nature structural & molecular biology

Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite *Plasmodium falciparum*

Zachary E R Newby[1], Joseph O'Connell III[1], Yaneth Robles-Colmenares[1], Shahram Khademi[1,2], Larry J Miercke[1] & Robert M Stroud[1]

The 2.05-Å resolution structure of the aquaglyceroporin from the malarial parasite *Plasmodium falciparum* (PfAQP), a protein important in the parasite's life cycle, has been solved. The structure provides key evidence for the basis of water versus glycerol selectivity in aquaporin family members. Unlike its closest homolog of known structure, GlpF, the channel conducts both glycerol and water at high rates, framing the question of what determines high water conductance in aquaporin channels. The universally conserved arginine in the selectivity filter is constrained by only two hydrogen bonds in GlpF, whereas there are three in all water-selective aquaporins and in PfAQP. The decreased cost of dehydrating the triply-satisfied arginine cation may provide the basis for high water conductance. The two Asn-Pro-Ala (NPA) regions of PfAQP, which bear rare substitutions to Asn-Leu-Ala (NLA) and Asn-Pro-Ser (NPS), participate in preserving the orientation of the selectivity filter asparagines in the center of the channel.

© 2008 Nature Publishing Group http://www.nature.com/nsmb

The single Aquaporin (AQP) of *Plasmodium falciparum* has the unusual property of conducting both glycerol and water, as efficiently as other glycerol channels and water water channels, respectively[1]. Thus, it provides a benchmark for determining what encodes high water conductance in the context of an efficient glycerol conductance. Plasmodia also are one of the few species known to date with not one, but two, sequence changes in the otherwise totally conserved NPA regions near the center of the channel.

Aquaporins fall into two subfamilies. One subfamily conducts water with a range of rates up to the diffusion limit for a pore that passes a single file of water. The other, aquaglyceroporins, conduct glycerol well at close to diffusion rates for glycerol and have low conductance of water. Aquaglyceroporins also pass small, uncharged solutes such as urea. They all accomplish this without leakage of any ions or protons. Whereas *Escherichia coli* has one of each subfamily, the high-conductance water channel AQPZ and the glycerol channel GlpF, which conducts water poorly, plasmodia have only one AQP.

We determined the crystal structure of PfAQP at 2.05-Å resolution. Glycerol and water molecules alternate in single file along the conduction channel, providing a detailed atomic view of the transport mechanism, in which each of the three hydroxyl groups of glycerol (CH_2OH-CHOH-CH_2OH) and each water molecule act both as an acceptor of a hydrogen bond and as a hydrogen bond donor. In comparison with GlpF, the structure offers a foundation to probe the distinction between a glycerol channel of poor water conductance and a glycerol channel of high water conductance. It also provides the atomic structure of an important protein in malarial pathology that may be relevant as an antimalarial drug target. To our knowledge, this represents the first atomic structure of a membrane protein from the malaria parasite and is in the small but growing category of a eukaryotic membrane protein structure that is expressed heterologously in fully functional form.

The importance of glycerol in the malarial life cycle was recognized when PfAQP, the sole protein from the aquaporin family present in *P. falciparum*, was cloned and characterized[1]. It was unexpected that there would only be one representative of the aquaporin family in the *P. falciparum* genome. Even *E. coli* contains two types of aquaporins, at least one specific for water transport (termed a canonical aquaporin) and at least one (termed an aquaglyceroporin) specific for transporting small, uncharged solutes. Humans have 4 glycerol channels among 13 AQPs.

Inside the mammalian host, the parasite invades the liver and undergoes asexual reproduction. Following rupture of the hepatocytes, the parasite is released back into the host's bloodstream where it invades erythrocytes. The blood stage of *P. falciparum* is marked by rapid metabolism in the course of a 32-fold multiplication that occurs within 2 d inside the erythrocyte. The rapid reproduction of *Plasmodium* in the host red blood cell requires massive biogenesis, including the synthesis of lipids for new membranes. The parasite accomplishes this by assembly from lipid precursors found in the host serum pool[2]. Glycerol from the host serum is incorporated into the lipids of newly

[1]Department of Chemistry and Chemical Biology Graduate Program, and Department of Biochemistry and Biophysics, Genentech Hall, School of Medicine, University of California in San Francisco, 600 16th Street, San Francisco, California 94158-2517, USA. [2]Present address: Department of Molecular Physiology and Biophysics, Carver College of Medicine, University of Iowa, 4-610 Bowen Science Building, Iowa City, Iowa 52242-1109, USA. Correspondence should be addressed to R.M.S. (stroud@msg.ucsf.edu).

Received 1 February; accepted 17 April; published online 25 May 2008; doi:10.1038/nsmb.1431

