Research Letter

A refined structure of human aquaporin-1

Bert L. de Groot[a], Andreas Engel[b], Helmut Grubmüller[a],*

[a] *Max Planck Institute for Biophysical Chemistry, Theoretical Molecular Biophysics Group, Am Fassberg 11, 37077 Göttingen, Germany*
[b] *M.E. Müller Institute for Microscopic Structural Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland*

Abstract **A refined structure of the human water channel aquaporin-1 is presented. The model rests on the high resolution X-ray structure of the homologous bacterial glycerol transporter GlpF, electron crystallographic data at 3.8 Å resolution and a multiple sequence alignment of the aquaporin superfamily. The crystallographic R and free R values (36.7% and 37.8%) for the refined structure are significantly lower than for previous models. Improved geometry and enhanced stability in molecular dynamics simulations demonstrate a significant improvement of the aquaporin-1 structure. Comparison with previous aquaporin-1 models shows significant differences, not only in the loop regions, but also in the core of the water channel.** © 2001 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

Key words: Protein structure; Electron microscopy; Water transport; Water channel; Membrane protein; Glycerol transporter; Aquaporin-1; GlpF

1. Introduction

The atomic structures of two members of the aquaporin family, human aquaporin-1 (AQP1) and the bacterial glycerol transporter GlpF, have recently been presented [1–3]. These structures have provided a wealth of information about the molecular mechanism of passive transport of neutral solutes across lipid bilayer membranes. Both proteins contain a long, narrow and relatively hydrophobic pore, and despite conservation of many amino acids facing the pore, AQP1 is a highly selective water channel [4] whereas GlpF also conducts glycerol and few other linear alcohols [5,6,3].

Two atomic models of the structure of human AQP1 have recently been published [1,2]. Both are based on cryo-electron crystallographic data sets, at 3.8 Å and 3.7 Å resolution, respectively. At more than 3 Å backbone root mean square deviation (RMSD), the structural differences between the two models are relatively large, even for the medium resolution available. The structure of GlpF was solved with X-ray crystallography at a resolution of 2.2 Å [3]. The two proteins share an overall sequence identity of 30.6%; the transmembrane region is better conserved than the helix-connecting loops. The two proteins are both active as tetramers [7], with each monomer contributing a channel [8]. The N- and

C-terminal halves of the monomers are sequence related due to an ancient gene duplication [9,10]. The two halves, which are connected by a long loop spanning the extracellular face of the protein, both contain three transmembrane-spanning helices [11]. Both the N- and the C-terminus are located on the cytoplasmic side of the membrane [12,11]. The fingerprint Asn-Pro-Ala motif, conserved throughout the aquaporin family, is located in loops B and E that connect the second and third helix in each half of the monomer. In both previous AQP1 models and in the GlpF structure, these NPA motifs meet in the central part of the channel, where the conserved B and E loops fold back into the protein, as predicted by the hourglass model [8]. The arrangement of the helices is in good agreement with lower resolution results [13,14] and with a detailed analysis of the amino acid sequences of the aquaporin family [15].

We and others have carried out molecular dynamics simulations starting from the two available AQP1 structures embedded in a solvated bilayer membrane, aimed at studying the molecular mechanism of the water flow through AQP1 (unpublished results; Zhu et al., this issue, [36]). These simulations indicated problems arising from both starting structures. In particular, the simulations of AQP1 deviated much faster and farther from the initial structure than simulations of GlpF. This finding, together with the medium resolution of the AQP1 structures, suggests that it should be possible to improve the AQP1 structure by utilizing the sequence similarity between the two proteins to build a homology model of AQP1 based on the high resolution GlpF structure. The similarity of the previous AQP1 models to the GlpF structure supports this idea (Fig. 1). In this paper we report the structure of AQP1 that was obtained after refinement of the homology model against electron crystallographic data at 3.8 Å resolution [1] and compare it to the previous AQP1 models.

