Water and glycerol permeation through the glycerol channel GlpF and the aquaporin family

John K. Lee, Shahram Khademi, William Harries, David Savage, Larry Miercke and Robert M. Stroud

Department of Biochemistry & Biophysics, School of Medicine, University of California, 600 16th Street, San Francisco, CA 94143-2240, USA.
E-mail: stroud@msg.ucsf.edu

The 2.2 Å resolution crystal structure of GlpF, an *E.coli* aquaporin that facilitates the flow of glycerol, water and other small solutes, provides much insight into the molecular function and selectivity of aquaporins. Using GlpF and its atomic structure as a paradigm for the ten highly conserved human aquaporins, site-directed mutagenesis has been used to mutate residues that are possibly integral to the structure and function of different aquaporins. X-ray crystallography and other biophysical and molecular simulation methods allows for assessment of these changes at the structural and functional level. Initial attempts to convert the glycerol specific properties of GlpF towards a water specific aquaporin resulted in the shifting of GlpF channel properties towards that of the water aquaporins. This result reveals the great possibility of emulating and deciphering the function of other aquaporins with GlpF via mutagenesis and investigation of structure and function.

Key words: aquaporin, GlpF, water channel, membrane protein, X-ray crystallography.

1. Introduction

The 'glycerol facilitator' GlpF was first identified by E.C.C. Lin and colleagues in the mid 1960s (Sanno *et al*., 1968). This led to the conducting specificity of GlpF (Heller *et al*., 1980), cloning and sequencing of the gene (Sweet *et al*., 1990), and characterization of the 281 amino acid 29,780 Dalton GlpF protein (Weissenborn *et al*., 1992). GlpF primarily conducts glycerol, but it also can conduct urea, glycine, and D,L-glyceraldehyde. It is both stereo- and enantio-selective in conductance of linear carbohydrates called alditols (Heller *et al*., 1980). The cyclized alditols are not conducted through the GlpF channel and the structure of the GlpF channel shows that it is too small to conduct the cyclic molecules and explains the aspects of the stereo and enantioselectivity at the molecular level (Fu *et al*., 2000; Nollert *et al*., 2001).

Agre first recognized and named the family of proteins that conduct water as aquaporins (Smith & Agre, 1991). With the conservation of amino acid sequence between any two AQPs in the range of 28-32%, they are expected to share highly conserved structural backbone fold. The evidence of this similarity was seen in the highly homologous structural folds in the molecualr models of AQP1 and GlpF.

GlpF is a member of the aquaporin family found in *E.coli* that conducts water at about 1/6 of the rate its water conducting *E. coli* homolog AqpZ. The tetramer seen in the GlpF crystal structure has close similarity to the tetramers that associate in reconstituted membranes to form two-dimensional crystals, as characterized by image reconstructions of AQP1 (Cheng *et al*., 1997; Hasler *et al*., 1998; Li *et al*., 1997; Mitra *et al*., 2002; Mitsuoka *et al*., 1999; Ringler *et al*., 1999; Walz *et al*., 1997). The structures of AQPs at increasing resolutions obtained from electron microscopy (Braun *et al*., 2000; Stahlberg *et al*., 2000) and culminating in atomic structures from X-ray crystallography at a uniform 2.2 Å resolution for GlpF (Fu *et al*., 2000) and bovine AQP1 (Sui *et al*., 2001) illustrate the basis of their selectivity and rates of conductance. The atomic structures of GlpF and of AQP1 also serve as templates for molecular mechanics calculations for the rates of glycerol conductance in quantitative terms (de Groot *et al*., 2001; Jensen *et al*., 2002). The simulations based on the high resolution GlpF structure reveals that the GlpF channel is a remarkably well behaved template that remains stable to molecular mechanics using current force fields (de Groot & Grubmuller, 2001). One reason for this stability is the degree of fidelity of the final refined structures to the actual protein, but another is perhaps that a channel that conducts molecules of neutral charge, maintained by an appropriate approximation to a lipid bilayer is stably restrained by the balance of hydrophobic forces in the membrane.

2. Three dimensional structure of GlpF with glycerol in transit

The structure of GlpF at 2.2 Å resolution (residues 6-259), with its primary permeant substrate glycerol, elucidates the mechanism of selective permeability (Fu *et al*., 2000). Three intermediate glycerol binding sites, as shown in figure 1, define an amphipathic transmembrane pathway that conducts glycerol and certain other linear carbohydrates. Glycerol bound close to the narrowest portion of the channel around the highly conserved residue Arg^{206} shows that glycerol molecules move through the channel in single file, with the alkyl backbone wedged against a two-sided hydrophobic corner. Two successive OH groups each form hydrogen bonds with a pair of acceptor and donor atoms on the opposite sides of a four-sided tripathic channel.