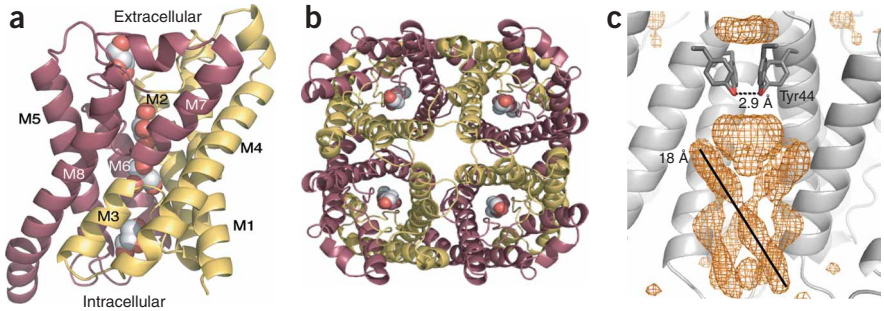

Figure 1 Channel architecture. Ribbon representations of the tetramer and monomer view of PfAQP. N- and C-terminal pseudo two-fold related portions of the channel are gold and maroon, respectively. (**a**) The PfAQP monomer constituting the asymmetric unit as viewed parallel to the membrane. Helices are labeled M1–M8. The pseudo two-fold axis protrudes from the page at the center of the membrane where helices M3 and M7 meet. The conduction pathway of glycerol is shown by sphere representations of the molecules observed in the crystal structure. (**b**) PfAQP tetramer viewed normal to the membrane down the crystallographic four-fold symmetry axis. Glycerol molecules contained in each conduction pore are shown in CPK representation. (**c**) Side view of the tetramer. Helices from two adjacent monomers are gray. Tyr44 residues from all four monomers are shown in stick representation. $F_o - F_c$ density around the four-fold axis is contoured at 3 σ and shown in orange mesh. The solid diagonal line represents the length of the density in the figure. All structural renderings were done with PyMOL (http://www.pymol.org).

© 2008 Nature Publishing Group http://www.nature.com/nsmb

synthesized parasite membranes[3]. Thus, glycerol is a potentially important molecule for the replication of the parasite.

Expression of the protein is limited to the surface of the parasite and is maximal during the later stages of the blood-stage parasite, when parasite growth is greatest[1]. Given the function and expression profile of PfAQP, it was speculated that this protein might have an important role in malarial biology by providing access to host serum glycerol pools and ensuring survival during osmotic stress[1,4]. Indeed, knocking out the ortholog of PfAQP in the rodent parasite *Plasmodium berghei* (PbAQP) impedes growth of the parasite, decreases parasitemia and increases the survival time of mice infected with the PbAQP-null parasites relative to mice infected with the wild-type parasite[5]. The same study also demonstrated that the mutant parasites have greatly diminished glycerol uptake when compared to wild-type *P. berghei*. Taken together, the results suggest that the malarial aquaglyceroporin is important in *Plasmodium* biology and pathology.

Correspondingly, glycerol uptake from the serum into erythrocytes is important for normal malarial virulence[6]. In murine red blood cells, the host membrane protein AQP9 is the major pathway for glycerol uptake into the red cell. AQP9 knockout increases the resistance of mice when challenged by wild-type *P. berghei*. After 2 weeks, 50% of infected wilde-type mice were dead, whereas none of the AQP9-null mice were dead, consistent with a requirement for efficient glycerol acquisition.

Plasmodia are pathogenic protozoa, transmitted by the bite of an *Anopheles* mosquito, that cause malaria in humans and mammals. The disease kills 1–3 million people each year and causes clinical illness in 300–500 million more. Of this total, *P. falciparum* accounts for the vast majority of death and illness in humans. This global health problem is compounded by the continuing emergence of resistance in *P. falciparum* to current therapeutics. The development of new drugs and the improvement of existing treatments rely on a detailed understanding of the molecular components involved in the complex malarial life cycle. PfAQP may thus be a plausible target for anti-malarial drug cocktails.

RESULTS

PfAQP expression of a synthetic gene in *E. coli*

A synthetic gene was key to expression of PfAQP. Attempts at heterologous expression of the native PfAQP gene in *E. coli* failed to yield protein detectable by western blot. This was not unexpected as expression of *Plasmodium* genes in heterologous systems is challenging because of the high A-T content in the genome of the parasite. The A-T content in the native gene encoding PfAQP is more than 70%, in agreement with the overall A-T content of genes in *P. falciparum*[7]. We designed a synthetic gene to match the codon usage to that of *E. coli* and to moderate the secondary structure from the 5′ end of the transcript that might otherwise impede translation of the gene. The gene was designed using GeMS[8]. Expression of the optimized gene in *E. coli* yielded ~1 mg of purified, correctly folded, tetrameric protein from 6 liters of culture.

PfAQP structure and organization

Tetrameric PfAQP crystallized in the I422 space group with one monomer per asymmetric unit (**Fig. 1a**). The tetrameric biological unit is symmetrically disposed around four-fold crystallographic symmetry axes (**Fig. 1b**). Each monomer contains an independent conduction pathway surrounded by a right-handed bundle of α-helices (M1–M8), as in all AQPs. The angle at which the transmembrane helices are oriented gives rise to funnel-shaped cytoplasmic and extracellular vestibules connected by the narrow, single file, ~25-Å conduction pore. Helices M1–M4 in the structure are related to M5–M8 by quasi two-fold symmetry in the plane of the membrane arising from an ancient gene-duplication event recognized throughout all aquaporins[9] (**Fig. 1a**).

In PfAQP, the external 25-amino-acid 'C loop' connecting helix M4 to M5 tucks into the protein core toward the conduction channel. In doing so, it defines a large portion of the extracellular vestibule. The C loop has a similar role in the water channel of erythrocytes, AQP1 (refs. 10,11). In PfAQP, the C loop begins with 1.5 turns of α-helix whereas the remainder lacks secondary structure, unlike the *E. coli* aquaglyceroporin, GlpF.