2. Methods

The multiple sequence alignment from Heymann and Engel [15] provided the basis for building a model of the AQP1 structure starting from the 2.2 Å GlpF structure with Protein Data Bank (PDB) entry 1FX8 [3]. The WHAT IF program [16] was used for this purpose. The longer A loop between the first and second transmembrane helix in AQP1 compared to GlpF was initially modeled by the FILGAP option in WHAT IF, and later manually fitted into the electron microscopy (EM) data at 3.8 Å resolution [1], kindly provided by K. Mitsuoka. One major difference between GlpF and AQP1 is that loops C and E are significantly longer in GlpF compared to AQP1. This difference did not pose a real problem during model building since these insertions in GlpF have their N- and C-terminal residues relatively

*Corresponding author. Fax: (49)-551-201 1089.
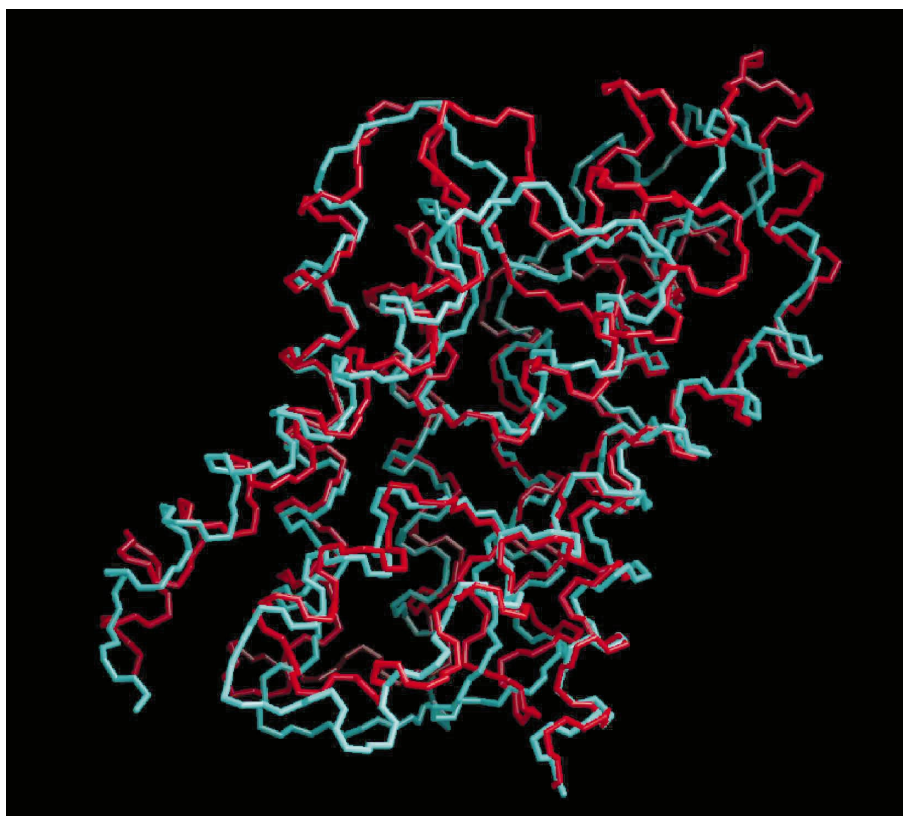*E-mail address:* hgrubmu@gwdg.de (H. Grubmüller).

Fig. 1. Superposition of the backbone traces of the X-ray structure of GlpF (red) and a previous EM structure of AQP1 (cyan, PDB entry 1FQY [1]).

close to each other such that also a shorter loop can easily connect the involved helices. The raw model was fitted to the experimental data by a rigid body search in reciprocal space using CNS [17]. The model was then refined against the EM data at 3.8 Å resolution, again using CNS. First, a rigid body refinement was performed, treating all α-helices and all loops as separate rigid bodies. A torsion angle molecular dynamics simulated annealing procedure [18,19] starting from 3000K with mild restraints towards the EM data was then employed to further drive the protein structure towards the experimental data. Finally, the structure was energy-minimized and optimal B factors were calculated. During refinement, the ϕ and ψ torsion angles in the helical region were restrained [20] since the experimental restraints at 3.8 Å resolution are too weak to preserve the secondary structure during the high temperature part of the simulated annealing. The EM data comprise phases derived from electron micrographs and amplitudes obtained from electron diffraction patterns. Except for the unit cell parameters, all EM parameters are identical to the ones reported by Murata et al. [1]. Instead of the unit cell lengths of $a = b = 96.0$ Å we applied $a = b = 99.58$ Å [2]. In CNS, the maximum likelihood 'MLHL' target was selected which includes an energy contribution from the deviation between experimental and model phases. Furthermore, an overall anisotropic B factor and a bulk solvent correction were applied.

