The structure of GlpF, a glycerol conducting channel

Dax Fu, Andrew Libson* and Robert Stroud*

Biology Department, Building 463, 50 Bell Avenue, Brookhaven National Laboratory, Upton, NY 11973 and *Macromolecular Structure Group, Department of Biochemistry and Biophysics, Room S-964, Box 0448, University of California San Francisco, San Francisco, CA 94143-0448, USA

Abstract. The passage of water or small neutral solutes across the cell membrane in animals, plants and bacteria is facilitated by a family of homologous membrane channels, variously known as aquaporins though perhaps more correctly as aquaglyceroporins. The glycerol facilitator (GlpF) is a 28 kDa aquaglyceroporin that catalyses transmembrane diffusion of glycerol and certain linear polyhydric alcohols in Escherichia coli. X-ray crystallographic analysis of GlpF to 2.2 Å resolution revealed an α-barrel structure, surrounded by six full-length transmembrane helices and two half-spanning helices that are joined head-to-head in the middle of the membrane. These helices are arranged in a quasi twofold manner relative to the central membrane plane, where highly conserved residues make helix-to-helix contacts that stabilize the relative position and orientation of the helices in the structure. This sequence–structure correlation suggests that the evolutionary divergence of aquaporins and aquaglyceroporins is constrained by a conserved structural framework within which specialized function may be developed. Three glycerol molecules were resolved in the central channel through the GlpF monomer, thereby defining a transmembrane channel for glycerol permeation. The structure of glycerol–GlpF complex provides insight into the chemical basis for transmembrane selective permeability.

2002 Ion channels — from atomic resolution physiology to functional genomics. Wiley, Chichester (Novartis Foundation Symposium 245) p 51–65

Fluid balance is a ubiquitous biological process in all living organisms. The molecular basis of the rapid transmembrane flux of water and small non-electrolytes remained elusive until the recent discovery of a superfamily of fluid-transporting membrane channels (Preston & Agre 1991, Preston et al 1992). On the basis of their permeability properties, these channels are subdivided into aquaporins — a subfamily of strict water-selective channels, and aquaglyceroporins — a homologous subfamily of less selective channels with permeability for small neutral solutes such as glycerol (Borgnia et al 1999a).
The fundamental importance of aquaporins and aquaglyceroporins is attested by their strong genetic conservation from bacteria to plants and humans (Park & Saier 1996, Heymann & Engel 2000). More than 180 aquaporins and aquaglyceroporins have been identified in organisms ranging from bacteria to mammals, including 10 human isoforms (AQP0–9). The *Escherichia coli* aquaporin AQPZ and aquaglyceroporin GlpF share 20–25% pair-wise sequence identity with each other, and with all human isoforms identified thus far (Unger 2000, Beitz & Schultz 1999). This extensive sequence similarity indicates that AQPZ and GlpF are part of the development that leads to the human evolutionary tree.

Functional studies of human aquaporins and aquaglyceroporins have been facilitated by the use of homologous *E. coli* isoforms. AQPZ and GlpF have been overexpressed, extracted from the membrane using non-ionic detergents and then purified to homogeneity in detergent micelles (Borgnia & Agre 2001). At the protein level, reconstitution of pure AQPZ or GlpF into liposomes makes it possible to examine transmembrane fluxes of water or glycerol by monitoring swelling or shrinkage of membrane vesicles in response to a transmembrane osmotic gradient (Agre et al 1999, Verkman 2000a). The rates of solute fluxes through these channels are diffusion limited, suggesting that AQPZ and GlpF function as channels rather than transporters. The channel selectivity was also examined, showing that AQPZ specifically increases water osmotic permeability (Calamita et al 1995) whereas GlpF increases glycerol (Maurel et al 1994) and water (Borgnia & Agre 2001) permeability. In addition, GlpF is permeable to certain polyols and antimonite (Heller et al 1980, Sanders et al 1997). At the organ level, the physiological significance of human aquaporins was implicated by phenotype analysis of aquaporin knockout mice and by pathophysiological studies of human aquaporin deficient diseases such as nephrogenic diabetes insipidus (Verkman et al 2000; Verkman 1999, Agre 1998). It is evident that human aquaporins can facilitate near iso-osmolar transepithelial fluid absorption and secretion, as well as rapid water movement driven by osmotic gradients (Verkman 2000b). Recent cloning of AQP8 from adipose tissue suggests that aquaglyceroporin is involved in glycerol release and body energy balance that may be relevant to obesity (Kishida et al 2000).

