

# The Structure of the Aquaporin-1 Water Channel: A Comparison between Cryo-electron Microscopy and X-ray Crystallography

Bert L. de Groot<sup>1\*</sup>, Andreas Engel<sup>2</sup> and Helmut Grubmüller<sup>1</sup>

<sup>1</sup>Theoretical Molecular Biophysics Group  
Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen Germany

<sup>2</sup>M. E. Müller Institute for Microscopic Structural Biology  
Biozentrum, University of Basel, Basel CH-4056 Switzerland

Three different medium-resolution structures of the human water channel aquaporin-1 (AQP1) have been solved by cryo-electron microscopy (cryo-EM) during the last two years. Recently, the structure of the strongly related bovine AQP1 was solved by X-ray crystallography at higher resolution, allowing a validation of the original medium-resolution structures, and providing a good indication for the strengths and limitations of state of the art cryo-EM methods. We present a detailed comparison between the different models, which shows that overall, the structures are highly similar, deviating less than 2.5 Å from each other in the helical backbone regions. The two original cryo-EM structures, however, also show a number of significant deviations from the X-ray structure, both in the backbone positions of the transmembrane helices and in the location of the amino acid side-chains facing the pore. In contrast, the third cryo-EM structure that included information from the X-ray structure of the homologous bacterial glycerol facilitator GlpF and that was subsequently refined against cryo-EM AQP1 data, shows a root mean square deviation of 0.9 Å from the X-ray structure in the helical backbone regions.

© 2003 Elsevier Science Ltd. All rights reserved

\*Corresponding author

**Keywords:** cryo-electron microscopy; X-ray crystallography; structure refinement; water channel; GlpF

## Introduction

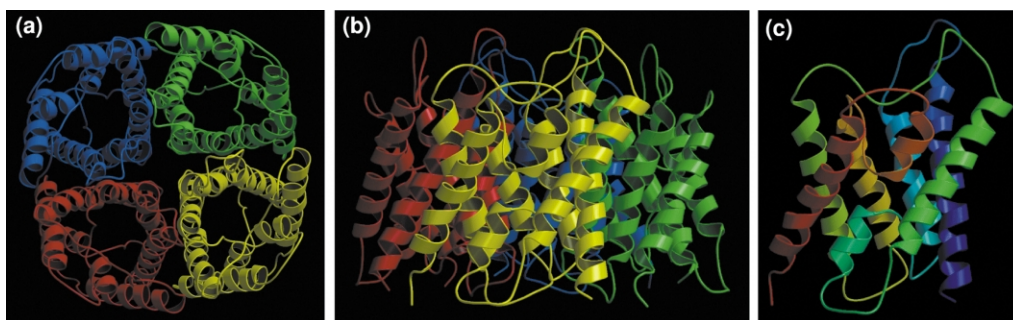
Aquaglyceroporins are integral membrane proteins responsible for passive permeation of water and other small neutral solutes across biological membranes.<sup>1–3</sup> Members of the family are found throughout nature, in species ranging from bacteria to humans, and more than 350 proteins have been identified and sequenced so far. In humans alone, more than ten different aquaporins with specialized functionality, expressed in tissues as diverse as kidney, red blood cells, and brain are known.<sup>1–5</sup> Since the discovery of the aquaporins as universal water channels a decade ago,<sup>6</sup> much effort has been invested to elucidate the structural basis for the remarkable efficiency of these channels, while maintaining strict selectivity, even against protons.

Phylogenetic sequence comparisons have identified two main subclusters in the aquaglyceroporin superfamily: the aquaporin (AQP) water channel cluster and the glycerol facilitator (GLP) cluster.<sup>7–11</sup> Model systems for both branches of the family are the water channel aquaporin-1 (AQP1)<sup>6</sup> and the bacterial glycerol facilitator GlpF.<sup>12,13</sup> AQP1 is a water-selective channel with a remarkably high water permeation rate,<sup>14–16</sup> whereas GlpF transmits glycerol and other linear alcohols,<sup>12,13,17</sup> and shows a reduced water permeability.<sup>18</sup>

Both AQP1 and GlpF are active as homotetramers, with each monomer contributing a channel.<sup>19–22</sup> About five years ago, the first structural information became available for the channel structure, when three groups independently presented low resolution ( $\approx 6$  Å) electron microscopy (EM) data.<sup>23–25</sup> These data confirmed the sequence-based hourglass model<sup>21</sup> and showed that the AQP1 structure consists of six transmembrane-spanning helices and two conserved loops B and E, containing the fingerprint Asn-Pro-Ala (NPA) motifs, that fold back into the protein and are in contact in the center of the channel.

Abbreviations used: AQP, aquaporin; GlpF, glycerol facilitator; hAQP1, human AQP-1; bAQP1, bovine AQP-1; RMSD, root-mean-square deviation; cryo-EM, cryo-electron microscopy; NPA, the Asn-Pro-Ala motif.

E-mail address of the corresponding author: [bgroot@gwdg.de](mailto:bgroot@gwdg.de)



**Figure 1.** Top-view (a) and side-view (b) of the tetrameric structure of hAQP1,<sup>29</sup> and a close-up of the monomeric structure (c). The pictures were generated with a modified version of Molscript<sup>41,42</sup> and Raster3D.<sup>42</sup>

With improving resolution, it was observed that the C-terminal parts of these two loops form short helices.<sup>26</sup> At this level of resolution, the three-dimensional fold could be unambiguously determined.<sup>27,28</sup>

Further improvement of the resolution of the cryo-EM data rendered the construction of atomic models of the human AQP1 (hAQP1) structure feasible. First, a 3.8 Å resolution† electron microscopy structure of hAQP1 (Figure 1)<sup>29</sup> was published, just before the X-ray structure of GlpF at 2.2 Å resolution became available,<sup>17</sup> which shares 30.6% sequence identity with hAQP1. After that followed an independent model of the hAQP1 structure, based on electron microscopic data at 3.7 Å resolution.<sup>30</sup> These AQP1 models provided invaluable insight into aquaporin function, but at the same time suffered from a number of problems concerning particular structural details, due to lack of resolution, which e.g. became evident through structural instability during molecular dynamics (MD) simulations.<sup>31,32</sup>

