of multiple molecules (Fig. 5B). It remains to be seen whether this pore is functional or primarily a structural necessity to facilitate monomer function.

Abbreviations:

Aquaporin 1 (AQP1)

Aquaporin Z (AqpZ)

Glycerol facilitator (GlpF)

n-Octyl- β -D-glucoopyranoside (OG)

Table 1. Data collection and Refinement Statistics

Data set	AqpZ
Wavelength (Å)	1.1
Resolution (Å)	50-2.5
Total reflections	58,536
Unique reflections	21,720
Redundancy	6,3
Completeness (last shell) (%)	90.0 (90.2)
R_{sym} (last shell) (%)	6.3 (44.3)
I/σ (last shell)	12.0 (3.4)

Refinement statistics	AqpZ
Reflections in working set	20,216
Reflections in test set (6.9%)	1,504
R_{cryst} (%)	22.8 %
R_{free} (%)	26.9 %
r.m.s.d. bonds (Å)	0.008
r.m.s.d. angles (°)	1.24
<i>Protomer A and Protomer B</i>	
Non-hydrogen protein atoms	1661 and 1641
Non-hydrogen heteroatoms	12 and 88
Solvent molecules	131
Average B-factors (Å ²)	35

r.m.s.d is the root-mean square deviation from ideal geometry.

$R_{sym} = \frac{\sum_{hkl} \sum_i |I_{hkl,i} - \langle I_{hkl,i} \rangle|}{\sum_{hkl} \sum_i I_{hkl,i}}$ where $\langle I_{hkl,i} \rangle$ is the average intensity of the multiple hkl, i observations for symmetry-related reflections.

$R_{cryst} = \frac{\sum |F_{obs} - F_{calc}|}{\sum |F_{obs}|}$. F_{obs} and F_{calc} are observed and calculated structure factors, R_{free} is calculated from a randomly chosen 6.9% reflections and R_{cryst} is calculated over the remaining 93.1% of reflections.

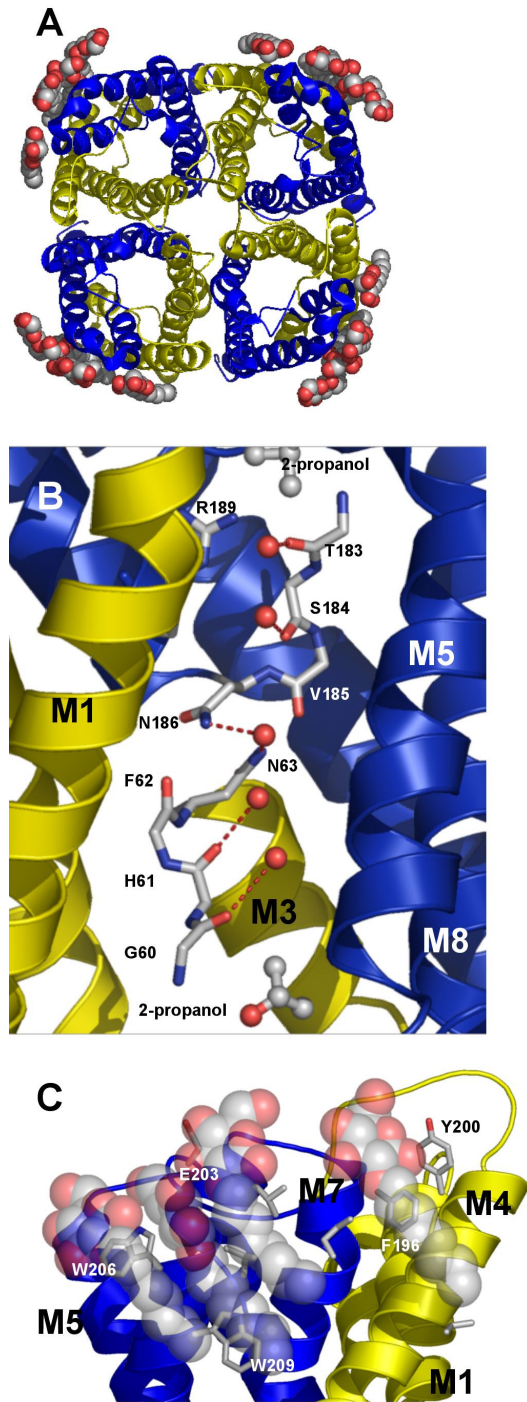


Fig.1. Structure of AqpZ. Three-dimensional fold of AqpZ with the quasi-twofold related segments in yellow (residues 1-117) and blue (residues 188-231). (A) Cartoon

representation of the AqpZ tetramer with n-Octyl- β -D-glucopyranoside (OG) detergent molecules represented as spheres; view is from the periplasmic side. Atoms are colored according to atom type (oxygen, red; carbon, gray; nitrogen, blue; sulfur, yellow). (B) Cartoon representation of the AqpZ monomer with M2 and M6 removed for ease of viewing. Single-file water is shown hydrogen bonding to carbonyls of main chain. Central water is shown accepting hydrogen bond from NH₂ group of N63 and N186. 2-propanol molecules located in density are shown as spheres, just outside the channel. (C) A view from the membrane plane of the OG micelle interactions with the periplasmic segment of AqpZ. OG molecules pack against the aromatic side chains, while making hydrogen bonds with the main chain carbonyls, E203, and each other. All figures were made with Pymol (Delano, 2003).

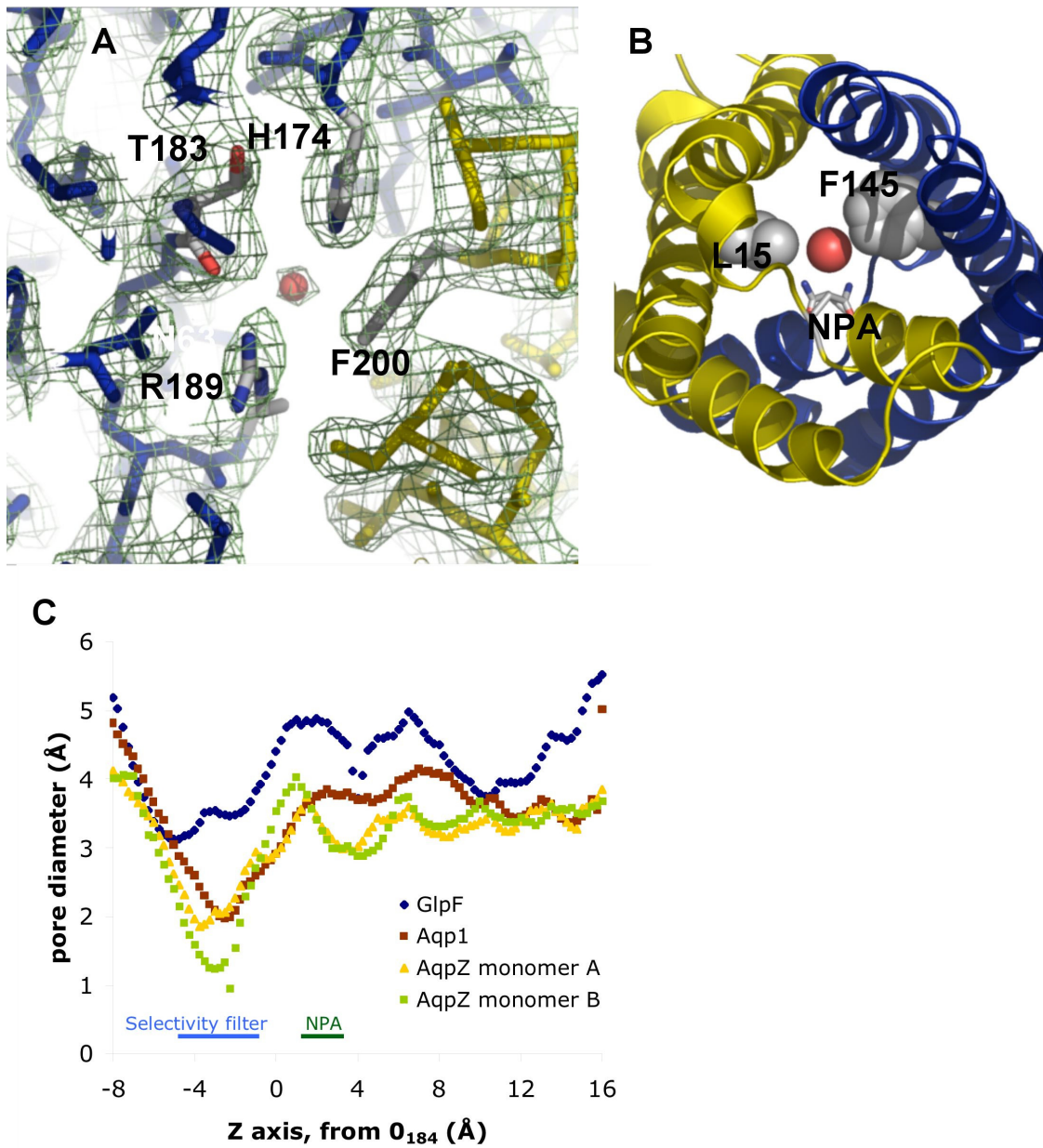


Fig. 2. Channel Constriction in Aquaporins. (A) A view of the aquaporin selectivity filter from the periplasmic side. Experimental electron density ($2F_o - F_c$) is contoured at 1.1σ . (B) Secondary constriction at the NPA due to F145 and L15 side chains. HOH1032 is shown hydrogen bonded to the NPA asparagines. (C) Pore diameters for the aquaporin X-ray structures, calculated with HOLE2.

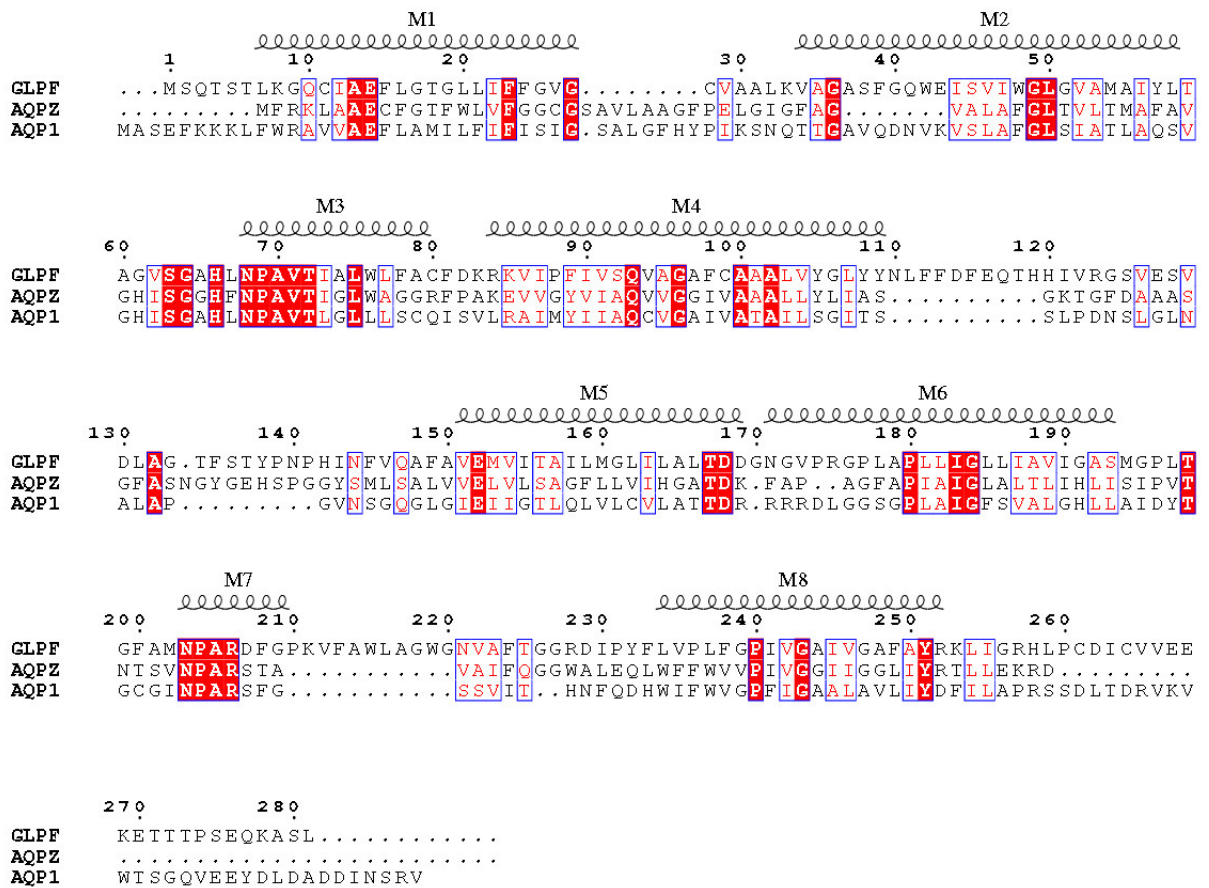


Fig. 3. Sequence Alignment of Aquaporins of Known Structure. Alignment of GlpF, AqpZ, and Aqp1 is numbered according to GlpF. Helices are shown and labeled M1-M8. Residue positions of similar chemical nature are shown in blue boxes; identical residues are shown in red.

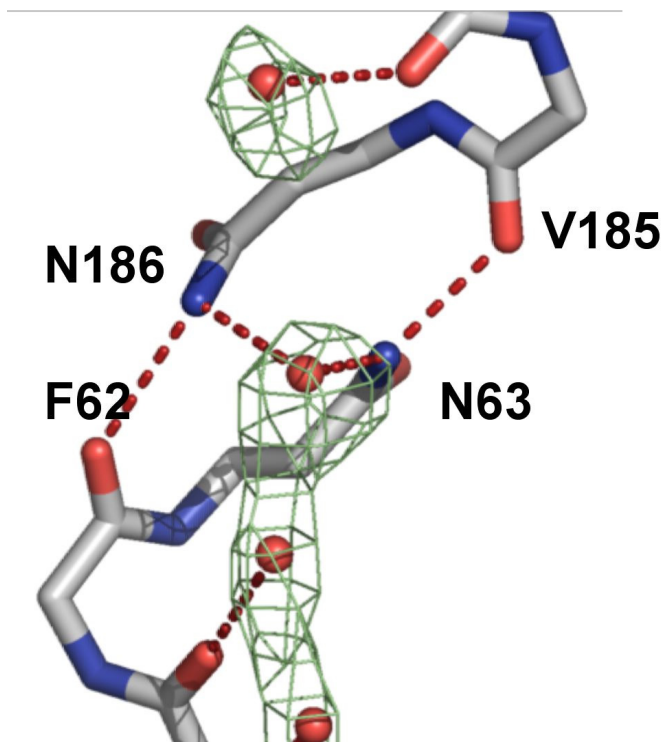


Fig 4. Water at the NPA Region. N63 and N186 donate hydrogen bonds to the central water by projecting their NH₂ moieties into the pore. This conformation is aided by a hydrogen bond from the adjacent carbonyls of V185 and F62 respectively. Experimental electron density ($2F_o - F_c$) is contoured at 0.7σ .

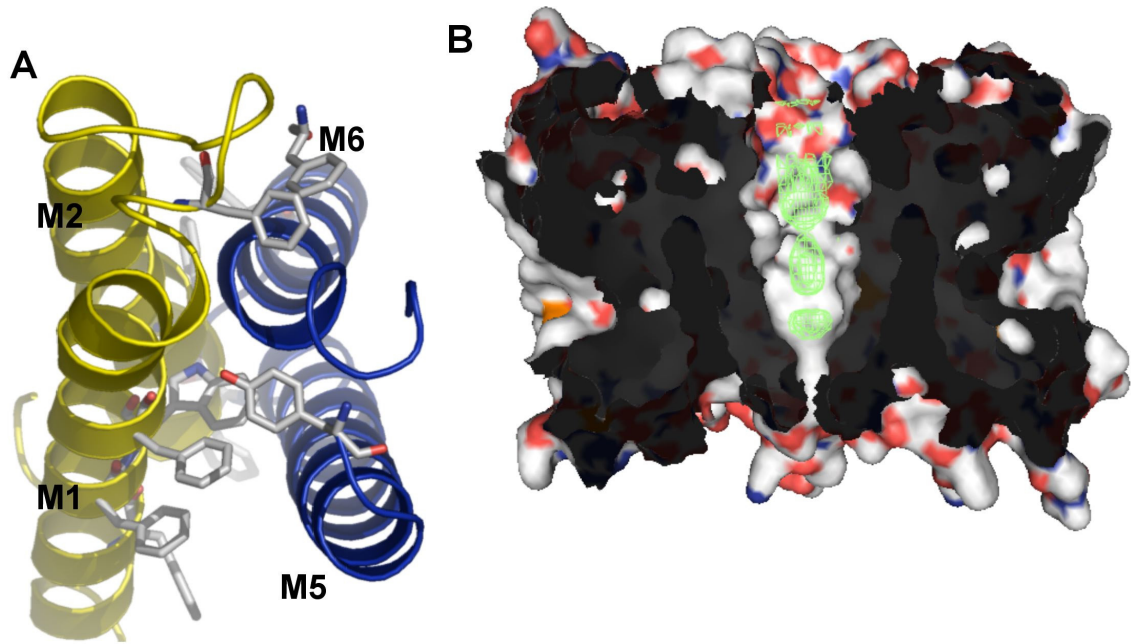


Fig. 5. Oligomerization of AqpZ (A) Protomer-Protomer interface. The interface is composed of helices M1 and M2 from one protomer (yellow) and M5 and M6 (blue) from a second protomer. The helices participate in knobs into holes packing, and the interface is extensive (3340\AA^2) due the large number of aromatic residues. (B) Abstruse electron density along the four-fold axis (colored in green) with two protomers displayed in surface rendering. The experimental electron density ($2F_o - F_c$) is contoured at 1.1σ .

References

Abramson, J., I. Smirnova, et al. (2003). Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* 301(5633): 610-5.

Alberts, B., A. Johnson. (2002). *Molecular biology of the cell*. New York: Garland Science. 615p.

Bernal, J.D and R.H Fowler (1933). A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. *J. Chem. Phys.* 1(8) 515-548

Borgnia, M. J. and P. Agre (2001). Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from *Escherichia coli*. *Proc Natl Acad Sci U S A* 98(5): 2888-93.

Borgnia, M. J., D. Kozono, et al. (1999). Functional reconstitution and characterization of AqpZ, the *E. coli* water channel protein. *J Mol Biol* 291(5): 1169-79.

Borgnia, M., S. Nielsen, et al. (1999). Cellular and molecular biology of the aquaporin water channels. *Annu Rev Biochem.* 68:425-458.

Brünger, A.T. X-PLOR Version 3.843 (Yale University, New Haven, Ct, 1996).

Calamita, G., W. R. Bishai, et al. (1995). Molecular cloning and characterization of AqpZ, a water channel from Escherichia coli. *J Biol Chem* 270(49): 29063-6.

Calamita, G., B. Kempf, et al. (1998). Regulation of the Escherichia coli water channel gene aqpZ. *Proc Natl Acad Sci U S A* 95(7): 3627-31.

Calamita, G., B. Kempf, et al. (1997). The aquaporin-Z water channel gene of Escherichia coli: structure, organization and phylogeny. *Biol Cell* 89(5-6): 321-9.

DeLano, W.L. The PyMOL Molecular Graphics System (2003) DeLano Scientific, San Carlos, CA, USA. <http://www.pymol.org>.

Doyle, D. A., J. Morais Cabral, et al. (1998). The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* 280(5360): 69-77.

Dutzler, R., Y. F. Wang, et al. (1996). Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway. *Structure* 4(2): 127-34.

Fu, D., A. Libson, et al. (2000). Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* 290(5491): 481-6.

de Grotthuss, C. J. T. Sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann. Chim.* LVIII (1806) 54-74

Heller, K. B., E. C. Lin, et al. (1980). Substrate specificity and transport properties of the glycerol facilitator of *Escherichia coli*. *J Bacteriol* 144(1): 274-8.

Heymann, J. B. and A. Engel (2000). Structural clues in the sequences of the aquaporins. *J Mol Biol* 295(4): 1039-53.

Marti-Renom, M. A., A. C. Stuart, et al. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29: 291-325.

Maurel, C., J. Reizer, et al. (1994). Functional characterization of the *Escherichia coli* glycerol facilitator, GlpF, in *Xenopus* oocytes. *J Biol Chem* 269(16): 11869-72.

Miroux, B. and J. E. Walker (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* 260(3): 289-98.

Muller, K. et al. *Bull. Soc. Chim. Belge.* 97, 655-667 (1988).

Murata, K., K. Mitsuoka, et al. (2000). Structural determinants of water permeation through aquaporin-1. *Nature* 407(6804): 599-605.

Neidhardt, F.C, et al. (1996) *Escherichia coli* and *Salmonella*. Washington D.C.: ASM Press. 14 p.

Otwinowski, Z., Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, 276, 307-326.

Park, J. H. and M. H. Saier, Jr. (1996). Phylogenetic characterization of the MIP family of transmembrane channel proteins. *J Membr Biol* 153(3): 171-80.

Preston, G. M., T. P. Carroll, et al. (1992). Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein. *Science* 256(5055): 385-7.

Ringler, P., M. J. Borgnia, et al. (1999). Structure of the water channel AqpZ from *Escherichia coli* revealed by electron crystallography. *J Mol Biol* 291(5): 1181-90.

Scheiner, S., T. Kar, et al. (2001). Strength of the C α H..O hydrogen bond of amino acid residues. *J Biol Chem* 276(13): 9832-7.

Senes, A., M. Gerstein, et al. (2000). Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J Mol Biol* 296(3): 921-36.

Senes, A., I. Ubarretxena-Belandia, et al. (2001). The C α —H...O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc Natl Acad Sci U S A* 98(16): 9056-61.

Sui, H., B. G. Han, et al. (2001). Structural basis of water-specific transport through the AQP1 water channel. *Nature* 414(6866): 872-8.

Tajkhorshid, E., P. Nollert, et al. (2002). Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* 296(5567): 525-30.

Valiyaveetil, F. I., Y. Zhou, et al. (2002). Lipids in the structure, folding, and function of the KcsA K⁺ channel. *Biochemistry* 41(35): 10771-7.

Van Gelder, P., F. Dumas, et al. (2002). Sugar transport through maltoporin of *Escherichia coli*: role of the greasy slide. *J Bacteriol* 184(11): 2994-9.

Wahl, M. C. and M. Sundaralingam (1997). C-HO hydrogen bonding in biology. *Trends Biochem Sci* 22(3): 97-102.

Walz, T., T. Hirai, et al. (1997). The three-dimensional structure of aquaporin-1. *Nature* 387(6633): 624-7.

Chapter 3

A structure-function study of the aquaporin selectivity filter

Research completed in collaboration with
Joseph D. O'Connell III, and Robert M. Stroud

*This chapter is currently in submission to the Proceedings of the National Academy of
Science*

Abstract

Aquaporins selectively facilitate the permeation water of and small molecules across the cell membrane. Aquaporins can functionally be divided into water selective channels and those which conduct both water and other small molecules including glycerol. It is thought the narrowest point in the channel, termed the selectivity filter, is the structural basis of selectivity. To test this idea, we sequentially engineered the selectivity filter of the glycerol facilitator (GlpF) in to the orthodox water channel aquaporin Z (AqpZ). Functional analysis shows a decrease in water permeability, but not the expected increase in glycerol conduction. X-ray structures of the AqpZ mutants show an unexpected decrease in pore size. Thus, the selectivity filter is the basis of selectivity, but the overall aquaporin architecture is also tuned to promote this selectivity.

Introduction

AQPs (AQPs) are integral membrane channels that selectively facilitate the permeation of water and small, amphipathic molecules across cellular membranes¹. They display rates of conduction at the diffusion limit, but are remarkably selective and exclude all ions, including hydroxide and hydronium. AQPs can functionally be divided into two subfamilies, the orthodox AQPs, which are selective for water and the more promiscuous aquaglyceroporins, which conduct both water and amphipathic molecules such as glycerol. Although AQPs are found throughout all kingdoms of life, this branching is most apparent in *E. coli*, whose genome only encodes aquaporin Z (AqpZ)², an orthodox AQP, and the glycerol facilitator (GlpF)³, an aquaglyceroporin.

Atomic resolution x-ray structures of both AQPs and aquaglyceroporins have been solved and a comparison between these structures reveals the conserved structural architecture characteristic to the family and particular motifs that modulate function⁴. The structure of GlpF (Figure 1) is a right-handed bundle of six transmembrane and two half-spanning helices (M1-M8) forming a central channel. Helices M1-M4 are related by pseudo-twofold symmetry to M5-M8. The channel diameter is less than 3 Å and roughly 20 Å long. At both the cytoplasmic and extracellular/periplasmic entrance there are conically shaped vestibules to the channel, which give the protein an hour-glass shape⁵. The channel is amphipathic, lined mostly by hydrophobic side chains and punctuated by two symmetry related sets of main chain carbonyls evenly spaced throughout the channel.

These carbonyls act as the water-coordinating motif and create a single-file chain of water running the length of the pore. The half-spanning helices M3 and M7 meet at the center of the channel and are capped by the signature Asn-Pro-Ala (NPA) motif. The asparagines side chains are located directly in the pore and are hydrogen bond donors. Despite being water-filled channels, AQPs do not conduct protons via the Grotthuss mechanism like similar channels such as gramicidin⁶. Molecular dynamics and quantum chemical simulations have located the likely barrier to proton conductance at the NPA region. This mechanism of exclusion is a combination of a bipolar water orientation by the NPA motif and the inherent two-fold symmetry, the helix dipoles of M3 and M7, and the desolvation penalty associated with the electrostatic differences between bulk water and the channel. The structure of the prototypical orthodox aquaporin 1 (AQP1) is nearly identical except the channel is of a narrower diameter⁷. The most significant difference occurs at a region known as the selectivity filter (SF).

The selectivity filter is the most narrow constriction point of the channel. In orthodox AQPs it is slightly smaller than the radius of water (1.4 Å), and the SF is about one Å larger in aquaglyceroporins. It is located roughly halfway between the NPA motif and the extracellular vestibule and composed of a strictly conserved arginine and three other amino acids that vary depending on selectivity. In water selective AQPs this region is more polar and contains a conserved histidine, while in aquaglyceroporins it is more hydrophobic with two conserved aromatic residues. One anomaly, the putative H₂S channel AqpM, has a more hydrophobic channel with a radius similar to orthodox AQPs⁸. Because of the simple clustering between sequence and function it is thought this

region forms the basis of selectivity in AQPs. Removal of the conserved aromatic residues in GlpF leads to an increase in water permeability and widening of the orthodox AQP1 pore through alanine mutation results in a significant increase in glycerol permeability. Simulation and mutational analysis also suggests that the SF arginine, and its conformation based on the local environment, plays a role in channel gating and selectivity for glycerol. Here, to further elucidate the structural barriers to glycerol conductance we engineer the GlpF SF into AqpZ.

AqpZ has been functionally characterized as an orthodox AQP⁹ and its x-ray crystal structure has been determined to 2.5 Å in our laboratory¹⁰. Therefore AqpZ and GlpF together constitute an excellent model system to probe structure-function relationships¹¹. Here, via mutation, we investigate the role of pore diameter and polarity on both conduction of water and glycerol in a reconstituted proteoliposome assay. Introduction of the GlpF SF results in a decrease in water permeability, but no increase in glycerol conduction. To explain this unexpected result we solved the x-ray crystal structures of all mutants. These structures surprisingly reveal a pore as, or more, constricted than wild-type AqpZ (WT). Thus, although size and polarity of the SF are the major determinants in AQP substrate preference, the overall architecture of the protein is also tuned to promote selectivity.

Results

The structures of GlpF and AqpZ show a remarkable degree of similarity (α -carbon root mean square deviation of 1.7 Å), but are quite different at the SF (Figure 1). To test the role of the SF in glycerol conduction we made the mutants F43W, H174G/T183F, and F43W/H174G/T183F to graft the GlpF SF onto AqpZ. These mutants were then expressed, solubilized in the detergent n-Octyl- β -D-glucopyranoside (OG), purified and then functionally and structurally characterized. WT expressed at ten mg / L of culture and the mutants expressed to roughly five mg / L of culture.

Water Permeability of AqpZ and GlpF. Flux through the AQP channel can be measured in osmotically-driven proteoliposome permeability experiments¹¹. AqpZ, AqpZ mutants, and GlpF were purified and reconstituted into liposomes. Permeabilities were measured in a stopped-flow device by challenging the proteoliposomes with a hyperosmolar reconstitution buffer to drive water efflux. The resulting proteoliposome shrinkage was measured by light scattering and the curves fit to an exponential equation with a single rate constant (k_{wat}). Raw light scattering curves can be seen in Figure 1A and the values for k_{wat} are listed in Table 1. Values are $102.9 \pm 1 \text{ s}^{-1}$, $56.8 \pm 3 \text{ s}^{-1}$, $34.8 \pm 0.2 \text{ s}^{-1}$, $12.3 \pm 0.1 \text{ s}^{-1}$, $20.2 \pm 0.4 \text{ s}^{-1}$, and $6.3 \pm 0.3 \text{ s}^{-1}$ for WT, F43W, H174G/T183F, F43W/H174G/T183F, GlpF, and control liposomes respectively. As expected sequential introduction of the GlpF SF into AqpZ results in a decrease in water permeability. F43W, with a larger aromatic side chain displays a large decrease in water permeability, and H174G/T183F shows an even larger decrease from the replacement of the polar

imidazole moiety of H174 with a phenyl ring of T183F. Finally, the triple mutant F43W/H174G/T183F has a rate just above that of empty liposomes. Though one would expect similar rates, the flux is also considerably less than GlpF.

Glycerol Permeability of AqpZ and GlpF. Glycerol permeability was measured in a similar manner as water. Proteins were reconstituted into liposomes and flux was measure in a stopped-flow device by challenging glycerol-loaded proteoliposomes with an isoosmolar buffer containing impermeant sucrose. Glycerol efflux was thus measure by light scattering and the curves fit to an exponential equation with a single rate constant (k_{gly}). Figure 2B shows the raw light scattering data and the values for k_{gly} are listed in Table 1. Values are $0.102 \pm 0.001 \text{ s}^{-1}$, $0.105 \pm 0.001 \text{ s}^{-1}$, $0.100 \pm 0.001 \text{ s}^{-1}$, $0.107 \pm 0.004 \text{ s}^{-1}$, $4.82 \pm .06 \text{ s}^{-1}$, and $0.094 \pm 0.002 \text{ s}^{-1}$ for WT, F43W, H174G/T183F, F43W/H174G/T183F, GlpF, and control liposomes respectively. AqpZ and its mutants have a negligible permeability above the membrane. Thus, despite the mutations, the values indicate there is no increase in glycerol permeability.

Structures of the engineered selectivity filter mutants. To explain the decrease in water flux coupled with a lack of increase in glycerol permeability we solved the x-ray crystal structures of the three AqpZ mutants. Following purification, the mutants were crystallized and atomic resolution diffraction data were collected (Table 2). All mutants crystallized in P4, the same space group as WT AqpZ, and had similar unit cell dimensions. The structures were solved by molecular replacement with the 1.9 Å L170C mutant structure (PDB CODE?) and with two monomers (chain A and chain B) in the

asymmetric unit. The final resolution cutoffs were 2.40 Å, 2.50 Å, and 3.10 Å and the R_{free} statistics for the refined structures were 21.8%, 22.0%, and 22.1% for F43W, H174G/T183F, and F43W/H174G/T183F respectively (Table 2).

Superposition of the structures indicates there is little main chain variation between WT and the mutants (Figure 3A). Electron density for the mutant SF residues is well-defined in $2F_o - F_c$ difference maps. An example is shown for mutant H174G/T183F in Figure 3B. For this mutant we were also able to locate two glycerol molecules (denoted Cry 1 and Cry2) just outside the selectivity filter in a similar location to glycerol G1 in the original GlpF structure. In the F43W and H174G/T183F mutant structures we were able to locate 136 and 94 solvent atoms, respectively, and there is a near continuous single-file chain of water present in the A monomers for both proteins. As in the WT structure we were able to locate 3 OG molecules clustered around the conserved aromatics of helix M8 near the periplasmic leaflet. These detergents molecules are in nearly the same conformation for all structures, indicative of a favorable binding site.

Figure 4A details the structural changes induced through mutagenesis. The van der Waals radii for the SF residues indicate that sequential mutation in fact decreases the pore size in comparison to WT AqpZ. Calculation of the channel (Z axis) radius using the program HOLE2 reveals that F43W and H174G/T183F mutants are extremely narrow at the SF ($Z = -10$ Å), with a radius under 1.0 Å. In contrast, the GlpF radius for this region is approximately 1.7 Å. Cross-sectional areas for the SF are 3.9 Å², 2.0 Å², 2.5 Å², 3.5 Å², and 9.3 Å² for WT, F43W, H174G/T183F, F43W/H174G/T183F, and GlpF further

underscoring the decrease in channel size. Thus, the series of mutations one would expect to dilate the pore actually decreases its size.

This surprising constriction is caused by the subtle variation in protein architecture between AqpZ and GlpF. A superposition of F43W/H174G/T183F on GlpF (Figure 5A) shows that the three SF residues projecting side chains into the pore are in different conformations. The indole ring of W43 is laterally shifted 0.8 Å into the pore, a result of helix M2 also being shifted roughly the same amount. A more drastic arrangement occurs at F183 (Figure 5B) resulting in a 1.2 Å shift of the phenyl ring into the pore.

This residue, unique among the SF residues in that it is also part of the water-coordinating carbonyl motif, is typically a small hydrophilic residue in orthodox AQPs. In AqpZ it is a threonine, while in the prototypic AQP1 it is a mercury-sensitive cysteine. In these AQPs, the main chain is oriented to position the side chain rotamer away from the pore and to make room for the strictly conserved water-selective histidine. In GlpF, however, the main chain is oriented in such a way as to position the phenyl ring tightly packed against helix M6 and the conserved glycine, which is a histidine in orthodox AQPs. Furthermore, the conformation of F200 in GlpF is in the unfavorable region of the Ramachandran plot and is preceded by a conserved glycine, which may aid in adopting this unique conformation. The main chain carbonyls of the water-binding motif end up rotating slightly away from the pore and are positioned as far away as 3.1 Å from their respective carbonyls in AqpZ (Figure 5B). Interestingly, in GlpF this carbonyl

conformation is not mirrored in the two-fold related set, and directly superposes on the AqpZ structure.

Finally, the conformation of R189 is also significantly different. In the AqpZ WT structure, the R189 C_{β} - C_{γ} - C_{δ} - C_{ϵ} dihedral angle is -101.0° , a state described as “down” (Figure 6A). This down state impinges on the channel and effectively closes it during in molecular dynamics simulations¹². In GlpF this angle is -179.8° , leading to an “up” state that positions the positively charged guanidinium moiety out of the SF. In all mutant structures we observe a similar down R189 rotamer and conclude that R189 conformation is due to the local environment independent of the particular amino acids present in the SF. The C loop (Figure 6B), which connects helices M4 and M5 and enters the periplasmic vestibule, is much longer in aquaglyceroporins and in GlpF projects deeper into the pore. In this arrangement, the carbonyl of F135 makes a hydrogen bond with R206 (distance of 2.9 Å) thereby creating the up state. In AqpZ R189 makes a hydrogen bond with the main chain of A117 (distance of 2.9 Å), leading to the down state. Furthermore, because the water coordinating carbonyls described above are directly in the pore, as opposed to the GlpF structure, R189 can also make a hydrogen bond (distance of 2.7 Å) with the carbonyl of T183 (or F183 in the double and triple mutants). The carbonyls of A117 and T183 are therefore 180° from each with respect to the guanidinium of R189 and position it directly in the pore oriented in the down state.