Structural analysis has led to the determination of the general channel architectures of AQP1 (Cheng et al 1997, Walz et al 1997, Murata et al 2000) and GlpF (Braun et al 2000, Fu et al 2000). The structure of human red cell AQP1 was determined to 3.8×6.5 Å resolution by electron crystallography (Murata et al 2000). More recently, we determined the structure of GlpF by X-ray crystallography to 2.2 Å resolution (Fu et al 2000). Arrangements of amino acid side chains and chemical groups along the GlpF transmembrane channel are revealed at this higher resolution, together with three bound glycerol molecules. AQP1 and GlpF are strikingly similar structures. Both crystallize as tetramers, with
FIG. 1. Ribbon representation of GlpF monomer (from a perspective viewed in parallel to the membrane plane) with two glycerol molecules bound in the middle of the channel.

Each monomeric subunit as an asymmetric unit. The monomers are folded as an $\alpha$-barrel structure, with a central channel surrounded by six full-length transmembrane helices and two half-spanning helices that are joined head-to-head in the middle of the membrane to give the appearance of a kinked transmembrane span (Fig. 1). In both AQP1 and GlpF, an internal quasi-twofold symmetry is evident. Superimposition of main chains of two halves of the GlpF
showed a close match with a root-mean-square deviation of 1.4 Å. The arrangement of the two halves in the structure creates an overall symmetrical relationship with respect to a quasi-twofold axis in the central membrane plane. This quasi-twofold structure corresponds to the two tandem repeats in the linear sequence that may have arisen from an intragenic gene duplication event (Wistow et al 1991).

**Channel permeability and selectivity**

Fluxes of water and polyols through AQPZ or GlpF channels have been monitored in real time using stopped-flow light scattering measurements (Borgnia et al 1999b, Borgnia & Agre 2001, Fu et al 2000). Permeability through the AQPZ channel was only observed for water, but not for any other neutral solutes. By comparison, GlpF was found to be permeable to a range of linear polyols of different chain lengths and to water at a slower rate compared with the rate through AQPZ (Borgnia & Agre 2001). In light of this flexibility for different sizes of polyols, it came as a surprise to find that the rate of ribitol flux was sevenfold faster than that of its stereoisomer xylitol (Heller et al 1980). The only difference between ribitol and xylitol is the stereo configuration of one OH group with respect to the alkyl backbone (Fig. 2). This result showed that GlpF is selective for one stereoisomer over another.

The GlpF structure explains how permeability to multiple solutes can be reconciled with stringent stereo-selectivity (Fu et al 2000). The channel lining is strongly amphipathic, with oxygens and nitrogens lined up on one side and carbons on the opposite side of the lumen surface. This amphipathic channel uniquely matches the chemical structure of glycerol and other polyols, which are composites of the polar hydroxyl group arranged on a non-polar alkyl backbone. Of particular interest is the exquisite hydrogen bonding network seen at the narrowest constriction of the channel, where two consecutive OH groups of a polyol molecule are oriented with ideal geometry with respect to the donors and acceptors from within the channel, thus positioning the donor-acceptor pairs at
about 120° in relation to each of two successive OH groups of the polyl (Fig. 3). As a polyl proceeds through the channel, each pair of successive OH groups has to satisfy this double H-bonding geometry. Thus, the narrowest constriction imposes a stereo selection upon the relative orientation of OHs, regardless of the chain length. This structure satisfactorily explains why ribitol is more permeable than xylitol, because all OHs in ribitol are on one side of the alkyl backbone (Fig. 3), matching with the amphipathic channel, whereas in xylitol one OH is on the opposite side of the backbone.