Molecular dynamics simulations of AQP1 and GlpF embedded in a solvated palmitoyloleoylphosphatidylethanolamine (POPE) bilayer membrane were carried out using the gromacs [21] simulation package. Both GlpF (PDB code 1FX8) and the three AQP1 models (PDB codes 1FQY and 1IH5, and our refined model) were simulated as a tetramer. Starting configurations were generated from an equilibrated POPE configuration [22], from which enough lipids were removed to fit the AQP1/GlpF tetramer. The resulting systems contained 271 POPE lipids and were solvated on both sides with SPC [23] water molecules. The total system contained approx. 100 000 atoms. The simulation systems were made neutral by adding eight and four chloride ions to AQP1 and GlpF, respectively. The lipids and water were equilibrated by simulating 200 ps with positional restraints on all

protein atoms. During the simulations, the lincs [24] and settle [25] algorithms were used to constrain bond lengths, allowing a time step of 2 fs. Pressure and temperature were kept constant by weak coupling to an external bath [26] (coupling constants of 1.0 and 0.1 ps, respectively). Long range electrostatic interactions were calculated with the Particle-Mesh Ewald method [27], except for the simulation starting from the 1FQY structure, which was simulated with a twin range cutoff radius of 10/14 Å. Figs. 1, 2 and 4 were made with a modified version of Molscript [28,29] and Raster3D [30].

3. Results

The starting R value of the raw AQP1 homology model with respect to the electron crystallographic data after a rigid body optimal fit in the density map was 46.1% (with a free R value of 46.8%). After rigid body refinement, using all loops and helices as separate α-helices, the R factor, and also the free R factor dropped to 43.0%. After simulated annealing, energy minimization and B factor optimization, the final R factor was 36.7%, with a free R factor of 37.8%, which is a significant improvement compared to the reported values for the two presently available AQP1 structures. The first structure, with PDB entry code 1FQY [1], was reported to have an R/R-free of 39.9/41.7%, and the second, with PDB entry code 1IH5, has an R/R-free of 36.1/45.8% [2].

Root mean square deviations calculated over the backbone atoms of helical residues (Table 1) show that after refinement, the AQP1 model deviates 1.36 Å from the GlpF structure. This value is smaller than the differences between all other pairs of structures. 1FQY and our structure were both refined against the same dataset. Nevertheless, the RMSD between
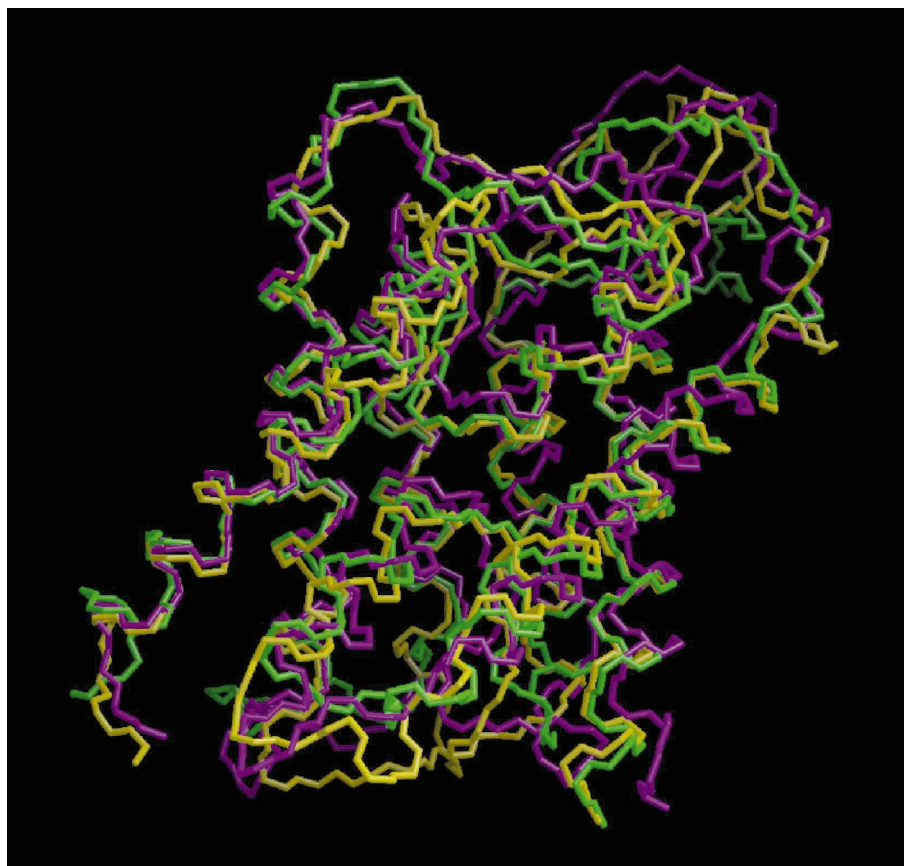
Fig. 2. Superposition of the backbone traces of the three AQP1 models, coloured in yellow (1FQY), magenta (1IH5) and green (our refined model).