This structure of the channel explains the preferential permeability for linear carbohydrates and absolute exclusion of ions and charged solute. The protein is constructed as two segments representing the genetic duplication seen in all the AQP sequences. These two segments are related by a quasi two-fold axis that would pass through the center of the membrane bilayer, and almost intersects the four fold axis of the tetramer. Channel specificity is delineated by three glycerol molecules (G1, G2, G3) with single water molecules between them and several water molecules that are bound in the entry vestibule and down to the selectivity filter in the center of the bilayer (figure 1). The O1-H of G1 remains hydrated by water molecules. G1 is found in the entry vestibule and is oriented by G1 O2-H as hydrogen bond donor to the O of Tyr^{138} (2.7 Å). G2 and G3 are both found in the selective region of the channel. G2 forms a hydrogen bond with one water molecule at its 1'-hydroxyl. Another bridging water molecule is hydrogen bonded both to G2 and G3, suggesting that glycerol and water are stoichiometrically co-transported.

The selectivity filter at G2 is strongly amphipathic, with two aromatic rings (Trp^{48}, Phe^{200}) forming a two-sided corner (figure 1), two N-H donors from the guanidinium sidechain of Arg^{206} on another side, and two main-chain carbonyl oxygens on the fourth side. G2 binds in the Trp-Phe-Arg triad leaving no free space around it, such that Van der Waals, hydrogen bond, and electrostatic forces each play a role. The OH1 and OH2 of G2 are each both hydrogen bond acceptors from NHs of the guanidinium group of Arg^{206} (2.9 Å, 2.7 Å), and hydrogen bond donors to the carbonyl oxygens of Gly^{199} (2.6 Å) and Phe^{200} (2.8 Å), respectively. The alkyl backbone of G2, tightly packed against the aromatic corner, leaves no space for any substitutions at the C-H hydrogen positions.

The conservation of the -NPA- sequences at 68-70 and 203-205 can be rationalized in both structural and functional terms.

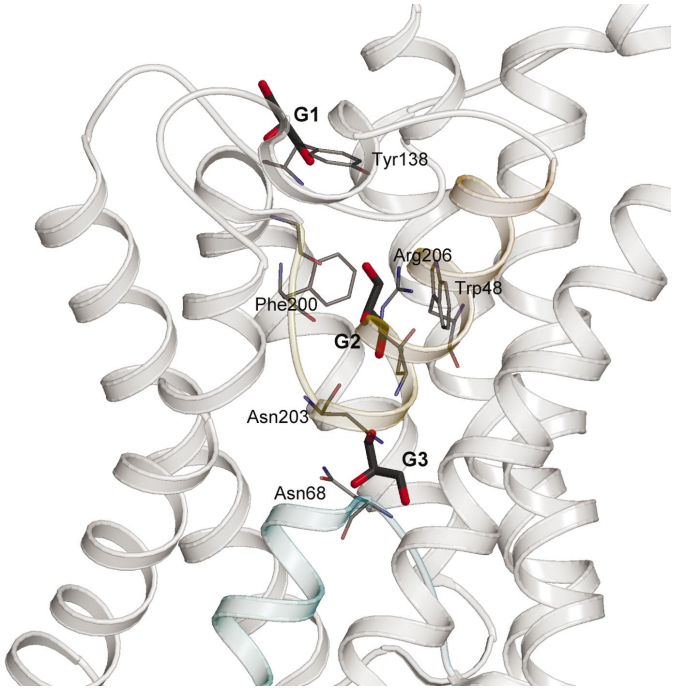

Figure 1
View of GlpF monomer with glycerol molecules G1, G2, and G3 highlighted. Also highlighted are residues of the channel wall that are hydrogen bond acceptors and donors that stabilize and orient the glycerol molecules in the channel and contribute to the specificity of the channel.

Structurally it preserves a key contact between M3 and M7. The structure explains why residues that are in contact around the quasi two-fold axis such as the pair of NPA motifs and the Gly^{49}-Gly^{184} pair would remain conserved, since each effectively resolves two packing situations at once. All the residues related by the quasi two-fold within GlpF are among the most conserved residues throughout the entire AQP protein family, suggesting that the GlpF structure is representative of all its AQP relatives.