Common to AQPZ and GlpF is the absolute exclusion of ions (Fu et al 2000, Murata et al 2000). The first barrier against ion entry is made by electrostatic repulsion between NHs of the guanidinium group of R206 and the carbonyl oxygens of G199 and F200 at the narrowest constriction of the channel. The negative charges on carbonyls are enhanced by a buried carboxyl of E152 that interacts with main chain nitrogens of G199 and F200. These charge pairs are lined up at a 4 Å-wide constriction, thus only allowing passage of dipoles that can interact with positive and negative charges simultaneously. Neither a cation nor an anion can fulfil this dual role of polarization. Both R206 and E152 are
invariant, suggesting that these residues are conserved across all aquaporins and aquaglyceroporins to exclude ions. The second barrier of ionic exclusion is imposed by an insurmountable activation energy for removing water of hydration from a passing ion in the hydrophobic sector of the membrane, where the cross-section of GlpF channel is only large enough to fit one fully dehydrated ion, but the hydrophobic side of the lumen provides no replacement for water of hydration on the entire hydrophobic face. By comparison, the K⁺-conducting channel (KcsA) contains a stack of carbonyl rings at its narrowest constriction, thus affording negatively charged oxygens to coordinate with a passing potassium ion and replace its water of hydration (Doyle et al 1998, Roux et al 1999).

The GlpF structure at 2.2 Å resolution provides the first structural and chemical clues as to the mechanism of channel-selective permeability. This crystal structure represents a snapshot in the permeation process that could be hypothesized as follows. Upon entry to the channel on the extracellular side, a glycerol molecule first meets with three highly constrained carbonyl oxygens and two nitrogens at the narrowest constriction, followed by two nitrogens in the centre of the channel, and then another set of three carbonyl oxygens on the cytoplasmic side. These two sets of carbonyl triplets and two central nitrogens are arranged in a closely spaced sequential order, serving as surrogate waters to form H-bonds with OH groups of a passing glycerol (Fig. 4). It is not clear how these H-bond acceptors and donors interact with waters before a glycerol molecule enters the channel. Furthermore, we do not know the sequential order of transmembrane events that occur in the permeation process. A glycerol molecule might move down the channel in a stepwise fashion by exchanging one set of H-bonds with the next in line. Alternatively, permeation may be carried out as a series of thermally activated jumps along several distinct binding sites in the channel (Lauger & Apell 1982). These hypotheses could be clarified by further crystallographic analysis of GlpF in complex with water and other polyols.

Conserved structure framework

The heritage of sequence conservation in the aquaporins and aquaglyceroporins is well preserved in protein structure from bacteria to mammals. Superposing peptide backbones of GlpF and human AQP1 shows a close alignment with a root-mean-square deviation of 1.8 Å, indicating that GlpF and AQP1 are essentially built upon a common structural framework. The GlpF structure reveals how tertiary interactions hold together the structural framework. The building block for GlpF is made of α helices with regular geometry as seen in many other proteins. What makes GlpF unique is the inter-helix crossing-over packing that determines the positions and orientations of helices in the structure. Crossing contacts in GlpF occur within helical bundles of M2–M1–M4–M3 and
FIG. 4. Atomic interactions in the GlpF channel reveal the structural basis of selective permeation.

M6–M5–M8–M7 (Fu et al 2000). These two bundles are joined by an anti-parallel interaction between M2 and M6 on one side of the channel. On the other side, the two half-spanning helices M3 and M7 close up the channel by a head-to-head interaction at the conjunction of two NPA motifs (Fig. 1). Furthermore, constrained by the quasi-twofold internal symmetry, all crossing contacts in GlpF are coplanar with the quasi-twofold axis in the central membrane plane.
FIG. 5. The conserved structure framework in aquaporin and aquaglyceroporin superfamily. All interhelical contacts in GlpF are localized to the central membrane plane. Highly conserved residues involved in helix–helix contacts are shown in CPK representation.