To improve the structure of hAQP1, information extracted from the higher resolution structure of the homologous GlpF protein<sup>17</sup> and from an extensive multiple sequence alignment<sup>11</sup> was used to build a modified hAQP1 structure.<sup>33</sup> The resulting structure was subsequently refined against the same EM data that had been used for constructing the original AQP1 model.<sup>29</sup> The crystallographic *R*-factor and free *R*-factor, as well as a number of geometrical properties indicated a significant improvement with respect to the original AQP1 models.<sup>33</sup> Additionally, this structure proved stable during MD simulations.<sup>34</sup>

The structure of bovine AQP1 (bAQP1) was recently solved to 2.2 Å resolution by X-ray crystallography.<sup>35</sup> Because of the high degree of sequence identity (90.6%, for residues 9–232, hAQP1 numbering) between the two proteins,

especially in the core region of the protein, we can assume that the structures of bAQP1 and hAQP1 should be nearly identical. This thus provides the rare opportunity to compare structural data independently derived from cryo-EM at moderate resolution and X-ray crystallography at higher resolution, and to judge the quality of the original cryo-EM models of hAQP1. Additionally, it enables us to assess the validity of the refinement approach of the hAQP1 model based on the GlpF X-ray structure, and, in particular, to address the question to about what extent this procedure introduced any bias towards the GlpF structure.

Here, we present a detailed structure comparison between all available structures of members of the aquaglyceroporin superfamily. The structural differences are characterized in terms of model quality (resolution), differences in experimental conditions (three-dimensional *versus* two-dimensional crystals), and substrate specificity. Additionally, to check whether the bAQP1 X-ray structure with higher resolution is actually a better model of the human AQP1 structure than the hAQP1 EM structures, the X-ray structure of bAQP1 was mutated to adopt the hAQP1 sequence and refined against EM data at 3.8 Å resolution. Finally, the structural integrity of each of these models was verified by subjecting them to extensive MD simulations.

## Results and Discussion

The refinement statistics of the different aquaglyceroporin structures in Table 1 show that the X-ray structure of GlpF (GlpF-X) has the lowest *R*-factor and free *R*-factor, followed by the X-ray structure of bAQP1 (bAQP1-X) and the EM structures of hAQP1 (hAQP1-EM1, hAQP1-EM2, hAQP1-EM3, and hAQP1-EM-X). For the hAQP1 EM structures, the structures that include structural information from the X-ray structures of GlpF and bAQP1 (hAQP1-EM3 and hAQP1-EM-X, respectively) show a better agreement to the EM data (see also Figure 2) than the hAQP1 models that were built *ab initio* in the experimental potential map (hAQP1-EM1 and hAQP1-EM2). It should be noted that *R*-factors at different levels of

† Due to the missing cone resulting from the limited tilt angle (60°) during image collection in EM, only the resolution in the plane of the membrane (*x*, *y*) is 3.8 Å; in the perpendicular (*z*) direction the resolution is  $3.8/\sin(60^\circ) = 4.4$  Å.<sup>29</sup> Throughout the rest of the manuscript the reported 3.8 Å resolution therefore actually refers to  $3.8 \times 3.8 \times 4.4$  Å.

**Table 1.** Refinement statistics for the different available aquaglyceroporin structures

| Structure  | PDB code           | Resolution (Å) | R-factor | Free R |
|------------|--------------------|----------------|----------|--------|
| GlpF-X     | 1FX8 <sup>17</sup> | 2.2            | 19.7     | 22.3   |
| bAQP1-X    | 1J4N <sup>35</sup> | 2.2            | 26.6     | 30.8   |
| hAQP1-EM1  | 1FQY <sup>29</sup> | 3.8            | 39.9     | 41.7   |
| hAQP1-EM2  | 1IH5 <sup>30</sup> | 3.7            | 36.1     | 45.8   |
| hAQP1-EM3  | 1H6I <sup>33</sup> | 3.8            | 36.7     | 37.8   |
| hAQP1-EM-X | –                  | 3.8            | 33.3     | 35.7   |

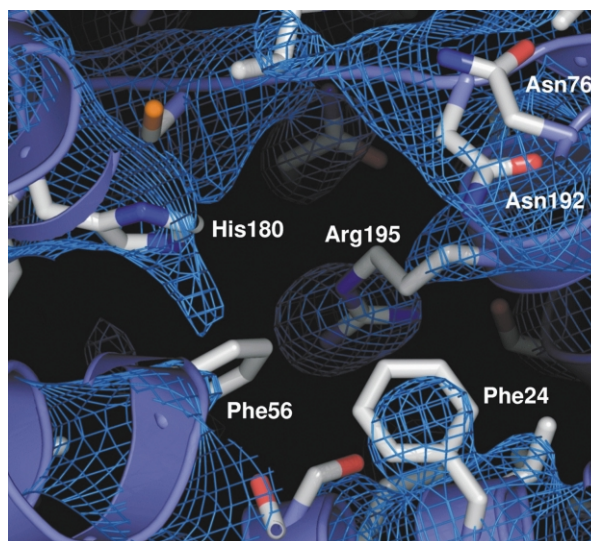
GlpF-X denotes the X-ray structure of GlpF with PDB code 1FX8<sup>17</sup>; bAQP1-X represents the X-ray structure of bovine aquaporin-1 (PDB code 1J4N<sup>35</sup>); hAQP1-EM1 and hAQP1-EM2 are the two original EM-based human aquaporin-1 (hAQP1) models, with PDB codes 1FQY and 1IH5, respectively,<sup>29,30</sup> hAQP1-EM3 denotes the GlpF-based refined hAQP1 EM structure with PDB code 1H6I,<sup>33</sup> hAQP1-EM-X denotes the bAQP1-X (1J4N) structure mutated to adopt the hAQP1 sequence and subsequently refined against hAQP1 EM data at 3.8 Å resolution. Note that all structures were refined to different experimental datasets. Only hAQP1-EM2, hAQP1-EM3, and hAQP1-EM-X were refined against the same EM dataset at 3.8 Å resolution.<sup>29</sup> Also note that an R-factor at 2.2 Å resolution cannot be directly compared to an R-factor at 3.8 Å resolution.

resolution cannot be directly compared to each other, because of the different data-to-parameter ratios. Additionally, compared to X-ray crystallographic data, the EM data contain an extra inherent noise level, caused by image merging and processing (most notably image alignment and unbending).