3. Simulation and rates of water passage through the GlpF (an AQP) channel

Molecular mechanics provides a method to assess the rates of events utilizing the structural model that identifies each residue that contributes to the specific activity of the channel. This method is often criticized for causing structures to diverge from the experimentally observed data. However, it provides an option for a more rational and quantitative assessment of protein mechanisms. Molecular mechanics calculations can also account for the experimental values of measured properties, and the results can be used as a guide for designing the next stage of experiments.

In the lumen of GlpF channel, 20 hydrogen bond acceptors and 23 hydrogen bond donors participate in the conductance of glycerol (Nollert *et al*., 2001; Jensen *et al*., 2002). The molecular mechanics calculations done with the atomic model of GlpF results in the model remaining within 1 Å rms of the crystal structure, after a 1ns equilibration time. The channel structure observed after the calculations is essentially identical to the bounds of motion and structure obtained by X-ray crystallography.

The passage of glycerol causes little if any change in the structure of the channel (Tajkhorshid *et al*., 2002). One of the driving forces of the glycerol transport is presented as a concentration gradient, so that the free energy seen by the glycerol molecule has a net down hill slope (Jensen *et al*., 2002). The potential of mean force allows for the mapping of the energy landscape seen by a glycerol molecule in transit through the conducting pathway, and the experimentally observed binding sites have been justified in molecular and thermodynamic terms. Barriers inside the channel could be associated with transition state free energies, and hence rates of conduction. The energy profile also reveals a low energy periplasmic vestibule suited for efficient uptake of glycerol from the environment.

4. Conduction of proton by GlpF

AQPs are highly insulating to the electrochemical potential across the cell membrane and absolutely excludes conduction of all charged molecules, including hydroxide and hydronium ions (Finkelstein, 1987) and protons. The structural data on GlpF and AQP1 along with molecular mechanics has been used to identify and evaluate the contributions of each factor to this highly preferential selectivity (Tajkhorshid *et al*., 2002). The agreement between simulation and the crystal structure indicates that the water molecules form a stable line of 7-9 water molecules through the entire channel (figure 2). This observation also suggests that the force fields and assumptions necessary in molecular mechanics are a good representation of the situation in this type of channel. The water molecules are also polarized, so that each one is a donor to the eight carbonyls that line the channel, with the exception of the central water molecule, which is oriented by the central asparagine side chain NH (figure 2) and polarizes the entire line of waters that face outward from that central position. The helix dipoles of helices M3 and M7 also contribute to the overall polarization of the central channel (de Groot & Grubmuller, 2001; Tajkhorshid *et al*., 2002)

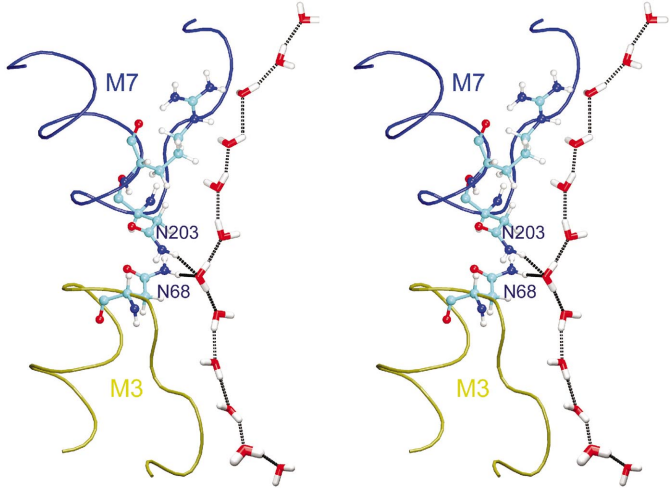

Figure 2
Image from a molecular mechanics simulation of water conduction through the channel. The central water molecule is constrained by the two conserved asparagines side chains (N68, N203) and is a hydrogen bond donor to the neighboring waters. The result is the polarization of the line of waters that prevents proton transfer along the line of waters. The water molecules are stabilized by the electrostatic fields of helices M3 and M7 while the central water molecule is stabilized by the side chains of N203 and N68. (figure adapted from Tajkhorshid, *et al*. 2002).

The conduction of protons through water can be remarkably fast, according to the Grotthus mechanism. The molecular mechanics observations of the contributors in the lumen of the GlpF channel show that the polarized central water, as implied by the crystallography, is key to this insulation against the Grotthus mechanism.