the two structures is 1.41 Å. All calculated RMSD values with respect to the 1IH5 structure are larger than 2 Å. This indicates that the 1IH5 structure deviates more from both the GlpF structure and the other AQP1 models than do the other structures among each other.

Three representative structure quality indices as calculated by the WHAT_CHECK module [31] are summarized in Table 2. All tabulated values are *Z*-scores, representing deviations from a database built from known, high resolution structures. Values close to zero are to be expected for high resolution protein structures, as is the case for GlpF. The larger values for the AQP1 structures indicate potential problems in these structures. The smaller values for all three tabulated properties for our refined structure, compared to the two other

AQP1 structures, confirm that this structure is indeed an improved model of the AQP1 structure.

As the RMSD values in Table 1 indicate, 1FQY and our refined model, initially based on the GlpF structure, are relatively similar. This is particularly true for the central helical region facing the pore. All helices are in register, and the backbone traces can be well superimposed (Fig. 2). However, there are significant differences in the loop regions.

The differences with respect to the 1IH5 structure are larger. There are no frameshifts in the helical regions, but helices 1, 2, 3, and 5 are shifted significantly with respect to their locations in both the structure with PDB entry 1FQY and our refined structure.

A particularly difficult region for model building is the re-

Table 1
RSMD between pairs of aquaglyceroporin structures, calculated over the heavy backbone atoms of all residues in helical conformation

Structure pair	Helical backbone RMSD (Å)
1FX8-1FQY	1.75
1FX8-1IH	2.67
1FX8-refi	1.36
1FQY-1IH5	2.45
1FQY-refi	1.41
1IH5-refi	2.24

1FX8 denotes the PDB code of the X-ray structure of GlpF [3], 1FQY and 1IH5 the two AQP1 models [1,2], and refi denotes our refined AQP1 structure.

Table 2
WHAT_CHECK scores [31] for the four discussed aquaglyceroporin structures

Structure	Packing	Backbone	Ramachandran
1FX8	1.52	−0.81	−0.51
1FQY	−4.93	−6.21	−6.94
1IH5	−5.40	−4.57	−6.26
refi	−2.96	−3.38	−2.79