Refinement of the X-ray structure of bAQP1 (after mutating the sequence to that of hAQP1) against the hAQP1 EM data at 3.8 Å resolution (denoted by hAQP1-EM-X) yields a lower R-factor (33.3 *versus* 36.7) and free R-factor (35.7 *versus* 37.8) with respect to the hAQP1-EM3 structure. This indicates that, after mutation to the hAQP1 sequence, the bAQP1 X-ray structure represents a better model for the hAQP1 EM data than the previous EM structures<sup>†</sup>. We therefore took the X-ray structure of bAQP1 as a reference and monitored the deviations for the various hAQP1 EM structures, to assess the quality of those medium-resolution structures.

An independent quality assessment can be obtained from a comparison of a number of geometrical properties to a distribution obtained from a selection of high-resolution protein structures. Table 2 lists Z-scores (deviations from the expectation value, in terms of multiples of the standard deviation) for the overall packing, backbone conformation, and Ramachandran plot appearance, as obtained from WHAT\_CHECK,<sup>36</sup> for the different aquaglyceroporin structures. The geometrical properties underscore the results of Table 1 and

<sup>†</sup> The structure does not change significantly during refinement: the RMSD, calculated over all backbone atoms, between bAQP1-X and hAQP1-EM-X, that arose during refinement against the EM data, amounts to 0.4 Å.



**Figure 2.** The central pore region of the hAQP1-EM3 structure (pdb code 1H6I<sup>33</sup>) together with the 3.8 Å resolution EM potential map<sup>29</sup> rendered at 1.0 $\sigma$ . A number of critical residues are labeled (see also Figure 7). The picture was generated with PyMOL<sup>‡</sup>.

show that for the listed properties, the GlpF X-ray structure has all scores close to zero, as would be expected for high resolution structures. In contrast, the original hAQP1 EM structures (hAQP1-EM1 and hAQP1-EM2) show large deviations from zero. The X-ray structure of bAQP1 (bAQP1-X) and the refined GlpF-based hAQP1 structure (hAQP1-EM3) show intermediate values.

An overlay of the different aquaglyceroporin structures (Figure 3) indicates that, apart from the loop regions, all structures show a very similar backbone conformation, with RMS deviations of less than 1.5 Å. Only the hAQP1-EM2 structure<sup>30</sup> deviates more than 2 Å from the other structures in the helical regions.

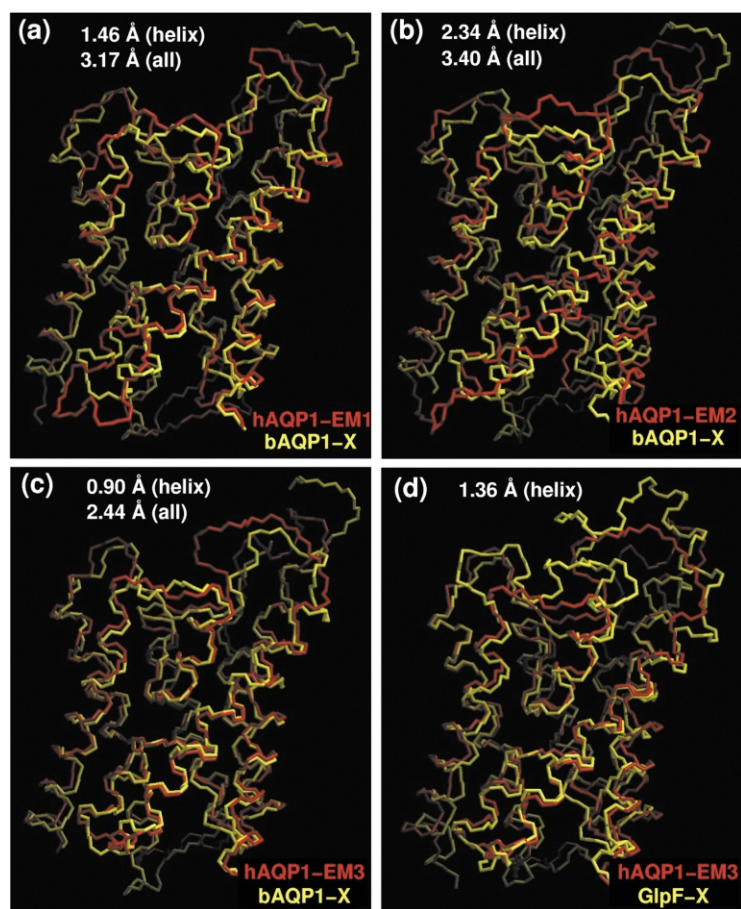
**Table 2.** WHAT\_CHECK scores<sup>36</sup> for the different aquaglyceroporin structures

| Structure  | Packing | Backbone | Ramachandran |
|------------|---------|----------|--------------|
| GlpF-X     | 1.52    | -0.81    | -0.51        |
| bAQP1-X    | 0.22    | -2.34    | -4.34        |
| hAQP1-EM1  | -4.93   | -6.21    | -6.94        |
| hAQP1-EM2  | -5.40   | -4.57    | -6.26        |
| hAQP1-EM3  | -2.96   | -3.38    | -2.79        |
| hAQP1-EM-X | -2.18   | -2.32    | -4.33        |

Three representative WHAT\_CHECK indices are tabulated: Packing denotes the overall packing Z-score; Backbone denotes the backbone conformation normality Z-score; Ramachandran denotes the Ramachandran Z-score. For a definition of the Z-scores, see Methods.