Tyrosine residues block the central four-fold axis

The symmetry axis in any homooligomeric transmembrane protein is always a potential, if not an actual, conducting transmembrane pathway as, in principle, no element of the protein can lie on the axis because it is represented symmetrically four times over. The four-fold axis of PfAQP is sterically occluded by a unique arrangement of four tyrosine side chains, Tyr44, which form a 'fireman's grip' arrangement, in which each ring forms a nearly 90° edge to π interaction with its neighbor. The tetrad is sealed by a square planar arrangement of ideally oriented hydrogen bonds between each terminal hydroxyl group group of each tyrosine, as hydrogen bond donor to the hydroxyl group of its neighbor in the cyclic array (**Fig. 1c**). The axial portal to the four-fold axis of the protein seems more open than that of other AQPs and shows density for what could be four ~20-Å aliphatic chains within the portal. These four aliphatic chains could possibly represent lipid chains of two phospholipids, or fatty acids that presumably collaborate to seal the cytoplasmic side of the portal.

© 2008 Nature Publishing Group http://www.nature.com/nsmb

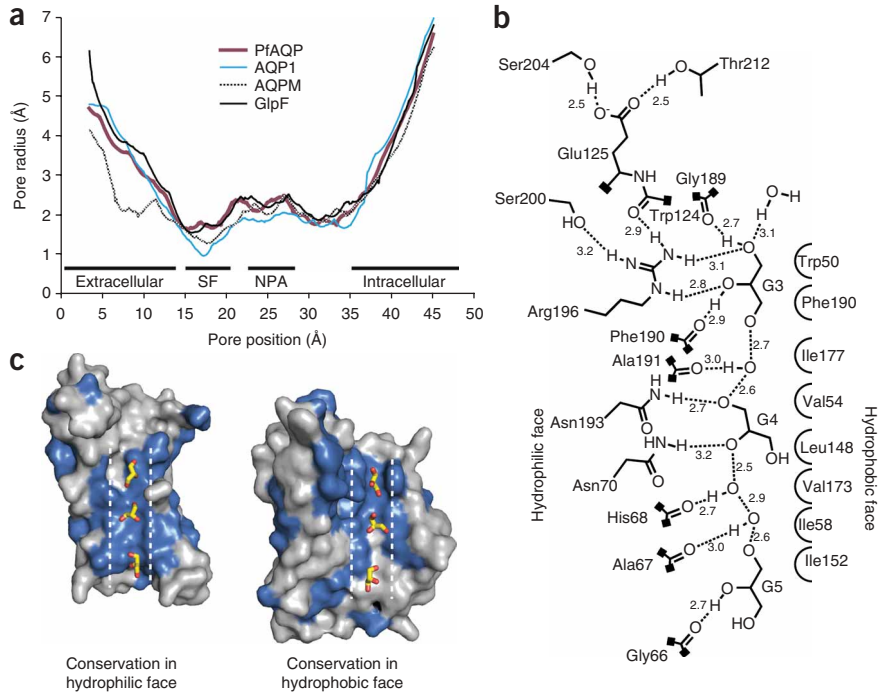

Figure 2 Details of the PfAQP conduction channel. (**a**) Plot of channel radius versus position along the conduction pore for the water-specific AQP1 and the aquaglyceroporins AQPM, GlpF and PfAQP. The regions of the channel comprising the extracellular vestibule, intracellular vestibule, selectivity filter (SF) and NPA motif region are labeled below. All calculations of pore dimensions were done with the program Hole2 (ref. 26). (**b**) Schematic of protein interactions with channel contents. Generated with ChemDraw Ultra (CambridgeSoft). (**c**) PfAQP is represented by a surface representation of two opened halves that contribute the predominantly hydrophobic (right) or hydrophilic (left) interactions to the conduction channel. The blue surface indicates residues that are identical to GlpF. Nonidentical residues are gray. Stick representations of the three glycerol molecules in the conduction channel are shown. The boundaries of the conduction channel are roughly defined by dashed lines.

Thus, overall, the four-fold axis is impenetrably sealed both by tyrosines of the protein and potentially by lipid chains.

The extracellular vestibule

There are five glycerol molecules associated with each monomer in the structure. Each extracellular vestibule contains two glycerol and seven water molecules, whereas each conduction pore contains three glycerol and four water molecules that alternate in lining up in single file. The vestibular glycerol molecules interact with the interhelical loops of the channel in the vestibule (**Supplementary Fig. 1** online). The O_2 of glycerol G1 is involved in a hydrogen bond (3.0 Å) with the amide backbone of Gly38 located on the loop connecting M1 and M2 in the neighboring monomer. Glycerol G2 donates a hydrogen bond through O_2 to the backbone carbonyl of Ser127 (2.8 Å) located on the C loop connecting M4 and M5 (**Supplementary Fig. 1**). G2 also comes into close contact with the loop connecting M6 to the end of the half-helix M7 that is closest to the quasi two-fold axis. The interaction of the channel with the glycerol molecules orients the polar alcohol moieties toward the protein surface and the hydrophobic backbones of the molecules toward the axis of the conduction channel.

The conduction pore

The conduction pore is ~25 Å long and constricts to ~3 Å at its narrowest region (**Fig. 2a**). It contains three glycerol and four water molecules arranged such that the glycerol molecules in the pore are always separated by at least one water molecule, as in GlpF. The hydroxyl groups of the molecules in transit are hydrogen bond donors to the eight carbonyls that line one stripe throughout the length of the channel, and acceptors of hydrogen bonds from Arg196 and from the signature asparagines of the two NPA motifs (NLA at the end of half-helix M3 and NPS at the end of half-helix M7) at the channel center (**Fig. 2b**). The pathway itself is amphipathic; the hydrophobic backbone of glycerol is oriented so that it contacts the hydrophobic side chains that constitute the almost vertical walls of the transmembrane pathway channel. There is a high degree of conservation of the residues constituting the pore of PfAQP and GlpF, especially on the hydrophilic face (**Fig. 2c**). This helps to frame the analysis of the difference between these two remarkably similar structures.