1FX8 denotes the PDB code of the X-ray structure of GlpF [3], 1FQY and 1IH5 the two AQP1 models [1,2], and refi denotes our refined AQP1 structure. Three representative WHAT_CHECK indices are tabulated: 'packing' denotes the packing quality control *Z*-score; 'Backbone' denotes the backbone conformation normality *Z*-score; 'Ramachandran' denotes the Ramachandran *Z*-score.
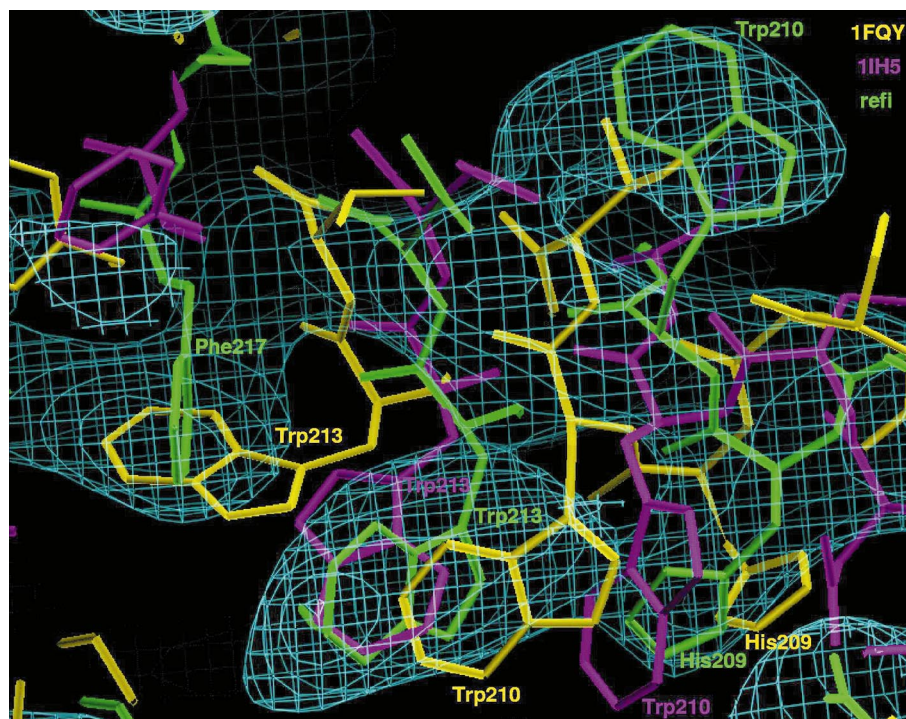
Fig. 3. Electron density of AQP1 at 3.8 Å resolution [1], rendered with O at 1.0 σ [35] around residue Trp210. The three AQP1 models are coloured in yellow (1FQY), magenta (1IH5) and green (our refined model). In this region, the differences between the models are dramatic, and our model seems to fit the experimental data most favourably.

gion around Trp210. From Phe206 to Phe217, there are six bulky residues out of 12, and the electron densities are not unambiguously interpretable. Fig. 3 shows a piece of the electron densities from this region, together with the three aquaporin models. After refinement, without further manual rebuilding, the GlpF-based model of AQP1 fits most favourably in this region of the map, compared to the other two models. In our refined model, the beginning of helix 6 is stretched slightly between 210Trp and 213Trp, which allows the side chain of 210Trp to lie on the other side of the helix from 213Trp, where it perfectly fits in a bulky patch of electron density, which is left unoccupied by the other two models.

Fig. 4 shows the main differences in the pore region between the four aquaglyceroporin structures. The conserved Arg just following the second NPA motif (residue 206 in GlpF and 195 in AQP1) deserves particular attention. This residue has a very similar side chain orientation in GlpF, 1IH5 and our refined AQP1 structure, in which it aligns with the pore axis. In contrast, in the 1FQY structure it forms a salt bridge with Glu142 in helix 4. In GlpF this Arg forms two hydrogen bonds with a bound glycerol molecule [3], indicating its crucial role for glycerol transport. The almost complete conservation of this residue throughout the aquaporin family [15] strongly suggests that also in AQP1 this residue is directly involved in solute transport. Like in GlpF, although this Arg is in relative proximity of the Glu in helix 4, there is no significant electron density that would indicate the presence of a salt bridge, leaving the two charged residues unpaired.

Two aromatic residues, Trp48 and Phe200, are facing this Arg in GlpF. In AQP1, this region is occupied by Phe56 and His180, which are smaller and less hydrophobic. A little bit farther down the pore, Phe24 is located in the narrowest region of the AQP1 pore [1], facing both ASN residues in the NPA motifs. In GlpF, there is a Leu at this position, making the pore in this region slightly larger compared to AQP1. Interestingly, in the 1IH5 structure, Phe24 is rotated away from the pore, resulting in a wider pore compared to the other two AQP1 models.