<sup>‡</sup> <http://www.pymol.org>

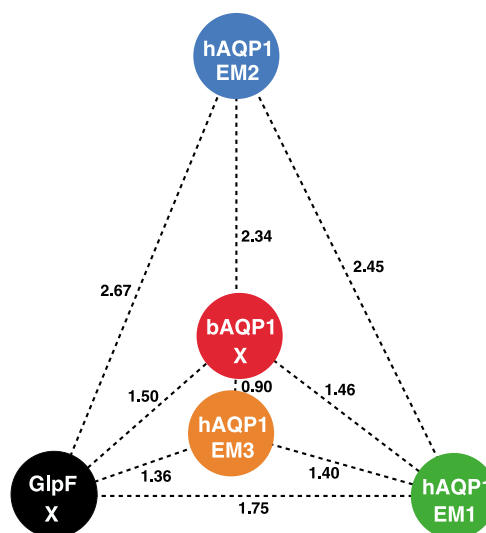




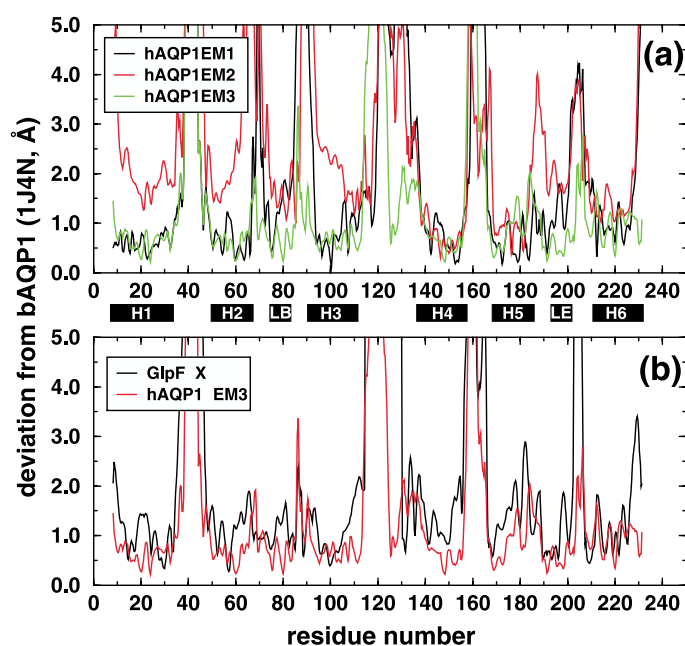
**Figure 3.** Comparison of backbone structures of different aquaglyceroporins. In (a)–(c), the EM models of hAQP1 (hAQP1-EM1, hAQP1-EM2, and hAQP1-EM3, in red) are compared to the X-ray structure of bAQP1 (bAQP1-X, yellow). (d) An overlay of the structures of hAQP1 (1H6L, red) to the X-ray structure of GlpF (GlpF-X, yellow). The view angle is chosen such that the two helices formed by the B and E-loops (which also contain the NPA motifs) are in the foreground. The numbers inside the graphs indicate the RMSD between the backbone atoms of the shown structure pairs, for the helical regions, and for the complete backbone, including the loops. The pictures were generated with a modified version of MOLSCRIPT<sup>41,42</sup> and Raster3D.<sup>43</sup>

Inclusion of structural data derived from the GlpF structure (GlpF-X) and re-refinement of the structure against the same EM data from which the hAQP1-EM1 structure was built,<sup>29</sup> improved the free *R*-factor from 41.7 to 37.8, resulting in the hAQP1-EM3 structure<sup>33</sup> (see Figure 2). Figure 3(c) shows that the hAQP1-EM3 structure is closer to the X-ray structure of bAQP1 (bAQP1-X) than the other hAQP1 structures (see also Figure 4), especially in the B and E-loops, which contain the NPA motifs. This indicates that indeed the inclusion of structural details from the GlpF structure helped in improving the hAQP1 structure. It is interesting to note that after refinement of the GlpF-based hAQP1 structure against the EM data at 3.8 Å resolution, the helical backbone RMSD from the GlpF structure (1.36 Å) was significantly larger than the RMSD to the bAQP1 X-ray structure (0.9 Å, Figures 3(c) and (d), and 4), which suggests that, if at all, only a mild bias towards the GlpF structure was introduced by inclusion of the GlpF X-ray data.

To obtain a more detailed picture of the observed differences, Figure 5(a) shows the deviations from the bAQP1 X-ray structure (bAQP1-X) for each residue. For hAQP1-EM1 and hAQP1-EM3, 27 and 30 residues in the helical regions deviate less than 0.5 Å from the bAQP1 X-ray structure (bAQP1-X), respectively. In the second half of the sequence, five residues show a similarly small deviation in



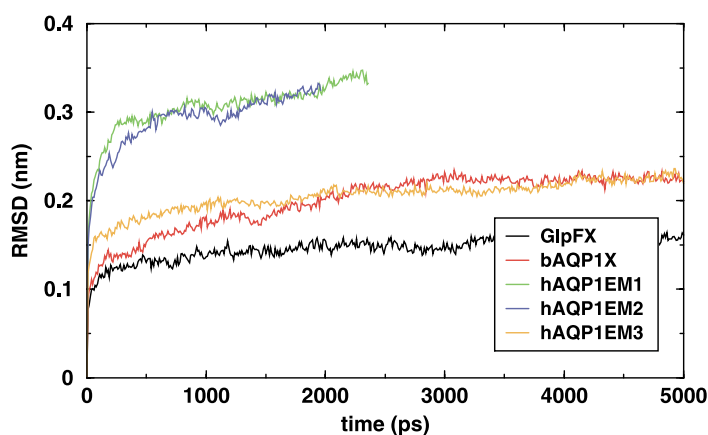
**Figure 4.** Graphical representation of the differences between the different aquaglyceroporin structures studied here. The positions of the structures were determined by a principal components analysis conducted over the backbone atoms in the helical regions of all aquaglyceroporin structures. The numbers in the graph indicate the RMSD between the different pairs of structures, calculated using the same atoms.