Discussion

The AQP family of channels can be divided into orthodox water-selective AQPs and aquaglyceroporins, selective for both water and other small amphipathic molecules. This functional division is thought to be rooted at the SF, the most narrow region of the channel, which is typically smaller and more polar in orthodox AQPs. This division is embodied in AqpZ and GlpF, and so to investigate the determinants of selectivity we sequentially engineered the GlpF SF into AqpZ, assayed the mutants for both water and glycerol conduction and solved their x-ray crystal structures.

Mutants F43W, H174G/T183F, and F43W/H174G/T183F display a serial decrease in water permeability, an expected result given the exchange of His for another Phe in the narrow constriction. Given that the size of the SF for the triple mutant and WT are essentially the same, we can say the polarity of the SF plays a key role in allowing for water conduction. Dissecting the single and double mutant is more complex, as the structures also show a significant and unexpected decrease in pore size. It has previously been shown in AQP1 that mutation of the SF residues to alanine, thereby increasing pore size, does very little to affect water permeability¹³. We therefore postulate two different states for the SF. In an arrangement where the pore is significantly larger (for example, 8 Å in the double Ala AQP1 mutant) than the radius of water, polarity and small changes in pore size do not affect permeability. Presumably, the SF environment is more like that of bulk water and changes in either property do not affect function. As the radius decreases to that of water, and conduction becomes single file, the protein itself must supply

hydrophilic moieties to lower the free energy penalty in desolvating a single water from the bulk. Furthermore, these hydrophilic moieties in the SF are balanced in such a way as to exclude ions, particularly protons, as removal of the conserved arginine allows for a small proton conductance in AQP1. In a SF the size of an orthodox AQP, best evidenced by F43W/H174G/T183F, a hydrophobic environment lowers water permeability to nearly that of a lipid bilayer.

In the case of glycerol, pore size seems to be the most important determinant. Despite the increased hydrophobic environment, none of the mutants display any glycerol conduction. The x-ray structures show that the SF size actually decreases for F43W and H174G/T183F and is nearly the same as AqpZ for the F43W/H174G/T183F mutant. In the AQP1 Ala mutants described above, opening of the SF also resulted in a significant increase in glycerol permeability. Finally, in a potential of mean force molecular dynamic simulation of AqpZ, glycerol conduction was only seen upon severe rearrangement of the SF¹⁴. Thus, the barrier to glycerol conduction in orthodox aquaporins is most likely rooted in a decreased SF size.

The lack of SF size increase, which allows us to draw the conclusions above, was unexpected. GlpF and AqpZ are highly homologous, but engineering of the SF actually led to a decrease in size from 3.9 Å² to 3.5 Å² instead of 9.3 Å² for GlpF. The structure of F43W/H174G/T183F superposed onto GlpF shows that differences in main chain conformation and an insertion into loop C are the reasons for the surprising shrinkage in the SF. These differences are most striking at the water coordinating carbonyls shown in

Figure 5B. Rearrangement of this motif in GlpF leads to carbonyls not pointed directly into the pore and F200 adopting an unfavorable rotamer to widen the SF. Superposition of this motif on both AqpZ and the twofold related carbonyls in GlpF shows that its conformation is unique. In AqpZ, where the carbonyls are directly in the pore, the motif can hydrogen bond to R189 and position it in the so-called down state, which may affect both water and glycerol conduction. This down state is also encouraged through hydrogen bonding by the C loop, which is significantly smaller in orthodox AQPs. In GlpF, this loop is extended and makes a hydrogen bond, which orients R206 out of the SF in the up state. The importance of this single interaction is underscored in the *Plasmodium* aquaglyceroporin ¹⁵, in which mutation E125S (S136 in GlpF), abolishes water conduction with no significant change in glycerol permeability.

Thus we have shown that engineering of the SF, even a point mutation, can modulate function. Interpretation of these changes often involves analyzing the subtle differences in protein folds and thus, functional results must be interpreted experimentally determined structure. Finally, answering these questions of selectivity benefits from the use of a simple model system such as we have described here. Understanding the basic of selectivity in these unique channels will require both *in vivo* and *in vitro* function, structure, and simulation.

Materials and Methods

Expression and Purification. Mutants of AqpZ were generated by site-directed-mutagenesis of the pET28b-AqpZ construct used in the original structure paper. The E. coli strain C43(DE3) was transformed, grown to .6-1 OD at 600 nm at 37°C in 2 x LB media, .5% glycerol (v/v), 1x M9 salts, and 25 mg / L kanamycin, and induced with 1mM isopropyl -D-thiogalactoside (Anatrace).

All purifications were carried out at 4°C or on ice. Cells from six L of culture were harvested and lysed by a microfluidizer in 20 mM Tris pH 7.4, 100 mM NaCl, .5 mM phenylmethylsulfonyl fluoride, and 5 mM BME. Membranes were recovered from supernatant by 100,000 x g centrifugation for 2 hr. Protein was solubilized from membranes by agitation in 20 mM Tris pH 7.4, 100 mM NaCl, 5 mM BME, 10% glycerol, and 270 mM OG (Anatrace) for 12-16 hours at 4°C. Solubilized protein was bound in batch to Ni-NTA resin (Qiagen) for 1 hour, washed with 25 resin volumes of 20mM Tris pH 7.4, 100 mM NaCl, 5 mM BME, 10% glycerol, 40 mM OG, and 20 mM imidazole, and eluted with 20 mM Tris pH 7.4, 100 mM NaCl, 5 mM BME, 10% glycerol, 40 mM OG, and 250 mM imidazole. Imidazole was removed using a Biorad Econo-Pac 10DG desalting column and the histidine tag was removed by digestion with 5 ug of trypsin for 12hr at 4°C. Trypsin was removed by passing over a benzamidine-sepharose matrix (GE Healthcare), and the protein sample was injected onto a Pharmacia Superose 12 gel filtration column running a mobile phase of 20 mM Tris pH 7.4, 100 mM NaCl, 2 mM dithiothreitol (DTT), 10% glycerol, and 40 mM OG. Except as noted, all materials were purchased from Sigma or Fisher. The sample was judged pure and

homogenous by both gel filtration chromatography and coomassie-stained denaturing gels.

Crystalization and Data Collection. Following gel filtration chromatography, the protein was concentrated to 25 mg / ml using a 30 kDa cutoff Amicon Ultra-15 Centrifugal Filter. Crystals were grown by hanging drop vapor diffusion at room temperature by 1:1 addition of protein and 25-30% polyethylene glycol monomethyl ether 2000 (Fluka), 100 mM sodium cacodylate pH 6.5, and 50-100 mM MgCl₂. Crystals grew to roughly 300 μm x 300 μm x 150 μm over the course of several days and were flash frozen in liquid nitrogen following a brief washing in the mother liquor plus 15% glycerol for cryoprotection. Diffraction intensities were collected on Advanced Light Source Beamline 8.3.1 using a ADSC Quantum-Q210 CCD detector.

Phasing and Model Refinement. Data were processed with Elves¹⁶ and CCP4¹⁷ (using MOSFLM¹⁸) and the structures were solved by molecular replacement with the 1.9 Å Leu170Cys mutant structure (Protein Data Bank Code 209F) using Phaser¹⁹. The models were refined with iterative cycles of manual building with Coot²⁰ and TLS restrained refinement, individual B-factor refinement, and tight non-crystallographic symmetry (NCS) restraints in Refmac5²¹. Mutations were verified and side chains positioned using omit maps in Refmac5.

Proteoliposome Reconstitution. Before removing the 6xHis tag with trypsin, aliquots of protein were set aside for proteoliposome reconstitution. E. coli polar lipids were sonicated to clarity and the reconstitution cocktail was prepared by sequentially adding

100 mM MOPS pH 7.5, 1.5% (wt/vol) OG, 50 ug / ul of purified protein, and 10 mg / ml *E. coli* polar lipids (Avanti)¹¹. To reduce oxidation, lipid stocks were stored in 2mM BME and all buffers were under argon atmosphere. Following cocktail incubation for 1 hour at room temperature (RT), proteoliposomes were formed by diluting the mixture 50-fold into a running buffer of 20 mM Hepes pH 7.5 and harvested by centrifugation at 100,000 x g for 2hr. Pelleted liposomes were resuspended into 1 ml of running buffer (RB) (20 mM Hepes pH 7.5) and stored at 4°C. Liposomes monodispersity was verified by dynamic light scattering with a mean diameter of 90 nm.

To analyze the kinetics of water conduction through the channel, we diluted the proteoliposomes 7.5 fold in RB and subjected them to an osmotic gradient by mixing 1:1 proteoliposomes (final AqpZ monomer concentration of .27uM) and RB with osmolyte (RBS) (20 mM Hepes pH 7.5, 570 mM sucrose; a 285 mosM final gradient) and measured water efflux (liposome shrinkage) by light scattering in a stopped-flow device (Applied Photophysics) at 440 nm. Resulting curves were fit to a single-exponential rate constant (k_{wat}) as a measure of permeability to use in comparison between proteins.

To analyze glycerol kinetics, pelleted proteoliposomes were resuspended in 1 ml of RB with 570 mM glycerol (RBG). They were then diluted 7.5 fold in RBG and mixed 1:1 with RBS in a similar manner as above. The impermeant sucrose then drives glycerol efflux and light scattering detects liposome shrinkage. Resulting curves were fit to a single-exponential rate constant (k_{gly}) as a measure of glycerol permeability.

Abbreviations

AQPs, aquaporins; GlpF, glycerol facilitator; AqpZ, aquaporin Z; AQP1, aquaporin 1;

NPA, asparagines– alanine –proline; SF, selectivity filter.

Table 1: Proteoliposome Permeability

	$k_{\text{wat}} \text{ (s}^{-1}\text{)}$	$k_{\text{gly}} \text{ (s}^{-1}\text{)}$
WT AqpZ	102.9 ± 1	0.102 ± 0.001
F43W	56.8 ± 3	0.105 ± 0.001
H174G / T183F	34.8 ± 0.2	0.100 ± 0.001
F43W / H174G / T183F	12.3 ± 0.1	0.107 ± 0.004
GlpF	20.2 ± 0.4	$4.82 \pm .06$
empty liposomes	6.3 ± 0.3	0.094 ± 0.002

Table 2: Crystallographic Data and Refinement Statistics

	F43W	H174G / T183F	F43W / H174G / T183F
Data Collection			
Space Group	<i>P4</i>	<i>P4</i>	<i>P4</i>
Unit Cell			
a (Å)	92.4	92.8	92.9
c (Å)	78.7	78.9	79.5
Resolution range (Å) ^a	20-2.40 (2.40-2.46)	20-2.50 (2.50-2.56)	20-3.10 (3.10-3.18)
Unique reflections	24848	21743	10963
Completeness ^a	95.4(86.6)	93.8(82.4)	88.0(90.4)
R _{sym} ^{a, b}	.074(.700)	.063(.645)	.115(.679)
I / σ ^a	15.4(1.3)	17.4(1.6)	10.6(1.5)
Refinement Statistics			
R _{work} / R _{free} (%)	18.0 / 21.8	18.3 / 22.0	19.8 / 22.1
RMSD ^c bonds (Å)	.014	.014	.025
RMSD ^c angles (Å)	1.45	1.46	1.54
Number of protein atoms	3356	3357	3359
Number of OG molecules	4	4	3

Number of solvent atoms	136	94	13
Average B-factor (Å ²)	49.7	49.7	63.7
<hr/>			
PDB code	XXXX	XXXX	XXXX

^a values in parenthesis refer to the highest-resolution shell

^b $\frac{\sum |I - \langle I \rangle|}{\sum I}$, where I equals observed intensity and $\langle I \rangle$ equals average intensity for symmetry-related reflections

^c Root-mean-square deviation of bond lengths and angles from ideal values

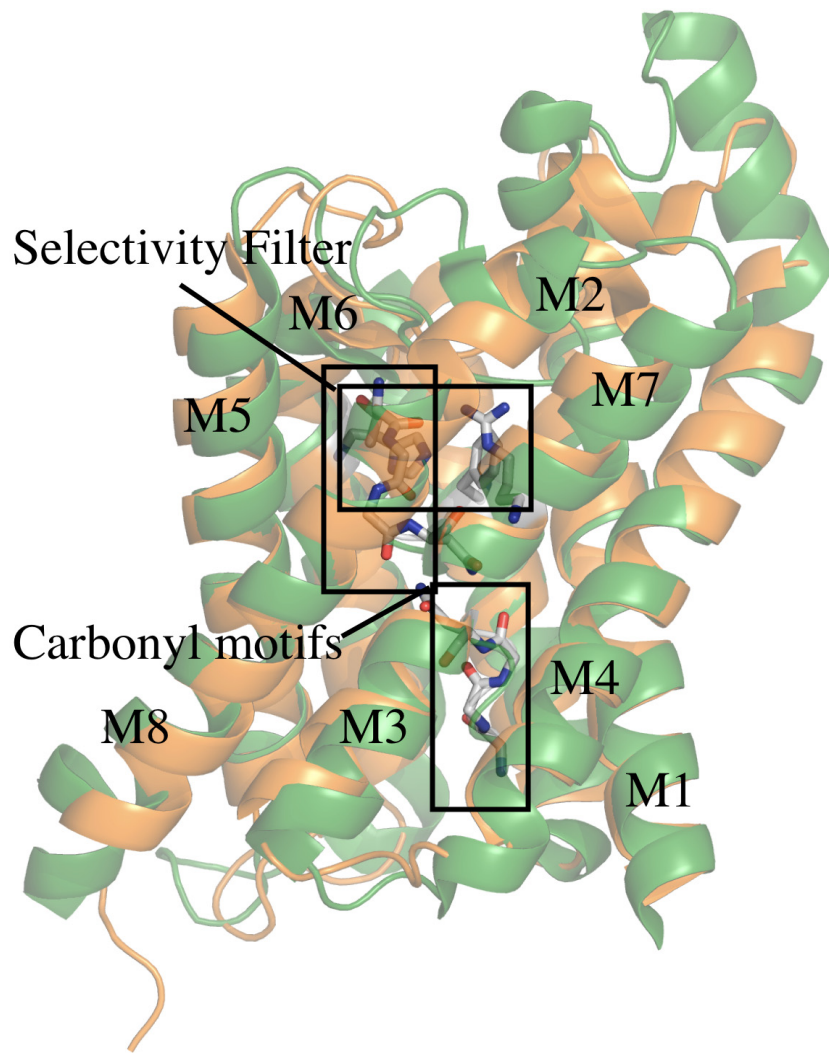


Figure 1. Superposition of AqpZ and GlpF. AqpZ (orange) is superposed on GlpF (green) and both structures are displayed in cartoon representation. Helices are labeled M1 through M8. The two symmetry related water-coordinating carbonyls are denoted by boxes, as is the selectivity filter near the periplasmic entrance to the channel.

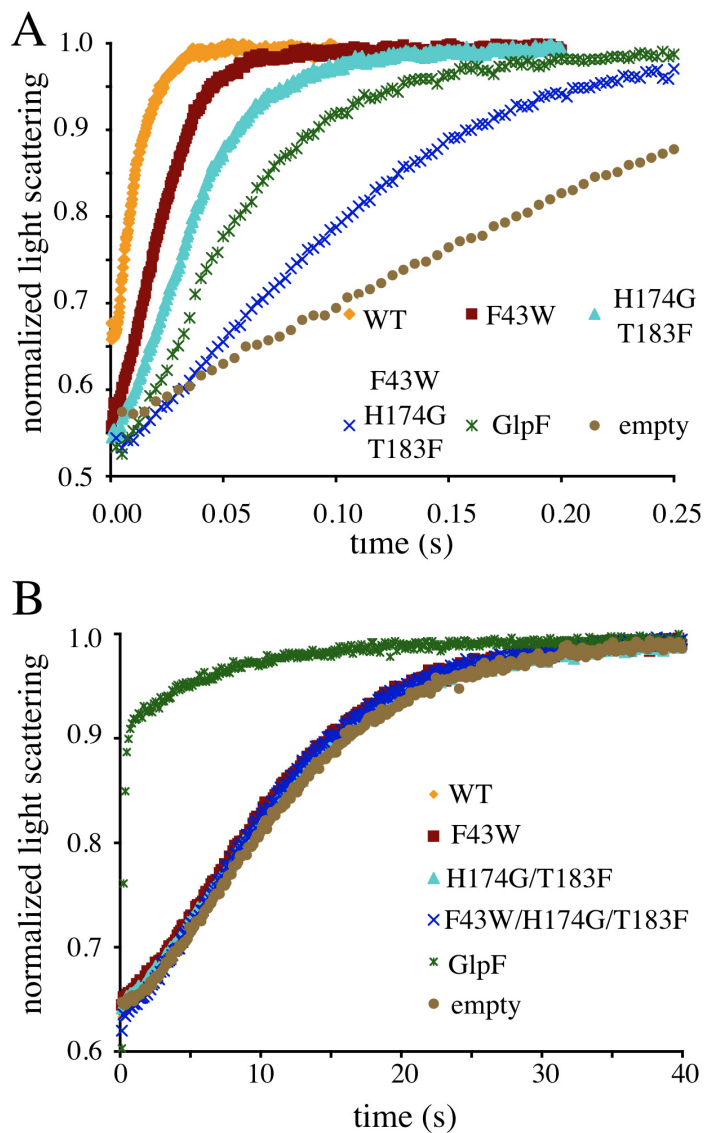


Figure 2. Proteoliposome conduction assays. (A) Curves of the raw data obtained in an osmotic challenge assay measuring water permeability. Colors code is WT (orange), F43W (maroon), H174G/T183F (cyan), F43W/ H174G/T183F (blue), GlpF (green), and empty liposomes (tan). These colors are used throughout the paper to denote their respective protein. (B) Raw data for glycerol conductivity.

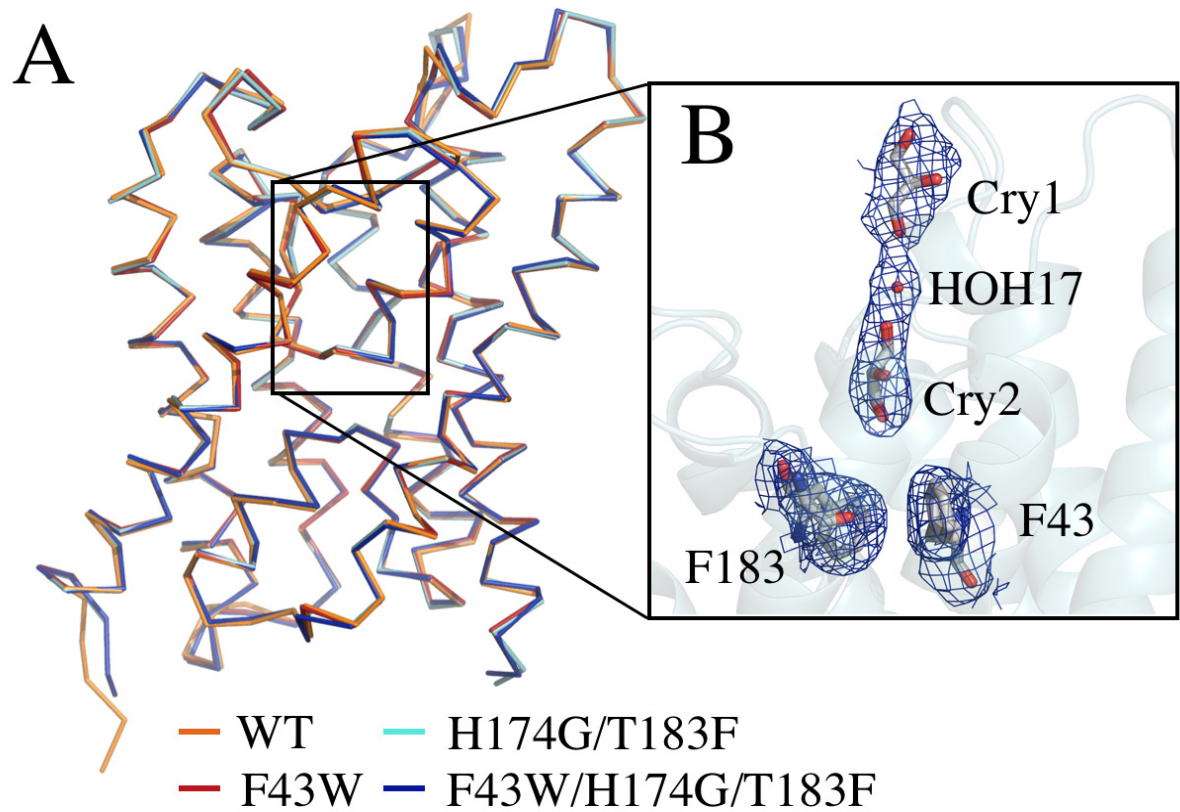


Figure 3. Superposition of the WT and mutant AqpZ structures. (A) Main chain overlay of all structures indicating little structural change in the mutants. (B) Example $2F_o - F_c$ density for mutant H174G/T183F at the SF. Also shown are two glycerol molecules and a water located just outside the SF.

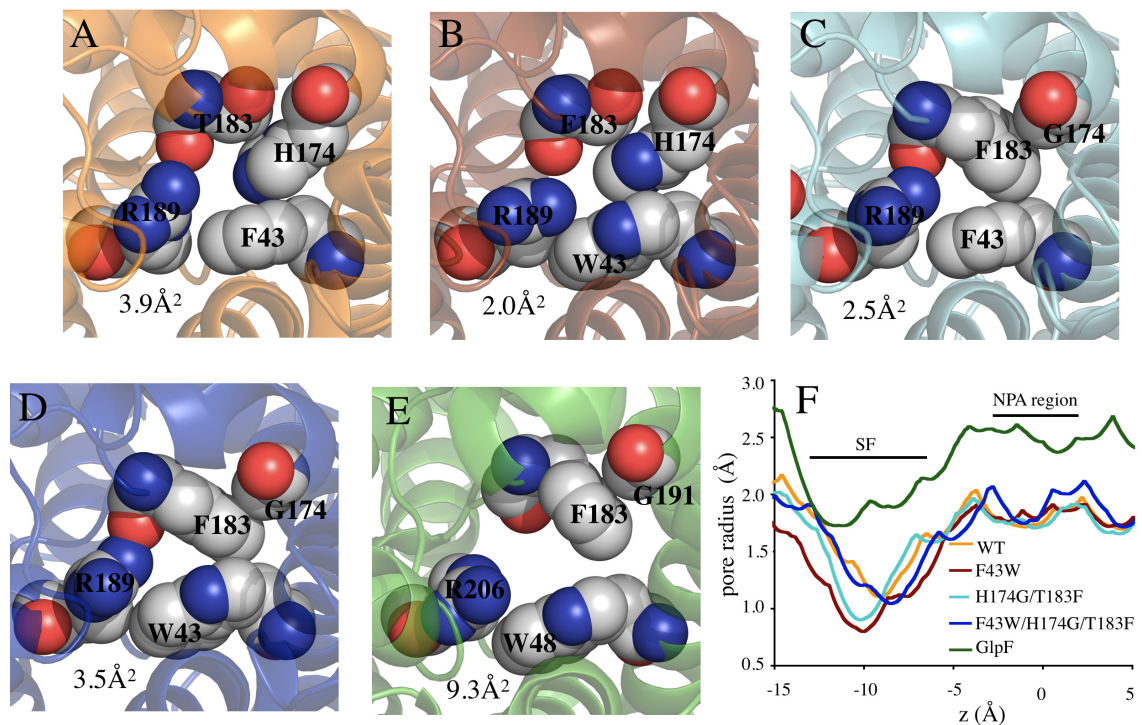


Figure 4. The SF of AqpZ, AqpZ mutants, and GlpF. (A) Structure of the WT AqpZ SF. Residues of the SF are drawn as spheres with van der Waals radii. The area of the selectivity filter measured with the program HOLE2 is shown in the lower left hand corner. (B) F43W SF. (C) H174G/T183F SF. (D) F43W/ H174G/T183F SF. (E) GlpF SF. (F) Channel radii determined by HOLE2 for all proteins plotted as a function of channel axis position. The NPA region is the origin.

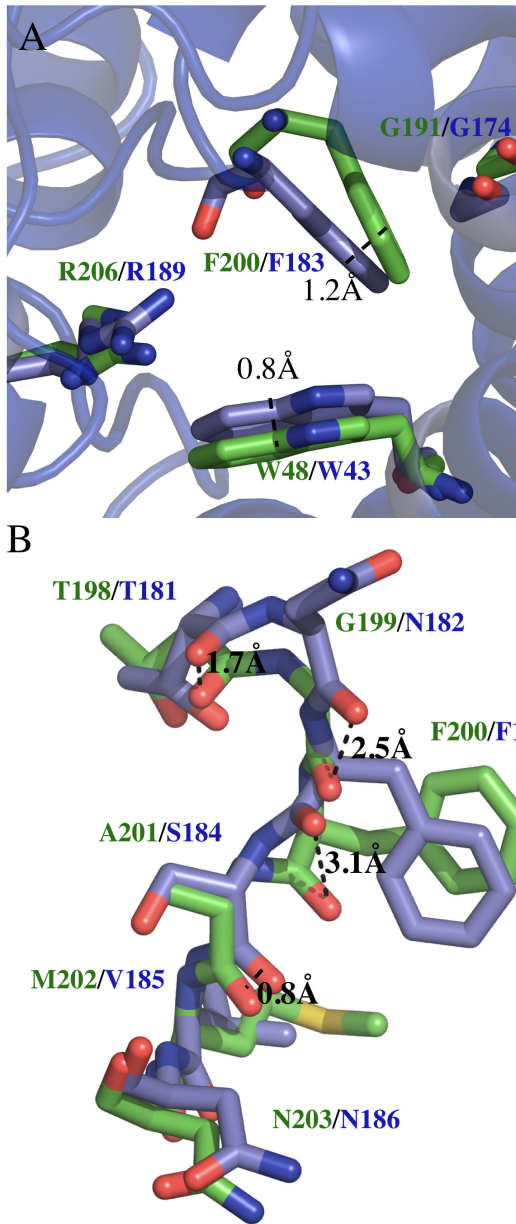


Figure 5. The F43W/ H174G/T183F SF is structurally different than GlpF. (A) Superposition of the mutant F43W/ H174G/T183F (blue) on GlpF (green) reveals significant differences despite the identical residues (shown in stick format). (B) Superposition of the C-terminal water-coordinating carbonyls from the two structures. Positional differences between carbonyls are indicated with a dashed line.

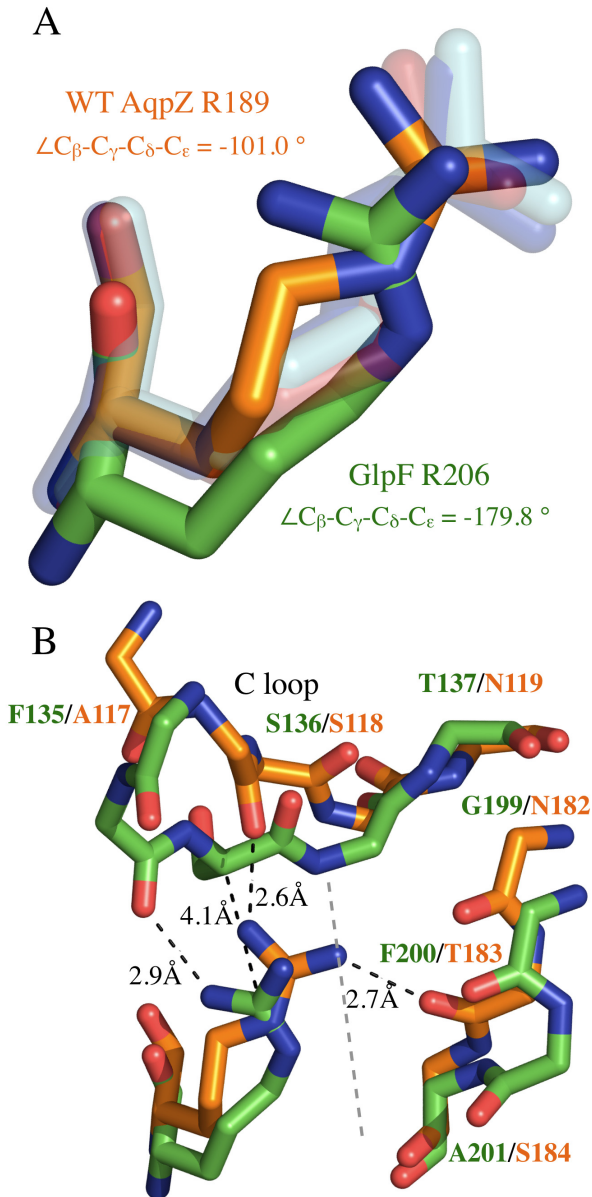


Figure 6. Structural differences in the SF arginine. (A) Superposition of the WT (orange) and GlpF (green) SF arginine conformations. In transparency are the other structures also superposed showing a consistent conformation between WT and all mutants. (B) Role of the C loop and carbonyls in positioning the arginine. Potential hydrogen bonding partners are shown as dashed lines. Pore axis is denoted as a gray dashed line.

References

1. Borgnia, M., Nielsen, S., Engel, A. & Agre, P. (1999). Cellular and molecular biology of the aquaporin water channels. *Annu Rev Biochem* **68**, 425-58.
2. Calamita, G., Bishai, W. R., Preston, G. M., Guggino, W. B. & Agre, P. (1995). Molecular cloning and characterization of AqpZ, a water channel from *Escherichia coli*. *J Biol Chem* **270**, 29063-6.
3. Heller, K. B., Lin, E. C. & Wilson, T. H. (1980). Substrate specificity and transport properties of the glycerol facilitator of *Escherichia coli*. *J Bacteriol* **144**, 274-8.
4. 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**, 481-6.
5. Walz, T., Hirai, T., Murata, K., Heymann, J. B., Mitsuoka, K., Fujiyoshi, Y., Smith, B. L., Agre, P. & Engel, A. (1997). The three-dimensional structure of aquaporin-1. *Nature* **387**, 624-7.
6. de Groot, B. L. & Grubmuller, H. (2005). The dynamics and energetics of water permeation and proton exclusion in aquaporins. *Curr Opin Struct Biol* **15**, 176-83.
7. 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**, 872-8.

8. Lee, J. K., Kozono, D., Remis, J., Kitagawa, Y., Agre, P. & Stroud, R. M. (2005). Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 Å. *Proc Natl Acad Sci U S A* **102**, 18932-7.
9. Borgnia, M. J., Kozono, D., Calamita, G., Maloney, P. C. & Agre, P. (1999). Functional reconstitution and characterization of AqpZ, the E. coli water channel protein. *J Mol Biol* **291**, 1169-79.
10. Savage, D. F., Egea, P. F., Robles-Colmenares, Y., O'Connell, J. D., 3rd & Stroud, R. M. (2003). Architecture and selectivity in aquaporins: 2.5 Å X-ray structure of aquaporin Z. *PLoS Biol* **1**, E72.
11. Borgnia, M. J. & Agre, P. (2001). Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from Escherichia coli. *Proc Natl Acad Sci U S A* **98**, 2888-93.
12. Jensen, M. O. & Mouritsen, O. G. (2006). Single-channel water permeabilities of Escherichia coli aquaporins AqpZ and GlpF. *Biophys J* **90**, 2270-84.
13. Beitz, E., Wu, B., Holm, L. M., Schultz, J. E. & Zeuthen, T. (2006). Point mutations in the aromatic/arginine region in aquaporin 1 allow passage of urea, glycerol, ammonia, and protons. *Proc Natl Acad Sci U S A* **103**, 269-74.
14. Wang, Y., Schulten, K. & Tajkhorshid, E. (2005). What makes an aquaporin a glycerol channel? A comparative study of AqpZ and GlpF. *Structure (Camb)* **13**, 1107-18.
15. Beitz, E., Pavlovic-Djuranovic, S., Yasui, M., Agre, P. & Schultz, J. E. (2004). Molecular dissection of water and glycerol permeability of the aquaglyceroporin

- from *Plasmodium falciparum* by mutational analysis. *Proc Natl Acad Sci U S A* **101**, 1153-8.
16. Holton, J. & Alber, T. (2004). Automated protein crystal structure determination using ELVES. *Proc Natl Acad Sci U S A* **101**, 1537-42.
 17. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3.
 18. Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**, 48-57.
 19. Read, R. J. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr D Biol Crystallogr* **57**, 1373-82.
 20. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32.
 21. 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**, 240-55.

Chapter 4

Structural basis of aquaporin inhibition by mercury

Research completed in collaboration with

Robert M. Stroud

This chapter is currently in submission to the Journal of Molecular Biology

Abstract

The aquaporin family of channels was defined based on the inhibition of water transport by mercurial compounds. Despite the important role of mercurials, little is known about the structural changes involved upon mercury binding leading to channel inhibition. To elucidate the mechanism we designed a mutant, T183C, of aquaporin Z (AqpZ) patterned after the known mercury-sensitive site of aquaporin 1 (AQP1) and determined the x-ray crystal structures of the unbound and mercury blocked states. Superposition of the two structures shows no conformational rearrangement upon mercury binding. In the blocked structure, there are two mercury sites – one bound to Cys183 and occluding the pore, and a second, also bound to the same cysteine but found buried in an interstitial cavity. To test the mechanism of blockade we designed a different mutant, L170C, to produce a more effective mercury block at the pore site. In a dose-response inhibition study, this mutant was 20 times more sensitive to mercury than wild-type AqpZ and 4 times more sensitive than T183C. The x-ray structure of L170C shows four mercury atoms at, or near, the pore site defined in the T183C structure and no structural change upon mercury binding. Thus, we elucidate a steric inhibition mechanism for this important class of channels by mercury.

Introduction

Aquaporins (AQPs) are integral membrane channel proteins that mediate the bidirectional flux of water and selected small amphipathic molecules across cellular membranes (Borgnia et al. 1999a). In the field of channel biology, selective inhibitors including naturally occurring toxins and organic molecules have played key roles in defining function (Hille 2001), but AQPs have no such specialized inhibitors. Mercurial compounds were found to reduce water transport in the red blood cell membrane to that of a bilayer and so defined the presence of a cellular water channel. (Macey 1984) This selective inhibition subsequently allowed for aquaporin isolation (Zeidel et al. 1992), cloning (Preston et al. 1992; Fushimi et al. 1993), membrane transport characterization (Javot and Maurel 2002), and mercury sensitivity mutational analysis (Preston et al. 1993; Kuang et al. 2001). The precise mechanism of AQP inhibition by mercury is still undetermined.