(Fig. 5). Outward from this plane, helices diverge to form the two vestibules on each side of the membrane (Fu et al 2000). The GlpF structure suggests that the transmembrane architecture of aquaporins and aquaglyceroporins are strongly constrained by a few key tertiary interactions that are localized to the central membrane plane.

Structure–sequence correlation shows that most inter-helix contacts in GlpF are made by invariant or highly conserved residues, including four glycines, G49, G96, G184 and G243. Far from the usual role in disrupting α helices, these glycines allow the closest approach of helices in the central membrane plane. Perpendicular to the membrane plane is the transmembrane channel, which contains seven functional
determinant residues, including G199, G64, N68, N203 and three charged residues, R206, E14 and E152. All these residues are extremely conserved from bacteria to mammals. These locations of highly conserved residues suggest that aquaporins and aquaglyceroporins are related and probably descend from a common evolutionary origin. The evolutionary divergence has been constrained by an almost absolute conservation in helix-to-helix contacts and in positions that are principal determinants for transmembrane function. Therefore, human isoforms may be built upon a conserved structural framework. Specialized functions may be developed through natural selection of mutational drifts of non-structural components. The structure of GlpF could be used as a template to model human isoforms with adequate treatments of side chain rotamers.

Future studies

The crystal structure of GlpF provides the first snapshot of the transmembrane permeation process in aquaporins, and unveiled a conserved structural framework within which specialized functions may be developed in human isoforms. Molecular dynamic simulations may reveal a detailed picture of atomic motions in the process of transmembrane selective permeation. Structure–function studies may lead to identification of functional determinants in human isoforms. The assignment of functional roles to key residues in sequences of human channels is crucial to understand the genetic consequences of human polymorphisms in aquaporins and aquaglyceroporins, thereby providing the structural basis for genetic diagnosis of disease or disease risk caused by genetic variations of certain human determinant residues.

Acknowledgements

We thank T. Earnest for help and support at the Advanced Light Source (ALS), Lawrence Berkley National Library, J. Krucinski for assistance in crystallization, L. Miercke for assistance in protein purification, and Drs P. Maloney, S. Sine and P. Freimuth for discussion. This work was supported by an NIH grant GM24485 to RMSDF and AL received postdoctoral support from NIH.

References


DISCUSSION

Mindell: You talked about the conservation of these channels. Are there specific sequence differences that would explain this dramatically different selectivity?

Fu: AQPs and GlpF have about 25% sequence similarity. But if we look at the channel-lining residues, they are highly conserved with the only exception at position 200. The GlpF has a Phe while AQP1 has a Cys at this position. Despite this strong similarity, there is no clear indication of the source of the water/glycerol selectivity.

Mitra: This is debatable. Different labs do experiments differently. Some labs have claimed that GlpF does not conduct water very well at all. If you compare it with the sequence of AQP3 in mammals (which is a glycerol transporter), it does not have a phenylalanine in that place, but a slightly polar tyrosine. In fact, at this position only GlpF has a phenylalanine. All the others have tyrosine there. We think His180 in AQP1 at the analogous site is important, but at that position, 180, it is glycine/valine for aquaglyceroporins. The additional polar residue there, in addition to phenylalanine versus cysteine, might act together to provide that occlusion of glycerol or transport of water.

Sansom: You said that it might be the extracellular mouth that contributes to the selectivity. Where in the sequence does this occur?

Fu: It is the only part that is not conserved. The selectivity is largely contributed by the linking group between the two halves of the protein.

Mitra: This is between helices 3 and 4.

Sansom: Has anyone tried making chimeras in which they switch these?