**Figure 5.** Deviation from the X-ray structure of bAQP1 (bAQP1-X) for the different hAQP1 EM structures (hAQP1-EM1 through hAQP1-EM3, (a)), and of the GlpF structure (GlpF-X) together with the refined hAQP1 structure (hAQP1-EM3, (b)), to illustrate the effect of the refinement. The locations of the transmembrane helices and the helices in B and E-loops are marked by black bars. The structure deviations were calculated over all backbone atoms after a least-squares fit on the helical backbone regions. The graphs are truncated at 5 Å since any difference above this level indicates an alternative (loop) conformation, rather than a local deviation.

the case of hAQP1-EM2, quite in contrast to the first half, where the deviations are significantly larger. This is mainly caused by shifts of transmembrane helices 1–3 along the helix axes with respect to the other structures (see also Figure 3(b)). The most noticeable difference between the deviations in the hAQP1-EM1 and hAQP1-EM3 structures from bAQP1-X is observed in loops B and E, which contain the highly conserved, functionally important NPA motifs. Structurally, these loops fold back towards the center of the pore, where both NPA motifs are in close contact. The C-terminal halves of these loops form two short alpha-helices. The GlpF-based refined hAQP1 structure (hAQP1-EM3) is in these regions significantly closer to the bAQP1 X-ray structure than the original hAQP1 structure based on the same EM data (hAQP1-EM1). This difference possibly explains why the MD simulations started from the hAQP1-EM1 structure<sup>31</sup> were not stable, but simulations that started from the hAQP1-EM3 structure were stable (see also Figure 6).<sup>34</sup>

The largest differences between the structures are seen in the loop regions (Figure 5(a)). There are a number of possible reasons for this. First, the crystal packing is different for the different crystals. The X-ray structures of GlpF and bAQP1 are derived from three-dimensional crystals (space group  $I422$ ), whereas the hAQP1 EM structures are based on two-dimensional crystals (space group  $P42_12$ ). The two-dimensional crystals used in EM are packed only in the plane of the membrane, with the loops that point in the perpendicular direction exposed to the solvent and thus relatively unaffected from crystal contacts, in contrast to X-ray structures, where exterior loops might be influenced by crystal packing. Second, because of this and from interactions with the supporting grid used in electron crystallography, these loops are relatively disordered in the two-dimensional crystals and, therefore, the resolution is worst for these parts of the hAQP1 EM structures, rendering their refinement problematic. Additionally, the missing cone effectively reduces



**Figure 6.** Structural stability during MD simulations, expressed as the RMSD from the starting structures as a function of time. The RMSDs were calculated over the alpha carbon atoms, for all simulated (tetrameric) aquaglyceroporin structures. RMSD values were calculated after a least-squares fit on the backbone atoms of the helical regions.

the resolution in the direction perpendicular to the membrane ( $z$ ). This effect uniformly blurs the map in the  $z$ -direction, but it most strongly affects the density of horizontal structures in the map, particularly the loops. Finally, the sequence conservation is lowest in these loops, both between GlpF and AQP1, but also between hAQP1 and bAQP1.

Figure 5(b) shows the effect of the refinement of the hAQP1 structure against the EM data at 3.8 Å resolution (hAQP1-EM3). That structure was initially built based on homology to the GlpF X-ray structure (GlpF-X). Virtually along the complete sequence, the refinement against the hAQP1 EM data drove the structure towards the bAQP1 X-ray structure, from an initial backbone RMSD of 1.50 Å (calculated over the helical regions) to a final value of 0.90 Å. After refinement, the final RMSD to the GlpF X-ray structure is 1.36 Å.

Root mean square deviations in MD simulations (Figure 6) with the different aquaglyceroporin structures as starting conformations, follow the same trend as the refinement statistics in Tables 1 and 2. The GlpF structure (GlpF-X) shows a remarkable stability over the complete time range (up to 5 ns). In contrast, the original hAQP1 EM models (hAQP1-EM1 and hAQP1-EM2) deviate very rapidly from their respective starting structures. The backbone RMSD values increase to about 3.5 Å within 2 ns, indicating structural instability. The GlpF-based refined hAQP1 EM structure (hAQP1-EM3) and the bAQP1 X-ray structure reach stable values between 2.0 Å and 2.5 Å after 2 ns. This suggests that the RMSD from the experimental structure during extended MD simulations is a sensitive tool to validate a protein structure.

It should be noted that there are many factors affecting the RMSD from the initial structure during an MD simulation. First, the simulation environment should be chosen such as to mimic the experimental conditions as closely as possible, to prevent any systematic structural deviations. Secondly, the parameters (or forcefield) used in the simulation should be compatible with the simulated system, such that the preferred conformation in the simulation is indeed the native conformation. Thirdly, proteins are inherently flexible, and this flexibility is reflected in the RMSD in conjunction with any systematic drifts. Therefore, in terms of stability, it is only meaningful to compare structurally similar proteins that share the same degree of flexibility. But, as the results demonstrate, when these precautions are considered, MD simulations can serve to assess the quality of a structure.

Due to its functional relevance, the packing of the central pore region deserves particular attention (Figure 7). Here, too, the GlpF-based hAQP1 EM structure (hAQP1-EM3) is closer to the bAQP1 X-ray structure (bAQP1-X) than the other two hAQP1 EM structures (hAQP1-EM1 and hAQP1-EM2). In hAQP1-EM1, the NPA region (including the B and E-loops and the part of helix 1 that faces