The selectivity filter

A single, well-resolved glycerol molecule is found in the selectivity filter (**Supplementary Fig. 2** online). The hydrophobic backbone of glycerol packs tightly against the V-shaped surface formed by the side chains of Trp50 and Phe190, whereas each of two successive hydroxyl groups of the glycerol substrate are hydrogen bond acceptors from the εNH (2.80 Å) and NH_2 (3.08 Å) of Arg196, and hydrogen bond donors to the backbone carbonyls of Gly189 and Phe190, respectively. The selectivity filter in PfAQP is identical in amino acid composition to that of GlpF. Together, these are the only aquaporin structures determined to date that have a glycerol molecule bound at the filter region[9], although glycerol was often present in the crystallization media of AQPZ, for example. The pore size at the selectivity filter in PfAQP and GlpF also match closely: 3.28 Å in PfAQP and 3.14 Å in GlpF. This is in agreement with conductance measurements showing that PfAQP and GlpF are the only aquaporins of known structure to have appreciable rates of glycerol conductance[1,12].

The NH_1 and NH_2 of Arg196, and the backbone carbonyls of Gly189 and Phe190 on the opposite polar surface, provide two key oriented donor and two acceptor hydrogen-bonding partners that select two adjacent amphipathic hydroxyl groups as essential to passage. Glycerol G3 interacts with this arginine in PfAQP exactly as it does with that of GlpF, even to the choice of prochiral species at this site. (**Supplementary Fig. 2**).

The side chain of Arg196 is a focus of mechanistic concern. The NH_1 of Arg 196 donates a hydrogen bond back onto its backbone carbonyl and another to the hydroxyl of Ser200 (**Fig. 3**). The NH_2 of Arg196 donates a hydrogen bond to the backbone carbonyl of Trp124. These hydrogen-bonding interactions, and other hydrophobic interactions of the side chain, constrain the arginine residue in the selectivity filter. This region of the channel is key to determining the chiro- and enantioselectivity of conductance and the preference for water versus glycerol[9,13–15].

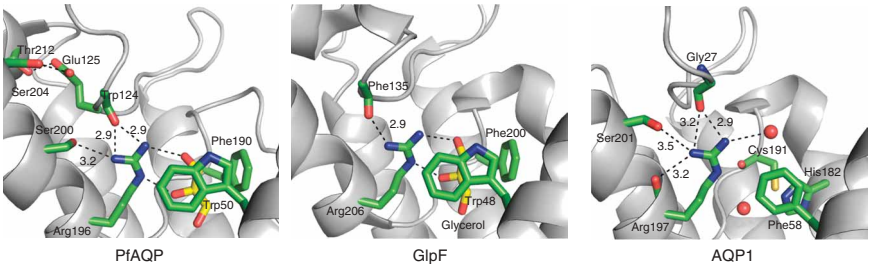

Figure 3 Selectivity in aquaporin channels. Ribbon representations of PfAQP, GlpF and AQP1. Residues comprising the selectivity filter of each channel are shown in stick representation. Glycerols are drawn as yellow sticks and waters as red spheres. The hydrogen bond network around each arginine residue is shown with measurements in angstroms. Helices M1 and M4 have been removed for easier viewing of the selectivity filter region.

© 2008 Nature Publishing Group http://www.nature.com/nsmb

Conductance of PfAQP in proteoliposomes

Water and glycerol conduction through PfAQP were assayed after reconstitution of the protein into proteoliposomes. Light-scattering measurements monitor vesicle shrinkage in response to either water or glycerol efflux. The rate for glycerol conductance in proteoliposomes containing PfAQP is 15.8 ± 0.2 compared to 0.066 ± 0.003 for liposomes. The rate for water conductance in proteoliposomes containing PfAQP is 21.5 ± 0.8 compared to 4.5 ± 0.1 for liposomes. These values are relative rather than absolute numbers, depending on the number of channels incorporated per liposome, the sizes of liposomes and so on. However, they serve as benchmarks for comparison between different permeants and can be compared directly with the ratios of glycerol to water conductance found for GlpF (**Fig. 4**). These assays also demonstrate that the eukaryotic protein heterologously expressed in *E. coli* is fully functional.

DISCUSSION

Extracellular vestibule recruits solute molecules

Glycerol in the extracellular vestibule increases the local concentration of glycerol at the channel opening and begins the process of desolvating the molecules for passage through the narrow conduction pore. This role has been proposed for other glycerol-conducting aquaglyceroporins[9,16]. The X-ray structure of the water-specific AQPZ does not have glycerol in its extracellular vestibule, despite the presence of a high concentration (~5% (v/v)) of the solute in the crystallization condition[17]. The crystal structure of GlpF has one glycerol molecule at the base of the extracellular vestibule above the selectivity filter[9]. This observation is explained and further supported through molecular dynamics studies of GlpF and AQPZ, which reveal a deep energy well for glycerol binding in the extracellular vestibule of GlpF but not AQPZ[14].