The AQP fold is a right-handed bundle of six transmembrane (TM) helices and two half-spanning helices (named M1-M8) with the pore running through the center of the helical bundle. There is a quasi two-fold symmetric relationship between the N-terminal portion of the polypeptide chain (M1-M4) and the C-terminal portion (M5-M8) from early gene duplication. In the membrane, AQPs occur as homotetramers of four monomer channels related by a four-fold symmetry axis (Figure 1A). The AQP family is divided into two subfamilies, the water selective channels and those with a more promiscuous selectivity for both water and other small amphipathic molecules such as glycerol

(aquaglyceroporins) (Park and Saier 1996). A comparison of GlpF, AQP1, and the other recently determined structures, shows that the AQP fold is conserved.(Savage et al. 2003) (Harries et al. 2004) (Tornroth-Horsefield et al. 2006) (Lee et al. 2005) AQPs are identified by two Asparagine-Proline-Alanine (NPA) sequence motifs located at the ends of the two quasi two-fold related half-spanning helices M3 and M7 (Figure 1B). The selectivity filter, a constricted region formed by four residues near the periplasmic/extracellular entrance, provides distinguishing features that identify the subfamilies (Figure 1B and 1C). In water selective AQPs this region is smaller and more polar and contains a conserved histidine, while in aquaglyceroporins it is larger and more hydrophobic with two conserved aromatic residues (Park and Saier 1996). Thus, the AQP architecture is conserved and it is the pore side chains that modulate specific functional differences (Figure 1C).

Mercurials can bind non-selectively to accessible cysteines, but in AQPs, due to a decrease in solvent accessibility from the membrane bilayer, mercury typically binds selectively to residues associated with the pore. In the mutational analysis of AQP1, removal of endogenous cysteines identified Cys189 as the one responsible for mercury sensitivity and predicted its pore location. Confirmation of Cys189 as a pore residue came from the first atomic resolution AQP structure, that of the *E. coli* glycerol channel (GlpF)(Fu et al. 2000), and the later AQP1 structure (Walz et al. 1997; Sui et al. 2001) showed precisely how the cysteine, not present in the GlpF sequence, was oriented. The structure of AQP1 was determined with protein obtained from natural sources so is not conveniently amenable to mutation and mercury was not used in the structural analysis.

To understand the mechanism of mercury inhibition in AQPs we focused our efforts on the bacterial homolog of AQP1, AqpZ (Calamita et al. 1995). AqpZ contains the water-selective sequence motif of AQP1 at the selectivity filter, has functionally been described as a water channel (Borgnia et al. 1999b), is not mercury sensitive, can be over-expressed and mutated, and its x-ray structure has previously been determined in our laboratory. AqpZ is thus a useful model system for probing the relationship between structure and function (Figure 1) (Savage et al. 2003) (Borgnia and Agre 2001).

Results

Structure of mutant T183C complexed with mercury. AqpZ and AQP1 are both functionally characterized as water-selective channels and have an identical selectivity filter except that AqpZ lacks the well-described mercury sensitive cysteine of AQP1 (Thr183 in AqpZ, Cys189 in AQP1). Based on this similarity, we postulated that a mutant of AqpZ lacking all endogenous cysteines but including the known mercury sensitive site of AQP1, would serve as a model for AQP1 (Figure 1C). The two endogenous cysteines of AqpZ were replaced with serine by mutagenesis, and the AQP1 mercury sensitive site was introduced via mutation Thr183Cys (protein hereby denoted T183C). This mutant, T183C, was expressed, purified, and crystallized in the presence or absence of HgCl₂ to determine the structural implications of mercury binding. The crystals diffracted to atomic resolution, and the two structures were solved by molecular replacement using the previously published structure of WT AqpZ (Protein Data Bank (PDB) code 1RC2) (Savage et al. 2003). The final resolution cutoffs were 2.30 Å and 2.20 Å, and the R_{free} statistics for the refined structures were 23.8% and 24.2% for the apo and mercury-bound forms respectively (Table 1). The overall structures of both the apo and complexed forms display the canonical AQP fold (Figures 2A and 2B) (Fu et al. 2000). Superposition of the structures shows there is no significant conformational change between the two forms, and the root mean square deviation (RMSD) of the main chain α-carbons is 0.27 Å.

A calculation of the channel surface (Figure 2C) using van der Waals radii with the program HOLE2 (Smart et al. 1993) reveals a 20 Å long pore resembling the WT structure. The most striking feature of the complex structure is the two large electron density peaks of the mercury atoms and their unexpected location – one is located directly in the pore and one is interstitially bound in a cavity just outside the pore (Figure 2C). T183C-Hg1 (nomenclature is mutant followed by mercury atom number), the mercury in the pore, is located roughly halfway between the NPA region and the narrow selectivity filter. It makes favorable electrostatic contacts with the main chain carbonyl of S184 (3.3 Å) and the imidazole ring of His174 (3.7 Å). Somewhat surprisingly, T183C-Hg1 is approximately 5.6 Å away from Cys183 and is 3.9 Å to the closest water. T183C-Hg2, the mercury outside of the pore, is bound to Cys183 (distance of 4.0 Å) and resides in a hydrophilic pocket formed by conserved Glu138 and Ser177 where it makes favorable electrostatic interactions of 2.6 Å and 3.1 Å respectively.

The thiol-mercury bond distances are considerably longer than the ideal length of 2.5 Å and noise in difference maps suggests that both mercury atoms are disordered and at partial occupancy. We therefore carried out joint occupancy and anisotropic B-factor refinement in SHELX (Sheldrick and Schneider 1997). In this calculation, T183C-Hg1 and T183C-Hg2 refine to occupancies of 0.24 and 0.32 respectively, suggesting a higher affinity or less disorder at the interstitial site. The anisotropic displacement parameters for T183C-Hg1 suggests disorder parallel to the channel axis, while T183C-Hg2 is disordered between residues Glu138 and Cys183. This disorder can also be observed in the $2F_o - F_c$ electron density (Figure 2C).

Based on these results, we hypothesized that the mercury site in the pore produces a steric block of the channel. To test this hypothesis we designed an optimized mercury-binding site in the pore, to provide a 'switch' as a probe of conductance in AqpZ. This new mutation, Leu170Cys in the cysteine-less background, was intended to bind mercury only at the site within the pore. We next biophysically characterized WT, T183C, and L170C to determine their relative mercury sensitivities.

Kinetics of Water Flux and Mercury Inhibition. Flux through the AQP channel was assayed in osmotically-driven liposome permeability experiments (Borgnia et al. 1999b). AqpZ is a water-selective channel with high rates of conduction and we first sought to determine the activity of WT, T183C, and L170C. Purified proteins were reconstituted into liposomes and kinetics assayed in a stopped-flow device by mixing proteoliposomes with a hyperosmolar reconstitution buffer to drive water efflux. The resulting proteoliposome shrinkage was measured by light scattering and the curve fit to an exponential equation with a single rate constant. Raw light scattering data are provided in Figure 3A, along with the fitted curves. We measured the rates of WT, T183C, and L170C to be 73.9 ± 0.4 (standard deviation) s^{-1} , $57.3 \pm 0.5 s^{-1}$ and $39.0 \pm 0.4 s^{-1}$ respectively, indicating that all proteins are functional water channels. Thus, WT has the fastest rates of conduction, T183C is slightly slower, and L170C much slower. Control liposomes without protein had rates of $4.6 \pm 0.1 s^{-1}$.

In order to ascertain the role of HgCl₂ as an inhibitor we determined the half maximal inhibitory concentration (IC₅₀) using a dose-response relationship. In this experiment proteoliposomes were incubated with various concentrations of HgCl₂ and then assayed for water conduction as described above. The results (Figure 3B) demonstrate that WT is less affected by mercury. The mutants, as expected, are inhibited at progressively lower concentrations with IC₅₀s of 345 μM, 84 μM, and 18 μM for WT, T183C and L170C respectively, (Borgnia et al. 1999b) and thus, L170C is the most sensitive to mercury. HgCl₂ did not affect control liposomes and mercury inhibition was reversible with the addition of 5 mM 2-mercaptoethanol (BME) following mercury incubation (data not shown).

Structure of mutant L170C blocked by mercury. Since L170C has a heightened sensitivity to mercury we determined the x-ray structure of the apo and complex forms. L170C was overexpressed, purified, and crystallized with and without mercury using HgCl₂ as an additive. Diffraction data were collected to 2.55 Å and 1.90 Å for the apo and complex proteins respectively, and the structures were solved by molecular replacement. As described in Materials and Methods, the model was built and refined iteratively to an R_{free} of 28.0% and 19.5% for the apo and complex forms respectively (Table 1). Like T183C, L170C displays the same AQP canonical fold, and also like T183C, shows very little structural difference between the apo and complex form. The superposition (RMSD 0.27 Å) is shown in Figure 4A.

As we predicted, introduction of a cysteine residue at position 170 increases the affinity at the T183C-Hg1 site, and we were able to locate four mercury atoms (named L170C-Hg1, L170C-Hg2, L170C-Hg3, and L170C-Hg4) near the NPA region between Cys170 and His174. These atoms are clustered together, as indicated in the channel pathway of Figure 4B and 4C. L170C-Hg2, L170C-Hg3, and L170C-Hg4 lie directly in the pore, while L170C-Hg1 is at the edge (Figure 4C). The occupancies of L170C-Hg1, L170C-Hg2, L170C-Hg3, and L170C-Hg4 were refined to 0.40, 0.23, 0.20, and 0.18 respectively. L170C-Hg2 is covalently bound to Cys170 at the distance expected for a sulfur-mercury bond length (2.6 Å). L170C-Hg2 is also at the proper Hg dinuclear complex distance from L170C-Hg1 (2.5 Å), L170C-Hg3 (2.3 Å), and L170C-Hg4 (2.4 Å) and there is continuous electron density between all mercury atoms at 1.5 σ in a $2F_o - F_c$ map. L170C-Hg2 may therefore mediate the binding of a second mercury in the pore at any of the three other positions. L170C-Hg3 lies directly in the pore at the same site as T183C-Hg1 (magenta double cross in Figure 4C) and may interact with the imidazole ring of H174 (3.8 Å) in a similar manner.

Discussion

Aquaporins and mercury inhibition. Even after the knowledge of the extreme toxicity of mercury and its various compounds, it continued to find use in disinfectants, cosmetics, and a suite of so-called medicines (Emsley 2001). Mercurials attack the reactive thiol moiety of cysteine found in nearly all proteins and are known to have a host of side effects including polyuria induced by AQP2 inhibition in the apical membrane of the kidney collection tubule (Zalups 2000). Due to its affinity for thiols, mercury has been useful in chemical probes of protein-mediated biological processes, as it was in defining the water channels as proteins. Mercury as a pore blocker has been instrumental in characterizing the AQP channel and revealing the role of AQPs in membrane transport for numerous tissues (Fushimi et al. 1993; Javot and Maurel 2002).

Steric inhibition by mercury. We located two mercury atoms in the T183C-mercury complex x-ray structure. These atoms, T183C-Hg1 and T183C-Hg2, are located in the pore and at an interstitial site respectively. Given that T183C is four times more sensitive to mercury than WT, we sought to establish the relevance of each site to this important functional result to determine the actual mechanism of inhibition.

The most obvious explanation for inhibition is that T183C-Hg1, due to its location, directly blocks the pore and inhibits through a steric mechanism. This is further supported by the fact that there is little structural change between the apo and mercury-complexed forms (Figure 2A). Lack of conformational change is surprising because

mercury binding requires the side-chain rearrangements necessary to accommodate both the binding of a mercury atom directly in the channel and one bound interstitially. This is particularly unexpected for mercury T183C-Hg2, which is bound to Glu138, a strictly conserved residue responsible for orienting the main-chain amides of residues 183-185. These amino acids are conformationally constrained to project their main chain carbonyls directly into the pore, but binding at this interstitial site seems to have no structural effect on this essential water-binding motif. The small ionic radius of Hg^{2+} , 1.10 Å, may explain the lack of structural perturbation and in the case of T183C-Hg2, its ability to bind deep within the protein. Thus, because there is no structural change to the protein, T183C-Hg1 most likely sterically blocks the pore.

There is also, however, an alternate mechanism. Occupancy refinement revealed that both T183C-Hg1 and T183C-Hg2 are present at significantly less than unity (0.24 and 0.32 respectively). Since the mercury sites are of low occupancy, they may occur (i) alternately in different channels, i.e they may be exclusive of one another, or (ii) they may reflect simple statistical occupancy at each site in an uncooperative manner. In either case, any changes in the protein around a partially occupied site would also be of low (~.3) occupancy, and so be difficult to refine as multiple conformers. This could be particularly important for T183C-Hg2, bound to the highly conserved Glu138.

Rearrangement of Glu138 could perturb the essential water-coordinating carbonyls and potentially disrupt water flux. However, inspection of $F_o - F_c$ difference maps reveals an increased noise level (Figures 6A and 6B), but no obvious alternate conformer. Also the potentially cleaner $F_o(\text{apo}) - F_o(\text{Hg})$ maps between observations per se were not possible

due to the degree of lack of isomorphism (Table 1). Thus, distinguishing between a steric mechanism and one involving a conformational change of the protein is difficult.

We therefore set out to demonstrate that the T183C-Hg1 site is sterically blocking the pore by creation of a new cysteine mutant. To differentiate between the steric and conformational mechanism, such a mutant should bind mercury at the pore site and not at the interstitial site. Furthermore, the lower occupancy of T183C-Hg1 and longer thiol-mercury distances (5.6 Å versus 4.0 Å) relative to T183C-Hg2 suggest the T183C-Hg1 interaction with Cys183 is not ideal. So, a structurally optimized mutant may also be more sensitive to mercury. In this region, the pore is formed mostly by main chain water-coordinating carbonyls and side chains from selectivity filter residues. Among the few side chains that project into the pore near T183C-Hg1, Leu170 is positioned closely to the site (Figure 4C). Thus, mutant L170C should be functionally sensitive to mercury and only bind mercury at the pore site.

L170C, with a cysteine at the proposed steric blockage site, is actually the most sensitive to mercury with respect to WT and T183C - its IC₅₀ is 20 times lower than WT.

Furthermore, this increase in sensitivity is also echoed in the structure. As with T183C, there is little evidence of conformational change upon mercury binding. Instead, there is complete occlusion of the pore. There are four mercury atoms in the pore clustered around the introduced cysteine. Importantly, there are no interstitially bound mercury atoms to disturb the water-coordinating carbonyls. These four atoms refine to occupancies of 0.40, 0.23, 0.20 and 0.18 for L170C-Hg1, L170C-Hg2, L170C-Hg3, and

L170C-Hg4 respectively. It is therefore possible these sites actually represent one statistically disordered mercury atom that is blocking the pore. Or, because of the low occupancy values, these sites may reflect mercury conformations that exist alternately in different blocked channels. In this case, based on bond lengths, L170C-Hg2 would be directly bound to Cys170 and mediate the binding of the other three mercury atoms. Therefore, we conclude that although there may be multiple conformations of the blocked channel, especially given the partial occupancies, the mechanism is a steric one. Furthermore, the increased mercury sensitivity of L170C and its more occluded channel at the same site as T183C-Hg1, indicates this site is most likely responsible for blockage.

Water Permeability of AqpZ reconstituted into liposomes. Using a proteoliposome-based assay we measured the water conduction rates and determined that both mutants were active at lower conductances than for WT. The decrease in conduction rates appears to correlate with the introduction of a more polar cysteine side chain into the hydrophobic pore. T183C shows a moderate decrease in rate and the cysteine rotamer points away from the pore, while L170C shows a much larger decrease and the side chain is in the pore (Figures 2C and 4C). AQPs most likely achieve high conduction rates by partially stabilizing the substrate, similar to an idea suggested for ion channels (Doyle et al. 1998). In this case water is stabilized as a line of molecules against eight carbonyls, the guanidinium group of Arg189, and N δ 2 moieties from the two conserved asparagines of the NPA motifs, in an otherwise hydrophobic channel. Introduction of a polar pore residue therefore increases channel water affinity and decreases flux.

Mercury and WT AqpZ. We unexpectedly discovered that WT AqpZ is also inhibited by HgCl₂. The WT protein contains two (Cys9 and Cys20) endogenous cysteines, which can presumably bind mercury but attempts at co-crystallizing WT with mercury to justify this were unsuccessful. Problems with co-crystallization suggests non-specific binding, so the pore itself may have some low affinity for mercury due to the presence of histidine and other polar side chains. We measured the IC₅₀ of a mutant in which both endogenous cysteines were mutated to serine to be roughly 1mM (data not shown). In light of this result it may be inappropriate to classify AQPs as either distinctively mercury sensitive or insensitive. The ionic radius of Hg²⁺ is even smaller than water (1.10 Å) and so regardless of the presence of a pore cysteine, it may become partially stabilized by the partial negative charge on the carbonyl oxygens and the imidazole ring of His174. Therefore, at higher (i.e. mM) mercury concentrations all AQPs may be inhibited. This also suggests there may be other charged inhibitors, possibly by the mechanism outlined above (Niemietz and Tyerman 2002; Yool et al. 2002).

Mercury and the tetrameric axis. A steric mechanism with no significant structural change also validates the use of mercury to distinguish between conduction through the tetramer axis and the monomer pore. As described above, the arrangement of the helical bundle monomer creates a four-fold symmetric channel running parallel to the monomer channel (Figure 4). This four-fold axis is hydrophobic, of larger dimensions than the monomer channel, and previous x-ray structures indicate the presence of some as yet undetermined molecules. Both experimental and computational studies have investigated possible substrates, including water, ions (Yu et al. 2006), and CO₂ (Cooper and Boron

1998). The mercury-bound structure of T183C, along with a simple steric blockage at the proposed site, suggests conduction studies using mercury are solely inhibiting the monomer channel.

Mercury dynamics in the pore. Mercury, due to its 80 electrons and aggressive thiol attacking nature, is one of the most common heavy atom derivatives for *de novo* phasing of x-ray crystal structures via isomorphous replacement and, to a lesser extent, anomalous diffraction (AD) (Blundell and Johnson 1976). The lack of successful mercury based AD experiments can be partially attributed to the success of other approaches, such as selenomethionine incorporation, but one emerging reason for the failure of mercury in AD experiments is the labile nature of the thiol-mercury bond under x-ray radiation exposure (Ji et al. 2001; Ramagopal et al. 2005).

It was our initial intent to solve the mercury-complex structures with unbiased experimental phases using either isomorphous replacement or AD methods, as this would allow unambiguous identification of mercury sites. Mercury bound crystals were not isomorphous to the native dataset (Table 1) so we adopted a multiwavelength AD strategy. We were unable to obtain interpretable experimental maps and subsequently solved the structure by molecular replacement. Mercury sites in both structures were located by a combination of anomalous difference and omit maps combined with thorough investigation of each putative site's chemical environment. During refinement of both mercury bound structures, it became clear there was significant motion in mercury atoms bound to the introduced cysteines, as anisotropic refinement of the

individual mercury sites, along with their occupancy, inevitably led to small (less than three electrons) “noisy” positive peaks in $F_o - F_c$ difference maps (Figure 6A and 6B). Based on their small size and proximity to adjacent Hg sites, we attribute these to alternate states of the protein as described above, and also partially to radiation-induced change of the mercury.

The difficulty in using mercury as an anomalous scatterer is due in part to the relative ease with which the mercury-thiol bond is cleaved, so we postulate that motion in the mercury peaks also reflects radiation damage induced cleavage of the mercury atoms' interaction with cysteine. A previous radiation damage study on the problems associated with using mercury has seen solvent exposed mercury atoms “escaping” over time, while those that are buried are more likely to remain bound to their respective sulfur moiety (Ramagopal et al. 2005). Using the L170C structure as a test case, we refine the mercury-bound structure against the data derived only from the first 45 frames (72.9% complete; scaled to the entire dataset) and the resulting omit map is shown in Figure 6C. A comparison with omit maps calculated from all of the data, shows there is indeed a change in the occupancies of sites. This change in occupancy can be seen as diffusion of mercury out of the solvent accessible channel while the more deeply buried L170C-Hg1 shows little motion.

Finally, it is interesting to hypothesize what effect mercury dynamics may have on inhibition. The unrealistic mercury-thiol bond lengths in the T183C structure can be explained by the dynamic nature of mercury during radiation exposure, but also by

“unoptimized” mercury binding sites. Mercury binding to L170C is clearly more favorable, but the unusual electron density seen in Figures 6B and 6C suggests that mercury is still quite dynamic and heavily influenced by the local chemical environment (i.e. the solvent filled channel). These observations hint that dynamics may influence the way mercury inhibits the channel and we propose such questions can be answered by higher-resolution structures allowing more accurate refinement of occupancies and distances and functional analysis of new mutants (e.g. double cysteine mutants). Furthermore, it will also be important to structurally characterize the role of larger mercurials such as p-chloromercuriphenylsulfonic acid, which may have a more complex inhibition mode. In combination, these experiments will elucidate the more subtle aspects of the simple steric inhibition mechanism we have defined here.

Materials and Methods

Expression and Purification. Mutants of AqpZ were generated by site-directed-mutagenesis of the pET28b-AqpZ construct used in the original structure paper. (Savage et al. 2003) Prior to mutagenesis, endogenous cysteines were removed via the mutations Cys9Ser and Cys20Ser. The *E. coli* strain C43(DE3) was transformed, grown to 0.6-1 OD at 600 nm at 37°C in 2 x LB media, 0.5% glycerol (v/v), 1x M9 salts, and 25 mg / L kanamycin, and induced with 1 mM isopropyl -D-thiogalactoside (Anatrace) (Miroux and Walker 1996; Mohanty and Wiener 2004).

All purification was carried out at 4°C or on ice as necessary. Cells from six L of culture were harvested and lysed by a microfluidizer in 20 mM Tris pH 7.4, 100 mM NaCl, .5 mM phenylmethylsulfonyl fluoride, and 5 mM BME. Membranes were recovered from supernatant by 100,000 x g centrifugation for 2 hours. AqpZ was solubilized from membranes by agitation in 20 mM Tris pH 7.4, 100 mM NaCl, 5 mM BME, 10% glycerol (v/v), and 270 mM n-Octyl-β-D-glucopyranoside (OG) (Anatrace) for 12-16 hours at 4°C. Solubilized protein was bound in batch to Ni-NTA resin (Qiagen) for 1 hour, washed with 25 resin volumes of 20mM Tris pH 7.4, 100 mM NaCl, 5 mM BME, 10% glycerol, 40 mM OG, and 20 mM imidazole, and eluted with 20 mM Tris pH 7.4, 100 mM NaCl, 5 mM BME, 10% glycerol, 40 mM OG, and 250 mM imidazole. Imidazole was removed using a Biorad Econo-Pac 10DG desalting column and the 6xHis tag was removed by digestion with 5 µg of trypsin for 12 hours at 4°C. Trypsin was removed by passing over a benzamidine-sepharose matrix (GE Healthcare), and the protein sample was injected onto a Pharmacia Superose 12 gel filtration column running

a mobile phase of 20 mM Tris pH 7.4, 100 mM NaCl, 2 mM dithiothreitol, 10% glycerol, and 40 mM OG. Except as noted, all materials were purchased from Sigma or Fisher.

The sample was judged pure and homogeneous by both gel filtration chromatography and coomassie-stained denaturing gels. Final yields were approximately ten, seven, and three mg of protein / L culture for WT, T183C, and L170C respectively.

Crystallization and Data Collection. Following gel filtration chromatography, the protein was concentrated to 25 mg / ml using a 30 kDa cutoff Amicon Ultra-15 Centrifugal Filter. Crystals were grown by hanging drop vapor diffusion at room temperature by 1:1 addition of protein and 25-30% polyethylene glycol (PEG) monomethyl ether 2000 (Fluka), 100 mM sodium cacodylate pH 6.5, and 50-100 mM MgCl₂. For co-crystallization, divalent mercury in the form of 1-3 mM HgCl₂ was added in batch to the protein sample before mixing with precipitant. In general, the best co-crystals were obtained at slightly lower (2-4%) PEG concentrations than in apo crystallization. Crystals grew to roughly 300 μm x 300 μm x 150 μm over the course of several days and were flash frozen in liquid nitrogen following a brief washing in the mother liquor plus 15% glycerol for cryoprotection. Diffraction intensities were collected on Advanced Light Source Beamline 8.3.1 using an ADSC Quantum-Q210 CCD detector. Prior to data collection on mercury-complex crystals, fluorescence energy scans of the L_{III} mercury edge were taken to verify mercury substitution.

Phasing and Model Refinement. Data were processed with Elves(Holton and Alber 2004) and CCP4(1994) (using MOSFLM(Leslie 2006)) and the structures were solved by

molecular replacement with the published WT structure (Protein Data Bank Code 1RC2) using Phaser (Read 2001). The models were refined with iterative cycles of manual building with Coot (Emsley and Cowtan 2004) and restrained refinement with individual B-factor refinement in Refmac5 (Murshudov et al. 1997). After several initial rounds of refinement, mercury atoms were located unambiguously by inspecting anomalous difference maps, $F_o - F_c$ omit maps (peaks of 15-7 σ), and the local chemical environment. Following refinement, the appearance of negative density in $F_o - F_c$ difference maps indicated that occupancy for mercury atoms was not unity. Occupancies for mercury atoms, along with their anisotropic B-factors were refined using least-squares refinement in SHELXL (Sheldrick and Schneider 1997).

Proteoliposome Reconstitution. Before removing the 6xHis tag with trypsin, aliquots of protein were set aside for proteoliposome reconstitution. *E. coli* polar lipids were sonicated to clarity and the reconstitution cocktail was prepared by sequentially adding 100 mM MOPS pH 7.5, 51.3 mM OG (1.5%, w/v), 50 $\mu\text{g} / \text{ml}$ of purified protein, and 10 mg / ml *E. coli* polar lipids (Avanti). (Borgnia and Agre 2001) To reduce oxidation, lipid stocks were stored in 2 mM BME and all buffers were under argon atmosphere. Following cocktail incubation for one hour at room temperature (RT), proteoliposomes were formed by diluting the mixture 50-fold into a running buffer of 20 mM Hepes pH 7.5 and harvested by centrifugation at 100,000 x g for two hours. Pelleted liposomes were resuspended into one ml of running buffer (20 mM Hepes pH 7.5) and stored at 4°C. Liposome monodispersity was verified by dynamic light scattering with a mean diameter of 90 nm.

To analyze the kinetics of water conduction through the channel, we subjected the proteoliposomes to an osmotic gradient by mixing 1:1 proteoliposomes (final AqpZ monomer concentration of 0.27 μM) and running buffer with osmolyte (20 mM Hepes pH 7.5, 570 mM sucrose) and measured water efflux (liposome shrinkage) by light scattering in a stopped-flow device at 440 nm. Resulting curves were fit to a single-exponential rate constant (k_1) as a measure of conduction to use in comparison between mutants and with inhibitor. Inhibition of the mutants was accomplished by incubating the resuspended proteoliposomes with the appropriate concentration of HgCl_2 for one hour at RT prior to stopped-flow analysis.

Abbreviations:

aquaporin 1, AQP1; aquaporin Z AqpZ; transmembrane, TM; glycerol facilitator, GlpF; wild-type, WT; half maximal inhibitory concentration, IC50; root mean square deviation, RMSD; 2-mercaptoethanol, BME; anomalous diffraction, AD; room temperature, RT

Table 1: Crystallographic Data and Refinement Statistics

	T183C	T183C- Mercury	L170C	L170C- Mercury
Data Collection				
Space Group	<i>P4</i>	<i>I4</i>	<i>P4</i>	<i>I4</i>
Unit Cell				
a (Å)	92.4	91.1	91.3	91.2
c (Å)	78.2	77.9	77.5	77.1
Resolution range (Å) ^a	50-2.30 (2.30-2.36)	50-2.20 (2.20-2.26)	50-2.55 (2.55-2.62)	50-1.90 (19.0-1.95)
Unique reflections	28897	16199	15206	23096
Completeness ^a	98.3 (93.8)	99.8 (100.0)	74.1 (65.5)	92.8 (66.4)
R _{sym} ^b (%) ^a	7.3 (59.5)	8.8 (59.5)	7.5 (46.2)	6.4 (33.4)
I / σ(I) ^a	15.3 (1.1)	15.5 (1.9)	10.3 (1.2)	15.4 (1.7)
Refinement Statistics				
R _{work} / R _{free} (%)	19.7 / 23.8	19.3 / 24.2	23.6 / 28.0	16.6 / 19.5
RMSD bonds (Å) ^c	0.016	0.015	0.012	0.014
RMSD angles (°) ^c	1.65	1.64	1.45	1.61
Number of protein atoms	3356	1671	3368	1696
Number of solvent atoms	109	68	43	139
Number of Hg	0	3	0	4

Average B-factor (Å ²)	37.7	36.4	24.9	23.2
PDB code	209D	209E	209F	209G

^a values in parenthesis refer to the highest-resolution shell

^b $\frac{\sum |I - \langle I \rangle|}{\sum I}$, where I equals observed intensity and $\langle I \rangle$ equals average intensity for symmetry-related reflections

^c Root-mean-square deviation of bond lengths and angles from ideal values

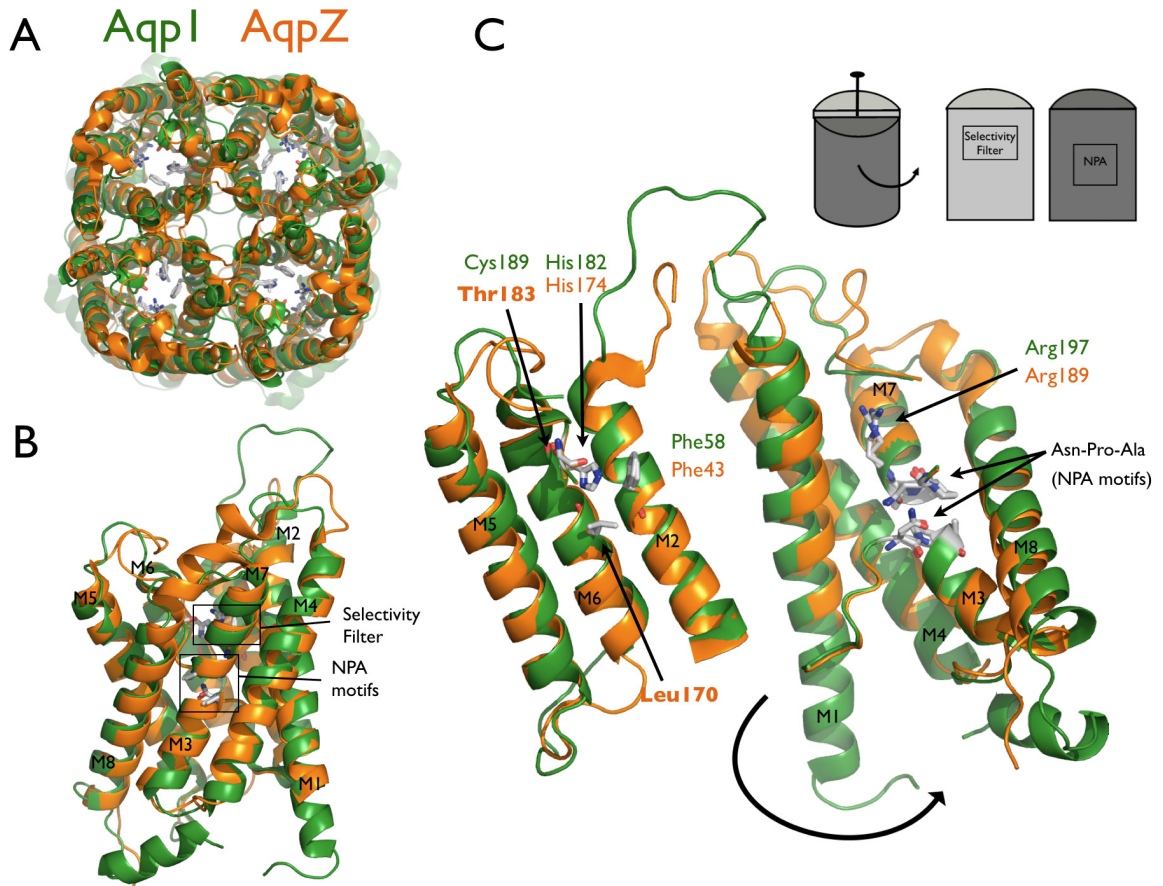


Figure 1. AqpZ is the bacterial homolog of AQP1. (A) Cartoon representation of the AqpZ (orange) and AQP1 (green) tetramers. Note the presence of the four monomer channels and the hypothetical channel down the tetramer axis. (B) Cartoon representation of the AqpZ and AQP1 monomers. Helices are labeled M1 through M8, and the selectivity filter and NPA motifs are designated with boxes. (C) Monomer opened up showing conservation of the water-selective motif. In this cartoon representation, the monomer is peeled open as shown in the inset schematic. The conserved selectivity filter and NPA motif are shown in sticks. Thr183 and Leu170 in AqpZ are the positions of cysteine mutants in this study. All molecular structure figures were made in Pymol (Delano Scientific).

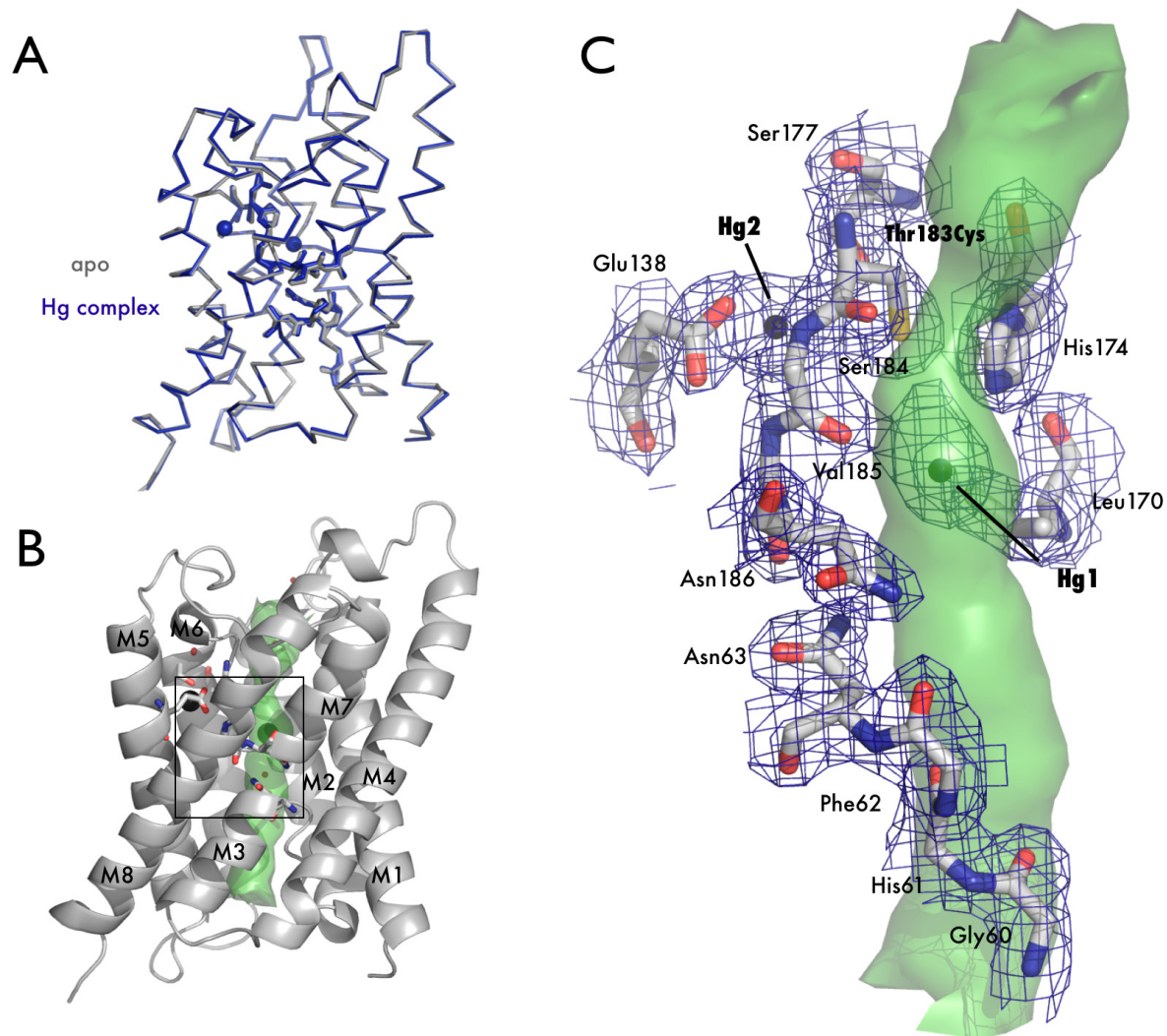


Figure 2. Crystal structure of apo T183C and mercury bound T183C mutants. (A) Main chain overlay of the apo (gray) and Hg-complex (blue) with an RMSD ($C\alpha$) of 0.27 Å. Bound Hg^{2+} atoms are displayed as spheres with a van der Waals radius of 1.10 Å. (B) Cartoon representation of T183C. Transmembrane helices are labeled M1-M8 and the interior surface of the channel is drawn as a green surface. The black square denotes the area of interest depicted in panel C. (C) Structure of the blocked channel. Amino acids involved with water binding in AQPs are shown as sticks and with $2F_o - F_c$ electron density

mapped contoured at 1.2σ drawn in blue. Mercury atoms are shown as spheres. In this orientation it can be seen that T183C-Hg1 sterically blocks the pore (green surface).