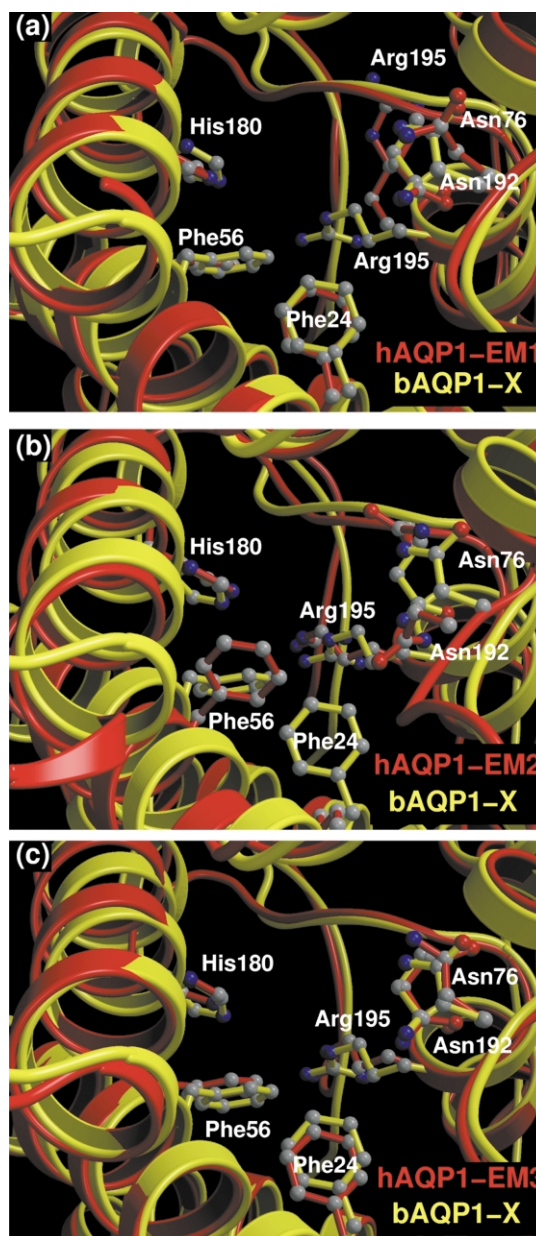
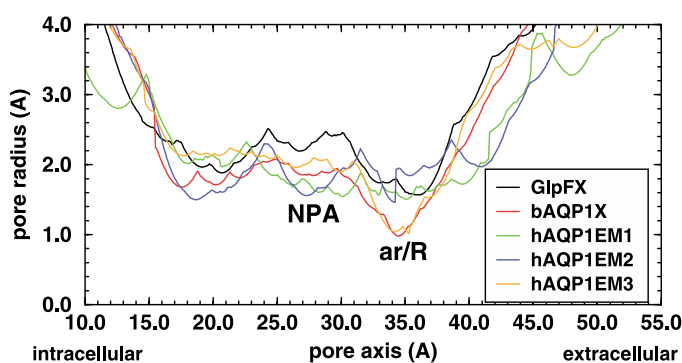


Figure 7. Comparison of the different hAQP1 EM structures to the bAQP1 X-ray structure in the pore region. The X-ray structure of bAQP1 (1J4N) is shown in yellow, the EM structures of hAQP1, 1FQY (a), 1IH5 (b), and 1H6I (c) are shown in red. Residue numbers are as in hAQP1. The pictures were generated with a modified version of MOLSCRIPT<sup>41,42</sup> and Raster3D.<sup>43</sup>

the Asn residues of both NPAs, here represented by Phe24) deviates on average by about 1 Å from bAQP1-X. In the constriction region further towards the extracellular side of the pore (residues Phe56, His180, and Arg195), the side-chain of the conserved Arg195 was modeled to form a salt bridge to Glu142 in helix 4 in this structure,<sup>29</sup> such that it does not align with the pore as in the other structures. In the hAQP1-EM2 structure, the deviations in the NPA region are even larger. The backbone deviation for the NPA motifs exceeds





**Figure 8.** Pore radii as a function of the pore axis for the different aquaglyceroporin structures as calculated by the HOLE program.<sup>44</sup>

1.5 Å, and the side-chain of Phe24 is rotated away from the pore. In the constriction region, the side-chain of Arg195 aligns with the pore, but the side-chain of Phe56 is rotated by about 90 degrees. Even though residues Phe24, Phe56 and His180 are not conserved between AQP1 and GlpF, all side-chains in the GlpF-based structure of hAQP1 (hAQP1-EM3) have similar conformations to bAQP1-X, the X-ray structure of bAQP1, both in the NPA region and in the constriction region. Also the side-chain of Cys189, the mercury-sensitive site in the pore,<sup>37</sup> is located at exactly the same position in hAQP1-EM3 and bAQP1-X (not shown).

Both in AQP1 and in GlpF the region around the conserved Arg195 (Arg197 in bAQP1 and Arg206 in GlpF) that is surrounded by aromatic residues (hence termed ar/R<sup>34</sup>) is the narrowest region in the pore and thus has also been termed the constriction region.<sup>17,35</sup> In Figure 8 the pore dimension for the different structures are compared to each other. Along the complete pore, the GlpF pore is wider than the AQP1 pore, in agreement with the observation that glycerol can permeate through GlpF but not through AQP1. In accord with the similar side-chain locations as illustrated in Figure 7, a very similar pore radius profile is observed in the constriction region in the hAQP1-EM3 and bAQP1-X structures. In this region the original models of hAQP1 (hAQP1-EM1 and hAQP1-EM2) are significantly wider. In the NPA region, where the two asparagine residues of the B and E-loops are in close contact, the differences between the AQP structures are smaller. Only the GlpF profile (black curve) clearly deviates from the AQP1 profiles. In GlpF, this region of the pore is wider, mainly because the phenylalanine residue facing the asparagine residues in AQP1 (Phe24) is replaced by a leucine in GlpF (Leu21). As previously noted, the glycerol facilitators exhibit mostly a leucine whereas aquaporins have mostly a phenylalanine at this narrowing.<sup>11</sup> Further towards the intracellular face of the channel (at around 20 Å in Figure 8), there is another narrowing of the AQP1 pore, formed by the carbonyl groups of the B-loop (Gly72, Ala73 and His74) surrounded by hydrophobic residues (Ile60, Ala64, Leu75, Val79, Leu83, Leu149, Val153 and Ile172). This narrowing is the most pronounced in the

hAQP1-EM2 and bAQP1-X structures and is positioned almost symmetrically to the constriction region with respect to the NPA region in the center. In MD simulations started from the hAQP1-EM3 structure, where the pore in this region is wider, this partly hydrophobic region forms the highest enthalpic barrier for water molecules to pass the pore.<sup>34</sup> A narrowing at this position in the pore as observed in the X-ray structure of bAQP1, therefore, may well be functionally relevant.