Compensatory mutations in the NPA region

Owing to the close interdigitation of the twinned NPA motifs with each other and the structural and functional role these motifs have, it would be expected that the tolerance for mutations in this region is low. This is reflected in the near-total conservation of the two motifs across all other aquaporin family members of known amino acid sequence. The two almost totally conserved NPA motifs in AQPs are found at the N-terminal ends of the half membrane–spanning helices M3 and M7, which meet at the center of the membrane plane. The NPA regions meet in an antiparallel manner such that the proline residues are in van der Waals contact with each other and also with the alanine of the opposite NPA motif. The side chain residues of the asparagines project toward the conduction channel center, and they are maintained in this position by the interactions of NPA motifs with each other and also by a hydrogen bond donated from the amide backbone of the alanine to the carbonyl group of the asparagine side chain within the same NPA motif.

All aquaglyceroporin sequences (from *P. falciparum, P. berghei, P. yoellii, P. vivax* and *P. knowlesi*) found on the *Plasmodium* genome database, PlasmoDB (http://plasmodb.org/plasmo/) are altered twice, to become NLA at position 70–72 at the end of helix M3 and NPS at position 193–195 at the end of helix M7. The only exception to this is the AQP sequence from *P. knowlesi*, which has a phenylalanine instead of a leucine substitution. PfAQP provides the first structure of an aquaporin containing this rare divergence in the NPA motifs. The loss of van der Waals interaction that occurs with the substitution of Leu71 for a proline in the NLA region is compensated for by a hydrogen bond donated by the amide nitrogen of Leu71 to the hydroxyl group of Ser195 (3.1 Å) in the NPS region (**Fig. 5a**). Furthermore, the position of the Ser195 side chain is restricted in its position by steric clashes with Asn193 or Leu96 in the two alternate rotamer states. Thus, the unique pair of substituents, one in each of the NPA regions, interdigitate against each other and explain the double mutation as a covariant alteration.

In addition to disrupting the proline-proline interaction between NPA motifs, the substitution of Leu73 for a proline introduces a bulkier hydrophobic residue in this region of the protein. The leucine side chain projects away from the channel pore, out toward the surface of the protein that would contact the lipid membrane. A series of mutations in this region of the protein compensate for the introduced bulk while maintaining the overall hydrophobicity of the surface (**Fig. 5b**). In particular, Cys228 accommodates for the bulk of Leu71 at a position occupied by the large hydrophobic residues of isoleucine, leucine or phenylalanine in nearly all other aquaporins. Therefore, the cost of the leucine alteration is compensated for by covariant changes not just in the NLA and NPS regions, but throughout the protein, reaching all the way out to the periphery.

In view of the rarity of this plasmodial sequence variant, it attracts attention as a source of the difference in selectivity from GlpF. However, the covariant mutation provides compensatory interactions that serve to maintain the overall structure and orientation of helices

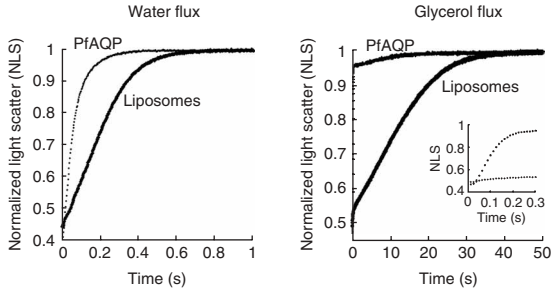

Figure 4 Water and glycerol conduction assay. Water (left) and glycerol (right) flux measurements in liposomes with PfAQP. Control measurements were made in empty liposomes containing no protein. Rates were determined from a fitting of a single exponential curve to the data ± s.d. from seven replicate measurements. The inset in the glycerol conduction plot is an expansion of the initial kinetics for glycerol conduction.

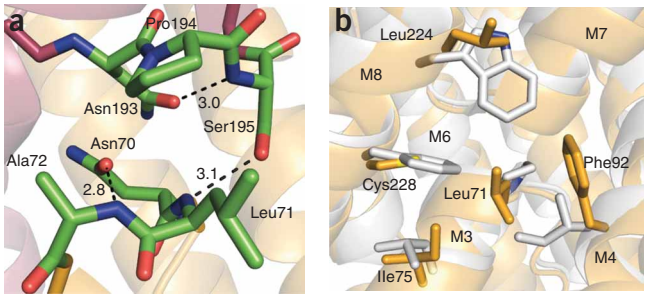

Figure 5 Alterations to the conserved NPA motif in PfAQP. (**a**) Ribbon representation of PfAQP showing the NLA and NPS regions of PfAQP in stick representation. Hydrogen bonds are indicated with dashed lines. (**b**) Structural alignment of PfAQP (gold) and AQP1 (silver). For both proteins, helices are shown in ribbon representation and residues around the Leu71 of PfAQP and corresponding proline in AQP1 are shown in stick representation.

M3 and M7. Furthermore, other conserved interactions in the NLA and NPS regions remain intact despite the covarient changes. As observed in all other aquaporins, the carbonyl of Asn70 acts as a hydrogen bond acceptor from the backbone amide of Ala72, and Asn193 acts as a hydrogen bond acceptor from the backbone amide of Ser195 (**Fig. 5a**), which would be an alanine in most other aquaporin family members. This maintains the presentation of two highly oriented N-H donors from each of the asparagine residues into the center of the conduction channel. Therefore, the signature interaction in AQPs that locks the N-terminal ends of the M3 and M7 helices together and precisely presents the asparagine residues into the channels is preserved despite the sequence variation. Thus, the variants are unlikely to provide a rationale for the difference in selectivity.