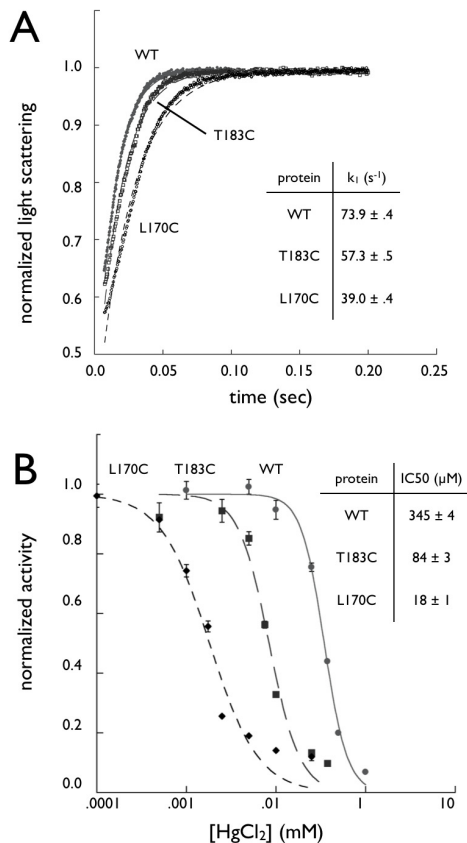


Figure 3. Kinetic Studies of Aquaporin Z. (A) Water conduction of WT, T170C, and T183C. Proteins were reconstituted in liposomes, challenged with a higher osmotic gradient in a stopped-flow device, and liposome shrinkage measured by light scattering at 440 nm. Plots were fit to a single exponential and the resulting rates are shown in the inset table. (B) Dose-response curve of proteoliposomes incubated with $HgCl_2$. After incubation with $HgCl_2$ proteoliposomes were assayed as in panel A and the rates were fit to a sigmoid dose-response curve in Kaleidagraph (Synergy Software). IC50 values are shown in the inset table.

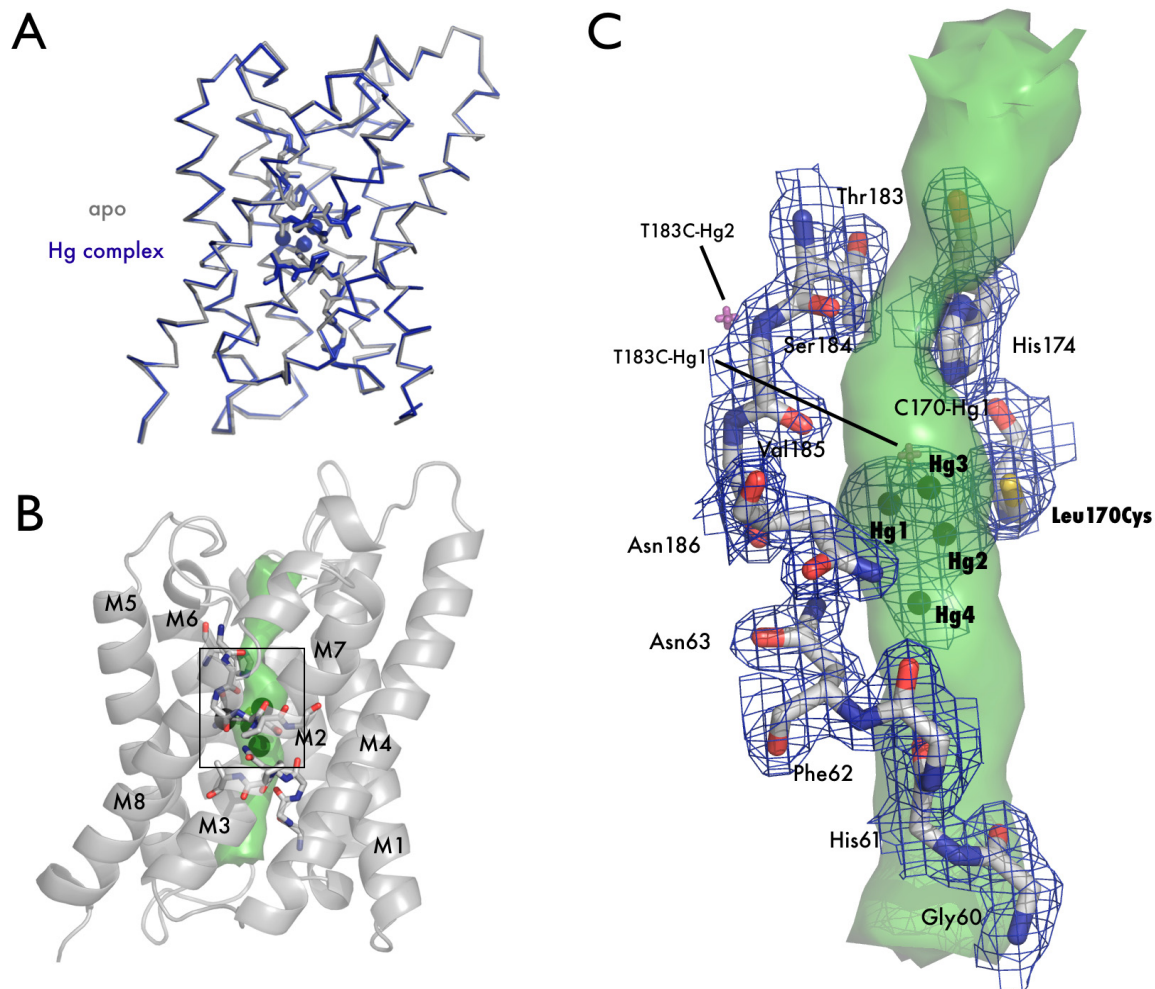


Figure 4. Crystal structure of apo L170C and mercury bound L170C. (A) Main chain overlay of the apo (gray) and Hg-complex (blue) with an RMSD of 0.27 Å. Bound mercury atoms are displayed as spheres with a van der Waals radius of 1.10 Å. (B) Cartoon representation of L170C. Transmembrane helices are labeled M1-M8 and the interior surface of the channel is drawn as a green surface. The black square denotes the area of interest depicted in panel C. (C) Structure of the blocked channel. Amino acids classically involved with water binding in AQPs are shown as sticks and with $2F_o - F_c$ electron density mapped contoured at 1.2σ drawn in blue. Mercury are shown as

spheres. Superposition of mercury atoms from the T183C structure are shown as magenta crosses. In this orientation it can be seen that all four mercury atoms sterically block the pore (green surface).

T183C with mercury
apo

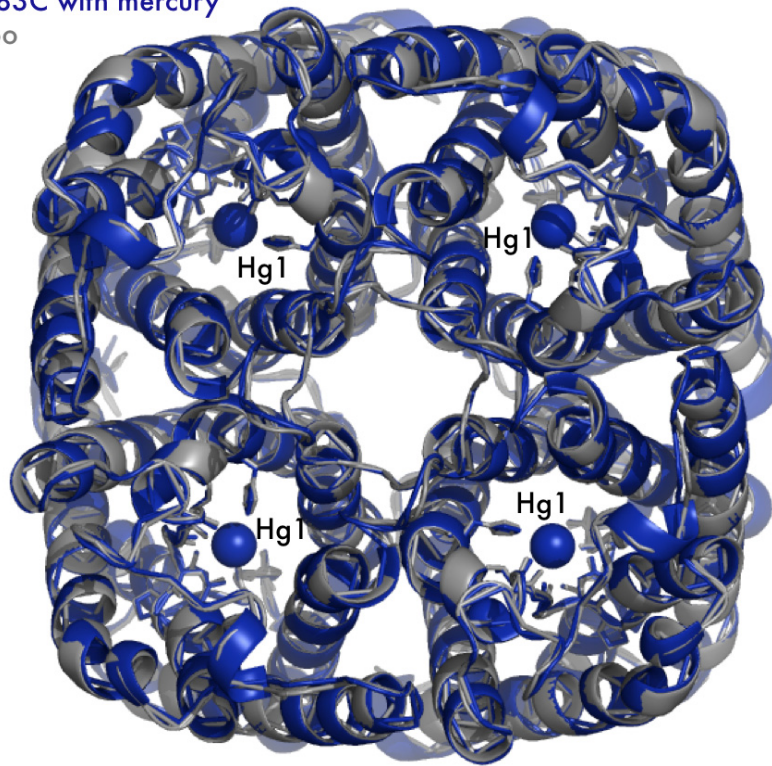


Figure 5. Mercury blocks the monomer channel. While AQPs are tetramers in the membrane, the monomer is the functional unit. By imposing crystal symmetry on both the apo (grey) and complex structures (blue), T183C is drawn as a tetramer in cartoon representation. Mercury-Hg1, with its proper van der Waals radius, is drawn as a sphere blocking the channel. Note there is almost no structural change to the tetrameric axis.

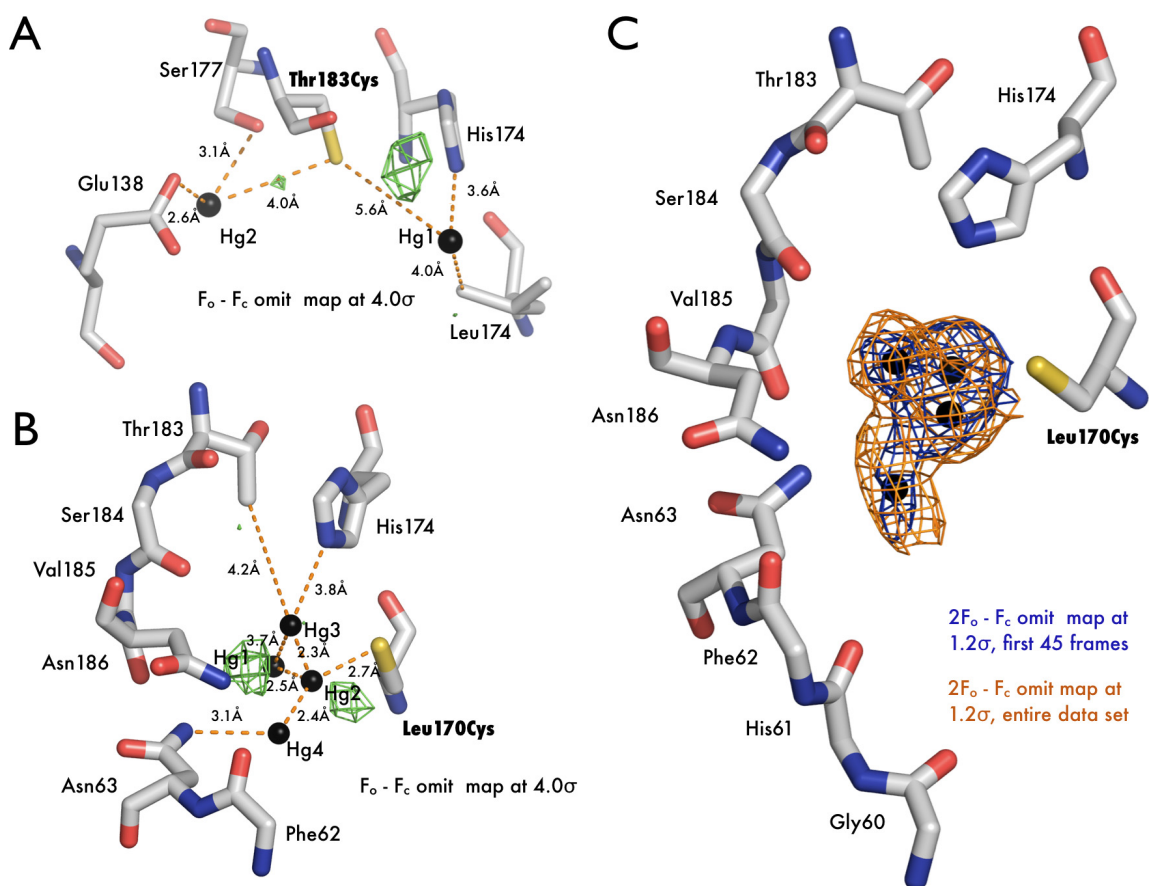


Figure 6. Mercury disorder in electron density maps. (A) $F_o - F_c$ electron density map (green) of mercury bound T183C structure contoured at 4σ . (B) $F_o - F_c$ electron density map (green) of mercury bound L170C structure contoured at 4σ . (C) $2F_o - F_c$ omit electron density map solved with the first 45 frames of data (blue) and the entire dataset (orange). Both maps are contoured at 1.2σ around the three mercury atoms.

References

1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**: 760-763.
- Blundell, T.L., and Johnson, L.N. 1976. *Protein crystallography*. Academic Press, New York, pp. xiv, 565 p.
- Borgnia, M., Nielsen, S., Engel, A., and Agre, P. 1999a. Cellular and molecular biology of the aquaporin water channels. *Annu Rev Biochem* **68**: 425-458.
- Borgnia, M.J., and Agre, P. 2001. Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from Escherichia coli. *Proc Natl Acad Sci U S A* **98**: 2888-2893.
- Borgnia, M.J., Kozono, D., Calamita, G., Maloney, P.C., and Agre, P. 1999b. Functional reconstitution and characterization of AqpZ, the E. coli water channel protein. *J Mol Biol* **291**: 1169-1179.
- Calamita, G., Bishai, W.R., Preston, G.M., Guggino, W.B., and Agre, P. 1995. Molecular cloning and characterization of AqpZ, a water channel from Escherichia coli. *J Biol Chem* **270**: 29063-29066.
- Cooper, G.J., and Boron, W.F. 1998. Effect of PCMBs on CO₂ permeability of Xenopus oocytes expressing aquaporin 1 or its C189S mutant. *Am J Physiol* **275**: C1481-1486.
- Doyle, D.A., Morais Cabral, J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., and MacKinnon, R. 1998. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**: 69-77.

- Emsley, J. 2001. *Nature's building blocks : an A-Z guide to the elements*. Oxford University Press, Oxford ; New York, pp. viii, 538 p.
- Emsley, P., and Cowtan, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126-2132.
- Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J., and Stroud, R.M. 2000. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**: 481-486.
- Fushimi, K., Uchida, S., Hara, Y., Hirata, Y., Marumo, F., and Sasaki, S. 1993. Cloning and expression of apical membrane water channel of rat kidney collecting tubule. *Nature* **361**: 549-552.
- Harries, W.E., Akhavan, D., Miercke, L.J., Khademi, S., and Stroud, R.M. 2004. The channel architecture of aquaporin 0 at a 2.2-Å resolution. *Proc Natl Acad Sci U S A* **101**: 14045-14050.
- Hille, B. 2001. *Ion channels of excitable membranes*, 3rd ed. Sinauer, Sunderland, Mass., pp. xviii, 814 p.
- Holton, J., and Alber, T. 2004. Automated protein crystal structure determination using ELVES. *Proc Natl Acad Sci U S A* **101**: 1537-1542.
- Javot, H., and Maurel, C. 2002. The role of aquaporins in root water uptake. *Ann Bot (Lond)* **90**: 301-313.
- Ji, X., Blaszczyk, J., and Chen, X. 2001. The absorption edge of protein-bound mercury and a double-edge strategy for HgMAD data acquisition. *Acta Crystallogr D Biol Crystallogr* **57**: 1003-1007.

- Kuang, K., Haller, J.F., Shi, G., Kang, F., Cheung, M., Iserovich, P., and Fischbarg, J. 2001. Mercurial sensitivity of aquaporin 1 endofacial loop B residues. *Protein Sci* **10**: 1627-1634.
- Lee, J.K., Kozono, D., Remis, J., Kitagawa, Y., Agre, P., and Stroud, R.M. 2005. Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 Å. *Proc Natl Acad Sci U S A* **102**: 18932-18937.
- Leslie, A.G. 2006. The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**: 48-57.
- Macey, R.I. 1984. Transport of water and urea in red blood cells. *Am J Physiol* **246**: C195-203.
- Miroux, B., and Walker, J.E. 1996. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**: 289-298.
- Mohanty, A.K., and Wiener, M.C. 2004. Membrane protein expression and production: effects of polyhistidine tag length and position. *Protein Expr Purif* **33**: 311-325.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255.
- Niemietz, C.M., and Tyerman, S.D. 2002. New potent inhibitors of aquaporins: silver and gold compounds inhibit aquaporins of plant and human origin. *FEBS Lett* **531**: 443-447.
- Park, J.H., and Saier, M.H., Jr. 1996. Phylogenetic characterization of the MIP family of transmembrane channel proteins. *J Membr Biol* **153**: 171-180.

- Preston, G.M., Carroll, T.P., Guggino, W.B., and Agre, P. 1992. Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein. *Science* **256**: 385-387.
- Preston, G.M., Jung, J.S., Guggino, W.B., and Agre, P. 1993. The mercury-sensitive residue at cysteine 189 in the CHIP28 water channel. *J Biol Chem* **268**: 17-20.
- Ramagopal, U.A., Dauter, Z., Thirumuruhan, R., Fedorov, E., and Almo, S.C. 2005. Radiation-induced site-specific damage of mercury derivatives: phasing and implications. *Acta Crystallogr D Biol Crystallogr* **61**: 1289-1298.
- Read, R.J. 2001. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr D Biol Crystallogr* **57**: 1373-1382.
- Savage, D.F., Egea, P.F., Robles-Colmenares, Y., O'Connell, J.D., 3rd, and Stroud, R.M. 2003. Architecture and selectivity in aquaporins: 2.5 Å X-ray structure of aquaporin Z. *PLoS Biol* **1**: E72.
- Sheldrick, G.M., and Schneider, T.R. 1997. SHELXL: High Resolution Refinement. *Methods in Enzymology* **277**: 319-343.
- Smart, O.S., Goodfellow, J.M., and Wallace, B.A. 1993. The pore dimensions of gramicidin A. *Biophys J* **65**: 2455-2460.
- Sui, H., Han, B.G., Lee, J.K., Walian, P., and Jap, B.K. 2001. Structural basis of water-specific transport through the AQP1 water channel. *Nature* **414**: 872-878.
- Tornroth-Horsefield, S., Wang, Y., Hedfalk, K., Johanson, U., Karlsson, M., Tajkhorshid, E., Neutze, R., and Kjellbom, P. 2006. Structural mechanism of plant aquaporin gating. *Nature* **439**: 688-694.

- Walz, T., Hirai, T., Murata, K., Heymann, J.B., Mitsuoka, K., Fujiyoshi, Y., Smith, B.L., Agre, P., and Engel, A. 1997. The three-dimensional structure of aquaporin-1. *Nature* **387**: 624-627.
- Yool, A.J., Brokl, O.H., Pannabecker, T.L., Dantzer, W.H., and Stamer, W.D. 2002. Tetraethylammonium block of water flux in Aquaporin-1 channels expressed in kidney thin limbs of Henle's loop and a kidney-derived cell line. *BMC Physiol* **2**: 4.
- Yu, J., Yool, A.J., Schulten, K., and Tajkhorshid, E. 2006. Mechanism of gating and ion conductivity of a possible tetrameric pore in aquaporin-1. *Structure* **14**: 1411-1423.
- Zalups, R.K. 2000. Molecular interactions with mercury in the kidney. *Pharmacol Rev* **52**: 113-143.
- Zeidel, M.L., Ambudkar, S.V., Smith, B.L., and Agre, P. 1992. Reconstitution of functional water channels in liposomes containing purified red cell CHIP28 protein. *Biochemistry* **31**: 7436-7440.

Chapter 5

Cell-free complements *in vivo* expression of the *E. coli* membrane proteome

Research completed in collaboration with

Corey L. Anderson, Yaneth Robles-Colmenares, Zach Newby, and Robert M. Stroud

This chapter is currently in submission to Protein Science

Abstract

Reconstituted cell-free (CF) protein expression systems hold the promise of overcoming the traditional barriers associated with *in vivo* systems. This is particularly true for membrane proteins, which are often cytotoxic and due to the nature of the membrane, difficult to work with. To evaluate the potential of cell-free expression, we cloned 120 membrane proteins from *E. coli* and compared their expression profiles in both an *E. coli in vivo* system and an *E. coli* derived cell-free system. Our results indicate CF is a more robust system and we were able to express 63% of the targets in CF, compared to 44% *in vivo*. To benchmark the quality of CF produced protein, five target membrane proteins were purified and their homogeneity assayed by gel filtration chromatography. Finally, to demonstrate the ease of amino acid labeling with CF, a novel membrane protein was substituted with selenomethionine, purified, and shown to have 100% incorporation of the unnatural amino acid. We conclude that CF is a novel, robust expression system capable of expressing more proteins than an *in vivo* system and suitable for production of membrane proteins at the milligram level.

Introduction

Integral membrane proteins (MPs), despite their biological importance, currently account for less than 1% of all known high resolution protein structures. MPs are notoriously difficult to work with and expression, detergent solubilization, purification, and crystallization all present unique challenges over their soluble counterparts (White 2004). MPs generally express at much lower levels than soluble proteins and when *in vivo* overexpression is successful, the protein can be cytotoxic or incorporated into insoluble inclusion bodies. Following successful MP expression, a suitable detergent condition must also be found that simultaneously extracts the protein from the membrane while retaining the native fold and function. This protein-detergent-complex (PDC) is often heterogeneous, creating numerous problems in purification and crystallization. Optimizing purification, assaying protein function, and crystallization all require milligram quantities of protein, and MP expression is therefore a limiting step in macromolecular structure determination (Dobrovetsky et al. 2005) (Eshaghi et al. 2005) (Korepanova et al. 2005) (Surade et al. 2006) (Columbus et al. 2006). One recognized alternative is cell-free (CF) expression (Klammt et al. 2004).

CF expression systems are reconstituted reactions based on cellular extracts that recapitulate the expression (i.e. transcription and translation) capabilities of a cell *in vitro*. CF has a lengthy history of small-scale studies (Zubay 1973), but there has only recently been a concerted effort to adapt these systems to larger (i.e. milligram) scales (Spirin et al. 1988). These adaptations include coupled transcription and translation, enzymatic subsystems to regenerate high-energy nucleotides, and continuous-exchange “feeding”

systems via dialysis, all of which allow the reaction to proceed at higher levels for a longer amount of time. Successful CF systems, both prokaryotic (Kigawa et al. 1999) (Klammt et al. 2004) and eukaryotic (Endo and Sawasaki 2003) (Tyler et al. 2005), have been described. Furthermore, as a testament to its success and promise, there are now a handful of nuclear magnetic resonance (NMR) (Koglin et al. 2006) and x-ray (Pornillos et al. 2005) structural studies of CF expressed MPs. Despite this potential, there are no in-depth studies comparing the use of CF and *in vivo* systems in expressing MPs.

Given the growing evidence for CF expression as a viable approach for producing MPs for structural studies, we decided to compare the *in vitro* and CF expression profiles of over 100 *E.coli* MPs. We analyzed the solubility properties of the successfully expressing proteins in two commonly used detergents, and purified five of these targets to homogeneity. Finally, to demonstrate the extensibility of our system and flexibility in easily labeling proteins, we successfully incorporate selenomethionine into one purified target. Here, we report the first large-scale attempt to compare the success rate of CF and *in vivo* systems in expressing MPs and demonstrate the value of CF as a robust complement to current *in vivo* methods.

Results

Target selection and cloning. Table 1 shows the 120 *E.coli* MPs selected for cloning.

These targets were selected based on potential for successful expression and structural impact. To achieve this, we selected *E.coli* MPs that are less than 30kDa, possess at least two transmembrane (TM) spanning helices, and if functionally annotated, not part of complex. 60% of the proteins are described as hypothetical membrane proteins. For a positive control, several proteins with known crystal structures and three with a single TM were included. Of the 120 genes targets, 117 and 116 were successfully cloned into the *in vivo* and CF expression vectors, respectively.

Cell-free and in vivo protein expression results. MP expression levels were determined using 2ml *E.coli* C43 (Miroux and Walker 1996) growths or 30 μ L CF reactions in batch. Several proteins on the list (AqpZ, GlpF, YidJ, and CcmG among others), which express at known levels, were used as positive controls to gauge the expression levels of the other proteins as noted in Table 1. By comparing band intensity on Western blots, we were able to qualitatively assign expression levels. A (-) indicates no protein expression detected on a Western blot. *In vivo* expressed proteins that had levels less than roughly 2 mg/L were assigned a (+), and all proteins higher were designated (++). Similar criteria were applied to CF expressed proteins; the (+) limit was roughly 200 μ g/ml and those proteins expressing higher were designated (++).

Expression profiles of each target are shown in Table 1 along with each protein's SwissProt ID, molecular weight (MW), predicted number of TM helices, and function.

The most striking result is the number of proteins successfully expressed in either system, summarized by a Venn diagram in Figure 1. Of the original 120 proteins, a total of 90 (75%) were expressed. 36 (30%) could be expressed in both systems, 38 (32%) in CF only, and 16 (13%) in *in vivo* only. Overall, 63% of proteins expressed in CF, while only 44% of proteins expressed *in vivo*. Thus, we can express the majority of *E. coli* MPs selected for this study, and combined use of the two systems results in increased coverage of “expression space” (Surade et al. 2006). Given the large number of successful expressers, we next sought to characterize the detergent solubility of these proteins and their potential for purification

Solubility of in vivo and cell-free produced proteins in DDM and OG. In general, a detergent must stabilize the hydrophobic nature of a MP without disrupting its native state or function. Further biochemistry, including solubilization, purification, and crystallization, therefore requires identifying the optimal detergent conditions for a given MP. With structure determination as an end goal, we focused on two mild nonionic detergents, which have the best history of success in x-ray crystallography, (summarized by Hartmut Michel’s MP structure database, <http://www.mpibp-frankfurt.mpg.de/michel/public/memprotstruct.html>) n-Octyl- β -D-glucopyranoside (OG) and n-dodecyl- β -D-maltopyranoside (DDM)

Solubilization efficiency in 270mM OG or 10mM DDM was assessed by western analysis of fractions from before and after a high-speed centrifugation step to pellet unsolubilized membrane material (see Materials and Methods). Two examples are

illustrated in Figure 2 where YiaA solubility in DDM is high and YagU solubility in DDM is low, and for all 120 we assigned them as not soluble (-), low solubility (+), and high solubility (++). In our nomenclature, (++) is essentially quantitative extraction from the membrane fraction. Table 2A summarizes the results obtained for the *in vivo* subset of proteins. Surprisingly, all were soluble in either OG or DDM to some level and the majority could be quantitatively solubilized in either one or both detergents.

Table 2B shows several of the CF produced proteins and their solubility in DDM or OG. 11 of the 15 proteins exhibited some level of solubility and roughly half could be quantitatively solubilized. This lesser result may be explained by the lack of proper translocation machinery in the CF reaction. These reactions are not supplemented with the addition of exogenous lipids or detergents and it is thought that expressed MPs form a non-traditional “precipitate” which can be solubilized after the reaction (Klammt et al. 2005). This hypothesis is supported by the fact that there is near 100% agreement between DDM and OG in the CF trials – the well-behaved MPs in non-traditional precipitates can generally be solubilized. As this calls into question the nature of the PDC for a successfully solubilized MP from CF, we next purified a subset of targets to verify homogeneity.

Protein purification of cell-free produced proteins. Given that many MPs have an unknown function, and assays for those with a known function are often complex, one must employ other methods to benchmark the quality of our CF produced proteins (Columbus et al. 2006). One such indicator of purity, homogeneity, and stability, is the

gel filtration chromatography profile, which can give some estimation of size and monodispersity. It has been our experience that a protein which elutes as a single symmetric peak on a gel filtration column and is stable over time, is correctly folded and functional. Furthermore, most of the MP targets in our laboratory, including colicin Ia, GlpF, AqpZ, AmtB, Aqp0, and AqpM among others, with this property were well behaved (i.e. crystallizable) in structural studies.

To verify the behavior of CF expressed proteins we chose five proteins that expressed well in CF and were soluble in OG. We purified the proteins using Ni²⁺ based immobilized metal affinity chromatography (IMAC) to show that they bound IMAC resin and eluted as a single, pure, sample. Figure 3A shows a coomassie-stained gel of DgkA, YidG, YijD, PsiE, and YiaA all having a single band at the correct molecular weight (MW). Eluted samples were then run on a gel filtration column to verify homogeneity (Figure 3B). Four of the five proteins display a well-resolved single peak, indicating a homogenous sample. Such a profile, in the absence of a more rigorous solution scattering analysis, is a good estimate of monodispersity and thus, the majority of targets in CF can be solubilized successfully. This, as demonstrated in other MP CF studies (Berrier et al. 2004; Elbaz et al. 2004), indicates that our expressed proteins (four of five) are well behaved and similar to their *in vivo* counterparts.

Selenomethionine labeling of CcmG with cell-free. One advantage of CF is its extensibility, which for example, makes amino acid labeling trivial. Labeling, something very important for structural studies, can be done simply replacing the natural amino acid

with the unnatural in the reaction mixture (Yokoyama 2003). This approach has shown promise in isotopically labeling proteins for NMR (Kigawa et al. 1995) (Koglin et al. 2006) (Kainosho et al. 2006), and in x-ray crystallography experiments (Kigawa et al. 2002) where the anomalous diffraction (AD) properties of selenomethionine (SeMet) can be used to solve the phase problem. To show the ease of labeling with CF, we set out to fully substitute a protein with SeMet.

CcmG is a thioredoxin protein with one TM helix involved in the maturation of cytochrome c (Thony-Meyer 2002). We have crystallized and collected native diffraction data on *in vivo* expressed CcmG but required unbiased experimental phases (data unpublished), and so decided to try an AD approach via SeMet labeling. CcmG was expressed in a large-scale CF reaction in which methionine had been replaced with L-selenomethionine. Even though we previously found CcmG to be only a (+) expresser in the CF system, we were still able to obtain 5mg of purified selenomethionine-labeled protein from a 22ml reaction. In order to verify labeling, we used matrix-assisted laser desorption/ionization (MALDI) mass spectroscopy (MS) on the natural *in vivo* expressed protein and compared it with the labeled CF expressed protein. Figure 4 shows an overlay of the two MS peaks from native and labeled CcmG, and a coomassie-stained gel of the two proteins. Given that there are 3 methionines in CcmG, we expected to see a shift of 141Da. The observed shift of 131Da is well within the error expected for MALDI and indicates 100% incorporation. Thus, CF can be conveniently used to express labeled MPs in the amounts necessary for structural purposes.

Discussion

Cell-free complements in vivo expression. Our study examines the potential of CF expression as a method for producing MPs for structural analysis. We compared the expression profiles of targets expressed with CF and *in vivo* methods and show the two can express 63% and 44% of the targets respectively (Figure 1). Thus, CF is a more robust system for expressing proteins. This result is interesting given that the CF system is ostensibly identical to the cell - it uses the same transcriptional, translational, and translocation machinery. However, the expression of MPs is often cytotoxic, and since a living cell must balance protein expression with its own viability, CF benefits from its reconstituted nature. It is also striking that the two systems can cumulatively express 75% of all targets. Lack of expression is the first bottleneck in structural studies of MPs and obtaining 75% of the desired targets is in line with the best current published reports (Eshaghi et al. 2005; Surade et al. 2006). Finally, 30% (36 out of 120) of the targets could be expressed in either system suggesting there is significant overlap in the abilities of the two systems to express targets. Given the union and intersection of these two large expression profile sets it is therefore important to evaluate the two expression systems.

When evaluating an expression system, many factors, such as cost, complexity, throughput, and protein quality, must be taken into account. In a structural biology setting, where the fundamental currency is high-resolution crystal diffraction data or well-resolved NMR spectra, the equation becomes pure, homogenous, and stable protein for the least amount of input. In this respect *in vivo* expression has many advantages. The protocols, from cloning to purification, are well defined, and nearly all biology

laboratories already have some system in place. It is relatively inexpensive, which allows for large-scale growths for milligram production of protein. Finally and most importantly, it has a long history of success. Albeit as seen in this study, this is often less than 50% of the time. CF, however, is nearly the exact opposite. CF protocols are under development, few laboratories have the expertise, and it is relatively expensive, all of which make obtaining milligram quantities of protein challenging. Nevertheless, our results indicate that CF is more successful than *in vivo* expression.

Size and TM complexity determines expression levels. From the expression profiles we have shown that CF complements *in vivo* expression. Due to the large number of proteins screened in this study, we can also categorize the proteins based on their physical properties. We have broken down the expression profiles for both CF and *in vivo* experiments by MW and number of TM helices, two related properties that have been hypothesized to affect expression levels. From these statistics, we can give some *a priori* estimation for the expression success of an unknown target.

MW is perhaps the single most important physical property when describing a protein, particularly for hypothetical ones. In essence, MW describes the “complexity” of a MP and thereby also affects the solubilization, purification, and biophysical characterization. In a closed system, such as CF, with limited energy and reactants, target protein size can drastically affect expression. Also, larger mRNA transcripts introduce losses from poor transcription efficiency and increased nuclease susceptibility. In Tables 3A and 3B we tabulate the results for CF and *in vivo* respectively. Since in this analysis we targeted

proteins of a limited size (roughly 10kDa to 30kDa) we clustered them into small (10-20kDa) and large (20-30kDa). The CF results in Table 3A clearly show that smaller proteins (74% success) express significantly better than larger proteins (51% success). Table 3B shows that the *E. coli in vivo* expression system, for the reasons cited above, is less dependent on MW.

The number of TM helices of an alpha helical MP, due to the two-dimensional constraints of a membrane bilayer, also determines its structural complexity. In Tables 4A and 4B we have organized the expression profiles for the CF and *in vivo* systems versus number of TM helices. Most strikingly, expression success decreases with TM number (most apparent where number of TMs is two to six). This is expected in light of the dependence on MW and the fact that the number of TMs also correlates with MW. What is surprising though is that CF, although clearly more robust, is much more sensitive on TM number than the *in vivo* system and shows a progressive decrease in success versus number of TMs. This is most likely due to a suboptimal translocation process for the CF reaction. We can also conclude that the targets most likely to express well are of lower MW with fewer TM helices. For structural genomics initiatives, where information content of a structure is important, these results suggest MPs with four to six TMs may be good targets to pursue. This is particularly significant for MPs, which often form homooligomers from low MW (less than 30kDa) monomers. Finally, we conclude that although CF is more robust than *in vivo* expression and shows a higher level of success across the board, it is more sensitive to increased MP size and complexity.