## Conclusions

The comparison of the different AQP1 structures in this study has shown that all structures share a high degree of similarity. The original cryo-EM models of hAQP1, however, also show a number of significant deviations from both the refined EM model of hAQP1 and the X-ray structure of bAQP1. Only after inclusion of structural data derived from the X-ray structure of the homologous GlpF could an hAQP1 structure be built from the cryo-EM data at 3.8 Å resolution that agrees with the X-ray structure of bAQP1 in all aspects, except for the loops. Additionally, refinement of the bAQP1 X-ray structure (after mutating it to adopt the hAQP1 sequence) against the cryo-EM data of hAQP1 yields improved *R* and free *R*-factors with respect to the previous hAQP1 structures. Together, these results indicate that cryo-EM data, even at a moderate resolution of 3.8 Å, can discriminate between different structures. Building a novel structure in such a medium-resolution map, however, still is a non-trivial task. Improvements in the refinement protocols or inclusion of additional data (like for example a high-resolution homologous structure) will facilitate the extraction of all structural information available in such medium-resolution maps.

## Methods

The coordinates of the structures used in this study were taken from the Protein Data Bank (PDB), with access codes 1FX8, 1FQY, 1IH5, 1H6I and 1J4N for the X-ray structure of GlpF from *Escherichia coli*,<sup>17</sup> the EM structures of hAQP1,<sup>29,30,33</sup> and the X-ray structure of bAQP1,<sup>35</sup> respectively. A model of hAQP1 was built

based on the bAQP1 structure, using the homology-modeling module of the WHAT IF package,<sup>38</sup> and subsequently refined against the EM data at 3.8 Å resolution<sup>29</sup> using the CNS package.<sup>39</sup> The refinement procedure is the same as described before,<sup>33</sup> with the only difference that the simulated annealing step was left out†. In short, a rigid-body refinement to optimally place the model in the map was followed by an energy minimization and *B*-factor optimization. The maximum likelihood target function *mlhl* was applied, which includes energy contributions for deviations between model and experimental amplitudes and phases.

The WHAT\_CHECK<sup>36</sup> module of the WHAT IF package<sup>38</sup> was used for structure validation. In particular, a number of representative normality *Z*-scores were calculated (the structural average packing *Z*-score, the backbone conformation *Z*-score, and the Ramachandran *Z*-score) indicating the deviation from a database of high-resolution structures in terms of the number of standard deviations from the mean. Values close to zero are to be expected for high-resolution, well-refined protein structures.

To check the structural integrity of the different aquaglyceroporin structures, all structures were subjected to MD simulations. All simulations were carried out on tetrameric protein structures embedded in an explicitly solvated palmitoyloleoylphosphatidylethanolamine (POPE) bilayer membrane system. The gromacs simulation package<sup>40‡</sup> was employed for these simulations, using the same parameters as described before.<sup>33,34</sup> The simulation lengths were 5 ns for structures GlpF-X, bAQP1-X and hAQP1-EM3, 2.5 ns for structures hAQP1-EM1 and hAQP1-EM2, respectively.

## Acknowledgements

We thank Henning Stahlberg for useful suggestions, and the referee for helpful comments. B.L.d.G. and H.G. were supported by the BIOTECH program of the EU, grants QLRT 2000/00778 and 2000/00504. A.E. was supported by the Swiss National Foundation, the M. E. Müller-Foundation of Switzerland, the European Union-Quality of Life and Management of Living Resources Project (grant QLRT-2000-00778 and QLRT-2000/00504) and the Human Frontier Science Program (grant RG0021/2000-M103).

## References

- Agre, P., Bonhivers, M. & Borgnia, M. J. (1998). The Aquaporins, blueprints for cellular plumbing systems. *J. Biol. Chem.* **273**, 14659–14662.
- Borgnia, M., Nielsen, S., Engel, A. & Agre, P. (1999). Cellular and molecular biology of the Aquaporin water channels. *Annu. Rev. Biochem.* **68**, 425–458.
- Engel, A. & Stahlberg, H. (2002). Aquaglyceroporins: channel proteins with a conserved core, multiple functions and variable surfaces. *Int. Rev. Cytol.* **215**, 75–104.
- Deen, P. M. T. & van Os, C. H. (1998). Epithelial Aquaporins. *Curr. Opin. Cell. Biol.* **10**, 435–442.
- Li, J. & Verkman, A. S. (2001). Impaired hearing in mice lacking Aquaporin-4 water channels. *J. Biol. Chem.* **276**, 31233–31237.
- Preston, G. M., Carroll, T. P., Guggino, W. B. & Agre, P. (1992). Appearance of water channels in *Xenopus* oocytes expressing red-cell CHIP28 protein. *Science*, **256**, 385–387.
- Reizer, J., Reizer, A. & Saier, M. H. (1993). The MIP family of integral membrane channel proteins—sequence comparisons, evolutionary relationships, reconstructed pathway of evolution, and proposed functional-differentiation of the 2 repeated halves of the proteins. *Crit. Rev. Biochem. Mol. Biol.* **28**, 235–257.
- Park, J. H. & Saier, M. H. (1996). Phylogenetic characterization of the MIP family of transmembrane channel proteins. *J. Membr. Biol.* **153**, 171–180.
- Froger, A., Tallur, B., Thomas, D. & Delamarche, C. (1998). Prediction of functional residues in water channels and related proteins. *Protein Sci.* **7**, 1458–1468.
- Heymann, J. B. & Engel, A. (1999). Aquaporins: phylogeny, structure, and physiology of water channels. *News Physiol. Sci.* **14**, 187–193.
- Heymann, J. B. & Engel, A. (2000). Structural clues in the sequences of the Aquaporins. *J. Mol. Biol.* **295**, 1039–1053.
- Heller, K. B., Lin, E. C. & Wilson, T. H. (1980). Substrate-specificity and transport-properties of the glycerol facilitator of *Escherichia coli*. *J. Bacteriol.* **144**, 274–278.
- Maurel, C., Reizer, J., Schroeder, J. I., Chrispeels, M. J. & Saier, M. H. (1994). Functional characterization of the *Escherichia coli* glycerol facilitator GlpF, in *Xenopus* oocytes. *J. Biol. Chem.* **269**, 11869–11872.
- Zeidel, M. L., Ambudkar, S. V., Smith, B. L. & Agre, P. (1992). Reconstitution of functional water channels in liposomes containing purified red-cell CHIP28 protein. *Biochemistry*, **31**, 7436–7440.
- van Hoek, A. N. & Verkman, A. S. (1992). Functional reconstitution of the isolated erythrocyte water channel CHIP28. *J. Biol. Chem.* **267**, 18267–18269.
- Zeidel, M. L., Nielsen, S., Smith, B. L., Ambudkar, S. V., Maunsbach, A. B. & Agre, P. (1994). Ultrastructure, pharmacological inhibition, and transport selectivity of Aquaporin channel-forming integral protein in proteoliposomes. *Biochemistry*, **33**, 1606–1615.
- Fu, D., Libson, A., Miercke, L. J., Weitzman, C., Nollert, P., Krucinski, J. & Stroud, R. M. (2000). Structure of a glycerol-conducting channel and the basis for its selectivity. *Science*, **290**, 481–486.
- Borgnia, M. J. & Agre, P. (2001). Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **98**, 2888–2893.
- Smith, B. L. & Agre, P. (1991). Erythrocyte Mr-28,000 transmembrane protein exists as a multisubunit oligomer similar to channel proteins. *J. Biol. Chem.* **266**, 6407–6415.
- Verbavatz, J. M., Brown, D., Sabolic, I., Valenti, G., Siello, D. A., Van Hoek, A. N. *et al.* (1993). Tetrameric assembly of CHIP28 water channels in liposomes