Mitra: Bai et al (1996) have done switching between AQP2 and AQP1 in the hydrophilic loops. These resulted in chimeras with permeability similar to the wild-types. On the other hand an AQP2 chimera in which a fragment of 3-4 loop of GlpF replaced a corresponding segment of AQP2 displayed significantly reduced water permeability.

Doyle: GlpF also has the conserved NPA residues. This probably has a role in the water movement, but does it have any role in the movement of glycerol?

Fu: Yes, the two NPA tripeptide sequences are located in the middle of the channel. The Asp206 and Asp68 form H-bonds with two OH groups of a glycerol molecule that is bound in the middle of the channel.

Choe: I would predict that the main chain phi, psi angles of glycine 199 are positive.
Fu: I have checked the geometries. All the main chain, with the exception of F200, is located in the allowed areas of the Ramachandran plot.

Sansom: So is your suggestion that the E152 is aiding that distortion? Or do you think that it might be polarizing the carbonyl oxygen?

Fu: The entire side chain of E152 is embedded in a highly hydrophobic environment. The only way that E152 could gain access to the aqueous channel is through hydrogen bonding interaction with the main chain nitrogen. This side chain–main chain interaction could strongly polarize carbonyl oxygens in the vicinity.

Jordan: I am having a hard time seeing what it is about the interior of this channel that makes it difficult for water to get inside. Otherwise I worry about glycerol getting in there and just being stuck. What is it that keeps the water out?

Fu: This is also a problem I have! People used to think that GlpF was only permeable to glycerol. I kept asking myself how this channel could be only permeable to glycerol but not water. Andy Libson and I came up with the following explanation. When a water molecule moves down the channel, the water itself has to be dehydrated because waters in aqueous solution are linked together by hydrogen bonds. Moving a water molecule through a narrow channel therefore requires removal of its hydration shell first. Further, to satisfy hydrogen donor and acceptor groups lining the hydrophilic face of the channel, at least two water molecules are required to be in a line. However, a string of water molecules does not have a non-polar backbone to satisfy the hydrophobic face of the channel.

Jordan: So what you are saying is that you think that there might be a genuine free energy barrier for water to enter, because you have to break up the water structure. The glycerol is nicely designed to be solvated by the channel.

Fu: Yes.

Sansom: We should remember that structure is solved at low temperature (100 K). Things might move a little bit.

Choe: I thought that the polarity of the molecule with respect to the membrane was strikingly clear. Is this the case for all aquaporin channels?

Fu: Like other membrane proteins in general, GlpF follows the positive inside rule. There are 16 excess positive charges on the cytoplasmic face. But I do not think that all these charges are conserved.

Choe: In other membrane proteins the polarity is not as distinct as in yours.

Fu: There are two rings of charged side chains, forming two layers of charges which sandwich two layers of aromatic side chains that are located on the polar–non-polar interface of the membrane. These aromatic rings seem to be a common feature in membrane proteins.

Choe: What is the membrane potential of the inner membrane of E. coli?
STRUCTURE OF GlpF

Fu: I would guess it is —30 to —70 mV.
Mindell: I have usually heard it quoted as —200 mV. It is a very high potential.
Choe: It could therefore be more pronounced for *E. coli* water channels to have that stability by creating higher polarity.

Roux: People say ‘stability’, but you don’t expect a membrane protein to tumble over (i.e. with the intracellular side pointing outward). The presence of positively charged amino acids on the intracellular side may have more to do with the folding and trafficking than the stability of the folded protein itself.

Fu: I think this dipole probably helps to orientate GlpF in the membrane. The protein is formed by two halves that have strong sequence similarities, and the dipole means that they can only organize in the membrane in the correct way.

Mitra: In many cases the charge on one leaflet of the bilayer is different from the other. This also acts to orient the membrane protein in a particular way.