Fig. 5 shows the root mean square deviation from the starting configuration during molecular dynamics simulations of GlpF and the three AQP1 models, carried out for 2 ns each, as described in Section 2. All simulations comprised a full protein tetramer embedded in a solvated POPE bilayer. As is to be expected, the GlpF simulation drifts least from the crystal structure during equilibration, yielding a backbone RMSD of about 1.5 Å after 2 ns simulation time. In contrast, the two published AQP1 structures, 1FQY and 1IH5, show a much larger drift and reach a RMSD of about 3 Å already after 1 ns. Our refined AQP1 structure is significantly more stable during MD, showing a 2 Å deviation after 2 ns of simulation.

The structure has been deposited in the Protein Data Bank, with entry code 1H6I.

4. Discussion

The sequence identity of 30.6%, together with the electron crystallographic evidence for structural similarity of the AQP1 and GlpF structures (Fig. 1 and Table 1), provides a solid basis for building a homology model of AQP1 based on the GlpF structure [32]. At the current resolution of 3.8 Å of the electron crystallographic AQP1 data, the two previous AQP1 models are likely to contain small imperfections. We have given evidence that after refinement against the EM data,
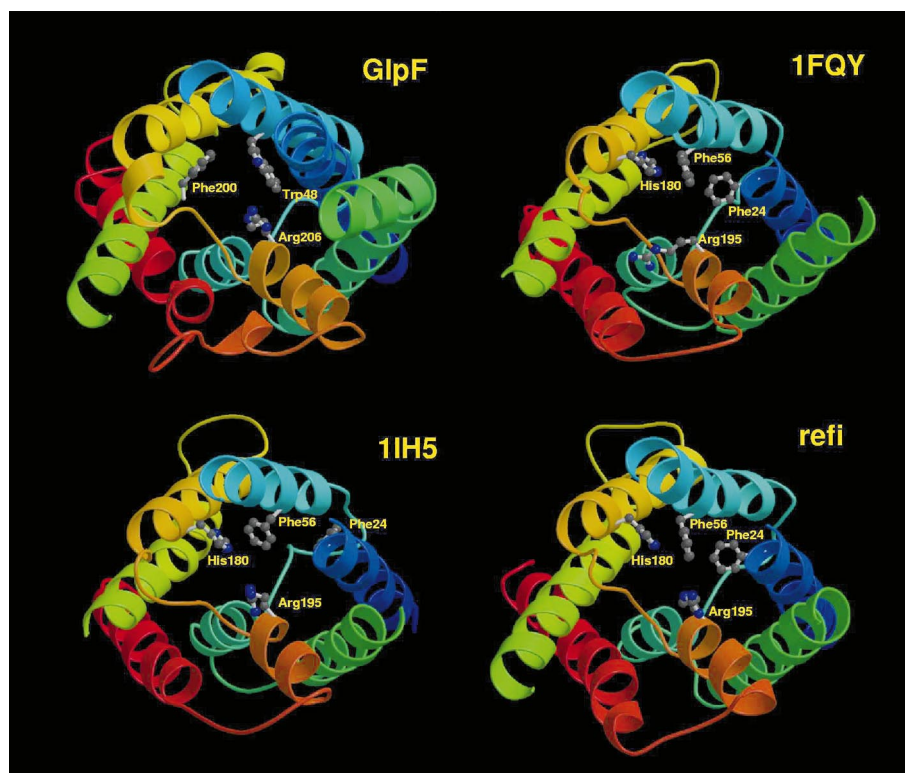
Fig. 4. Structural details in the pore region of GlpF and AQP1, viewed from the extracellular side through the pore. For clarity, the A and C loops, as well as the last few residues of helix 1 are not shown.

the new structure, with many features inherited from the GlpF structure determined at higher resolution, is superior in regions of the experimental map which are poorly resolved, or which for another reason are difficult to thread. A good example of this is the region around Trp210 (Fig. 3). Additionally, residues that are conserved between the two proteins, or
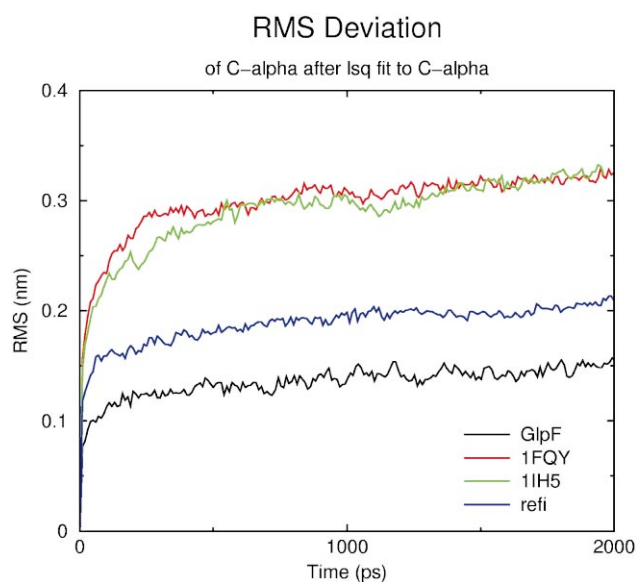