Cell-free proteins behave like their in vivo counterparts. Given CF's increasing prevalence and its complementation of *in vivo* expression outlined here, it is important to verify that CF produced proteins behave similarly to their *in vivo* counterparts. Certainly, this has been shown for soluble proteins (Yokoyama 2003) (Tyler et al. 2005) but with the importance of translocation and proper folding within the membrane, this is not so obvious for MPs. Much of the initial CF work on MPs, however, was showing activity for channels and transporters (Elbaz et al. 2004) (Berrier et al. 2004). Furthermore, the NMR and functional studies of Dötsch and colleagues on isotopically labeled CF produced MPs in a variety of detergent conditions has shown them to be well behaved and properly folded (Klammt et al. 2005) (Koglin et al. 2006). We show here that MPs can be expressed in the absence of any exogenous lipid or detergent and can be solubilized post-reaction, a technique still being explored. The S30 *E. coli* extract (see Materials and Methods) is a concentrated mixture containing membranes and translocation machinery, but for some proteins it may be useful to include additives (e.g. natural or synthetic lipids) to assist the translocation process. Addressing questions of foldedness and characterization, especially since many are presumed MPs without a known function or assay, will require the use of other biophysical methods (Columbus et al. 2006). We propose here that gel filtration chromatography, which gives a rough measure of size and monodispersity, can be used as a simple benchmark to validate purity, homogeneity, and stability. It is quick, reliable, and useful for both purification and as a quality-check before structural studies.

Cell-free is an extensible system. The addition of exogenous lipids or detergents highlights one of the most powerful features of the CF system. CF is reconstituted and can be viewed as modular, where subsystems can be added, deleted, and modified. One such example is the energy subsystem responsible for regenerating nucleoside triphosphates (NTPs) from high-energy compounds. A coupled transcription-translation CF system requires a highly concentrated pool of NTPs, so CF systems often use an additional enzymatic (i.e. kinase) system to recycle NTPs via other phosphate compounds such as phosphoenolpyruvate or phosphocreatine that are added to the reaction. In our hands such systems are interchangeable (data not shown) and furthermore, this modularity can be extended to other parts of the reaction. One can imagine the use of chaperones to aid in processes such as folding or disulfide bond formation, both of which can be important for proper membrane incorporation. Finally, as demonstrated by the SeMet substitution of CcmG, extensibility is an easy way to label proteins. The reconstituted nature of the system allows for replacement of any of the twenty natural amino acids with an unnatural (provided they can be loaded by the appropriate aminoacyl tRNA synthetase) amino acid resulting in 100% substitution. With respect to *in vivo* systems, it is more efficient and requires significantly less unnatural amino acid.

Conclusion and Perspectives. Given these successful expression results, we therefore suggest (Figure 5) a general strategy for the production of MPs for structural studies in both systems. Following target selection, constructs can be generated via ligation independent cloning (LIC), which allows for simultaneous cloning into multiple destination vectors from a single polymerase chain reaction (PCR) product without the

need for proprietary enzyme mixtures (Aslanidis and de Jong 1990). Small-scale growths and reactions cull expressers from non-expressers, followed by larger-scale experiments with solubility screens to identify well-behaved PDCs. For those targets that can only be expressed/solubilized in one system or the other, the choice is clear and the target can be transitioned to scaling up, characterization, and crystallization. For those expressed in both (30% in our case) however, the choice is more complex. Based on the relative ease outlined above, the most prudent decision is to transition the *in vivo* expressed target through the pipeline. Then, due to the importance of labeling in both NMR and x-ray experiments, CF expression should be used as necessary in the pipeline once characterization and purification is well defined. Thus, from our results showing CF is a more robust expression system yet requiring more input, we believe CF complements *in vivo* expression.

Materials and Methods

Cloning of genes. Coding sequences for 120 genes were obtained from Ecocyc using the E.coli K-12 dataset (Keseler et al. 2005). Predicted number of TM helices and function (Table 1) are from the Swiss-Prot annotation. For cloning, we used a ligation independent cloning strategy that allows for PCR products containing LIC overhangs to be directly cloned into any of our LIC expression vectors (Aslanidis and de Jong 1990). For *in vivo* expression of proteins, genes were cloned into pET3a based LIC vectors, which contained either a N-terminal TEV protease cleavable 6xHis tag with or without a Maltose Binding Protein (MBP) tag. For CF expression of proteins, a high-copy number plasmid was desired so the LIC site of the pET3a based expression vector was subcloned into pcDNA3.1 (Figure 6). Primers were designed using the Express Primer Tool for High Throughput Gene Cloning and Expression (<http://tools.bio.anl.gov/bioJAVA/jsp/ExpressPrimerTool/>) containing the appropriate LIC overhangs. PCR reactions were done using Phusion polymerase (New England Biolabs) with *E. coli* genomic DNA and LIC cloned into our expression vectors.

S30 extract. S30 extract was made based on the method of Kigawa et al. 2003, as modified by Liu et al. 2005. Briefly, a starter culture picked from a single colony of BL21 (DE3) cells (Invitrogen) transformed with the Rosetta II plasmid (Novagen) was used to inoculate at least 6L of 2xYT media. Cells were grown to an OD₆₀₀ of 2.0 and then harvested. Cells were then washed with S30A buffer (10mM Tris-acetate, pH 8.2, 14mM Mg(OAc)₂, 60mM KOAc, 1mM dithiothreitol (DTT), 7mM β-mercaptoethanol (BME)) and then pelleted. The washed cells were then frozen in liquid N₂ and stored at -

80°C for no more than 3 days. Cells were thawed in S30B buffer (10mM Tris-acetate, pH 8.2, 14mM Mg(OAc)₂, 60mM KOAc, 1mM DTT) and lysed with a C5 EmulsiFlex (Avestin). Lysed cells were centrifuged at 30,000xg for 30min and the supernatant was centrifuged again at 30,000xg for 30min. The supernatant was then concentrated to ~10-15ml in a 10kDa MWCO Amicon Ultra Centrifugal Filter. The concentrate was incubated for 80min at 37°C and then dialyzed with 2L of S30B buffer using a 14kDa dialysis bag (Spectrum) for 1hr. The dialyzed extract was then centrifuged at 30,000xg for 30min with the supernatant being the final S30 extract used in CF. Extract was frozen in liquid N₂ and stored for up to 6 months at -80°C. The activity of each S30 extract was tested by expressing green fluorescent protein (GFP) using various concentrations of magnesium chloride. The GFP levels were quantified by measuring the fluorescence intensity with a Fluoromax-3 Spectrofluorometer (HORIBA-Yobin Yvon) and compared to a pure sample of known concentration.

T7 RNA polymerase purification. A starter culture from a single colony of BL21 (DE3) cells (Invitrogen) transformed with pT7-911Q plasmid (Ichetovkin et al. 1997) containing a 6xHis-tagged T7 polymerase was used to grow 12 L of cells. Once cells reached an OD₆₀₀ of 0.4-0.6, they were induced for 4hr with 1mM isopropyl-D-thiogalactoside. Cells were then pelleted and washed in 50mM Tris-HCl, pH 7.4, 150mM NaCl buffer. Washed cells were then resuspended in lysis buffer (50mM Tris-HCl pH 7.4, 150mM NaCl, 5mM BME, 5% glycerol, 1mM imidazole, 100µM phenylmethylsulfonyl fluoride (PMSF) and lysed with a C5 EmulsiFlex. Cellular debris was pelleted at 20,000xg for 30min and the supernatant was incubated with IMAC Ni-NTA resin (Qiagen) for 30min. Ni-bound

protein was then washed with lysis buffer and washed again with lysis buffer containing 10mM imidazole. Ni bound protein was then eluted with lysis buffer containing 100mM imidazole. Eluted protein was dialyzed overnight in storage buffer (50mM Tris-HCl, pH 7.7, 100mM NaCl, 1mM EDTA, 50% glycerol, 10mM DTT) in a 14kD dialysis bag. All procedures were carried out at 4°C or on ice. The T7 was stored at -20°C for up to 1 year.

In vivo protein expression and solubility. For small-scale *in vivo* protein expression, BL21 (DE3) C43 cells (Avidis) (Miroux and Walker 1996) were transformed with the *in vivo* expression constructs. Single colonies were used to grow 2ml of cells overnight at 37°C in auto induction media (0.5 % glycerol, 0.5% glucose, 0.2% a-lactose, 25mM Na₂HPO₄, 25mM M KH₂PO₄, 50mM NH₄Cl, 5mM Na₂SO₄, and 2mM MgSO₄) (Studier 2005). Cells were harvested and resuspended in 100mL lysis buffer (20mM Tris- HCl, pH 7.4, 150mM NaCl, 1mg/ml lysozyme (EMD Biosciences), complete protease inhibitor cocktail EDTA-free (Roche) and 10U/ml Benzonase (Novagen) for 1hr at 4⁰C. Sodium dodecyl sulfate (SDS) was added to a final concentration of 2% and incubated for an additional hr at 4°C. Cellular debris was pelleted at 16,000xg and supernatant containing SDS soluble protein was diluted with an equal volume of 2x SDS loading buffer (125mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 5mM BME, and 0.005% Bromophenol Blue) and detected by Western Blot analysis using an anti-6xHis horseradish peroxidase conjugated antibody (Santa Cruz Biotech).

For large scale protein expression, 250ml-12L of auto-induction media growth from a single BL21(DE3) C43 colony was harvested and resuspended in lysis buffer (20mM Tris-HCl, pH 7.4, 100mM NaCl, 100 μ M PMSF, and 4mM BME). Cells were lysed with the EmulsiFlex and undisturbed cells were pelleted at 10,000xg for 30min. The supernatant was pelleted at 200,000xg for 1hr to collect membranes. Membranes were solubilized in OG or DDM solubilization buffer (20mM Tris pH 7.4, 150mM NaCl, 10% glycerol, 4mM BME, 100 μ M PMSF, and 270mM OG or 10mM DDM) at 4°C overnight, which we call the before-spin. Soluble protein was collected from the supernatant of a 200,000xg spin for 30min, which we call the after-spin. A qualitative analysis of protein solubility in DDM and OG was done by comparing the intensity of the before-spin and after-spin band on a Western blot.

Cell-free protein expression and solubility. Small-scale 30 μ L test reactions were carried out on all targets for 3hr at 37°C in parallel (see column one of Table 5 for reaction ingredients). An equal amount of 2x SDS loading buffer was added and run on a SDS-polyacrylamide electrophoresis gel (SDS-PAGE). Expression levels were qualitatively determined by Western blot analysis.

Large-scale reactions were done in 25kDa dialyzers (Spectrum) bathed in feeder solution of 10-20 times the reaction volume (Table 5) (Klammt et al. 2004). Reactions ranging from 2 to 22ml were run overnight at 37°C with gentle shaking. Solubility of the reaction products was tested by either first pelleting at 100,000xg for 20min and then washing with 20mM Tris-HCl pH 7.4, 150mM NaCl before solubilizing in DDM or OG buffer or

by diluting the reaction 5X in DDM or OG solubilization buffer. Solubility of each protein in DDM or OG was done in the same before-spin/after-spin manner as described above.

Protein purification. Proteins soluble in OG were incubated at 4°C with IMAC Ni-NTA resin for 30min. Protein bound beads were washed with OG size exclusion buffer (SEC) (20mM Tris-HCl pH 7.4, 100mM NaCl, 10% glycerol, 2mM DTT, 40mM OG) containing 15-25mM imidazole and eluted with OG SEC buffer containing 300mM imidazole. The elution was then desalted into OG SEC buffer using a 10DG disposable desalting column (Biorad) and concentrated to 1ml in a 10kDa MWCO Ultra Centrifugal Filter (Amicon). The concentrated protein was injected on a Superdex 200 gel filtration column at 0.33ml/min running the SEC mobile phase. Peak fractions were collected and assayed by SDS-PAGE. The single peak fraction containing purified protein was collected and concentrated to 10mg/ml for crystal screens.

Selenomethionine labeling and mass spectrometry. Cell-free synthesized protein was purified as described above with the exception of using L-selenomethionine (Avanti) in the reaction mixture. For MS, an aliquot of 10 mg/ml selenomethionine labeled CcmG was diluted 10:1 in water to lower salt concentration and then mixed 1:1 with a solution of saturated cinnamic acid, 50% acetonitrile, and 0.1% trifluoroacetic acid for matrix formation. The sample was analyzed with a MALDI MS (Applied Biosystems).

Abbreviations: CF, cell-free; MP, membrane protein; NMR, nuclear magnetic resonance; PDC, protein-detergent complex; TM, transmembrane; MW, molecular weight; OG, octyl glucoside; DDM, dodecyl maltoside; IMAC, immobilized metal affinity chromatography; NTP, nucleoside triphosphates; PCR, polymerase chain reaction; LIC, ligation independent cloning

Table 1. Target list and expression results.

ID No.	Protein	SwissP	kDa	TMs	Function	C43	CF
1	AqpZ	P60844	23.7	6	aquaporin	+	+
2	UppP	P60932	29.8	7	undecaprenyl diphosphatase	-	-
3	CrcB	P37002	13.8	4	hypothetical MP	+	+
4	CypA	P08550	17.9	4	colicin V production	+	-
5	CysZ	P0A6J3	29.3	4	sulfate transport	+	+
6	DedA	P0ABP6	24.5	5	hypothetical MP	-	-
7	DgkA	P0ABN1	13.2	3	diacylglycerol kinase	++	++
8	CcmG	P0AA86	20.1	1	cytochrome c biogenesis	++	+
9	EmrE	P23895	12.0	4	multidrug transporter	+	++
10	FxsA	P37147	17.7	2	suppressor of F exclusion of T7	+	+
11	GlpF	P0AER0	29.8	6	aquaporin	++	+
12	GspO	Q2M700	25.0	6	prepilin leader peptidase	+	+
13	HdeD	P0AET5	20.9	6	unknown/ acid resistance	+	++
14	LspA	P00804	18.1	4	lipoprotein signal peptidase	+	-
15	MarC	P0AEY1	23.6	6	multiple antibiotic resistance	-	-
16	MreD	P0ABH4	18.8	5	rod shaped determining	+	-
17	MscL	P0A742	15.0	2	large mechanosensitive channel	++	++
18	PgpA	P18200	19.4	3	phosphatase	++	++
19	PgpB	P0A924	29.0	6	phosphatase	++	+
20	PgsA	P0ABF8	20.7	4	PGP synthase	+	-

21	PppA	Q46836	29.6	6	prepilin peptidase	-	+
22	SieB	P38392	19.2	2	phage superinfection exclusion	-	+
23	SugE	P69937	10.9	4	multidrug transporter	+	-
24	UspB	P0A8S5	13.0	2	universal stress protein	-	-
25	YaaH	P0AC98	20.1	6	ammonium transporter	++	+
26	YbbJ	P0AAS3	16.8	3	hypothetical MP	N	N
27	YbjM	P64439	14.2	4	hypothetical MP	+	+
28	YcfZ	P75961	28.9	4	homolog of a virulence factor	+	-
29	YdgC	P0ACX0	12.3	3	hypothetical MP	-	-
30	YebN	P76264	20.1	5	terpenoid synthesis like	-	-
31	YfdG	P77682	13.2	4	hypothetical MP	-	-
32	YgdD	P67127	26.1	7	hypothetical MP	-	-
33	YhgN	P67143	21.5	6	hypothetical MP	-	-
34	YicG	P0AGM2	22.0	6	hypothetical MP	-	-
35	YiiR	P0AF34	16.5	4	hypothetical MP	+	++
36	YjiH	P39379	23.8	4	hypothetical MP	+	+
37	YabI	P30149	28.2	6	hypothetical MP	-	N
38	YbbM	P77307	28.1	7	metal reistance protein	-	++
39	YbjO	P0AAZ0	18.5	3	hypothetical MP	-	++
40	YchE	P25743	23.5	6	hypothetical MP	+	-
41	YdgK	P76180	16.3	4	oxido reductase	-	+
42	YecN	P64515	15.2	1	hypothetical MP	++	-

43	YfeZ	P76538	17.1	4	hypothetical MP	-	++
44	ArgO	P11667	23.1	6	arginine outward transport	+	-
45	YhhN	P0ADI9	23.8	6	hypothetical MP	-	-
46	YidG	P0ADL6	13.8	3	hypothetical MP	-	++
47	YijD	P0AF40	13.0	4	hypothetical MP	+	++
48	YjjB	P0ADD2	11.9	3	hypothetical MP	+	-
49	YadS	P0AFP0	22.1	7	hypothetical MP	-	+
50	YbcI	P45570	19.5	4	hypothetical MP	-	+
51	YcaP	P75839	26.2	3	hypothetical MP	-	-
52	YchQ	Q46755	14.6	4	hypothetical MP	-	++
53	YdjM	P64481	22.8	3	hypothetical MP	-	-
54	YecS	P0AFT2	24.8	3	permease	+	-
55	YfiK	P38101	21.2	4	permease	-	+
56	YggT	P64564	21.1	5	reistance protein	-	+
57	YhiD	P0AFV2	23.2	5	transport ATPase	++	-
58	YidH	P0ADM0	12.8	3	hypothetical MP	++	++
59	PsiE	P0A7C8	15.6	4	hypothetical MP	+	++
60	YjjP	P0ADD5	28.0	4	structural protein	-	+
61	YafU	P77354	12.1	2	hypothetical MP	-	+
62	YbfB	P0AAU5	12.6	3	hypothetical MP	-	+
63	YcbC	P0AB01	28.7	2	putative enzyme	-	-
64	YciB	P0A710	20.8	5	septation protein	-	+

65	YdjX	P76219	26.1	5	hypothetical MP	-	+
66	YedR	P76334	13.8	2	hypothetical MP	-	++
67	YgaH	P43667	12.0	3	transport protein	+	+
68	YghB	P0AA60	24.1	4	hypothetical MP	-	-
69	YiaA	P0ADJ8	16.1	4	hypothetical MP	-	++
70	YidI	P31446	15.7	3	hypothetical MP	++	+
71	YjcH	P0AF54	11.7	2	hypothetical MP	+	++
72	YkgB	P75685	21.9	3	hypothetical MP	-	++
73	YagU	P0AAA1	23.0	3	hypothetical MP	-	++
74	YbhL	P0AAC4	25.9	7	transport protein	+	-
75	YccA	P0AAC6	23.4	7	hypothetical MP	-	+
76	YciC	P21365	26.4	6	hypothetical MP	-	+
77	YdjZ	P76221	26.1	5	hypothetical MP	+	+
78	YeiU	P76445	26.8	5	permease	N	N
79	YgaP	P55734	18.6	2	phosphatase	++	++
80	ZupT	P0A8H3	26.5	8	Zn transport	++	-
81	YiaB	P11286	12.6	4	hypothetical MP	+	+
82	YieI	P31468	16.7	4	hypothetical MP	+	+
83	YjdF	P39270	23.4	5	hypothetical MP	-	+
84	YkgH	P77180	25.6	2	hypothetical MP	+	++
85	YahC	P77219	17.3	5	hypothetical MP	-	+
86	YbhM	P75769	26.1	7	hypothetical MP	-	+

87	YccF	P0AB12	16.3	3	hypothetical MP	+	+
88	YciS	P0ACV4	11.4	2	hypothetical MP	-	-
89	YeaL	P0ACY6	15.3	4	hypothetical MP	N	N
90	YfbJ	P76474	14.1	4	transport receptor	-	+
91	YgaW	P64550	16.9	4	hypothetical MP	-	-
92	YgiH	P60782	22.2	5	hypothetical MP	+	+
93	YiaD	P37665	22.2	2	OmpA-OmpF Porin family	++	-
94	YigF	P27842	14.5	2	hypothetical MP	-	++
95	YjeO	P39284	12.6	3	hypothetical MP	++	++
96	YlaC	P0AAS0	18.3	2	hypothetical MP	+	++
97	YaiY	P0AAP7	11.4	2	hypothetical MP	-	++
98	YbhQ	P0AAW5	15.4	4	hypothetical MP	-	+
99	PgaD	P69432	16.1	2	polysaccharide synthesis	-	++
100	YdcZ	P76111	15.9	5	transport protein	-	+
101	YeaS	P76249	23.2	6	hypothetical MP	-	-
102	YfbV	P0A8D9	17.2	2	hypothetical MP	+	-
103	YgaZ	P76630	26.1	6	hypothetical MP	-	++
104	YgiZ	Q46867	13.2	3	hypothetical MP	+	++
105	YiaW	P0ADK4	12.4	2	hypothetical MP	-	++
106	YigG	P27843	15.8	4	hypothetical MP	-	-
107	YjfL	P0AF80	14.2	4	hypothetical MP	+	-
108	YmcD	P75885	10.1	2	hypothetical MP	-	++

109	YbaN	P0AAR5	14.8	4	hypothetical MP	-	++
110	YbjG	P75806	22.4	6	permease	-	++
111	YcdZ	P75916	15.9	6	transport protein	-	++
112	YddG	P46136	29.4	9	export of methyl viologen	-	-
113	YebE	P33218	23.6	1	hypothetical MP	+	-
114	YfcA	P0AD30	28.6	6	hypothetical MP	-	-
115	YgdD	P0ADR2	14.3	3	hypothetical MP	-	-
116	YgjV	P42603	20.5	5	hypothetical MP	-	-
117	YibI	P32108	13.9	2	hypothetical MP	-	++
118	RhtB	P0AG34	22.4	6	homoserine efflux transporter	-	-
119	YjiG	P0AEH8	16.2	4	hypothetical MP	-	++
120	YmfA	P75962	17.4	2	hypothetical MP	-	-

Table1. Target list and expression results. This table includes the expression target list including protein name, SwissProt ID, molecular weight, number of transmembrane helices, and function. The final two columns are qualitative expression results for C43 *in vivo* and CF systems. Results are tabulated as: N, not tested; (-) no signal detected on Western blot; (+) weak Western blot detection (++) strong Western blot detection.

Table 2A. *in vivo* solubility screening

ID No.	Protein	DDM	OG
5	CysZ	+	++
7	DgkA	++	++
8	CcmG	++	++
14	LspA	++	+
16	MreD	N	+
17	MscL	++	++
18	PgpA	++	++
25	YaaH	++	+
35	Yiir	++	+
47	YijD	+	++
54	YecS	++	N
58	YidH	+	++
59	PsiE	+	++
71	YjcH	+	+
74	YbhL	-	+
77	YdjZ	N	-
79	YgaP	++	++
81	YiaB	-	+
84	YkgH	-	+
93	YiaD	-	+

96	YlaC	+	++
102	YfbV	++	++
104	YgiZ	++	++
107	YjfL	-	++

Table 2B. CF solubility screening

ID No.	Protein	DDM	OG
13	HdeD	++	++
38	YbbM	-	++
43	YfeZ	+	+
46	YidG	++	++
64	YciB	++	++
69	YiaA	++	++
73	YagU	+	+
94	YigF	-	-
95	YjeO	-	-
97	YaiY	+	+
99	PgaD	-	-
108	YmcD	++	++
109	YbaN	-	-
110	YbjG	+	+
111	YcdZ	+	+

Table 2. Detergent solubility screening. (A) Successful *in vivo* expressers were screened for OG and DDM solubilization as described in Materials and Methods. Briefly (-) indicates partial, (+) low, and (++) high solubility. (B) Successful CF expressers were screened for detergent solubilization as in (A).

Table 3A. CF expression profile versus MW

	10-20kDa	20-30kDa	Total
Total Number	61	55	116
% (-)	26	49	37
% (+)	28	38	33
% (++)	46	13	30
% (+ and ++)	74	51	63

Table 3B. *In vivo* expression profile versus MW

	10-20kDa	20-30kDa	Total
Total Number	61	56	117
% (-)	52	61	56
% (+)	35	27	31
% (++)	13	12	13
% (+ and ++)	48	39	44

Table 3. Expression profiles versus MW. (A) The CF expression profiles binned for the 10-20kDa and 20-30kDa ranges. First row is the total number of targets screened and the remaining rows are percentages. (B) Same as (A), except results are from *in vivo*.

Table 4A. CF expression profile versus number of TMs

TMs	1	2	3	4	5	6	7	8	9	Total
Total Number	3	21	19	31	13	20	7	1	1	116
% (-)	67	29	32	32	38	45	43	100	100	37
% (+)	33	14	21	39	62	35	43	0	0	33
% (++)	0	57	47	29	0	20	14	0	0	30
% (+ and ++)	33	71	68	68	62	55	57	0	0	63

Table 4B. *In vivo* expression profile versus number of TMs

TMs	1	2	3	4	5	6	7	8	9	Total
Total Number	3	21	19	31	13	21	7	1	1	117
% (-)	0	62	48	48	69	62	86	0	100	56
% (+)	33	24	26	52	23	24	14	0	0	31
% (++)	67	14	26	0	8	14	0	100	0	13
% (+ and ++)	100	38	52	52	31	38	14	100	0	44

Table 4. Expression profiles versus number of TM helices. (A) The CF expression profiles binned by number of transmembrane helices (from one to nine). First row is the total number of targets screened and the remaining rows are percentages. (B) Same as (A), except results are from *in vivo*.

Table 5. Cell-free reaction components.

Reagent	Concentration
Mg acetate	~ 15mM
NaN ₃	0.05%
PEG8000	4%
HEPES buffer pH 7.5	55mM
Potassium glutamate	270mM
Folinic acid	0.068mM
Ammonium acetate	27.5mM
DTT	1.7mM
NTP	1mM
Creatine phosphate	80mM
Amino acids	1mM
3,5 cAMP	0.64mM
tRNA-Ecoli	0.175mg/ml
Creatine kinase	0.25mg/ml
T7 RNA polymerase	0.2mg
Plasmid	6.7ug/ml
S30 Extract	35%
RNase inhibitor	0.3U/uL

Table 5. Cell-free reaction components. This table lists the reactants present in our CF reaction. All reagents (including grey box) are present in the reaction mixture. All

except grey box are include in the feeding mixture as in the methods (Klammt et al. 2004). The feeder and reaction mixture were separated by a 25kDa cutoff dialysis membrane. Optimal magnesium concentrations vary with the S30 extract preparation.

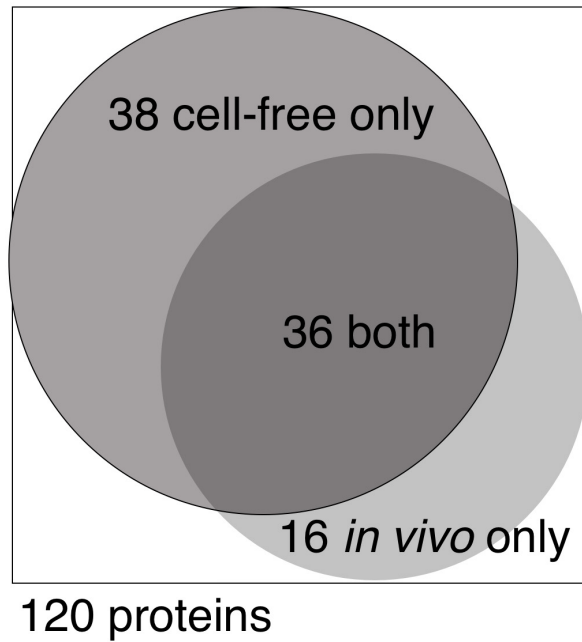


Figure 1. Venn diagram of expression results. This figure is a Venn diagram to scale showing the expression success for the 120 proteins. Cumulatively, 90 proteins (75%) are expressed. 74 are expressed by CF (38 only in CF) and 52 are expressed *in vivo* (16 only in *in vivo*). There is an overlap between the two systems of 36 proteins.

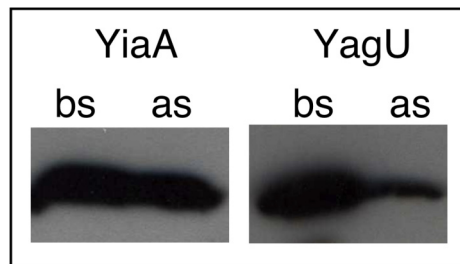


Figure 2. Solubilization example. Two examples of detergent solubilization assigned as (++) (left panel) and (+) (right panel). Examples are for #69, YiaA and #73, YagU in DDM. BS and AS are the supernatants of before-spin and after-spin as described in Materials and Methods.

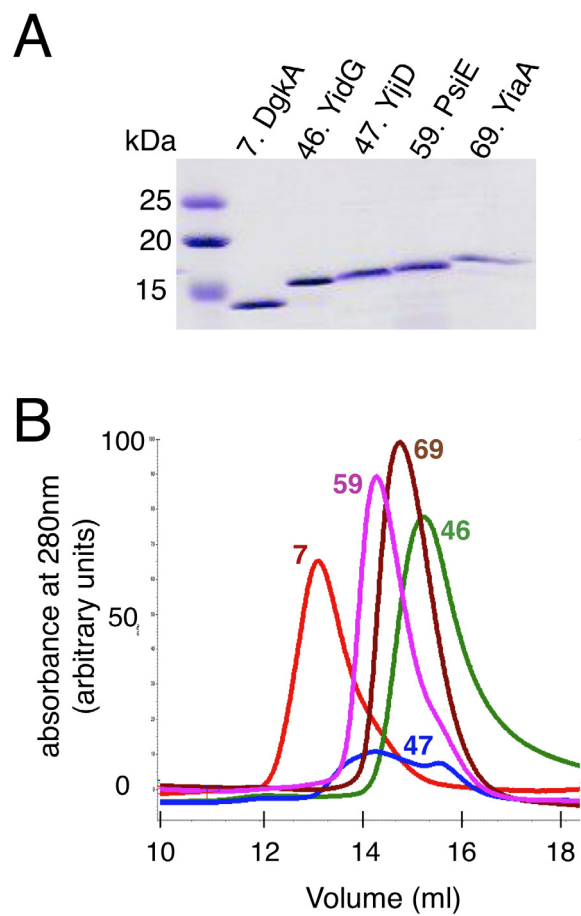


Figure 3. Purification of CF products. To verify protein behavior, five proteins (IDs 7, 46, 47, 47, 59, and 69) were expressed at the milligram scale, purified, and homogeneity assayed by gel filtration profile. (A) Coomassie-stained gel of IMAC purified proteins. (B). Gel filtration profile showing that four of the five targets are homogeneous.

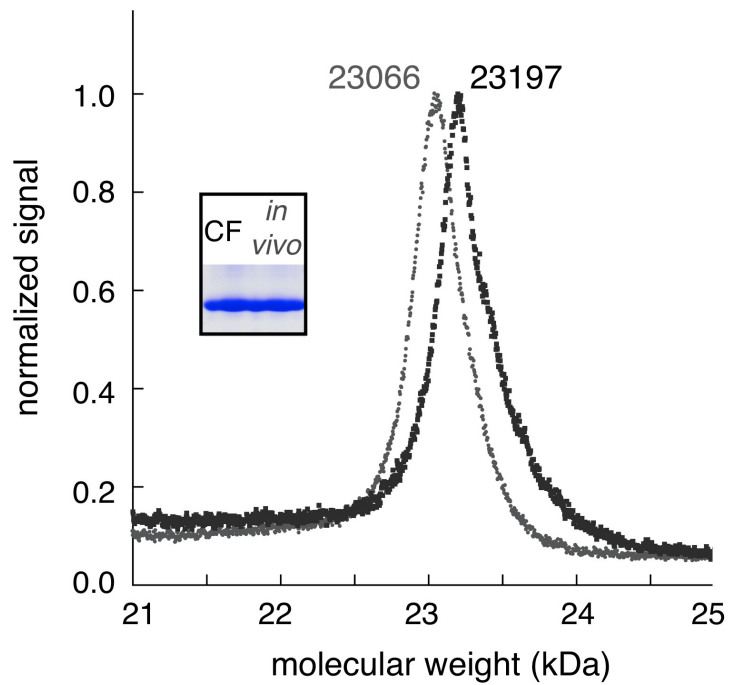


Figure 4. SeMet incorporation. CcmG was labeled by replacing methionine with L-selenomethionine in the CF reaction. Incorporation was assayed by MALDI-MS. Grey denotes native protein (MW 23066kDa) and black is the labeled (23197kDa) showing a difference of 131Da. A coomassie-stained gel of the two products is shown in the inset.

Figure 5. Membrane protein structure determination pipeline. This figure details our proposed pipeline for the expression, solubilization, purification, and structural study of MPs. Black arrows and boxes indicate the traditional *in vivo* pathway, while grey arrows indicate the most gainful and complementary use of cell-free expression.



Figure 6. Sequence of N-terminal tags for ligation independent cloning (LIC) vectors. Modified pet3a and pcDNA3.1 (-) vectors were used for in vivo and cell-free expression, respectively.

References

- Aslanidis, C., and de Jong, P.J. 1990. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* **18**: 6069-6074.
- Berrier, C., Park, K.H., Abes, S., Bibonne, A., Betton, J.M., and Ghazi, A. 2004. Cell-free synthesis of a functional ion channel in the absence of a membrane and in the presence of detergent. *Biochemistry* **43**: 12585-12591.
- Columbus, L., Lipfert, J., Klock, H., Millett, I., Doniach, S., and Lesley, S.A. 2006. Expression, purification, and characterization of *Thermotoga maritima* membrane proteins for structure determination. *Protein Sci* **15**: 961-975.
- Dobrovetsky, E., Lu, M.L., Andorn-Broza, R., Khutoreskaya, G., Bray, J.E., Savchenko, A., Arrowsmith, C.H., Edwards, A.M., and Koth, C.M. 2005. High-throughput production of prokaryotic membrane proteins. *J Struct Funct Genomics* **6**: 33-50.
- Elbaz, Y., Steiner-Mordoch, S., Danieli, T., and Schuldiner, S. 2004. In vitro synthesis of fully functional EmrE, a multidrug transporter, and study of its oligomeric state. *Proc Natl Acad Sci U S A* **101**: 1519-1524.
- Endo, Y., and Sawasaki, T. 2003. High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. *Biotechnol Adv* **21**: 695-713.
- Eshaghi, S., Hedren, M., Nasser, M.I., Hammarberg, T., Thornell, A., and Nordlund, P. 2005. An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins. *Protein Sci* **14**: 676-683.
- Ichetovkin, I.E., Abramochkin, G., and Shrader, T.E. 1997. Substrate recognition by the leucyl/phenylalanyl-tRNA-protein transferase. Conservation within the enzyme

- family and localization to the trypsin-resistant domain. *J Biol Chem* **272**: 33009-33014.
- Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Mei Ono, A., and Guntert, P. 2006. Optimal isotope labelling for NMR protein structure determinations. *Nature* **440**: 52-57.
- Keseler, I.M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I.T., Peralta-Gil, M., and Karp, P.D. 2005. EcoCyc: a comprehensive database resource for Escherichia coli. *Nucleic Acids Res* **33**: D334-337.
- Kigawa, T., Muto, Y., and Yokoyama, S. 1995. Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J Biomol NMR* **6**: 129-134.
- Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., and Yokoyama, S. 1999. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett* **442**: 15-19.
- Kigawa, T., Yamaguchi-Nunokawa, E., Kodama, K., Matsuda, T., Yabuki, T., Matsuda, N., Ishitani, R., Nureki, O., and Yokoyama, S. 2002. Selenomethionine incorporation into a protein by cell-free synthesis. *J Struct Funct Genomics* **2**: 29-35.
- Klammt, C., Lohr, F., Schafer, B., Haase, W., Dotsch, V., Ruterjans, H., Glaubitz, C., and Bernhard, F. 2004. High level cell-free expression and specific labeling of integral membrane proteins. *Eur J Biochem* **271**: 568-580.
- Klammt, C., Schwarz, D., Fendler, K., Haase, W., Dotsch, V., and Bernhard, F. 2005. Evaluation of detergents for the soluble expression of alpha-helical and beta-

- barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *Febs J* **272**: 6024-6038.
- Koglin, A., Klammt, C., Trbovic, N., Schwarz, D., Schneider, B., Schafer, B., Lohr, F., Bernhard, F., and Dotsch, V. 2006. Combination of cell-free expression and NMR spectroscopy as a new approach for structural investigation of membrane proteins. *Magn Reson Chem* **44 Spec No**: S17-23.
- Korepanova, A., Gao, F.P., Hua, Y., Qin, H., Nakamoto, R.K., and Cross, T.A. 2005. Cloning and expression of multiple integral membrane proteins from *Mycobacterium tuberculosis* in *Escherichia coli*. *Protein Sci* **14**: 148-158.
- Miroux, B., and Walker, J.E. 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**: 289-298.
- Pornillos, O., Chen, Y.J., Chen, A.P., and Chang, G. 2005. X-ray structure of the EmrE multidrug transporter in complex with a substrate. *Science* **310**: 1950-1953.
- Spirin, A.S., Baranov, V.I., Ryabova, L.A., Ovodov, S.Y., and Alakhov, Y.B. 1988. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **242**: 1162-1164.
- Studier, F.W. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**: 207-234.
- Surade, S., Klein, M., Stolt-Bergner, P.C., Muenke, C., Roy, A., and Michel, H. 2006. Comparative analysis and "expression space" coverage of the production of prokaryotic membrane proteins for structural genomics. *Protein Sci* **15**: 2178-2189.