Choe: Do you also have a similar level of polarity in your model?
Mitra: I have not checked it in detail.
Ashcroft: How long a molecule can go through the channel?
Fu: Probably 20 Å.
Ashcroft: Is it possible for something to span the channel?
Fu: Yes, if you have a polyol of 5 or 6 carbons, this will be much longer than the pore.

Choe: But these sugars are rings, not long chains. If you build that stereochemistry, the hydroxyl groups are all on one side. In one of your stereotypical variants, some are on the other side, but it had only sevenfold less permeability.

Fu: This is because we were depicting things in two dimensions. If you view a three-dimensional molecule, the stereo arrangement of the OH groups is not so much different.

Choe: It occurs to me that your sugar may linearize in order to cross over. As a ring you can’t force it through the pore.

Fu: People have used ring sugars to examine the permeability of the GlpF channel.

Mindell: My impression was that those weren’t sugars, they were polyalcohols.
Fu: Yes. They don’t form rings.
Wallace: Physiologically, what is the main role of this channel? Is it to transport polyalcohols, glycerol, water or something else?

Fu: These proteins can be induced in the presence of glycerol. This is an inducible glycerol-permeable channel. People believe that this channel is involved in glycerol uptake. When the channel is knocked out the bacteria grow much slower.

Ashcroft: At the beginning of your presentation you showed four pores, one in each monomer, and you also had a channel in the centre, which you haven’t said anything about.
Fu: That putative channel was surprisingly empty along the fourfold axis. We do not know the physiological function of this fourfold pore. This channel has a small hydrophobic constriction on the cytoplasmic side, which is too small to allow permeation of any type of ion or water.

Choe: When you isolate your protein, is it a stable tetramer? Does it ever monomerize or dimerize?

Fu: When I purify the protein it is not stable. We have spent a long time trying to work out why.

Perozo: What do you mean when you say that it isn’t stable?

Fu: This is a typical problem with membrane proteins. Then one day, by accident we found that if we added Ca\textsuperscript{2+} or Mg\textsuperscript{2+}, it made the GlpF more stable. After the structure was solved, it became clear that Ca\textsuperscript{2+} or Mg\textsuperscript{2+} could help in stabilizing the formation of tetramer. In the absence of divalent cations, GlpF may be in either a tetrameric or monomeric form.

Choe: At low concentrations do you ever see monomers? They have to go to monomers at some point if you keep diluting.

Fu: But how can we determine this? The protein is in the detergent and it is very hard to determine the precise molecular weight.

Choe: It is hard, but it is probably doable if you know how much detergent is bound in the complex.

Schrempf: Has anyone done mutagenesis to prevent the contacts among the monomers? This would address the question.

Mitra: There are some unpublished reports from the French group in Rennes that the amount of detergent in the purification cocktail can also modulate the oligomeric state. At low concentration of octyl glucoside (\(~\)1%), they claim that more of GlpF is in monomeric state, and as the detergent concentration is increased (\(~\)10%) the proportion of tetrameric structure increases. This is somewhat counterintuitive.

Fu: I suspect that tetramer formation may be essential for the stability of the protein. Two transmembrane-spanning helices in GlpF are located close to the tetramer fourfold axis, each is approximately 20 Å in length. They are apparently not long enough to cross the membrane if not forming the tetramer. By forming the tetramer, GlpF shields the short helices from the hydrophobic core of the membrane.

Mitra: Couldn’t this be done by dimers?

Fu: That is possible.

Papazian: What about the results of something simple such as sucrose gradients? Does the protein migrate as a single species?

Fu: This will depend on how many detergent molecules are binding to a protein. The protein can go just about anywhere. One way to estimate detergent number per micelle would be to use radioactive detergent. This will allow one to calculate
how many detergent molecules are bound per protein, and then it is possible to estimate their contribution to the apparent molecular weight of the protein–detergent micelle.

Choe: Even with the analytical centrifuge you need to account for how much detergent is present in the complex. The lipid density is significantly different from protein density.

Reference