† Refinement was also carried out with simulated annealing but in this case it was found to improve the fit to the experimental data only marginally. We therefore decided to skip this step in order to remain as close as possible to the starting structure.

‡ <http://www.gromacs.org>



- and cell-membranes—a freeze-fracture study. *J. Cell Biol.* **123**, 605–618.
21. Jung, J. S., Preston, G. M., Smith, B. L., Guggino, W. B. & Agre, P. (1994). Molecular structure of the water channel through Aquaporin CHIP—the hourglass model. *J. Biol. Chem.* **269**, 14648–14654.
  22. Braun, T., Philippsen, A., Wirtz, S., Borgnia, M. J., Agre, P., Kühlbrandt, W. *et al.* (2000). The 3.7 Ångstrom projection map of the glycerol facilitator GlpF: a variant of the Aquaporin tetramer. *EMBO Rep.* **1**, 183–189.
  23. Walz, T., Hirai, T., Murata, K., Heymann, J. B., Mitsuoka, K., Fujiyoshi, Y. *et al.* (1997). The three-dimensional structure of Aquaporin-1. *Nature*, **387**, 624–627.
  24. Cheng, A. C., Van Hoek, A. N., Yeager, M., Verkman, A. S. & Mitra, A. K. (1997). Three-dimensional organization of a human water channel. *Nature*, **387**, 627–630.
  25. Li, H. L., Lee, S. & Jap, B. K. (1997). Molecular design of Aquaporin-1 water channel as revealed by electron crystallography. *Nature Struct. Biol.* **4**, 263–265.
  26. Mitsuoka, K., Murata, K., Walz, T., Hirai, T., Agre, P., Heymann, J. B. *et al.* (1999). The structure of Aquaporin-1 at 4.5 Å resolution reveals short  $\alpha$ -helices in the center of the monomer. *J. Struct. Biol.* **128**, 34–43.
  27. de Groot, B. L., Heymann, J. B., Engel, A., Mitsuoka, K., Fujiyoshi, Y. & Grubmüller, H. (2000). The fold of human Aquaporin 1. *J. Mol. Biol.* **300**, 987–994.
  28. Ren, G., Cheng, A., Reddy, V., Melnyk, P. & Mitra, A. K. (2000). Three-dimensional fold of the human AQP1 water channel determined at 4 Ångstrom resolution by electron crystallography of two-dimensional crystals embedded in ice. *J. Mol. Biol.* **301**, 369–387.
  29. Murata, K., Mitsuoka, K., Walz, T., Agre, P., Heymann, J., Engel, A. & Fujiyoshi, Y. (2000). Structural determinants of water permeation through Aquaporin-1. *Nature*, **407**, 509–605.
  30. Ren, G., Reddy, V. S., Cheng, A., Melnyk, P. & Mitra, A. K. (2001). Visualization of a water-selective pore by electron crystallography in vitreous ice. *Proc. Natl Acad. Sci. USA*, **98**, 1398–1403.
  31. Zhu, F., Tajkhorshid, E. & Schulten, K. (2001). Molecular dynamics study of aquaporin-1 water channel in a lipid bilayer. *FEBS Letters*, **504**, 212–218.
  32. Kong, Y. & Ma, J. (2001). Dynamic mechanisms of the membrane water channel aquaporin-1 (AQP1). *Proc. Natl Acad. Sci. USA*, **98**, 14345–14349.
  33. de Groot, B. L., Engel, A. & Grubmüller, H. (2001). A refined structure of human Aquaporin-1. *FEBS Letters*, **504**, 206–211.
  34. de Groot, B. L. & Grubmüller, H. (2001). Water permeation across biological membranes: mechanism and dynamics of Aquaporin-1 and GlpF. *Science*, **294**, 2353–2357.
  35. Sui, H., Han, B.-G., Lee, J. K., Walian, P. & Jap, B. K. (2001). Structural basis of water-specific transport through the AQP1 water channel. *Nature*, **414**, 872–878.
  36. Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). Errors in protein structures. *Nature*, **381**, 272.
  37. Preston, G. M., Jung, J. S., Guggino, W. B. & Agre, P. (1993). The mercury-sensitive residue at cysteine-189 in the CHIP28 water channel. *J. Biol. Chem.* **268**, 17–20.
  38. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics*, **8**, 52–56.
  39. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system (cns): a new software system for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
  40. Lindahl, E. & Van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Modell.* **7**, 306–317.
  41. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
  42. Esnouf, R. M. (1997). An extensively modified version of MOLSCRIPT that includes greatly enhanced coloring capabilities. *J. Mol. Graphics Model.* **15**, 132–134.
  43. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photo-realistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
  44. Smart, O. S., Neduelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. P. (1996). HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graphics*, **14**, 354–360.

Edited by W. Baumeister

(Received 4 July 2002; received in revised form 7 October 2002; accepted 31 October 2002)