Selectivity of PfAQP

The dual specificity of PfAQP for glycerol and water separate it from other aquaporin family members whose structures have been determined. Although the overall protein architecture is highly conserved in all aquaporin family members, there are significant selectivity differences between aquaporins and aquaglyceroporins. The narrowest constriction of the conduction pore has attracted attention as a possible determinant of water versus glycerol selectivity. This region, the selectivity filter, is comprised of amino acid side chains and backbone carbonyl groups that project toward the center of the pore. In the water-specific aquaporin channels, the exclusion of glycerol and other small solutes can be explained by the diameter of the conduction pore at the selectivity filter, which is perhaps too small for the passage of solutes, and possibly by the more aliphatic hydrophobic side chains, which may provide less ideal contact with hydrophobic solutes (**Fig. 2a** and **Supplementary Fig. 2**). The mechanism by which aquaglyceroporin channels pass solute molecules through a larger-diameter selectivity filter while permitting water passage at much lower rates than aquaporins is still unresolved.

PfAQP transports water as efficiently as some water-specific channels, yet also passes glycerol well[1]. To accommodate glycerol molecules, the selectivity filter in PfAQP, as in GlpF, has a larger diameter and a more hydrophobic composition than water-specific channels. However, the fact that the selectivity filter of PfAQP and GlpF, an aquaglyceroporin with lower water conductance than PfAQP, are nearly identical (**Supplementary Fig. 2**) indicates that other factors besides just amino acid composition and arrangement influence the water conduction in this family of channels.

The arginine residue in the selectivity filter is highly conserved in both aquaporin and aquaglyceroporin channels. The hydrogen bond network that this arginine is involved in provides insight into how the aquaglyceroporin PfAQP might achieve water conductivity similar to aquaporin channels. In all of the X-ray structures of water-specific channels, the NH_1 and NH_2 of this arginine together donate at least one hydrogen bond each to a protein moiety. In GlpF, only the NH_1 is involved in hydrogen-bonding interactions with the protein (**Fig. 3**). The NH_2 that is most proximal to the conduction channel is left to interact only with solvent molecules and not with any protein moieties. The greater number of hydrogen bonds formed between the arginine donors and acceptors on the protein in water-specific channels serves to decrease the desolvation energy of arginine, allowing for more transient interactions of water molecules with this arginine. In PfAQP, the NH_1 and NH_2 of the selectivity filter arginine donate three hydrogen bonds to the protein, as in most other highly water-permeable channels (**Fig. 3**). Thus, PfAQP has the pore diameter and amino acid composition to facilitate passage of glycerol but also establishes a hydrogen-bonding network around the conserved arginine, similarly to water-specific channels that decrease the desolvation energy of the guanidinium cation, allowing for more rapid passage of water compared to GlpF.

A glutamate mutant that abrogates water conductance

Previous conduction assays performed on PfAQP mutants support the arginine hypothesis[4]. Glu125 was previously identified as contributing to PfAQP selectivity during mutational analysis of residues predicted to be in the vicinity of the extracellular pore. The residue was identified as part of an effort to determine amino acids that participate in the conduction of water and glycerol through the channel. Predictions were based on a homology model of PfAQP using GlpF as a template. The E125S mutation was identified as disrupting water conduction, but not glycerol conduction, through the channel[4], and is therefore useful in probing the basis of selectivity for glycerol versus water in AQPs.

Glu125 is located on the C loop, the long loop that dips into the extracellular vestibule to come into close contact with Arg196 in the selectivity filter. Glu125 serves to anchor this loop in place through two shorter hydrogen bonds from the donor hydroxyl group of Ser202 (2.5 Å) and that of Thr212 (2.5 Å) (**Fig. 3**). The backbone carbonyl group of the residue that neighbors Glu125, Trp124, is the acceptor of a hydrogen bond from Arg196 NH_2 in the selectivity filter. This is a crucial interaction in maintaining the water permeability of the channel. The E125S mutation eliminates this stabilization of the C loop and therefore would disrupt the hydrogen-bonding of Trp124 with Arg196, resulting in the NH_2 of Arg196 having increased solvent interactions and a higher desolvation penalty, so presenting a higher barrier for passage of water. The hydroxyl groups of glycerol, on the other hand, have lower polarity than in water, and thus are not held back to the same degree.

METHODS

Protein expression and purification. We cloned a codon-optimized gene (Kosan Biosciences) encoding the PfAQP protein into a pET28b expression vector with an N-terminal hexahistidine (His_6)-affinity tag and thrombin cleavage site (Novagen). *Escherichia coli* C43(DE3) cells were transformed and plated on LB agar with 15 μg ml^{-1} kanamycin. All cultures were grown with LB media plus 15 μg ml^{-1} kanamycin at 37 °C in Fernbach flasks shaking at 200–225 r.p.m. For expression, each liter of growth was inoculated with 10 ml of an overnight culture started from a freshly transformed colony and grown to an optical density at 600 nm (OD_{600}) of 0.4–0.6. Protein expression was induced with 1 mM IPTG, and cultures were harvested by centrifugation for 15 min at 4 °C 5,000*g* 4 h after induction. Pellets were washed once with

© 2008 Nature Publishing Group http://www.nature.com/nsmb

Table 1 Data collection and refinement statistics

	PfAQP
Data collection	
Space group	*I*422
Cell dimensions	
a, *b*, *c* (Å)	92.4, 92.4, 193.2
Resolution (Å)	2.05 (2.11–2.05)
R_{sym}	8.4 (12.1)[a]
$I / \sigma I$	28.8 (2.3)
Completeness (%)	95.7 (69.9)
Redundancy	15.0 (8.1)
Refinement	
Resolution (Å)	46.2–2.05
No. reflections	24,179
R_{work} / R_{free}[b]	18.1/19.8
No. atoms	
Protein	1,880
Glycerol	30
Water	56
B-factors, Å^2	
Protein	36
Glycerol	68
Water	57
R.m.s. deviations	
Bond lengths (Å)	0.018
Bond Angles (°)	1.63

Diffraction data from two crystals were used to determine the structure.
[a]Values in parentheses are for the highest resolution shell. [b]R_{free} is calculated from 5% of reflections chosen randomly.