- Thony-Meyer, L. 2002. Cytochrome c maturation: a complex pathway for a simple task?
Biochem Soc Trans **30**: 633-638.
- Tyler, R.C., Aceti, D.J., Bingman, C.A., Cornilescu, C.C., Fox, B.G., Frederick, R.O.,
Jeon, W.B., Lee, M.S., Newman, C.S., Peterson, F.C., et al. 2005. Comparison of
cell-based and cell-free protocols for producing target proteins from the
Arabidopsis thaliana genome for structural studies. *Proteins* **59**: 633-643.
- White, S.H. 2004. The progress of membrane protein structure determination. *Protein Sci*
13: 1948-1949.
- Yokoyama, S. 2003. Protein expression systems for structural genomics and proteomics.
Curr Opin Chem Biol **7**: 39-43.
- Zubay, G. 1973. In vitro synthesis of protein in microbial systems. *Annu Rev Genet* **7**:
267-287.

Chapter 6

Development of a Membrane Protein Structure Pipeline

Research completed in collaboration with

Zachary Newby, Corey Anderson, Yaneth Robles Colmenares, and Robert Stroud

Abstract

Membrane protein structure determination has lagged significantly behind soluble protein structure determination. Membrane proteins are inherently more difficult to work with, which complicates the structure process. Here, we present an empirically derived set of protocols to quickly identify successful structure targets and highlight the important aspects of working with this class of proteins in the framework of our proposed pipeline. We then choose a model membrane protein, CcmG, discovered in the initial stages of this pipeline, and subject it to our standardized methods. This test was successful and we determined the structure of CcmG to 2.3 Å resolution. Thus, the pipeline we describe is useful for membrane protein structure determination on a large scale.

Introduction

Membrane proteins (MPs) account for 20-30% of all proteins in a typical cell and play fundamental roles in cellular processes such as signaling, transport, bioenergetics, and spatial organization. Despite this important role, and being important pharmaceutical targets, MP structural information is severely lacking - MPs account for less than 1% of all structures in the Protein Data Bank (PDB). The lack of structural information can be attributed to the many barriers that exist along a typical MP structure determination pipeline. The two-dimensional nature of the membrane and an involved translocation process leads to lower expression levels of well-behaved protein. Next, detergent conditions must be empirically determined that can simultaneously extract the protein from its native membrane and retain its active state. Once in “soluble” form, the protein can then be purified in a detergent-containing buffer, but properties associated with the protein-detergent-complex (PDC) can severely affect protein stability, affinity tag purification, affinity tag cleavage, and heterogeneity. Lastly, even once the protein is pure, homogeneous, and stable (PHS), the nature of the PDC can affect crystallization. The PDC introduces a larger than usual amount of heterogeneity to the sample and limits the protein surface area available for forming crystal contacts. Crystals of MPs are often mosaic, have high solvent contents, and diffract poorly. Thus, MP structure determination is notoriously difficult (White 2004).

In order to overcome the barriers outlined above, we have instead proposed a simple protocol, based on empirical results, to streamline MP structure determination. First and

foremost, we believe that because of the paucity of structural information, any MP structure is inherently interesting. At every step in the pipeline we can therefore expect significant attrition. Positively, the remaining lower maintenance, more robust proteins lead to increased throughput. Secondly, we have focused the pipeline protocols and checkpoints (Figure 1) towards those methods which have historically been the most successful. To determine the success of this proposed pipeline, we therefore injected a moderate number, 120, of MPs from *E. coli* and determined their profile for expression through solubilization as described in the cell-free expression results of Chapter 5. Briefly, we observed that 44% of the proteins were overexpressed and successfully incorporated into the membrane. Furthermore, in solubilizations, 16 out of 24 (67%) were quantitatively solubilized in either n-Octyl- β -D-glucopyranoside (OG) or n-dodecyl- β -D-maltopyranoside (DDM). However, we wished to assay the second half of our proposed pipeline, so we next picked a model MP, CcmG, to pursue x-ray structure determination.

CcmG is a periplasmic thioredoxin-like protein involved in reducing the disulfide bond of apocytochrome-c proteins in preparation for heme ligation (Thony-Meyer 2002). CcmG contains the canonical Cys-X-X-Cys motif of thioredoxins, but is unique in that it is a selective redox partner in only cytochrome-c maturation as opposed to other thioredoxins involved in disulfide bond formation. This specificity is believed to originate from the unique groove near the redox active site and the site's conserved acidic environment. The specific role of CcmG within cytochrome-c maturation is unknown, although it is postulated CcmG exists within a membrane associated supramolecular protein complex.

Several structures of the soluble domain of CcmG have been determined, including those from *E. coli* (Ouyang, Gao et al. 2006), *Bradyrhizobium* (Edeling, Guddat et al. 2002), and *Mycobacterium* (Goulding, Apostol et al. 2004), which make solving the structure easy via molecular replacement (MR). Sequence analysis of CcmG shows it to be monotopic membrane protein, where the first 25 amino acids are part of transmembrane (TM) alpha helix. Finally, activity of CcmG can be assayed in a simple redox assay (Fabianek, Huber-Wunderlich et al. 1997). Therefore, CcmG is a model MP, with which we can test the pipeline protocol.

CcmG was determined to have excellent overexpression properties from the aforementioned expression screening results and was soluble in both OG and DDM. CcmG was purified and crystallized with a limited amount of work input. Following crystal optimization we were able to obtain crystals that diffracted to 1.4 Å and solved the crystal structure via MR. Thus, we believe the pipeline approach is well suited to MP structure determination.

Results

Cloning and Expression. As described in Chapter 5, 120 E. coli proteins were cloned into LIC vectors and their expression profiles were determined. CcmG (ID # 8) was determined to be among the best expressers, and qualitatively estimated to express at greater than ten mg / L culture. The original gel can be seen as lane Exp in Figure 2A. CcmG, along with the other expressers, was then tested for detergent solubilization. CcmG can be quantitatively solubilized in 270 mM OG at 4°C overnight as shown in the lanes marked BS and AS, which are before and after a spin to pellet the unsolubilized membrane fraction.

Purification. CcmG was scaled up to 4 L growths, from which we were able to obtain up to 30 mg / L of autoinduction media culture. Using the N-terminal 6xHis affinity tag, we purified CcmG with Ni-based immobilized metal affinity chromatography. The resulting product is seen in the Ni lane of the gel in Figure 2A and is over 95% pure. The upper band (near 37 kDa) is detected in a western blot with anti-6xHis antibody and may be a homoligomeric state, as seen with other MPs. The affinity tag was cleaved and the final product was injected on a Pharmacia Superdex 200 gel filtration column. The gel filtration profile is indicative of CcmG being PHS. The protein was also analyzed by mass spectroscopy, which indicated a pure sample at the expected molecular weight.

Crystallization and Data Collection. Purified protein (with 6xHis tag) was concentrated to ~15 mg / mL and screened for crystallization conditions in 96 well hanging drop

format with 100 nL drops (1:1 protein and precipitant). A number of promising conditions containing polyethylene glycol 4000 (P4K) were identified as containing crystals, microcrystals, or crystalline precipitate. Of these, 30% (w / v) P4K and 0.2 M ammonium sulfate (Nextal Classics #91) yielded small rods and two-dimensional plates (Figure 3A). This hit was reproducible by hand using the Nextal purchased solution and yielded crystals up to several hundred μm in the longest dimension (Figure 3B).

Unfortunately, attempts at reproducing the crystals with Fluka P4K or Hampton Research P4K was unsuccessful (Figure 3C). After measuring the pH of the crystallization solutions we determined that the P4K from the various vendors was of variable quality and buffering capacity. Specifically, the crystals could only be reproduced with a final of pH 5.4. Using a pH'ed solution of P4K we were able to reproduce and improve the crystals as seen in Figure 4D. Buffering agents as well as additives did not improve crystal quality. By increasing protein quality / purity we were able to grow crystals of roughly $500\ \mu\text{m} \times 500\ \mu\text{m} \times 20\ \mu\text{m}$ (Figure 3E). Crystals were extremely difficult to loop as they tended to grow from a central skin and had to be broken off from this region.

Crystals were looped and flash frozen in liquid nitrogen following a brief soak in mother liquor plus 15% ethylene glycol for cryoprotection. Diffraction intensities were collected on the Advanced Light Source Beamline 8.3.1 using an ADSC Quantum-Q210 CCD detector. Crystal mosaicity was extremely high (typically greater than 2°) and is evident in the diffraction pattern of Figure 4, which is nearly a full projection of reciprocal space onto the detector plate. This diffraction pattern belongs to the data set of the best crystal, which is shown in Figure 3E. This particular crystal was thicker than most and had a

mosaicity of roughly 1.4° . The intensity data were also less anisotropic (i.e. extremely weak in the c direction) for this crystal than in any other crystal screened, which is most likely related to the increased thickness, higher resolution, and lower mosaicity.

Phasing and Model Refinement. Data were processed using Elves (Holton and Alber 2004) and CCP4 (1994) (using MOSFLM (Leslie 2006)). The space group is I222 with a unit cell of $a = 52.7 \text{ \AA}$, $b = 67.4 \text{ \AA}$, and $c = 159.4 \text{ \AA}$. Analysis of the solvent content with molecular weight of 23 kDa is consistent with one molecule per asymmetric unit. The structure was solved by MR with the published 1.14 \AA Bradyrhizobium structure (Protein Data Bank Code 1KNG) using Phaser (Read 2001). The solution, positioned within the unit cell, is shown in Figure 5A. The Bradyrhizobium construct was missing the first 38 amino acids (containing the single TM helix) and this can be seen in Figure 5B where the empty region suggests the TM helix forms crystal contacts between sheets of monomers. Following MR, the model was refined with iterative cycles of manual building within Coot (Emsley and Cowtan 2004) and restrained refinement with individual B-factor refinement in Refmac5 (Murshudov, Vagin et al. 1997). The R_{free} for the refined structure is 28.9%.

Most strikingly, despite a resolution of 2.3 \AA , we were unable to locate the first 27 amino acids of the protein. The extent of the model (orange) is shown in Figure 6A, superposed on the Bradyrhizobium (green) structure. We were, however, able to locate an additional 15 amino acids that extended away from the soluble thioredoxin domain (Figure 6B), presumably in the direction of the putative helix. The electron density for the new

region, calculated with the terms $2F_o - F_c$, is (Figure 6C) is continuous and allowed for positioning of all side chains. There is no density prior to Ala28, indicating disorder of the region. Given that the first 25 amino acids is predicted to be a TM helix and a proline at position 33 indicates the end of a helix, we next tried to locate a helix of varying length. A MR search with helices from 7-25 amino acids failed to produce any solution with signal above background. The CCP4 program fffear also failed to produce any solutions for helices from 7-25 amino acids in a brute force real space search.

Discussion

Structure of CcmG. The monotopic MP CcmG was identified as a test case for the pipeline strategy and injected into the system. Briefly, CcmG was cloned and shown to express at greater than 10 mg / L of culture. CcmG was then solubilized, purified, crystallized, and atomic resolution diffraction data were collected at a synchrotron light source. Using the known structure, phases were determined via MR.

After the MR solution was determined, the initial model was iteratively built and refined to an R_{free} value of 28.9%, indicating a near complete model. We were able to identify an additional 15 amino acids N-terminal to the start of the initial model (Figure 6B and 6C), but surprisingly, unable to locate the initial 27 amino acids containing the single TM helix. This is most surprising given the crystal packing (Figure 5B), in which the TM must make the essential crystal contact in the *c* direction of the unit cell. One positive result for the location of the helix is shown in Figure 7. Indication of protein/solvent contrast was discovered by turning off the initial solvent b-factor correction in Refmac5, which may flatten a smaller signal. Furthermore, calculation of lower resolution maps (using data from 50 to 5-8 Å) also indicated the presence of electron density in this region. Unfortunately, neither map was interpretable. Further attempts at locating a helix using both MR programs and real space-methods failed.

It is possible, then, that the helix is extremely disordered in this direction, an idea supported by the inability of the crystals to grow any larger than 20 μm thick.

Furthermore in screening crystals we noticed an extremely variable range in the c dimension of the unit cells – values fluctuated from 140 to 170 Å. Thus the helix may be disordered, creating heterogeneity in microcrystalline environments, which poisons crystal growth. Or, it is also possible that the first 25 amino acids are not actually a helix and instead form a disordered loop, which participates in the essential crystal contact. Indeed, the structures of other monotopic membrane proteins (Picot, Loll et al. 1994) (Kurumbail, Stevens et al. 1996), do not form long membrane spanning helices in the manner of the prototypic single crosser glycoporphin A (MacKenzie, Prestegard et al. 1997). However, the primary sequence analysis programs we used (TMHMM and online tools at www.expasy.org) all indicated the presence of a helix from residues 1 to 25-27.

Although we have solved the structure of CcmG and added new information to the structure in the form of 15 additional amino acids, we are unable to locate the essential TM helix. This is unfortunate as one of the reasons CcmG was selected was because of functional data suggesting it functions as part of an oligomeric complex (Thony-Meyer 2002) and from the work of Engleman et al., it is well known single TM helices can participate in complex formation. Furthermore, we have secondary evidence from both denaturing gels and mass spectroscopy the protein may form a homooligomer. The two-fold symmetry operator for I222 and the position of CcmG within the unit cell also makes this possible. Thus, it will be essential to locate the helix. Given the problems associated the current crystallization conditions this may be accomplished by changing the protein construct or finding new crystal conditions. There are currently only six x-ray crystal structures of monotopic MPs. More structures of this class of proteins will elucidate the

regulatory and oligomeric role of single TM helices and further the definition of monotopic membrane proteins.

Pipeline Foundations. Besides solving the structure of a novel MP, the purpose of this study was to benchmark the efficacy of our proposed membrane pipeline. This pipeline, outlined in Figure 1, was created based on empirical knowledge culled from the literature and our laboratory. Essentially, the aim of this pipeline is to solve the highest number of membrane protein structures with the least amount of inputted work.

Ligation independent cloning (LIC) was chosen as the means of cloning as it yields multiple constructs for a given target in one polymerase chain reaction (PCR). Its efficiency is on par with other types of cloning and requires no proprietary enzyme mixtures. Our proposed pipeline is *E. coli*-centric, but in a LIC regime, one could also create constructs for multiple expression systems from one PCR product. It has been shown that increasing the number of fusion tags for a give target yields little improvement in expression space, where expression space is defined as the set of proteins which can be expressed in various systems for a global set of targets (Eshaghi, Hedren et al. 2005) (Surade, Klein et al. 2006). We therefore suggest a very small tag dimension (e.g. only an N-terminal and C-terminal 6xHis tag).

We next screen very small cultures and discard all non-expressers. It is our opinion that expression is the largest barrier to MP structure determination, so this step is perhaps the most essential. By creating an extremely strict cutoff one can increase the throughput of

those proteins that have the best chance of structure determination. We suggest expression of greater than 1 mg / L of culture (assayed qualitatively on gels) makes a good empirical cutoff.

Following the large culling of expressers from non-expressers, we transition towards the more labor intensive steps of purification. Membrane preparations from 250 ml cultures provide enough material for solubilization screens and pilot rounds of biochemistry. We propose, as a first pass, to only assay for solubility in the detergents OG and DDM. This is also a very strict cutoff, but informed by strong empirical evidence. One, based on the evidence presented in Chapter 5, most proteins passing the expression cutoff are also soluble in either OG or DDM. Secondly, OG and DDM are the two most successful detergents in MP structure determination (summarized by Hartmut Michel's database, <http://www.mpibp-frankfurt.mpg.de/michel/public/memprotstruct.html>). Lastly, and most importantly, the Stroud lab has generated a of body of empirical evidence that MPs which can be stably solubilized in either detergent are well-behaved in structure studies (at the very least crystallizable). This includes the MPs colicin Ia, GlpF, AqpZ, AqpM, Aqp0, AmtB, MerC, SecYEG, CcmG, AcrB, and Vglut homologs, among others. Thus, solubility in OG or DDM predisposes a target towards structure.

Once a protein has been expressed and solubilized it can be purified in the traditional manner as a soluble protein. The two strict barriers, expression and solubility, bias the remaining targets towards those that are well behaved and therefore require less input than other MPs. We estimate, based on the results of Chapter 5, 10-20% of all targets

will reach the purification phase. Upon reaching this stage, the targets can be scaled up (growths of, say, 6 l) and purified by immobilized metal affinity chromatography.

Affinity tags can also be removed from the protein at this point. Finally, as both a polishing step and a measure of monodispersity, the protein should be purified on a gel filtration column. Gel filtration gives some measure of molecular weight and homogeneity and is therefore quite useful before crystallization.

During purification and the steps following it, care must be taken to maintain a constant detergent concentration. Typically, solubilization screens are undertaken at roughly 10 times the critical micelle concentration and biochemistry at one to two times. After gel filtration, the protein will be concentrated and transitioned to crystallization trials.

During this concentration step, depending on the size of the molecular weight cut-off of the concentrator, the detergent will also be concentrated. Concentrated detergent must be dialyzed away after this step, and if not, will often negatively affect crystal trials. Most often, high detergents concentration will create large phase separations during trials, leading to a new source of error and irreproducibility. Light scattering data (Strop and Brunger 2005), suggests the micelle of DDM is roughly 70 kDa while our own data (unpublished) indicate an OG size of roughly 30 kDa. Therefore, one must be conscious of the micelle presence at every step (gel filtration, concentration, dialysis), where particle size is used to discriminate between protein and contaminants.

Finally, following the concentration / dialysis described above the MP can enter crystal trials. Ideally, crystal trials are carried out at protein concentrations above 10 mg / mL.

Crystal trials can be carried out in conditions similar to soluble proteins (hanging drop, 96 well, etc.), however, care must be taken to deal with the decreased surface tension due to the detergent. Also, crystal screens should be biased towards polyethylene glycol precipitating agents. Historically, such screens are the most successful for MPs. Following discovery of initial hits, the crystals can be optimized in the traditional manner and transitioned to data collection. Diffraction quality of crystals is the final barrier. Poor diffraction is perhaps the most frustrating aspect of structure determination due to the amount of work invested for a target at this point, however, our pipeline is suited to both circumvent and deal with this. OG is the single most successful detergent in terms of number of well-diffracting crystals, hence the bias towards it. Furthermore, the availability of other detergents and multiple tags for targets which have made it this far, increase the possibility of finding a well diffracting crystallization conditions for a given target. Finally, our pipeline is essentially optimized so that only the proteins most likely to diffract well have made it to the final step. One would expect a majority of them to be solvable, but current numbers do not allow us to determine statistics thus far.

We have described a set of protocols focused on a MP structure determination pipeline. This pipeline is empirically optimized to cull proteins at the essential barriers to increase throughput and quickly identify successful targets. We picked the model MP CcmG and injected it into the pipeline and the structure of CcmG was determined to 2.3 Å. Furthermore, as a statement to throughput possibility, the total amount of work inputted on CcmG from cloning to initial crystals was less than 20 hours. The pipeline is biased towards discovering well behaved proteins and similar results can be expected for other

targets. Thus, the pipeline is a viable approach to MP structure determination on a large scale.

Materials and Method

The following protocols can be used for the pipeline as described in Figure 1.

Cloning. We used a ligation independent cloning strategy that allows for PCR products containing LIC overhangs to be directly cloned into any of our LIC expression vectors (Aslanidis and de Jong 1990). For in vivo expression of proteins, genes were cloned into pET3a based LIC vectors, which contained an N-terminal TEV protease cleavable 6xHis tag. PCR reactions were done using Phusion polymerase (New England Biolabs) with *E. coli* genomic DNA and LIC cloned into our expression vectors.

In vivo protein expression and solubility. For small-scale in vivo protein expression, BL21 (DE3) C43 cells (Avidis) (Miroux and Walker 1996) were transformed with the in vivo expression constructs. Single colonies were used to grow 2ml of cells overnight at 37°C in auto induction media (0.5 % glycerol, 0.5% glucose, 0.2% a-lactose, 25mM Na₂HPO₄, 25mM M KH₂PO₄, 50mM NH₄Cl, 5mM Na₂SO₄, and 2mM MgSO₄) (Studier 2005). Cells were harvested and resuspended in 100mL lysis buffer (20mM Tris- HCl, pH 7.4, 150mM NaCl, 1mg/ml lysozyme (EMD Biosciences), complete protease inhibitor cocktail EDTA-free (Roche) and 10U/ml Benzonase (Novagen) for 1hr at 4⁰C. Sodium dodecyl sulfate (SDS) was added to a final concentration of 2% and incubated for an additional hr at 4°C. Cellular debris was pelleted at 16,000xg and supernatant containing SDS soluble protein was diluted with an equal volume of 2x SDS loading buffer (125mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 5mM BME, and 0.005%

Bromophenol Blue) and detected by Western Blot analysis using an anti-6xHis horseradish peroxidase conjugated antibody (Santa Cruz Biotech).

For large scale protein expression, 250ml-12L of auto-induction media growth from a single BL21(DE3) C43 colony was harvested and resuspended in lysis buffer (20mM Tris-HCl, pH 7.4, 100mM NaCl, 100 μ M PMSF, and 4mM BME). Cells were lysed with the EmulsiFlex and undisturbed cells were pelleted at 10,000xg for 30min. The supernatant was pelleted at 200,000xg for 1hr to collect membranes. Membranes were solubilized in OG or DDM solubilization buffer (20mM Tris pH 7.4, 150mM NaCl, 10% glycerol, 4mM BME, 100 μ M PMSF, and 270mM OG or 10mM DDM) at 4°C overnight, which we call the before-spin. Soluble protein was collected from the supernatant of a 200,000xg spin for 30min, which we call the after-spin. A qualitative analysis of protein solubility in DDM and OG was done by comparing the intensity of the before-spin and after-spin band on a Western blot.

CcmG purification. CcmG from 4 L of culture was obtained as described above. Protein soluble in OG was incubated at 4°C with IMAC Ni-NTA resin for 30min. Protein bound beads were washed with OG size exclusion buffer (SEC) (20mM Tris-HCl pH 7.4, 100mM NaCl, 10% glycerol, 2mM DTT, 40mM OG) containing 15-25mM imidazole and eluted with OG SEC buffer containing 300mM imidazole. The elution was then desalted into OG SEC buffer using a 10DG disposable desalting column (Biorad) and concentrated to 3 ml in a 10 kDa MWCO Ultra Centrifugal Filter (Amicon). The concentrated protein was injected on a Superdex 200 gel filtration column at 0.33ml/min

running the SEC mobile phase. Peak fractions were collected and assayed by SDS-PAGE. The single peak fraction containing purified protein was collected and concentrated to 15 mg/ml for crystal screens. The 6xHis affinity tag was cleaved off by the addition of 1 / 30 (w/w) TEV protease to the sample and incubated overnight at room temperature. Ccmg was concentrated to 15 mg / ml in a 10 kDa MWCO Ultra Centrifugal Filter and dialyzed overnight against SEC mobile phase for 2 days at 4°C to remove detergent. It was crystallized and the structure was solved as described in Results.

Table 1: Crystallographic Data and

Refinement Statistics

CcmG

Data Collection

Space Group	<i>I</i> 222
Unit Cell	
<i>a</i> (Å)	52.7
<i>b</i> (Å)	67.4
<i>c</i> (Å)	159.4
Resolution range (Å) ^a	50-2.1 (2.21- 2.10)
Unique reflections	94618
Completeness ^a	98.8 (95.8)
R _{sym} ^b (%) ^a	15.2 (91.7)
Average I / σ(I) ^a	14.2 (2.4)
Mosaicity (°)	1.4

Refinement Statistics

Resolution range (Å)	50-2.3
R _{work} / R _{free} (%)	23.9 / 28.9
Number of protein atoms	1231
Number of solvent atoms	32

Average B-factor (\AA^2) 30.5

^a values in parenthesis refer to the highest-resolution shell

^b $\Sigma |I - \langle I \rangle| / \Sigma I$, where I equals observed intensity and $\langle I \rangle$ equals average intensity for symmetry-related reflections

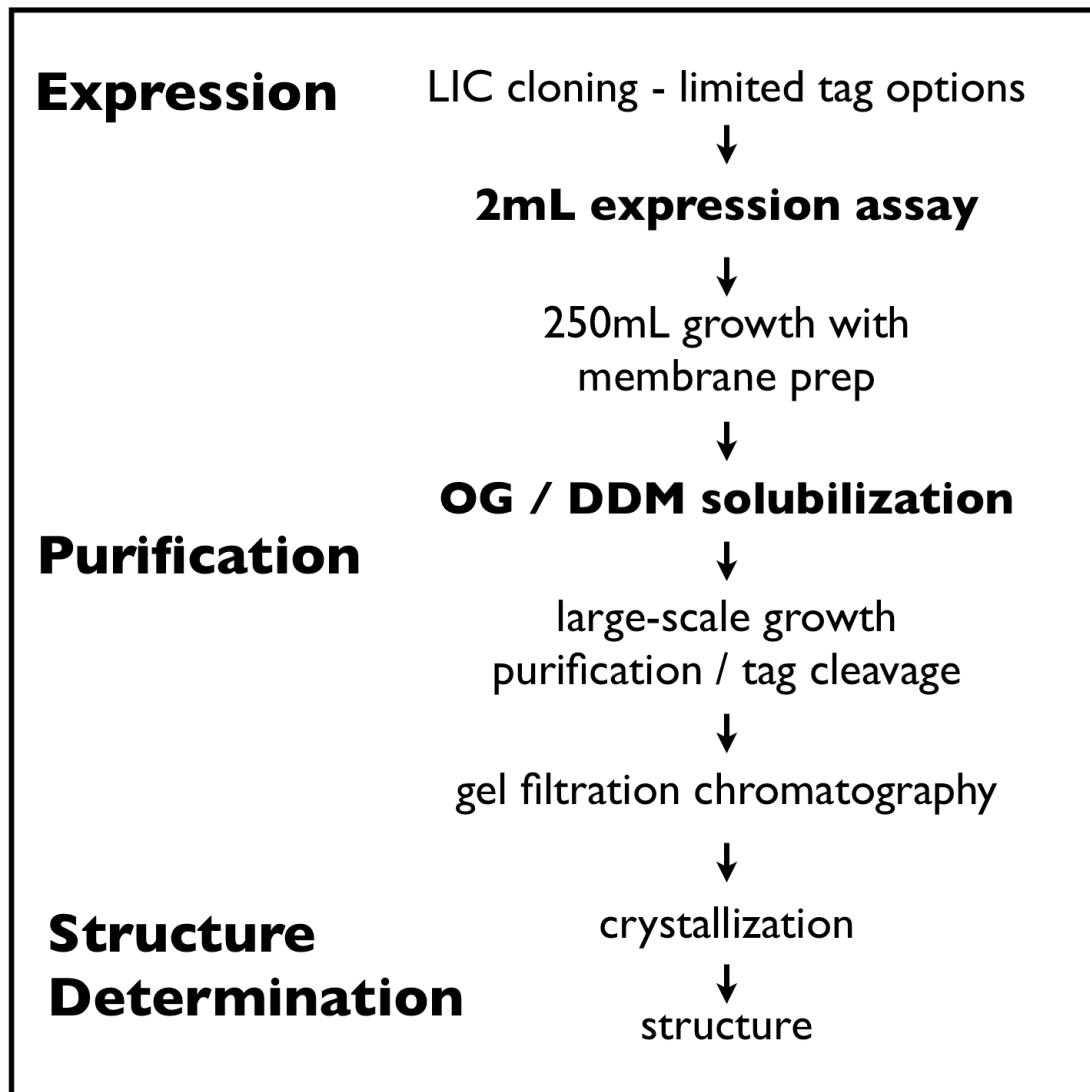


Figure 1. Membrane protein structure determination pipeline. The general steps are labeled in the left column, while specific steps in the right. The two rigorous checkpoints, expression and solubilization, are denoted in bold.

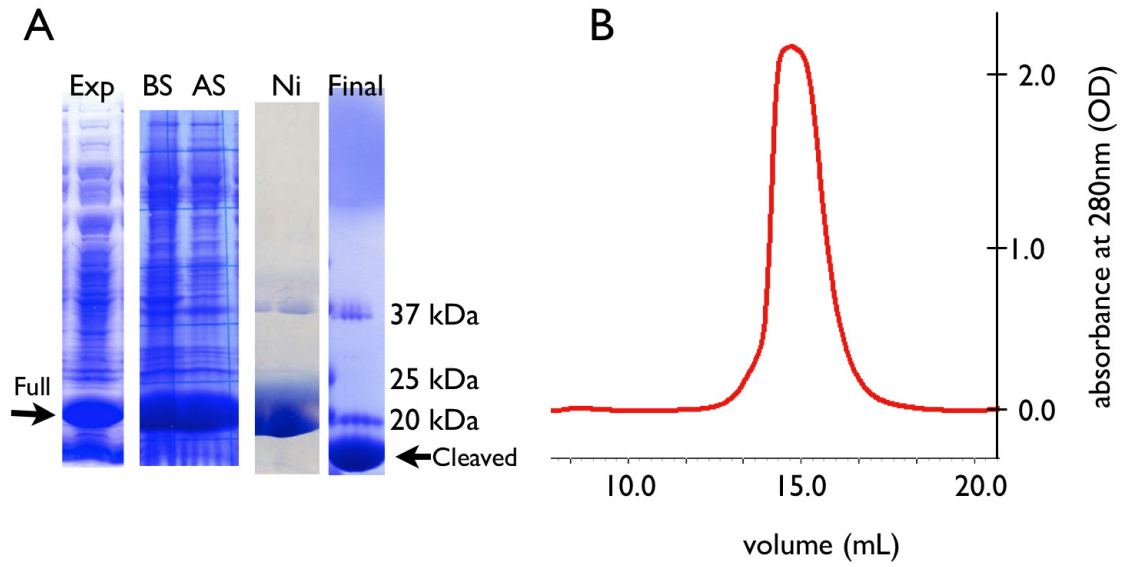


Figure 2. Purification of CcmG. (A) Comassie-stained denaturing gel of the purification of CcmG. Labels are for crude expression in the membrane (Exp), before solubilization (BS), after solubilization (AS), after Ni column (Ni), and final cleaved product after gel filtration (Final). CcmG with tag is 23 kDa. (B) Gel filtration profile of CcmG. CcmG was purified on a Superdex 200 gel filtration column – the peak is indicative of being PHS.

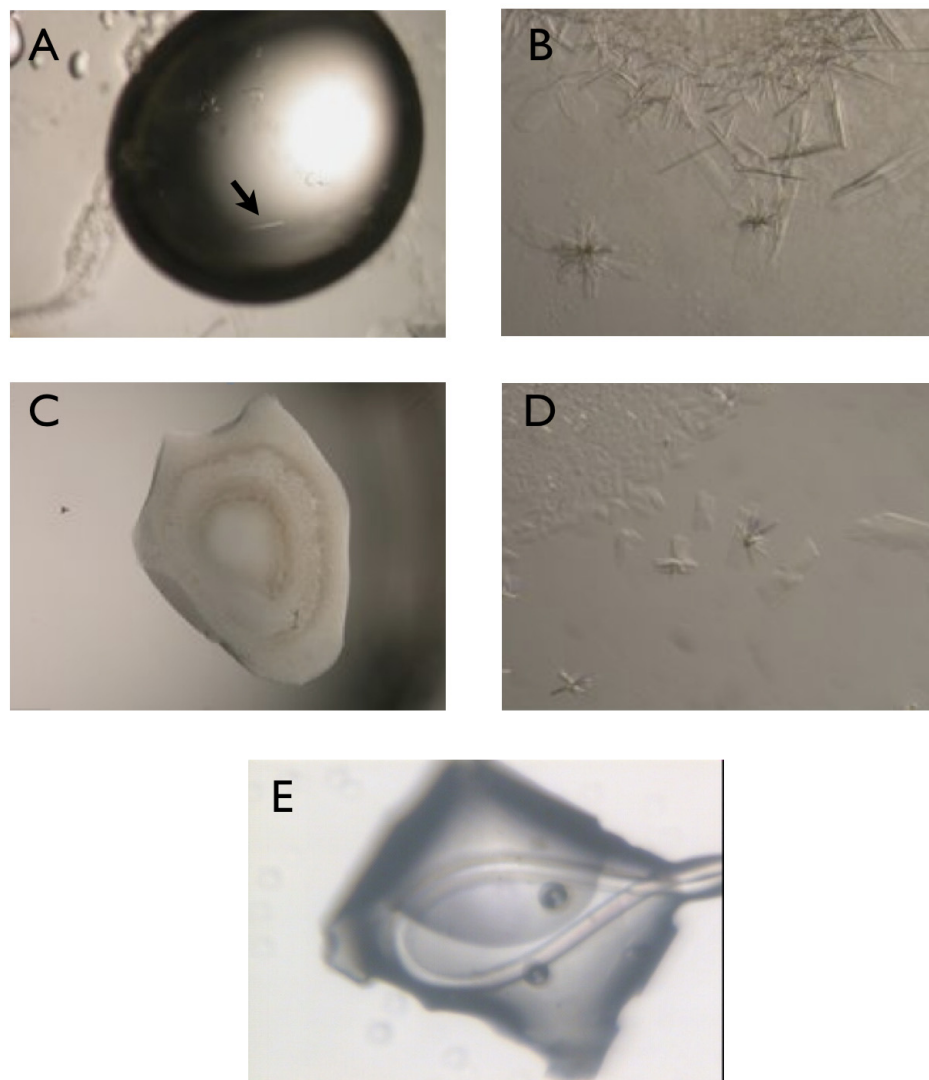


Figure 3. Crystallization of CcmG. (A) Initial hit found in Nextal Classics #91 with 30% (w / v) P4K and 0.2 M ammonium sulfate. (B) Zoomed in photograph of the condition in A. (C) Failed crystal reproduction using Fluka P4K. (D) Successful reproduction of crystals using a pH'ed solution of P4K. (E) Crystal used to solve structure under data collection at ALS Beamline 8.3.1.