5. Utilization of mutagenesis to mimic structure and function of other aquaporins

The structural and functional observations reported about GlpF and other aquaporins are being tested by mutagenesis. Several residues in the lumen of GlpF have been mutated in attempt to convert the glycerol properties of GlpF toward those of the water aquaporins, negate the natural diploes of helices M3 and M7, and to mimic the structure and function of other aquaporins. Stopped-flow vesicular re-swelling assay of proteoliposomes formed with the first mutant GlpF, W24F/F200T, showed reduced glycerol efflux rates while the water efflux rate increased, demonstrating the shifting of GlpF channel properties towards that of the water aquaporins (Tajkhorshid *et al*. 2002). This result is remarkably consistent with expected results based on structural data and initial molecular dynamics simulations. Preliminary data on other mutants continue to support the initial results and suggest the possibilities of harvesting deeper insights into this large and physiologically relevant family of transmembrane channel proteins by applying the strategy of mutagenesis to emulate the function of other aquaporins within GlpF.

This work was carried out with support from NIH grant No. Gm 24485 to rms and support by the Kidney and Urology Foundation of America to jkl.

References

Braun, T., Philippsen, A., Wirtz, S., Borgnia, M. J., Agre, P., Kuhlbrandt, W., Engel, A., Stahlberg, H. (2000). *EMBO Rep*, **1**(2), 183-189.
Cheng, A., van Hoek, A. N., Yeager, M., Verkman, A. S., Mitra, A. K. (1997). *Nature* **387**(6633), 627-630.
de Groot, B. L., Engel, A., Grubmuller, H. (2001). *FEBS Lett* **504**(3), 206-211.
de Groot, B. L., Grubmuller, H. (2001). *Science* **294**(5550), 2353-2357.
Finkelstein, A. 1987. *Water movement through lipid bilayers, pores and plasma membranes, theory and reality.*, Wiley, New York, NY.
Fu, D., Libson, A., Miercke, L. J., Weitzman, C., Nollert, P., Krucinski, J., Stroud, R.M. (2000). *Science* **290**(5491), 481-486.
Hasler, L., Heymann, J. B., Engel, A., Kistler, J., Walz, T. (1998). *J Struct Biol* **121**(2), 162-171.
Heller, K. B., Lin, E. C., Wilson, T. H. (1980). *J Bacteriol* **144**(1), 274-278.
Jensen, M. O., Park, S., Tajkhorshid, E., Schulten, K. (2002). *Proc Natl Acad Sci U S A* **99(**10), 6731-6736.
Li, H., Lee, S., Jap, B. K. (1997). *Nat Struct Biol* **4**(4), 263-265.
Mitra, A. K., Ren, G., Reddy, V. S., Cheng, A., Froger, A. (2002). *Novartis Found Symp* **245**, 33-46.
Mitsuoka, K., Murata, K., Walz, T., Hirai, T., Agre, P., Heymann, J. B., Engel, A., Fujiyoshi, Y. (1999). *J Struct Biol* **128**(1), 34-43.
Nollert, P., Harries, W. E., Fu, D., Miercke, L. J., Stroud, R. M. (2001). *FEBS Lett* **504**(3), 112-117.
Ringler, P., Borgnia, M. J., Stahlberg, H., Maloney, P. C., Agre, P., Engel, A. (1999). *J Mol Biol* **291**(5), 1181-1190.
Sanno, Y., Wilson, T. H., Lin, E. C. C. (1968). *Biochem Biophys Res Comm* **32**(2), 344-349.
Smith, B. L., Agre, P. (1991). *J Biol Chem* **266**(10), 6407-6415.
Stahlberg, H., Braun, T., de Groot, B., Philippsen, A., Borgnia, M. J., Agre, P., Kuhlbrandt, W., Engel, A. (2000). *J Struct Biol* **132**(2), 133-141.
Sui, H., Han, B. G., Lee, J. K., Walian, P., Jap, B. K. (2001). *Nature* **414**(6866), 872-878.
Sweet, G., Gandor, C., Voegele, R., Wittekindt, N., Beuerle, J., Truniger, V., Lin, E. C., Boos, W. (1990). *J Bacteriol* **172**(1), 424-430.
Tajkhorshid, E., Nollert, P., Jensen, M. O., Miercke, L. J., O'Connell, J., Stroud, R. M., Schulten, K. (2002). *Science* **296**(5567), 525-530.
Walz, T., Hirai, T., Murata, K., Heymann, J. B., Mitsuoka, K., Fujiyoshi, Y., Smith, B. L., Agre, P., Engel, A. (1997). *Nature* **387**(6633), 624-627.
Weissenborn, D. L., Wittekindt, N., Larson, T. J. (1992). *J Biol Chem* **267**(9), 6122-6131.