20 mM Tris, pH 8 at room temperature, 20 mM EDTA (EDTA), 4 mM β-mercaptoethanol and 1 mM PMSF and pelleted as described above. Cells were resuspended in 20 mM Tris, pH 7.5 at room temperature, 500 mM NaCl, 10 mM EDTA, 4 mM β-mercaptoethanol and 1 mM PMSF, and lysed with 3–5 passages at 10,000–15,000 ψ in an EmulsiFlex-C5 high-pressure homogenizer (Avestin). *Escherichia coli* membranes were pelleted at 4 °C and 160,000*g* for 1 h. Pellets were resuspended in 20 mM Tris, pH 7.4 at room temperature, 100 mM NaCl, 20% (v/v) glycerol and 4 mM β-mercaptoethanol (1ml buffer per 0.5g membrane pellet). At this point, membranes were either stored at –80 °C or processed. Protein was solubilized by adding 270 mM n-octyl-β-D-glucopyranoside (β-OG, Anatrace) to the resuspended membranes and stirring at 4 °C for 1 h. Unsolubilized material was pelleted at 4 °C and 26,500*g* for 1 h. The supernatant was loaded onto a nickel–nitrilotriacetic acid (Ni-NTA) resin (Qiagen), washed consecutively with 20 mM Tris, pH 7.4 at room temperature, 100 mM NaCl, 5% (v/v) glycerol, 40 mM β-OG, 4 mM β-mercaptoethanol containing 10 mM, 25 mM, 40 mM and 45 mM imidazole, and then eluted with 300 mM imidazole. Imidazole was removed using an Econo-Pac DG10 desalting column (Bio-Rad). The His_6 tag was removed using 4 units of thrombin (Novagen) per milligram of PfAQP at 4 °C for 12–16 h. Thrombin was removed with benzamidine Sepharose resin (GE Healthcare), before the final protein-purification step achieved by size-exclusion chromatography on a TSK3000 column (Tosoh Biosciences) in 20 mM HEPES, pH 7.4, 100 mM NaCl, 10% (v/v) glycerol and 2 mM DTT.

Crystallization. Purified protein was concentrated in a 50,000 molecular weight cutoff Amicon Ultra Centrifugal Filter Device (Millipore) to 6 mg ml^{-1} and crystallized in 17% (w/v) polyethylene glycol monomethyl ether 2000 (PEG2000MME), 100 mM sodium citrate, pH 5.8, 15% (v/v) glycerol, 40 mM β-OG, 10 mM nickel sulfate and 10 mM DTT in hanging-drop plates (Nextal Biotechnologies) by vapor diffusion at 18 °C. Crystals grew to ~100 μm in the longest axis within 2 d and were harvested within 1 week. An initial crystal-dehydration solution was made by mixing 80 μL of the mother liquor with 20 μL of 50% (v/v) glycerol. After soaking for 15 min, crystals were transferred to a solution with a 2% increase in PEG2000MME. This was repeated up to 27% (w/v) PEG2000MME. Crystals were flash frozen from this solution and stored in liquid nitrogen.

Data collection and model building. Diffraction data were collected at 1.11587 Å under cryoconditions at Beamline 8.3.1 (Lawrence Berkeley National Laboratory). Data were processed using DENZO/SCALEPACK[18]. Molecular replacement was performed with Phaser[19] using the transmembrane portions of GlpF[9] as a search model. Initial automatic model building was done with ARP/wARP[20]. Incomplete portions of the model were subsequently built with iterative cycles of manual building using Coot[21] and restrained refinement with translation, libration and screw-rotation displacement (TLS)[22] using Refmac5 (ref. 23) in the CCP4 suit of programs[24]. In the Ramachandran plot, 96.7%, 3.3% and 0% of the residues fell in favored, allowed and outlier regions according to MolProbity[25]. Data collection and refinement statistics are summarized in **Table 1**.

Proteoliposome assay. *Escherichia coli* polar lipid extracts acetone/ether preparation (Avanti Polar Lipids) were used to generate liposomes for assays of conductivity. All buffers were purged with argon. Lipid stocks were made at 50 mg ml^{-1} in water plus 4 mM β-mercaptoethanol and stored at –80 °C. Before the assay, stocks were thawed and 300 μL lipid and 100 μL water were aliquoted into 16 × 125 mm glass culture tubes (VWR), overlaid with argon and sonicated to clarity in a bath sonicator (Laboratory Supplies Company, Inc.). Cocktails for reconstitution were formed by mixing (in order): 100 mM MOPS (pH 7.5), 43 mM β-OG, protein (final concentration 125 μg ml^{-1}) and sonicated lipids (final concentration 125 μg ml^{-1}). The total volume of the mixture was 1.5 ml. The cocktails were dialyzed against assay buffer (20 mM HEPES, pH 7.4) in a 25,000 molecular weight cut off Spectra/Por Float-a-lyzer (Spectrum Laboratories, Inc.) for 24–72 h. Liposomes were then harvested by centrifugation at 75,000*g* and 4 °C for 1 h. Liposomes for water conduction assays were resuspended into 1 ml of assay buffer. Liposomes for glycerol conduction assays were resuspended into 1 ml of 20 mM HEPES, pH 7.3, and glycerol. The concentration of glycerol was adjusted to make the solution match the osmolarity of a 20 mM HEPES, 570 mM sucrose solution. Vesicle shrinkage in either the water of the glycerol conduction assays was induced by mixing 100 μl of liposome solutions and 100 μl of the sucrose solution in a stopped-flow apparatus. Light scattering at 440 nm was monitored over time. The resulting curves were fitted to a single exponential and the rate was extracted from these curves.