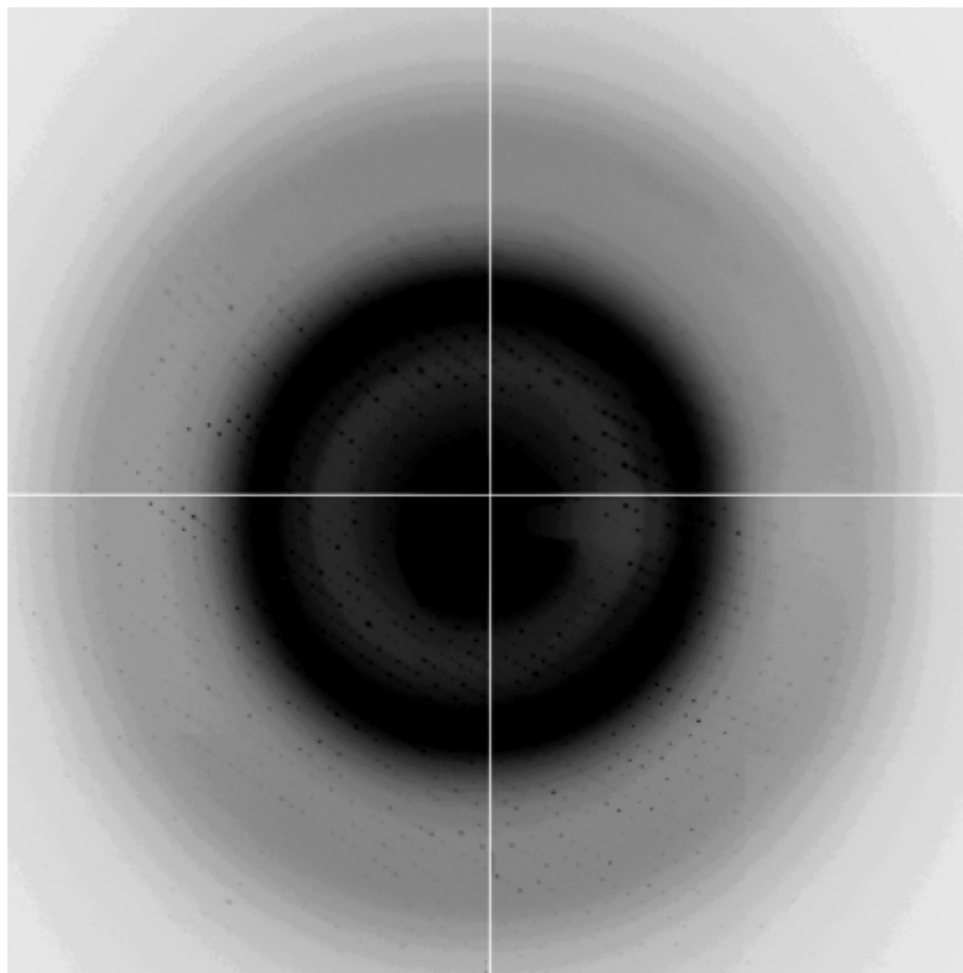


Figure 4. Diffraction pattern from crystal in Figure 3E. Image is at high contrast to show high resolution spots. Dark circle denotes solvent ring. Spots were measured out to 1.4 Å and data set it complete to 2.1 Å.

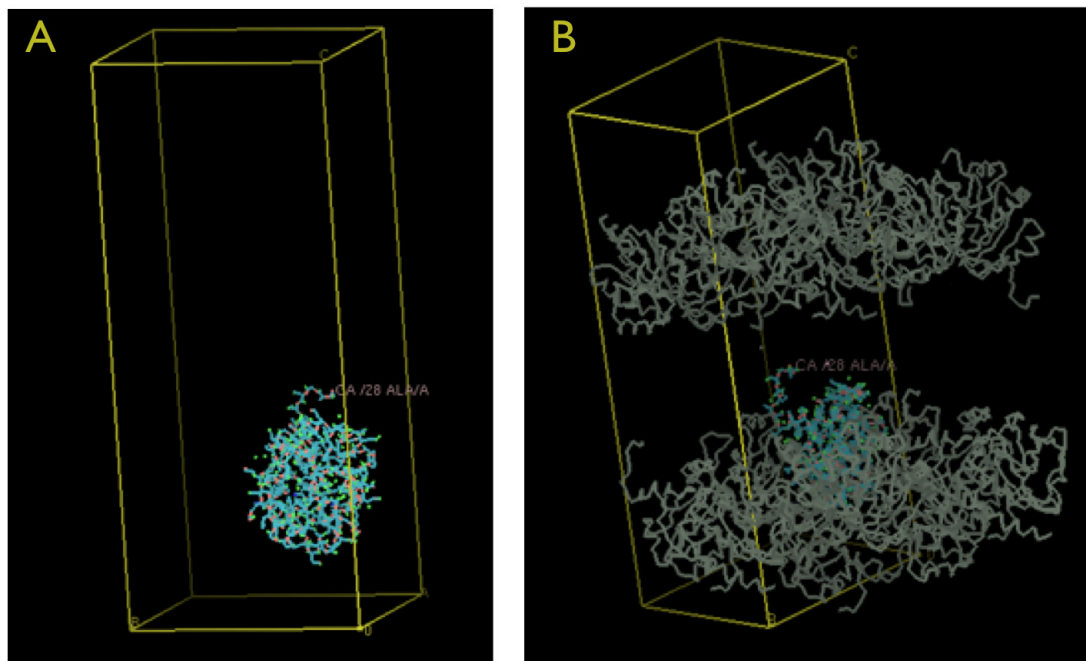


Figure 5. MR solution for CcmG. (A) CcmG monomer positioned in the unit cell. Space group is $I222$ where $a = 52.7 \text{ \AA}$, $b = 67.4 \text{ \AA}$, and $c = 159.4 \text{ \AA}$. (B) Packing within unit cell. The missing helix is located somewhere in the solvent region between the sheets of protein.

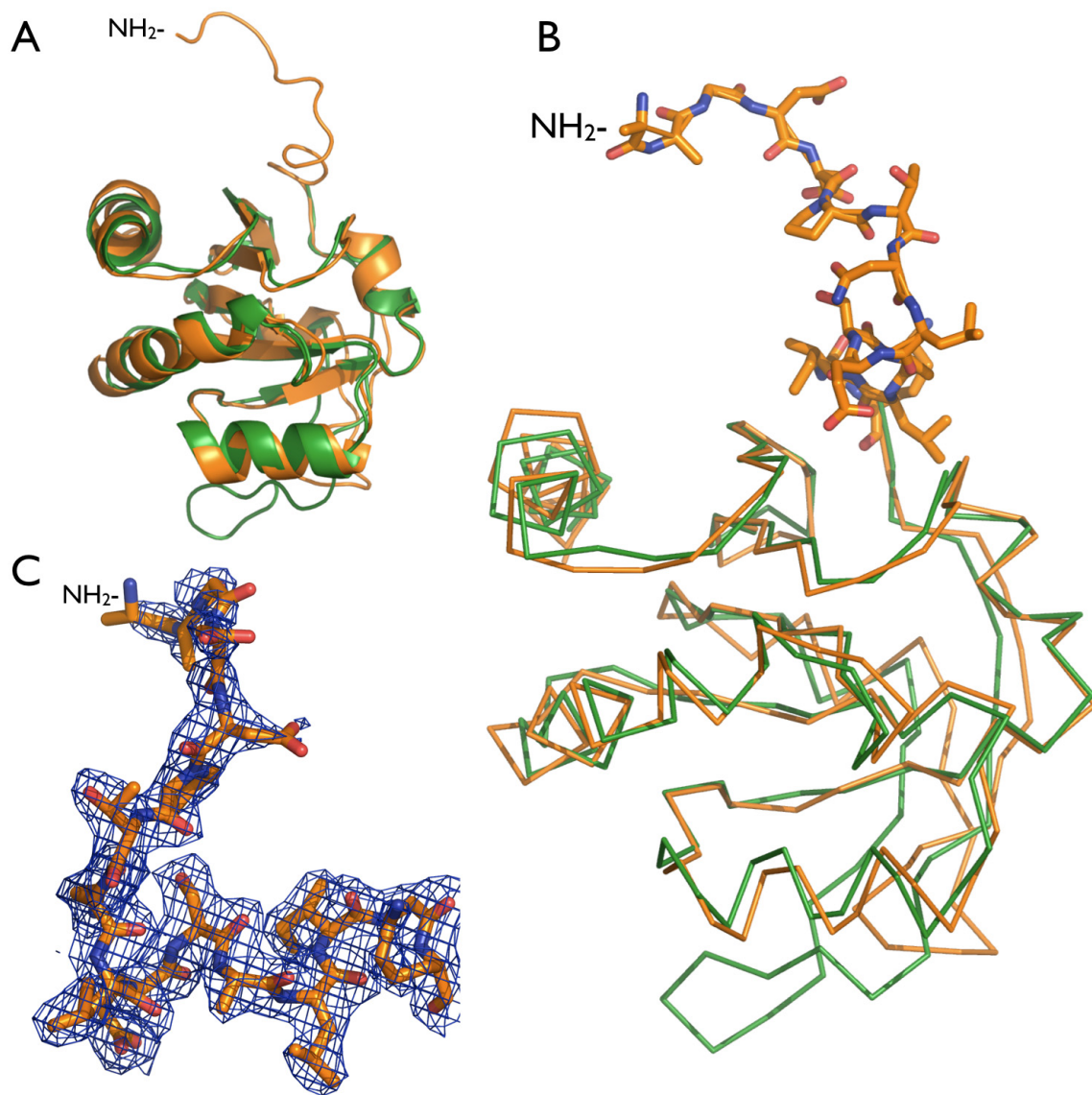


Figure 6. Structure of CcmG. (A) Cartoon overlay of CcmG (orange) with the *Bradyrhizbium* (green) structure (1KNG) used for MR. (B) Ribbon overlay with similar color scheme. The 15 additional amino acids we were able to locate are shown in sticks. (C) $2F_o - F_c$ electron density map contoured at 1σ for the 15 new residues.

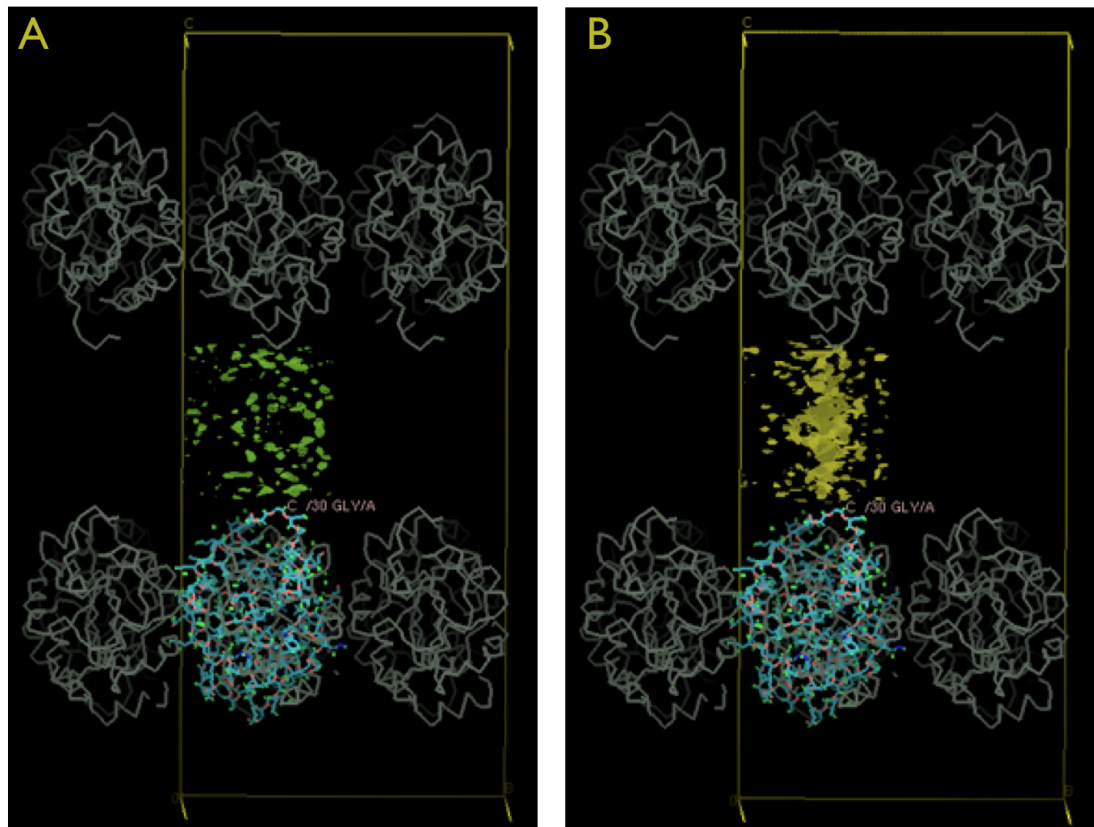


Figure 7. Possible solvent / protein contrast for the missing helix. (A) $2F_o - F_c$ density map contoured at $.8 \sigma$ using the simple (exponential scaling) bulk solvent correction in Refmac. (B) Same map as A, but calculated without solvent correction. This map shows features of solvent / protein contrast at the specific region displayed.

References

- (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-3.
- Edeling, M. A., L. W. Guddat, et al. (2002). "Structure of CcmG/DsbE at 1.14 Å resolution: high-fidelity reducing activity in an indiscriminately oxidizing environment." Structure **10**(7): 973-9.
- Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.
- Eshaghi, S., M. Hedren, et al. (2005). "An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins." Protein Sci **14**(3): 676-83.
- Fabianek, R. A., M. Huber-Wunderlich, et al. (1997). "Characterization of the *Bradyrhizobium japonicum* CycY protein, a membrane-anchored periplasmic thioredoxin that may play a role as a reductant in the biogenesis of c-type cytochromes." J Biol Chem **272**(7): 4467-73.
- Goulding, C. W., M. I. Apostol, et al. (2004). "Gram-positive DsbE proteins function differently from Gram-negative DsbE homologs. A structure to function analysis of DsbE from *Mycobacterium tuberculosis*." J Biol Chem **279**(5): 3516-24.
- Holton, J. and T. Alber (2004). "Automated protein crystal structure determination using ELVES." Proc Natl Acad Sci U S A **101**(6): 1537-42.
- Kurumbail, R. G., A. M. Stevens, et al. (1996). "Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents." Nature **384**(6610): 644-8.

- Leslie, A. G. (2006). "The integration of macromolecular diffraction data." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 48-57.
- MacKenzie, K. R., J. H. Prestegard, et al. (1997). "A transmembrane helix dimer: structure and implications." Science **276**(5309): 131-3.
- Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-55.
- Ouyang, N., Y. G. Gao, et al. (2006). "Crystal structures of E. coli CcmG and its mutants reveal key roles of the N-terminal beta-sheet and the fingerprint region." Proteins **65**(4): 1021-31.
- Picot, D., P. J. Loll, et al. (1994). "The X-ray crystal structure of the membrane protein prostaglandin H2 synthase-1." Nature **367**(6460): 243-9.
- Read, R. J. (2001). "Pushing the boundaries of molecular replacement with maximum likelihood." Acta Crystallogr D Biol Crystallogr **57**(Pt 10): 1373-82.
- Strop, P. and A. T. Brunger (2005). "Refractive index-based determination of detergent concentration and its application to the study of membrane proteins." Protein Sci **14**(8): 2207-11.
- Surade, S., M. Klein, et al. (2006). "Comparative analysis and "expression space" coverage of the production of prokaryotic membrane proteins for structural genomics." Protein Sci **15**(9): 2178-89.
- Thony-Meyer, L. (2002). "Cytochrome c maturation: a complex pathway for a simple task?" Biochem Soc Trans **30**(4): 633-8.

White, S. H. (2004). "The progress of membrane protein structure determination." Protein Sci **13**(7): 1948-9.

Appendix 1

AQUAPORINS: INTEGRAL MEMBRANE CHANNEL PROTEINS

Research completed in collaboration with

Robert M. Stroud, William E. C. Harries, John Lee, and Shahram Khademi

This chapter is adapted from a review originally appearing in Structural Biology of Membrane Proteins (Grisshammer, G. and Buchanan, S. eds), pp. 195, Royal Society of Chemistry, Cambridge.

INTRODUCTION

The first of what today are called ‘aquaporins’ (AQPs) were bacterial glycerol channels. Alfred Fischer recognized the phenotype over 100 years ago. He described pathogenic bacteria that when placed in hyperosmotic glycerol solutions failed to undergo lysis¹. He concluded that the membranes had to be highly permeable to glycerol in these organisms, but not in the organisms that shriveled up and died. This stimulated interest in glycerol channels, but since the substrates are uncharged the field lagged behind that of ion channels where electrical phenomena were evident. The glycerol channel from *E. coli*, has a special place in AQP history. Genetic analysis pioneered by E.C.C. Lin and colleagues identified the ‘glycerol facilitator’ GlpF². This work led to the cloning and sequencing of the gene³, followed by characterization of the 281 amino acid 29,780 Dalton GlpF protein⁴. The crystal structure of GlpF became the first AQP structure to be determined at atomic resolution⁵.

Early functional studies of GlpF characterized⁶ it as a highly selective transmembrane channel that conducts water, glycerol, and other small uncharged organic molecules such as, urea, glycine, and D,L-glyceraldehyde. Inside the cell, glycerol is rapidly phosphorylated by glycerol kinase to produce glycerol-3-phosphate, which is no longer a substrate for any back flow, and so retains the substrate and maintains the gradient of the substrate glycerol from outside to inside. Glycerol-3-phosphate proceeds by dehydrogenation to dihydroxyacetone phosphate (DHAP), or onward to phospholipid synthesis, where glycerol provides the attachment base for fatty acid chains and

phosphatidyl headgroups in $\sim 2/3$'s of the cellular phospholipids. Stimulating *E. coli* growth on glycerol is the glp regulon, which is inducible by glycerol 3-phosphate. GlpF is also stereo and enantio selective in conductance of linear carbohydrates, alditols^{2, 5}. Aldoses or sugars -- the cyclized alditols -- are not conducted through the GlpF channel. The structure of the GlpF channel shows that it is indeed too small to conduct cyclic molecules, illustrating the basis of stereo and enantioselectivity^{5, 7}.

GlpF conducts water at about 1/6 of the rate of its *E. coli*, homolog, AQPZ, which conducts water but not glycerol. The presence of two AQPs even in *E. coli* illustrates the necessity for differentiation of function to cover the spectrum of activities within the AQP family. In mammals, 13 AQPs are recognized termed AQP0 up to AQP12⁸. MIP26 from the eye lens was one of the first mammalian protein to be recognized as a water channel in 1974^{9, 10}. Water channels were recognized in plants in the early 1980s. In 1990, the genetic similarity between GlpF, the major intrinsic protein (MIP) of the eye lens that was well known to act as a water channel, and a soybean nodulin-26 had been described by Saier and Baker¹¹. The red blood cell contains a water channel protein, that was called CHIP28 (abbreviated from channel-like integral membrane protein, a 28 kiloDalton protein). These now all fall into the 'Aquaporin family' so named by Peter Agre in 1992. Peter Agre shared the Nobel Prize in 2003 for his pioneering work in the ensuing decade that elaborated properties of the 'AQP' family¹².

AQPs are comprised of three functionally distinct subgroups that include transmembrane water conducting channels (aquaporins), and channels that conduct glycerol, perhaps the

most relevant physiological substrate in humans (i.e., GlpF), called aquaglyceroporins, and most recently, aquaamminoporins¹³. These channels conduct water, but also variously conduct urea, D,L-glyceraldehyde, linear polyalcohols (called alditols), other small organic molecules¹⁴⁻¹⁶ and ammonia⁸. In humans, AQP-3, AQP-7, AQP-9 and AQP-10 are in the aquaglyceroporin subclass¹⁷. The AQPs can conduct their substrates at close to the diffusion-limited maximum rate through a pore of this cross section. Many eukaryotic AQPs are regulated by phosphorylation, pH, osmolarity or the binding of other proteins or ligands^{18,19}.

The amino acid sequence between any two AQPs typically shows conservation in the range 28–32%. They are all constructed around a highly conserved structural backbone fold of six trans-membrane and two half membrane-spanning helices numbered M1 to M8 that surround a central water filled channel (Figure 1). The family arose by tandem intragenic duplication²⁰ such that the N-terminal segment displays ~20% conservation with the C-terminal segment²¹. This duplication occurred early in evolution since bacteria contain both an aquaglyceroporin (GlpF) and an aquaporin (AQPZ). Near the center of each segment there is a conserved -Asn-Pro-Ala- signature sequence (-NPA-).

A key question is how AQPs exclude conduction of all charged molecules and ions including hydroxide, hydronium ions²², and protons. The structures of now over 5 AQPs instruct as to the determinants of AQP selectivity^{5, 23-25}. Mutational analysis, and molecular mechanics seek to evaluate the contributions of each factor to this property²⁶. There is only one exception to the absolute insulation against ions in human AQPs:

AQP6. AQP6, which conducts ions at low pH. This adaptation most probably controls its action in membranes of intracellular organelles²⁷.

In this chapter we focus on the information that derives from the 5 structures of AQPs at atomic resolution. The amino acid sequences of all five AQPs are aligned in Figure 2A. The residues that line the pore are colored and represented by spheres in Figure 2B to indicate where in the structure they lie. Each AQP has different conductivity, substrate specificity, and role in the biology of the organism. Thus, they serve to extract features of the molecular structures that encode these properties. The common elements of structure serve to illustrate the elements of function that are preserved in the family. We begin by focusing on the mechanisms by which the AQPs that are normally filled with a continuous line of hydrogen bonded water molecules, remain insulating to protons and ions. This is followed by what we learn about the selectivity of AQPs for what they do conduct. Finally we focus on functional associations between AQPs and their role in higher levels of cellular organization in the eye lens.

THE EXCLUSION BARRIER TO IONS AND PROTONS IN AQUAPORINS

The electrochemical gradient across biological membranes, established by active transport, is the basis of energy generation and storage in most organisms. In bacteria the gradients are usually protonic. In mammals they are generally Na⁺/K⁺ ion gradients. These gradients must be maintained and any disruption can induce cellular distress. The cell, however, also requires the passive and active coincident transport of many compounds. Thus, the membrane and the proteins therein, must balance both selectivity and permeability. Aquaporins demonstrate this balance by selectively conducting water and small amphipathic molecules with complete exclusion of ions (particularly protons)²⁸.²⁹ It has long been known that bulk water and channels able to conduct water, such as gramicidin, can also conduct protons by the Grothuss “hop and turn” mechanism where hydrogen bonds among a chain of waters are rapidly realigned resulting in a proton transfer with no net water movement³⁰. Though aquaporins contain a similar single-file chain of water, they display no such conductance of protons²⁹. The explanation of this phenomenon has interested both theoretical and experimental scientists alike and there are three current explanations, all of which are probable components in varying magnitudes to the proton exclusion barrier⁵.

Global Orientational Tuning by the NPA motif

The first concept for how this might be encoded is ‘global orientational tuning’, that centers on the conserved asparagine-proline-alanine (NPA) region near the middle of the pore²⁶. This conserved sequence, present in both the N-terminus and C-terminus of all

aquaporins, is seen as the signature motif of aquaporins and high-resolution x-ray structures suggest a possible role in proton exclusion (see Figure 3). The two NPA motifs cap the N-terminal end of helices M3 and M7, which meet in the center of the membrane along the quasi-twofold axis. At the twofold, the proline rings are in Van der Waals contact with each other, while the asparagine side chain is constrained by two hydrogen bonds in a way that orients the amide chemical moiety into the pore toward hydrogen bond acceptors on the permeant substrate⁵.

In the case of water, the central water's two lone electron pairs accept two hydrogen bonds from the two asparagine amide groups and is therefore aligned to donate two hydrogen bonds to the neighboring water molecules on either side of the central one. These neighboring water molecules each are hydrogen bond donors first to one of a line of equally spaced carbonyl oxygen atoms that run outward from the NPA region to the external surfaces of the pore. In turn it is therefore oriented to present its second hydrogen outward toward the next water molecule, and so on throughout the length of the conduction pathway. While this 'global orientation' is deduced from the crystal structure, the resolution of the X-ray structures is not quite adequate to 'see' the hydrogen orientation. This alignment has been reiterated multiple times in molecular mechanics simulations and supports the conclusion from the structure. This ordering of the line of water molecules implies that as a water molecule enters the pore its dipole will be oriented generally towards the entrance of the channel, it moves through the pore until it reaches the center where it rotates to interact with the NPA motifs, it then continues to

rotate until its dipole points towards the exit, and it progresses through the rest of the pore^{26, 31}.

The hypothesis is that this ordering makes it difficult for the central water molecule to receive a proton from its nearest neighbor to become a hydronium ion (H_3O^+) since its hydrogen points toward the source of the proton. Viewed a different way, if a proton were to leave the central water toward to the inside, that water would donate a proton to its neighbor. However that water would then need to rotate in its position to regenerate the hydrogen bonded alignment. The covariant alignment of the line of waters might make this difficult and so provide a barrier. This mechanism is bolstered by multiple molecular dynamics simulations that reiterate the co-alignment of the entire line of water molecules. However, what is needed, is a simulation where the energetic cost of moving a proton or a hydronium ion along the axis of the channel is calculated. That kind of simulation is much harder to accomplish since it involves the movement of an ion, and the energetic costs of partial dehydration of the ionic species must be built in correctly. There is no good estimate of this effect. One simulation described the barrier as centered at the NPA region and was able to assign a value of nearly 2 kcal/mol to the barrier, by selectively ‘turning off’ specific coulombic interactions in silico³².

Helix Dipole

Recent structural work has elucidated the chemical toolkit of transmembrane channels and how these proteins function with the chemical implications of their location in a low dielectric medium. One such tool is the use of helix dipoles to interact, both favorably

and unfavorably, with charged species³³. In aquaporins, the half-membrane spanning helices M3 and M7 meet at the quasi-twofold axis and project their N-termini into the pore⁵. These two helix dipoles work in synergy creating an electrostatic field that opposes entry of positive charge into the channel. Electrostatic calculations in which one can switch the dipole effect off, have set the barrier height at roughly 3 kcal/mol. Furthermore, by calculating the static field of proton conductance, Roux and colleagues predicted that mutating one of the two NPA motifs to DPA would be enough to negate the dipole effect and therefore the channel would conduct protons³². In our hands, the N68D point mutant, while functional for water conductance, still does not conduct protons in proteoliposome assays (unpublished results). This argues against the role of the helix dipoles in maintaining the ordering of water, in so far as, the alignment is responsible for proton exclusion.

Electrostatic desolvation penalty

A third hypothesis to explain the insulation to proton conductance is based on the electrostatic free energy of transferring charge from a high dielectric medium such as bulk water to the low dielectric protein channel. The enthalpy of hydration for most ions is on the order of 100 kcal/mol, and thus, dehydration is an extremely unfavorable process. Ion conducting channels, like the potassium channels have evolved an elaborate four-fold axis of mainchain carbonyls that can favorably “solvate” potassium, yet do not act as binding sites to slow conduction all together³⁴. The exclusion of ions from channels that do not compensate for the water of hydration is the rule and conduction is the carefully constructed exception.

In the aquaporins, electrostatic simulations compare the channel with a simple macroscopic hydrophobic channel and find very little difference, suggesting that it is simply the electrostatic nature of the pore and an inability to solvate charge that is the barrier to proton conduction³⁵. This is the mechanism we suggested based on the structure that would prevent leakage of ions such as Na⁺ through the channel⁵. It clearly must apply to the bulk transfer of a hydronium or a Zundel ion (H₅O₂) through a channel. This mechanism differs in principle from the hypothetical ‘hopping’ of protons between waters along the line, which one might imagine could even occur in a concerted fashion such that there would never be a formal charge within the line or in the center of the channel. It is this latter mechanism that the orientation dependent proposals seek to address.

SELECTIVITY IN THE AQUAPORIN FAMILY

The extraordinary permeation rate of more than one billion molecules per second makes AQP one of the fastest membrane channels. The high rate of permeation along with the strict selectivity for water and polyols raises fascinating questions concerning the structural basis for substrate selectivity in AQP family.

GlpF from *E. coli* was the first member of the AQP family whose structure was determined by X-ray crystallography to a high resolution of 2.2Å⁵. GlpF facilitates passive and selective permeation of water and small-uncharged organic molecules, such as glycerol, across plasma membranes of the cell. The next water channel whose crystal

structure was determined was AQP1 from bovine red blood cells²³. To date, the atomic resolution crystal structures of five members of the AQP superfamily have been determined, namely GlpF, AQP1, AQPZ²⁴, AQP0²⁵ and AQPM (unpublished). Comparisons of the structures of aquaporins with aquaglyceroporins may explain the basis for their different profile of selectivity. All AQPs form a homotetramer of four AQP monomers that contain single channels arranged around a four-fold symmetry axis. The tetramers are generally stable even in the presence of detergents as shown by ultracentrifugation, size exclusion chromatography, and mass spectrometry^{5,36}.

The AQP channel pathway, defined by bound water or glycerol, is bounded by extracellular and intracellular vestibules and connected by a $\sim 28\text{\AA}$ long amphipathic channel⁷. The channel contains two highly conserved regions, the selectivity filter, which is the narrowest point in the entire channel, and the NPA region^{5,36}. The selectivity filter in aquaporins is generally narrower than in the aquaglyceroporins, matching the difference between a water, and a carbon backboned alcohol (Figure 3).

The selectivity filters for AQP1 and AQPZ, water channels from bovine and *E. coli* respectively, have 'diameters' (defined in spherical terms) of $\sim 2.0\text{\AA}$ ^{23,24}, while GlpF, an aquaglyceroporin from *E. coli*, has the widest selectivity filter of 3.4\AA ⁵. To a first approximation, the preference of aquaporins for water and of aquaglyceroporins for glycerol can be explained by the size of the selectivity filter. The larger glycerol molecule, compared to the size of a water molecule, needs a wider selectivity filter in GlpF. AQP0, a water channel from eye lens, has the narrowest selectivity filter ($\sim 1.4\text{\AA}$)

of all the AQPs with high resolution crystal structures, which is consistent with AQP0 being a very poor water channel²⁵.

The selectivity filter in GlpF is strongly amphipathic, with the planes of two perpendicular aromatic rings (W48 and F200) forming a hydrophobic corner (Figure 3). The alkyl backbone of glycerol is tightly packed against this corner, leaving no space for any substitution at the C-H hydrogen positions. Each of the three hydroxyl groups of glycerol are hydrogen bond acceptors from successive NHs of the guanidinium group of R206 and hydrogen bond donors to the carbonyl oxygens of G199 and F200, respectively. The buried carboxyl group of conserved E152, orienting the three adjacent carbonyls of G199, F200, and A201 in the extracellular/periplasmic vestibule, may also increase the negative charge on the carbonyl oxygens of F200 and A201⁵. The binding of permeant molecules in the selectivity filter makes it possible for the negative charge of E152, acting through the amides of 199-201, to form an electrostatic interaction with positively charged R206, through the substrate; The amide carbonyls of F200 and A201 act as hydrogen bond acceptors from successive hydroxyl OHs of the substrate, that in turn accept hydrogen bonds from each of two N-H groups of the positively charged R206. This implies that permeant molecules should be polarizable in cross-section, such as glycerol OHs and water.

In AQPZ, the selectivity filter is formed by three hydrophilic residues of H174(191, the residue abbreviations and numbers in the parentheses refer to the GlpF amino acid sequence), and R189(206) and T183(200), and one hydrophobic residue of F43(48). The

AQP1 selectivity filter is almost identical, with cysteine substituted for threonine T183. This cysteine explains the inhibition of water transport in AQP1 by mercury.

The selectivity filters in AQPZ and AQP1 have three polar residues and one non-polar residue sidechain, whereas in GlpF, the ratio is inverted; the selectivity filter contains only one polar residue. This might partially explain the higher glycerol conductivity and lower water conductivity in GlpF.

A microbial MIP from *Lactococcus lactis*, named Glal_{Lac}, has been demonstrated by expression in oocytes, to be permeable to glycerol, at the same rate as *E. coli* GlpF, and permeable to water at the same rate as in *E. coli* AQPZ³⁷. In the aquaglyceroporin Glal_{Lac} compared to AQP1, small non-polar residues, V223 and P232, replace H182 and C191 respectively. Interestingly, F58 is replaced by Y49 a polar residue. Consequently, the selective filter of Glal_{Lac} is made of two polar and two non-polar residues. These substitutions with small residues, would result in an enlargement of the constriction region with a potential aperture larger than the aquaporin one, a size compatible with a glycerol channel, and the switch of F58 to Y49 provides the necessary polar environment for an efficient water channel³⁸.

PERMEATION OF SUBSTANCES OTHER THAN WATER AND GLYCEROL

Conductance of Other molecules

Despite the similarity in the general architecture across the aquaporin family, *in vivo* and *in vitro* conductance assays show the selectivity spectra of aquaporins extending to the

permeation of small molecules such as CO₂ (AQP1³⁹, tobacco aquaporin NtAQP1)^{40,41}, nitrate (AQP6)⁴², and urea (AQP3, AQP7⁴³, AQP9⁴⁴, AQP10)⁴⁵. For example, cells such as red blood cells or plant leaves have situational demands for the specific and efficient conductance of a gas such as CO₂, and the characterization of aquaporins with this functional capacity supports the observations that certain aquaporins may be gas channels.

The archaeal aquaporin AQPM probably conducts CO₂ or H₂S, and the structure of AQPM, especially at the selectivity filter, suggests an adaptation for the conductance of a permeant that is larger and less polar than water (Figure 4). The hydrophobicity and size of the channel, especially at the selectivity filter, leads to the speculation that the differences in AqpM compared to that of AQP1 may be an adaptation that enables it to function as a multi-functional channel that conducts either H₂S or CO₂ in addition to water. *M. marburgensis* relies on H₂S as the terminal electron acceptor in its energy production pathway, and due to the structural similarity of H₂S to H₂O, the mechanism of selectivity for these molecules must necessarily be very similar and points to H₂S as a likely candidate for conductance by AQPM.