Fig. 5. Root mean square deviations of C_α carbon atoms (computed over the entire tetramer) during MD simulations of GlpF and the three AQP1 models, after a least-squares fit to the starting conformation.

even throughout the whole superfamily, can be expected to have similar conformations in both proteins. Deviations, like the one observed for the conserved Arg195 (Arg206 in GlpF, see Fig. 4), therefore probably point at problems with the medium resolution models. Nevertheless, both previously published AQP1 models were built without prior knowledge of the GlpF structure. The fact that these structures correspond well to the GlpF structure shows that this was a major achievement that illustrates the power of cryo-electron microscopy of two-dimensional (membrane protein) crystals.

The narrowest region in the GlpF pore is surrounded by the conserved Arg206, and the aromatic residues Trp48 and Phe200, and is occupied by a glycerol molecule in the X-ray structure [3]. In AQP1, two smaller residues occupy the positions of these aromatic residues, namely Phe56 and His180, respectively (Fig. 4). Phe24, which is a leucine in GlpF, is facing the two asparagines of the conserved NPA motifs in AQP1, thereby forming the smallest constriction of the AQP1 pore. Interestingly, in contrast to the 1FQY structure and our refined AQP1 model, this phenylalanine is rotated away from the pore in the 1IH5 structure, yielding a larger pore (Fig. 4). The functional significance of this, the effect on water passage through the central pore, will have to await higher resolution studies.

The crucial role of Phe24 for the specificity of AQP1 is further confirmed by sequence analyses of the aquaporin family [15,33]. Together with Leu149, it forms a hydrophobic pair in AQP1. In the majority of the glycerol facilitators, this is a Leu-Leu pair. In contrast, most aquaporins have an aromatic residue in one of the two positions. Interestingly, some aquaporins, such as AQP0 and AQP6, have a Tyr instead of a Phe

at position 24. These proteins have been reported to have peculiar transport properties, compared to AQP1 [34].

5. Conclusions

We have presented a refined structure of human aquaporin-1, based on the X-ray structure of the bacterial glycerol transporter GlpF. The significantly lower crystallographic *R* factor and free *R* factor with respect to cryo-electron crystallographic data at 3.8 Å resolution demonstrate a better agreement to the electron microscopic data as compared to the two previous AQP1 models. Moreover, backbone and side chain packing normality indices as well as the global Ramachandran distribution appearance are significantly improved in the refined model. Additionally, the drift from the starting structure during MD simulations of our new AQP1 model is only slightly larger than that observed in an MD simulation of GlpF. These drifts were found to be much larger in simulations started from the previously published AQP1 models. Taken together, these results strongly suggest that the newly refined AQP1 model is a significant improvement of the AQP1 structure and thus represents an important step towards the molecular understanding of specific transport of water molecules across biological membranes.

References

[1] Murata, K., Mitsuoka, K., Walz, T., Agre, P., Heymann, J., Engel, A. and Fujiyoshi, Y. (2000) Nature 407, 509–605.
[2] Ren, G., Reddy, V.S., Cheng, A., Melnyk, P. and Mitra, A.K. (2001) Proc. Natl. Acad. Sci. USA 98, 1398–1403.
[3] Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J. and Stroud, R.M. (2000) Science 290, 481–486.
[4] Preston, G.M., Carroll, T.P., Guggino, W.B. and Agre, P. (1992) Science 256, 385–387.
[5] Heller, K.B., Lin, E.C.C. and Wilson, T.H. (1980) J. Bacteriol. 144, 274–278.
[6] Maurel, C., Reizer, J., Schroeder, J.I., Chrispeels, M.J. and Saier, M.H. (1994) J. Biol. Chem. 269, 11869–11872.
[7] Verbavatz, J.M., Brown, D., Sabolic, I., Valenti, G., Siello, D.A., Van Hoek, A.N., Ma, T. and Verkman, A.S. (1993) J. Cell Biol. 123, 605–618.
[8] Jung, J.S., Preston, G.M., Smith, B.L., Guggino, W.B. and Agre, P. (1994) J. Biol. Chem. 269, 14648–14654.