Accession codes. Protein Data Bank: Coordinates for PfAQP have been deposited with accession code 3C02.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS
Research was supported by the US National Institutes of Health (NIH) grant GM24485 to R.M.S. and by the NIH Roadmap center grant P50 GM073210. The authors thank the reviewers for their helpful suggestions. The authors thank F.A. Hays, P. Egea, J. Lee and D. Savage in the Stroud laboratory for helpful advice throughout the project and with preparation of the manuscript. The authors thank T.P. Kearney III and M. Miller for advice on the manuscript. We thank J. Holton for his assistance at the Advanced Light Source (ALS) beamline 8.3.1, supported by NIH grant GM074929 (R.M.S.).

AUTHOR CONTRIBUTIONS
Z.E.R.N., J.O. III, Y.R.-C. and S.K. carried out experiments; Z.E.R.N. collected diffraction data and determined the structure; R.M.S. and L.J.M. supervised the research; Z.E.R.N. and R.M.S. wrote the manuscript.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Hansen, M., Kun, J.F., Schultz, J.E. & Beitz, E. A single, bi-functional aquaglyceroporin in blood-stage *Plasmodium falciparum* malaria parasites. *J. Biol. Chem.* **277**, 4874–4882 (2002).

© 2008 Nature Publishing Group http://www.nature.com/nsmb

2. Vial, H.J. & Ancelin, M.L. Malarial lipids. An overview. *Subcell. Biochem.* **18**, 259–306 (1992).
3. Vial, H.J., Ancelin, M.L., Thuet, M.J. & Philippot, J.R. Phospholipid metabolism in *Plasmodium*-infected erythrocytes: guidelines for further studies using radioactive precursor incorporation. *Parasitology* **98**, 351–357 (1989).
4. Beitz, E., Pavlovic-Djuranovic, S., Yasui, M., Agre, P. & Schultz, J.E. Molecular dissection of water and glycerol permeability of the aquaglyceroporin from *Plasmodium falciparum* by mutational analysis. *Proc. Natl. Acad. Sci. USA* **101**, 1153–1158 (2004).
5. Promeneur, D. *et al.* Aquaglyceroporin PbAQP during intraerythrocytic development of the malaria parasite *Plasmodium berghei*. *Proc. Natl. Acad. Sci. USA* **104**, 2211–2216 (2007).
6. Liu, Y. *et al.* Aquaporin 9 is the major pathway for glycerol uptake by mouse erythrocytes, with implications for malarial virulence. *Proc. Natl. Acad. Sci. USA* **104**, 12560–12564 (2007).
7. Gardner, M.J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
8. Jayaraj, S., Reid, R. & Santi, D.V. GeMS: an advanced software package for designing synthetic genes. *Nucleic Acids Res.* **33**, 3011–3016 (2005).
9. Fu, D. *et al.* Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486 (2000).
10. Sui, H., Han, B.G., Lee, J.K., Walian, P. & Jap, B.K. Structural basis of water-specific transport through the AQP1 water channel. *Nature* **414**, 872–878 (2001).
11. Fujiyoshi, Y. *et al.* Structure and function of water channels. *Curr. Opin. Struct. Biol.* **12**, 509–515 (2002).
12. Kozono, D. *et al.* Functional expression and characterization of an archaeal aquaporin. AqpM from *Methanothermobacter marburgensis*. *J. Biol. Chem.* **278**, 10649–10656 (2003).
13. Fu, D., Libson, A. & Stroud, R. The structure of GlpF, a glycerol conducting channel. *Novartis Found. Symp.* **245**, 51–61; Discussion 61–5, 165–8 (2002).
14. Wang, Y., Schulten, K. & Tajkhorshid, E. What makes an aquaporin a glycerol channel? A comparative study of AqpZ and GlpF. *Structure* **13**, 1107–1118 (2005).
15. Jensen, M.O., Park, S., Tajkhorshid, E. & Schulten, K. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci. USA* **99**, 6731–6736 (2002).
16. Lee, J.K. *et al.* Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 Å. *Proc. Natl. Acad. Sci. USA* **102**, 18932–18937 (2005).
17. Savage, D.F., Egea, P.F., Robles-Colmenares, Y., O'Connell, J.D. III & Stroud, R.M. Architecture and selectivity in aquaporins: 2.5 X-ray structure of aquaporin Z. *PLoS Biol.* **1**, e72 (2003).
18. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
19. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 458–464 (2005).
20. Perrakis, A., Morris, R. & Lamzin, V.S. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463 (1999).
21. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
22. Winn, M.D., Isupov, M.N. & Murshudov, G.N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 122–133 (2001).
23. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
24. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
25. Davis, I.W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
26. Smart, O.S., Goodfellow, J.M. & Wallace, B.A. The pore dimensions of gramicidin A. *Biophys. J.* **65**, 2455–2460 (1993).

© 2008 Nature Publishing Group http://www.nature.com/nsmb