The wider and more hydrophobic selectivity filter of the AQPM channel (2.54 Å), in comparison to that of AQP1 (1.86 Å), is well structured to accommodate the passage of H₂S (Figure 3), which is larger and less polar than water. In the selectivity filter of AQP1, a passing water molecule would be coordinated by the hydrogen bond donor Nε of R197(206) and the carbonyl oxygen of C191(200). The hydrogen-bond acceptor

H182(191) facilitates the passage of the polar water molecule by providing a secondary hydrogen-bond partner for the passing H₂O molecule. H₂S, which is significantly less polar and larger (dipole moment (μ) = 0.97 Debye, Van der Waals diameter (d) = 3.1 Å) than water (μ = 1.85 Debye, d = 2.8 Å), requires a channel that is larger and less charged than that required for the efficient selection and passage of water. The presence of a hydrogen-bond partner, such as the histidine residue found opposite across the channel from the arginine residue at the selectivity filter in AQP1, facilitates the hydrogen bonding requirements of a polar water molecule. But it may be repulsive to the effectively non-polar hydrogen atoms of an H₂S molecule. Therefore, an aliphatic residue such as I187(191), instead of a histidine residue is favorable for the passage of H₂S although inhibitory for that of H₂O. Initial permeation experiments using AQPM proteoliposomes indeed indicate the conductance of H₂S by AQPM (unpublished data) and also indicate that AQPM and possibly other aquaporins are gas-conducting channels.

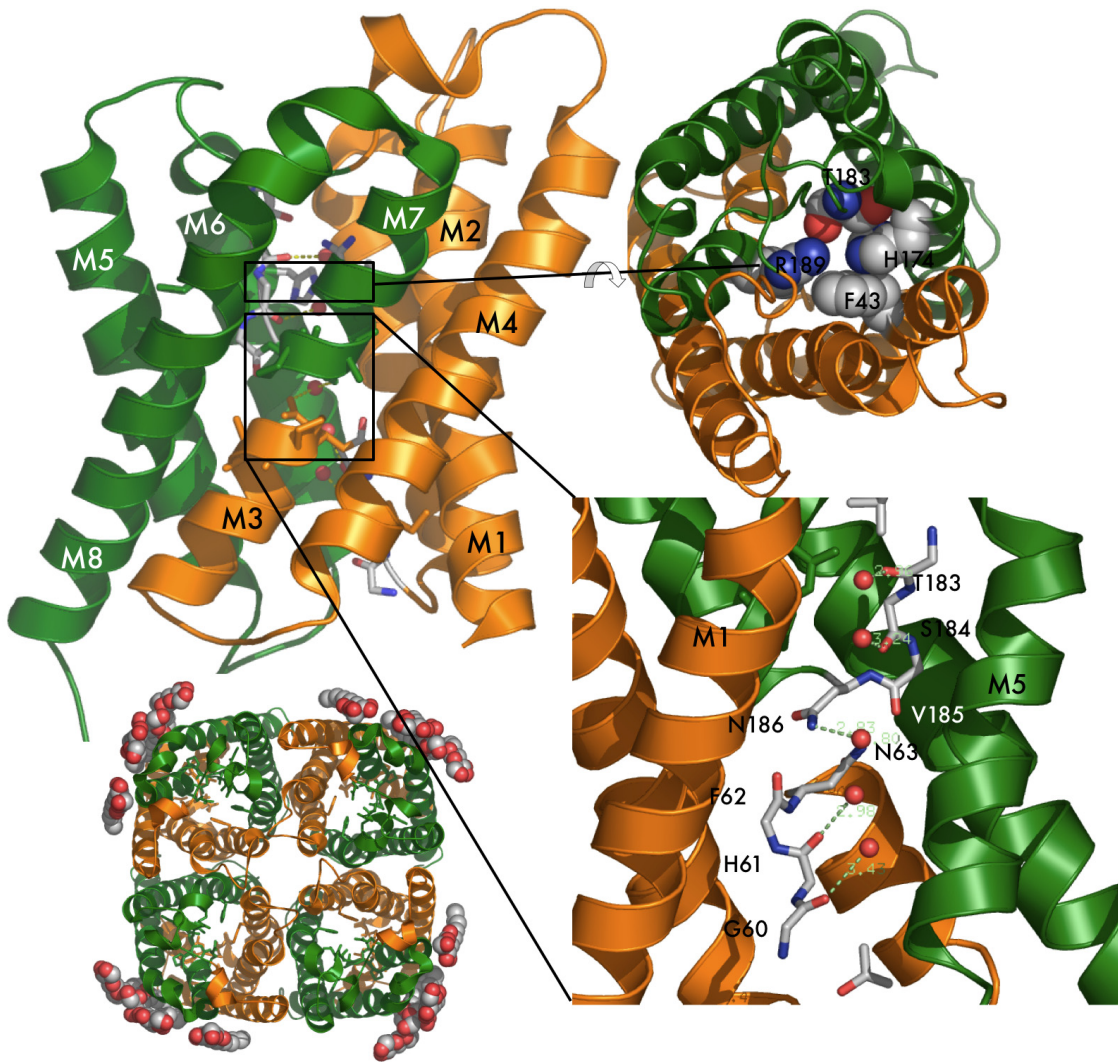
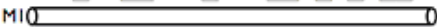
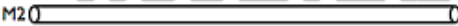




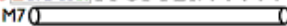



Figure 1 (A) Each aquaporin monomer is composed of six transmembrane helices and two half-length helices (M3, M7) that meet in the center of the bilayer. They are numbered M1 to M8. (B) The expansion of the upper rectangle shows the ‘selectivity filter in AQPs in which conserved R206 is a key player. (C) Aquaporin monomers associate in the plane of the membrane to form tetramers. (D) In the expansion from the lower rectangle in (A) focus is on the region where the two NPA regions meet. Residues that contribute to the orientation of the central water molecules include N63 and N186 of the NPA regions.

GlpF	1	...MSQTSTLKGQCI [.] AEFLGTGL [.] L [.] IFF [.] GVGC [.] V [.] AALKVAGAS.....
Aqp1	1	MASEFKKKLFWRAVV [.] AEFLAMIL [.] F [.] IF [.] IS [.] IGSALGFHYPIKSNQ.....
AqpZ	1MFRKLA [.] AE [.] CFGT [.] FW [.] LV [.] VF [.] GC [.] GS [.] AV [.] LAAG.....
Aqp0	1	..MWELRSASFWR [.] AI [.] CA [.] EFFAS [.] LF [.] Y [.] V [.] FF [.] GL [.] GAS [.] LRWAP.....
aqpM	1MVSLTKRCIA [.] E [.] FIG [.] T [.] FI [.] LV [.] FF [.] GA [.] GS [.] AA [.] AV [.] TLMIASGGTSPNPFNI
		M1 
GlpF	39FGQWE [.] IS [.] .VI [.] W [.] GL [.] GV [.] AMA [.] I [.] Y [.] LT [.] AG [.] V [.] SG [.] AHL [.] NP [.] AV [.] TIAL [.] WL [.] PAC
Aqp1	44	...TTGAVQDNV [.] KV [.] SL [.] AF [.] GL [.] SI [.] AT [.] LA [.] Q [.] SV [.] GH [.] IS [.] GA [.] HL [.] NP [.] AV [.] TL [.] G [.] LL [.] LS [.] CQ
AqpZ	29	...FP [.] EL [.] G [.] IG [.] F [.] AG [.] V [.] ALA [.] F [.] GL [.] T [.] V [.] LT [.] MA [.] FA [.] V [.] GH [.] IS [.] GG [.] H [.] FN [.] PA [.] VT [.] I [.] GL [.] W [.] AG [.] GR
Aqp0	37G [.] PL [.] H [.] V [.] L [.] Q [.] VAL [.] A [.] F [.] GL [.] AL [.] AT [.] LV [.] Q [.] AV [.] GH [.] IS [.] GA [.] H [.] V [.] NP [.] AV [.] TFA [.] FL [.] V [.] GS [.] Q
aqpM	45	G [.] IG [.] LL [.] G [.] EL [.] GD [.] W [.] VA [.] I [.] GL [.] AF [.] GF [.] AI [.] AA [.] SI [.] Y [.] AL [.] GN [.] IS [.] GC [.] H [.] IN [.] PA [.] VT [.] I [.] GL [.] W [.] SV [.] KK
		M2  M3 
GlpF	81	F [.] DK [.] R [.] K [.] V [.] IP [.] FI [.] VS [.] Q [.] V [.] AG [.] AF [.] CA [.] AA [.] LV [.] Y [.] GL [.] Y [.] N [.] L [.] FF [.] DF [.] E [.] Q [.] TH [.] H [.] I [.] VR [.] GS [.] VE [.] SV [.] D
Aqp1	91	I [.] SV [.] L [.] RA [.] IM [.] Y [.] II [.] A [.] Q [.] CV [.] GA [.] IV [.] ATA [.] IL [.] SG [.] ITS.....SL [.] P
AqpZ	76	F [.] PA [.] KE [.] V [.] V [.] GY [.] V [.] IA [.] Q [.] V [.] VG [.] GI [.] V [.] AA [.] ALL [.] Y [.] LI [.] AS.....G [.] K [.] T
Aqp0	81	M [.] SL [.] LR [.] AI [.] CY [.] M [.] VA [.] Q [.] LL [.] G [.] AV [.] AG [.] AA [.] V [.] LY [.] SV [.] TP.....P [.] AV [.]
aqpM	95	F [.] PG [.] RE [.] V [.] VP [.] Y [.] II [.] A [.] Q [.] LL [.] G [.] AA [.] FG [.] S [.] FI [.] FL [.] QC [.] AG.....I [.] G [.] AA [.] T
		M4 
GlpF	131	L [.] AG [.] T [.] F [.] ST [.] Y [.] PN [.] P.....H [.] IN [.] F [.] V [.] Q [.] AF [.] AV [.] EM [.] VI [.] T [.] AIL [.] M [.] GL [.] I [.] L [.] AL [.] T [.] DD [.] G [.] N
Aqp1	123	D [.] NS [.] L [.] GL [.] N [.] AL [.] AP.....G [.] V [.] NS [.] G [.] Q [.] GL [.] GI [.] E [.] I [.] IG [.] TL [.] Q [.] L [.] V [.] LC [.] V [.] L [.] ATT [.] .D [.] RR
AqpZ	108	G [.] FD [.] AA [.] AS [.] GF [.] AS [.] NG [.] Y [.] GE [.] HS [.] PG [.] GY [.] S [.] ML [.] S [.] AL [.] V [.] VEL [.] V [.] L [.] S [.] AG [.] FL [.] LV [.] I [.] H [.] G [.] AT [.] .D [.] K [.] F
Aqp0	113	R [.] GN [.] L [.] AL [.] N [.] TL [.] HP.....G [.] V [.] SV [.] G [.] Q [.] AT [.] IV [.] E [.] IF [.] L [.] T [.] L [.] Q [.] F [.] V [.] LC [.] I [.] F [.] AT [.] Y [.] .D [.] ER
aqpM	129	V [.] GG [.] L [.] G [.] AT [.] AP [.] FP.....G [.] IS [.] Y [.] W [.] Q [.] AML [.] AE [.] V [.] VG [.] T [.] FL [.] LM [.] IT [.] I [.] M [.] GI [.] AV [.] DER
		M5 
GlpF	172	G [.] V [.] PR [.] G [.] PL [.] AP [.] LL [.] I [.] GL [.] L [.] IA [.] V [.] IG [.] AS [.] M [.] G [.] PL [.] T [.] G [.] F [.] AM [.] NP [.] AR [.] DF [.] G [.] PK [.] V [.] FA [.] W [.] LAG [.] W [.] GN
Aqp1	163	R [.] RD [.] L [.] GG [.] SG [.] PL [.] AI [.] GF [.] S [.] VAL [.] G [.] H [.] LL [.] AI [.] DY [.] T [.] G [.] C [.] GIN [.] PAR [.] S [.] FG.....S
AqpZ	157	AP...AG [.] F [.] API [.] AI [.] GL [.] AL [.] TL [.] I [.] HL [.] IS [.] IP [.] VT [.] NT [.] SV [.] NP [.] AR [.] STA.....V
Aqp0	153	R [.] NG [.] RL [.] GS [.] VAL [.] AV [.] GF [.] S [.] LT [.] L [.] GH [.] L [.] FG [.] MY [.] YT [.] G [.] AG [.] M [.] NP [.] AR [.] S [.] FA.....P
aqpM	170	.AP [.] KG [.] .F [.] AG [.] II [.] I [.] GL [.] TV [.] AG [.] IT [.] TL [.] GN [.] IS [.] GS [.] SL [.] NP [.] AR [.] T [.] FG [.] P [.] YL [.] N.....D
		M6  M7 
GlpF	222	V [.] A [.] FT [.] G [.] GR [.] DI [.] PY [.] FL [.] V [.] PL [.] F [.] GP [.] I [.] V [.] GA [.] I [.] V [.] GA [.] F [.] AY [.] R [.] K [.] LI [.] GR [.] HL [.] PC [.] DI [.] CV [.] VE [.] E [.] KE [.] T
Aqp1	202	SV [.] ITH...NF [.] Q [.] D [.] HW [.] IF [.] FW [.] VG [.] PF [.] IG [.] AA [.] LAV [.] LI [.] Y [.] DF [.] IL [.] AP [.] RS [.] SDL [.] T [.] DR [.] V [.] K [.] V [.] W [.] TS
AqpZ	194	AI [.] F [.] Q [.] GG [.] W [.] ALE [.] QL [.] W [.] FF [.] W [.] V [.] PI [.] VG [.] GI [.] IG [.] GL [.] I [.] Y [.] RT [.] L [.] LE [.] K [.] RD.....
Aqp0	192	AIL [.] TR...N [.] FT [.] N [.] H [.] W [.] V [.] Y [.] W [.] VG [.] P [.] VI [.] G [.] AG [.] L [.] GS [.] LL [.] Y [.] DF [.] LL [.] FP [.] RL [.] K [.] SV [.] SER [.] L [.] S [.] IL [.] KG
aqpM	211	M [.] I [.] F [.] AG [.] T [.] N [.] L [.] W [.] N [.] Y [.] YP [.] I [.] Y [.] Y [.] IG [.] P [.] IV [.] G [.] AV [.] L [.] AA [.] L [.] TY [.] Q [.] Y [.] LT [.] SE.....
		M8 
GlpF	272	T [.] TP [.] SE [.] Q [.] K [.] AS [.] L.....
Aqp1	250	G [.] Q [.] VE [.] E [.] Y [.] DL [.] D [.] ADD [.] IN [.] SR [.] VE [.] M [.] K [.] PK...
AqpZ	
Aqp0	240	S [.] RP [.] SE [.] S [.] NG [.] Q [.] PE [.] VT [.] G [.] EP [.] VEL [.] K [.] T [.] Q [.] AL
aqpM	

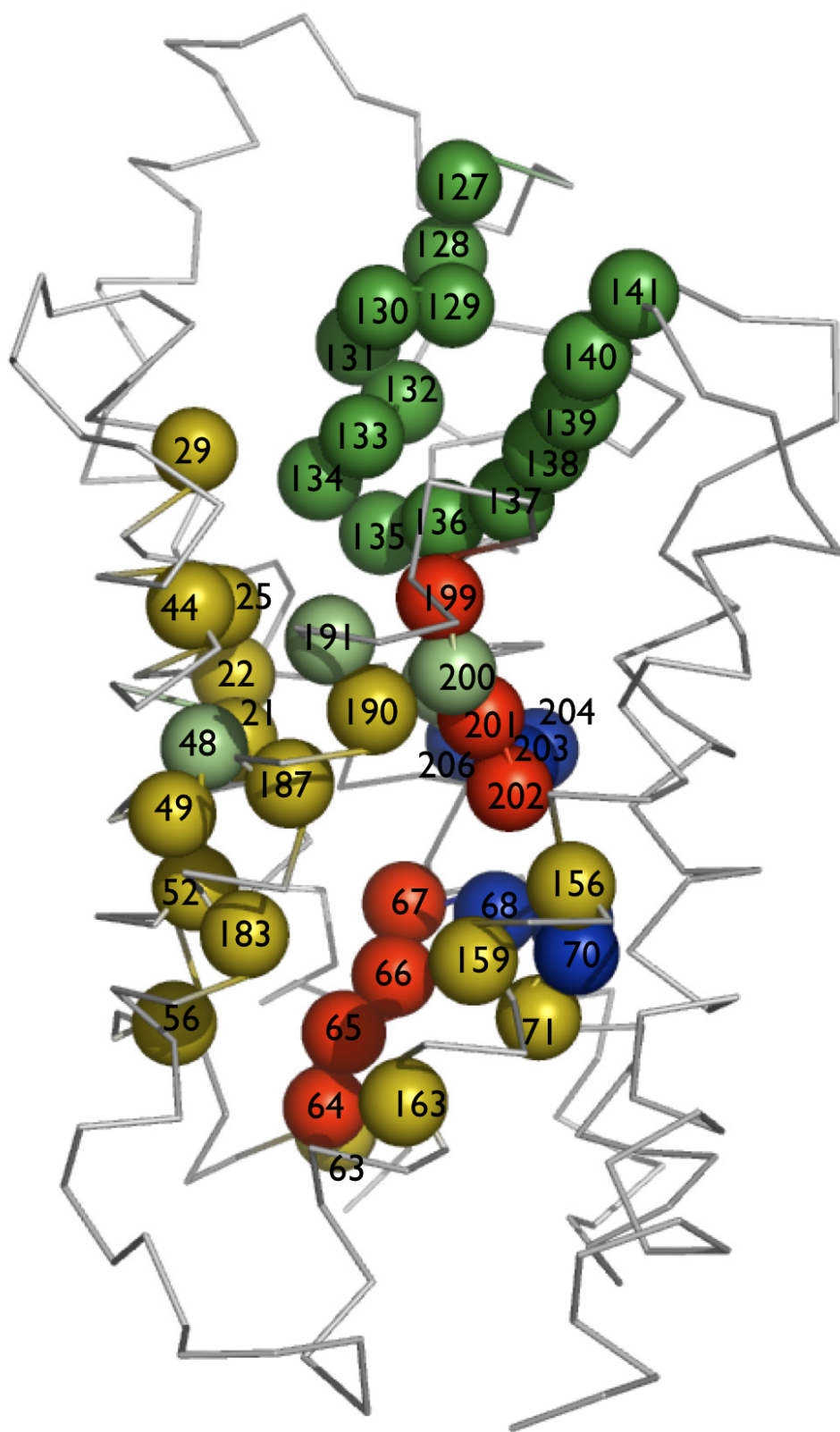
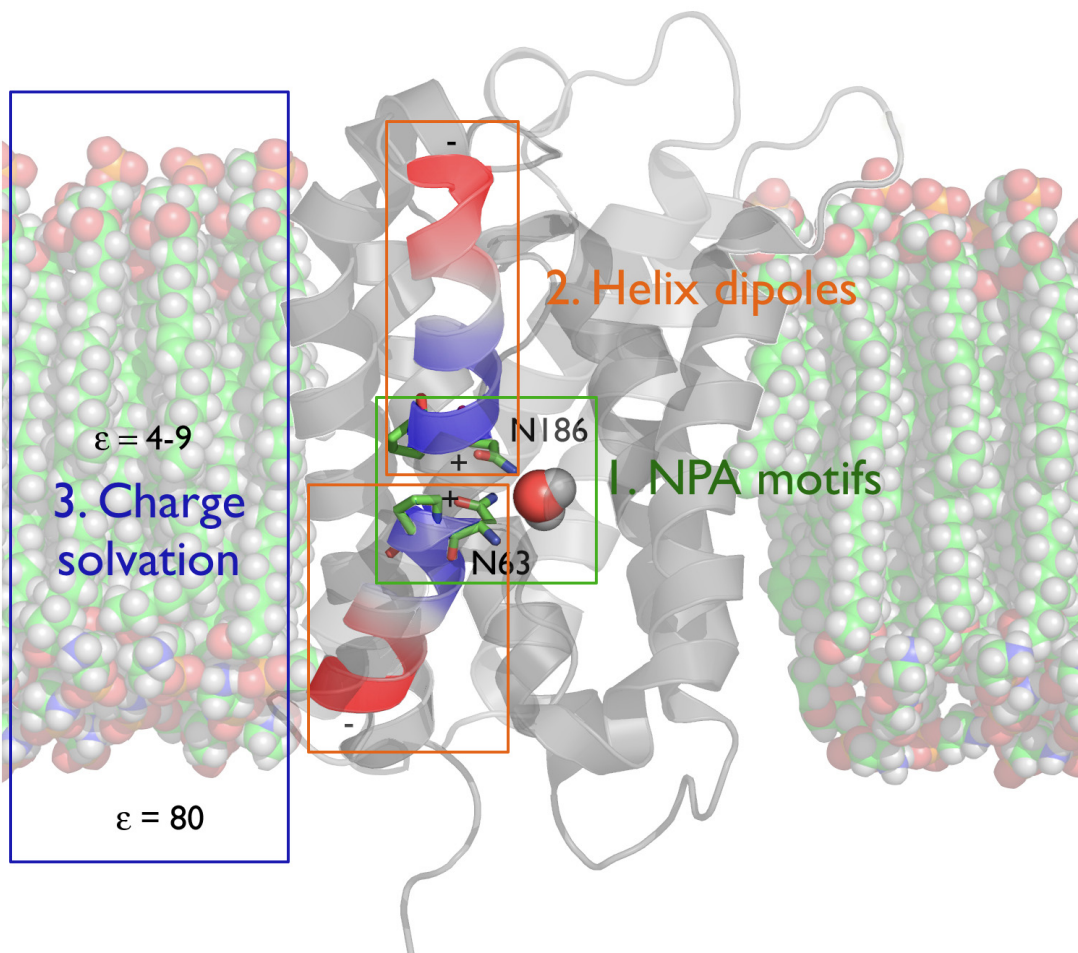


Figure 2 Alignment of Aquaporin sequences Mapped to Pore-Lining Residues. (A) Alignment of the 5 aquaporins whose structure is known. The ruler on the top is GlpF numbering at 10 residue intervals, as is the secondary structure information. Color scheme is: grey boxes, conserved residues; orange: pore residues; red: residues whose mainchain carbonyls project into pore; blue: NPA residues; lime: selectivity filter residues; dark green: inter-repeat loop. These colors also map on figure 3B. (B) Structure of GlpF with residues mapped from alignment as detailed above.



This figure did not appear in the RSC review due to space constraints. It corresponds to the section entitled “THE EXCLUSION BARRIER TO IONS AND PROTONS IN AQUAPORINS” and serves as an outline to the three barriers.

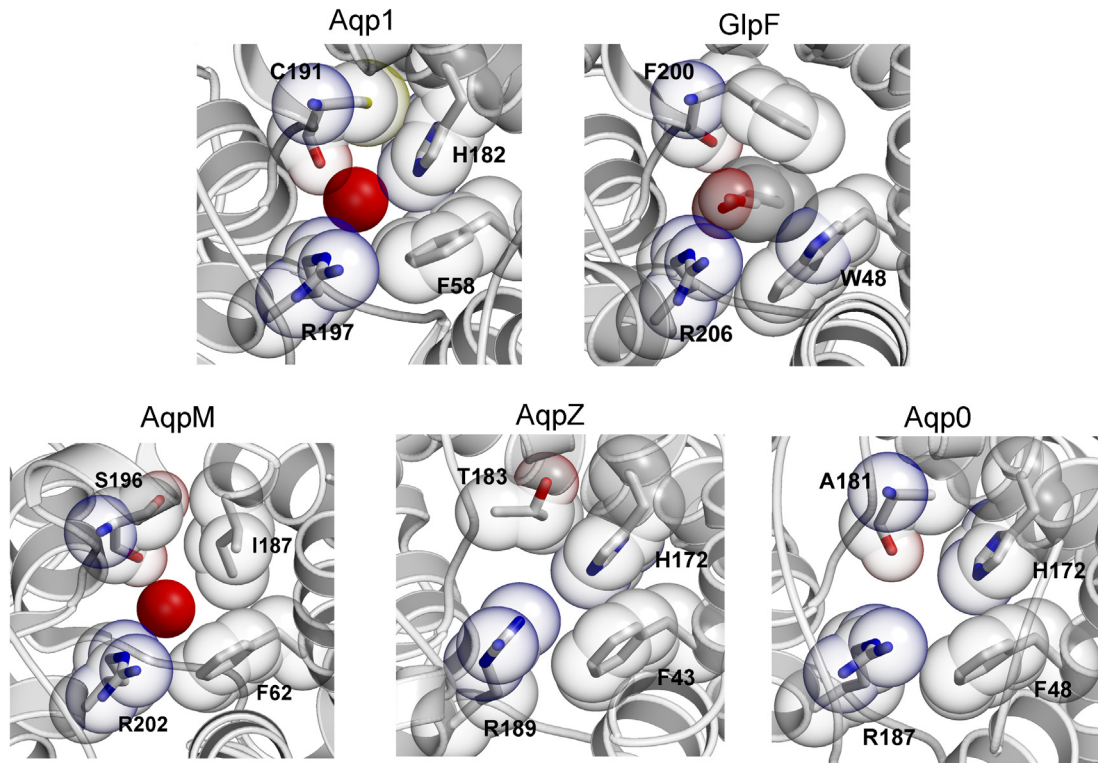


Figure 3 The selectivity filters are compared. The view is down the channel in the glycerol and water channel GlpF from *E.coli*, in the water channel AQP1 from bovine red blood cells, AQP1M from the archaebacterium *Methanothermobacter marburgensis* which may serve to conduct H₂S as well as water, AQPZ the bacterial water conducting AQP from *E.coli*, and the AQP from the bovine eye lens AQP0. The conserved Arg 206 is shown at the left side of each one. While water and glycerol in GlpF are seen throughout the channels only those substrate molecules seen within the slab shown are shown.

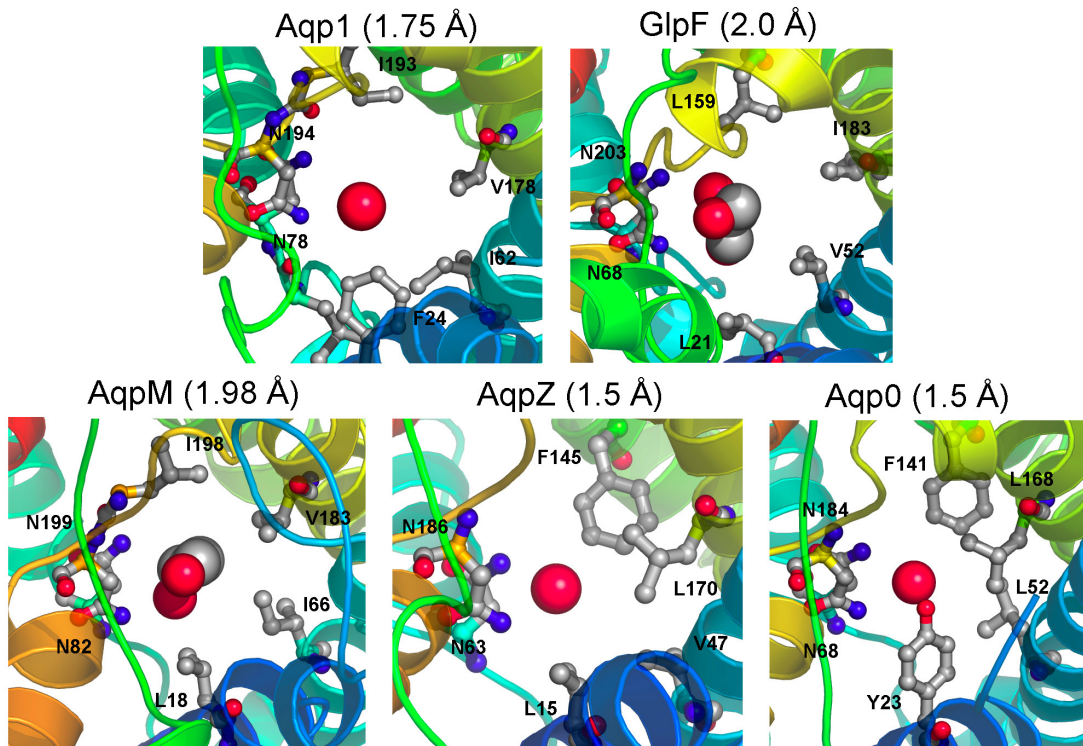


Figure 4. The center of the channel in AQPs is formed by two NPA motifs that focus hydrogen bond donors onto the central water or substrate molecule. The nitrogens of the asparagines side chains are shown at the left side in the same orientation as in Figure 3, but further down the channel.

REFERENCES

- ¹ A. Fischer, 'Vorlesungen uber Bakterien', 1903.
- ² K. B. Heller, E. C. Lin, and T. H. Wilson, *J Bacteriol*, 1980, **144**, 274.
- ³ G. Sweet, C. Gandor, R. Voegelé, N. Wittekindt, J. Beuerle, V. Truniger, E. C. Lin, and W. Boos, *J Bacteriol*, 1990, **172**, 424.
- ⁴ D. L. Weissenborn, N. Wittekindt, and T. J. Larson, *J Biol Chem*, 1992, **267**, 6122.
- ⁵ D. Fu, A. Libson, L. J. Miercke, C. Weitzman, P. Nollert, J. Krucinski, and R. M. Stroud, *Science*, 2000, **290**, 481.
- ⁶ M. O. Eze and R. N. McElhaney, *J. Gen Microbiol*, 1978, **105**, 233.
- ⁷ P. Nollert, W. E. Harries, D. Fu, L. J. Miercke, and R. M. Stroud, *FEBS Lett*, 2001, **504**, 112.
- ⁸ K. Liu, H. Nagase, C. G. Huang, G. Calamita, and P. Agre, *Biol Cell*, 2005.
- ⁹ J. Alcala, N. Lieska, and H. Maisel, *Exp Eye Res*, 1975, **21**, 581.
- ¹⁰ R. M. Broekhuysse and E. D. Kuhlmann, *Exp Eye Res*, 1974, **19**, 297.
- ¹¹ M. E. Baker and M. H. Saier, Jr., *Cell*, 1990, **60**, 185.
- ¹² P. Agre, *Angew Chem Int Ed Engl*, 2004, **43**, 4278.
- ¹³ L. M. Holm, T. P. Jahn, A. L. Moller, J. K. Schjoerring, D. Ferri, D. A. Klaerke, and T. Zeuthen, *Pflugers Arch*, 2005, **450**, 415.
- ¹⁴ G. M. Preston, T. P. Carroll, W. B. Guggino, and P. Agre, *Science*, 1992, **256**, 385.
- ¹⁵ C. Maurel, J. Reizer, J. I. Schroeder, M. J. Chrispeels, and M. H. Saier, Jr., *J Biol Chem*, 1994, **269**, 11869.

- ¹⁶ J. H. Park and M. H. Saier, Jr., *J Membr Biol*, 1996, **153**, 171.
- ¹⁷ A. S. Verkman, L. B. Shi, A. Frigeri, H. Hasegawa, J. Farinas, A. Mitra, W. Skach, D. Brown, A. N. Van Hoek, and T. Ma, *Kidney Int*, 1995, **48**, 1069.
- ¹⁸ T. L. Anthony, H. L. Brooks, D. Boassa, S. Leonov, G. M. Yanochko, J. W. Regan, and A. J. Yool, *Mol Pharmacol*, 2000, **57**, 576.
- ¹⁹ A. Engel, Y. Fujiyoshi, and P. Agre, *Embo J*, 2000, **19**, 800.
- ²⁰ G. M. Pao, L. F. Wu, K. D. Johnson, H. Hofte, M. J. Chrispeels, G. Sweet, N. N. Sandal, and M. H. Saier, Jr., *Mol Microbiol*, 1991, **5**, 33.
- ²¹ G. J. Wistow, M. M. Pisano, and A. B. Chepelinsky, *Trends Biochem Sci*, 1991, **16**, 170.
- ²² A. Finkelstein, 'Water movement through lipid bilayers, pores and plasma membranes, theory and reality.' Wiley, New York, NY, 1987.
- ²³ H. Sui, B. G. Han, J. K. Lee, P. Walian, and B. K. Jap, *Nature*, 2001, **414**, 872.
- ²⁴ D. F. Savage, P. F. Egea, Y. C. Robles, J. D. O'Connell, and R. M. Stroud, *PLoS Biology*, 2003, **in the press**.
- ²⁵ W. E. Harries, D. Akhavan, L. J. Miercke, S. Khademi, and R. M. Stroud, *Proc Natl Acad Sci U S A*, 2004, **101**, 14045.
- ²⁶ E. Tajkhorshid, P. Nollert, M. O. Jensen, L. J. Miercke, J. O'Connell, R. M. Stroud, and K. Schulten, *Science*, 2002, **296**, 525.
- ²⁷ M. Yasui, A. Hazama, T. H. Kwon, S. Nielsen, W. B. Guggino, and P. Agre, *Nature*, 1999, **402**, 184.
- ²⁸ M. L. Zeidel, S. V. Ambudkar, B. L. Smith, and P. Agre, *Biochemistry*, 1992, **31**, 7436.

- ²⁹ M. L. Zeidel, S. Nielsen, B. L. Smith, S. V. Ambudkar, A. B. Maunsbach, and P. Agre, *Biochemistry*, 1994, **33**, 1606.
- ³⁰ R. Pomes and B. Roux, *Biophys J*, 2002, **82**, 2304.
- ³¹ B. L. de Groot and H. Grubmuller, *Science*, 2001, **294**, 2353.
- ³² N. Chakrabarti, B. Roux, and R. Pomes, *J Mol Biol*, 2004, **343**, 493.
- ³³ B. Roux and R. MacKinnon, *Science*, 1999, **285**, 100.
- ³⁴ D. A. Doyle, J. Morais Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon, *Science*, 1998, **280**, 69.
- ³⁵ A. Burykin and A. Warshel, *Biophys J*, 2003, **85**, 3696.
- ³⁶ B. L. Smith and P. Agre, *J Biol Chem*, 1991, **266**, 6407.
- ³⁷ A. Froger, J. P. Rolland, P. Bron, V. Lagree, F. Le Caherec, S. Deschamps, J. F. Hubert, I. Pellerin, D. Thomas, and C. Delamarche, *Microbiology*, 2001, **147**, 1129.
- ³⁸ D. Thomas, P. Bron, G. Ranchy, L. Duchesne, A. Cavalier, J. P. Rolland, C. Raguene-Nicol, J. F. Hubert, W. Haase, and C. Delamarche, *Biochim Biophys Acta*, 2002, **1555**, 181.
- ³⁹ B. Yang, N. Fukuda, A. van Hoek, M. A. Matthay, T. Ma, and A. S. Verkman, *J Biol Chem*, 2000, **275**, 2686.
- ⁴⁰ A. S. Verkman, *J Physiol*, 2002, **542**, 31.
- ⁴¹ G. V. Prasad, L. A. Coury, F. Finn, and M. L. Zeidel, *J Biol Chem*, 1998, **273**, 33123.
- ⁴² M. Ikeda, E. Beitz, D. Kozono, W. B. Guggino, P. Agre, and M. Yasui, *J Biol Chem*, 2002, **277**, 39873.

- ⁴³ K. Ishibashi, K. Yamauchi, Y. Kageyama, F. Saito-Ohara, T. Ikeuchi, F. Marumo, and S. Sasaki, *Biochim Biophys Acta*, 1998, **1399**, 62.
- ⁴⁴ K. Ishibashi, M. Kuwahara, Y. Gu, Y. Tanaka, F. Marumo, and S. Sasaki, *Biochem Biophys Res Commun*, 1998, **244**, 268.
- ⁴⁵ K. Ishibashi, T. Morinaga, M. Kuwahara, S. Sasaki, and M. Imai, *Biochim Biophys Acta*, 2002, **1576**, 335.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

A handwritten signature in black ink, appearing to be 'P. S. M.', written over a horizontal line.

Author Signature

2.21.07

Date