[9] Pao, G.M., Wu, L.F., Johnson, K.D., Hofte, H., Crispeels, M.J., Sweet, G., Sandal, N.N. and Saier, M. (1991) Mol. Microbiol. 5, 33–37.
[10] Wistow, G.J., Pisano, M.M. and Chepelinsky, A.B. (1991) Trends Biochem. Sci. 16, 170–171.
[11] Preston, G.M., Jung, J.S., Guggino, W.B. and Agre, P. (1994) J. Biol. Chem. 269, 1668–1673.
[12] Smith, B.L. and Agre, P. (1991) J. Biol. Chem. 266, 6407–6415.
[13] de Groot, B.L., Heymann, J.B., Engel, A., Mitsuoka, K., Fujiyoshi, Y. and Grubmüller, H. (2000) J. Mol. Biol. 300, 987–994.
[14] Ren, G., Cheng, A., Reddy, V., Melnyk, P. and Mitra, A.K. (2000) J. Mol. Biol. 301, 369–387.
[15] Heymann, J.B. and Engel, A. (2000) J. Mol. Biol. 295, 1039–1053.
[16] Vriend, G. (1990) J. Mol. Graph. 8, 52–56.
[17] Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, N., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921.
[18] Brünger, A. (1988) J. Mol. Biol. 203, 803–816.
[19] Rice, L.M. and Brünger, A.T. (1994) Proteins Struct. Funct. Genet. 19, 277–290.
[20] Kleywegt, G.J. and Jones, T.A. (1997) Methods Enzymol. 277, 208–230.
[21] Van der Spoel, D., Berendsen, H.J.C., Van Buuren, A.R., Apol, E., Meulenhoff, P.J., Sijbers, A.L.T.M. and Van Drunen, R. (1995) Nijenborgh 4, 9747 AG, Groningen, The Netherlands. Internet: http://md.chem.rug.nl/∼gmx.
[22] Tieleman, D.P. and Berendsen, H.J.C. (1998) Biophys. J. 74, 2786–2801.
[23] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. and Hermans, J. (1981) In: Intermolecular Forces (Pullman, B., Ed.), pp. 331–342. D. Reidel Publ. Co., Dordrecht.
[24] Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) J. Comp. Chem. 18, 1463–1472.
[25] Miyamoto, S. and Kollman, P.A. (1992) J. Comp. Chem. 13, 952–962.
[26] Berendsen, H.J.C., Postma, J.P.M., DiNola, A. and Haak, J.R. (1984) J. Chem. Phys. 81, 3684–3690.
[27] Darden, T., York, D. and Pedersen, L. (1993) J. Chem. Phys. 98, 10089–10092.
[28] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.
[29] Esnouf, R.M. (1997) J. Mol. Graph. 15, 132–134.
[30] Merritt, E.A. and Bacon, D.J. (1997) Methods Enzymol. 277, 505–524.
[31] Hooft, R.W.W., Vriend, G., Sander, C. and Abola, E.E. (1996) Nature 381, 272.
[32] Martin, A.C.R., MacArthur, M.W. and Thornton, J.M. (1997) Proteins Struct. Funct. Genet. 1 (Suppl.), 14–28.
[33] Engel, A., Fujiyoshi, Y. and Agre, P. (2000) EMBO J. 19, 800–806.
[34] Yasui, M., Hazama, A., Kwon, T.H., Nielsen, S., Guggino, W.B. and Agre, P. (1999) Nature 402, 184–187.
[35] Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Acta Crystallogr. A 47, 110–119.
[36] Zhu, F., Tajkhorshid, E. and Schulten, K. (2001) FEBS Lett. 504, 212–218.