FEBS 28566 FEBS Letters 570 (2004) 41–46

On the origin of the electrostatic barrier for proton transport in aquaporin

Anton Burykin, Arieh Warshel*

Department of Chemistry, University of Southern California, 3620 McClintock Ave. SGM 418, Los Angeles, CA 90089-1062, USA

Received 4 May 2004; revised 2 June 2004; accepted 2 June 2004

Available online 19 June 2004

Edited by Maurice Montal

Abstract The nature of the electrostatic barrier for proton transport in aquaporins is analyzed by semimacroscopic and microscopic models. It is found that the barrier is associated with the loss of the generalized solvation energy upon moving from the bulk solvent to the center of the channel. It is clarified that our solvation concept includes the effect of the protein polar groups and ionized residues. The nature of the contributions to the solvation barrier is examined by using the linear response approximation. It is found that the residues in the NPA region contribute much less than what would be deduced from calculations that do not consider the protein reorganization. It is clarified that the contributions of different structural or electrostatic elements to the solvation barrier can be established by removing these elements and examining the corresponding effect on the barrier height. Using this definition and "mutating" the NPA residues to their non-polar analogues establishes that these residues do not provide the major contribution to the solvation barrier.

1. Introduction

Aquaporins are membrane proteins that allow efficient transfer of water molecules through cell membrane [1,12], while preventing the transport of protons. The origin of this water/proton selectivity is a subject of a major current interest [45] both as a special feature of aquaporin and as a possible clue and guide for studies of the general issue of proton translocations (PTR) in biological systems (for a reviews on general studies of biological PTR, see [11]). Early studies (e.g., [9,37]) suggested that the barrier for PTR in aquaporin may be due to orientational effects that disrupt the structure of the water chain from the optimal arrangement needed presumably for an efficient proton transport. Such proposals were inspired by the major role played by the so-called Grotthuss mechanism in PTR in bulk water [8,13,39,46] and by the proposal that such mechanisms also control biological PTR [29]. However, recent works [6,7,10,17,18] appear to conclude that this is due to the electrostatic barrier of the transferring proton, in agreement with the earlier general proposal [31,40] which argued that PTR in proteins is controlled by electrostatic barriers. What remains somewhat controversial, however, is the origin of the electrostatic barrier and its magnitude. Some workers have attributed the barrier to special structural elements [10,18], and in particular to the so-called NPA motif [17,18,45]. On the other hand, Burykin and Warshel [6] concluded that although the electrostatic barrier reflects all the electrostatic contributions of the channel (polar and non-polar) the barrier will remain very high even when these contributions are removed.

A part of the problem in analyzing the origin of the barrier might be due to the use of different methods, including some that have not been validated in studies of known electrostatic contributions and pK_a calculations. The problem also reflects the fact that most of the reported studies have not examined the actual contributions of different residues by removing these residues.

The validation of different proposals has been perhaps slowed because of the implicit assumption that PTR processes in proteins are mainly controlled by the orientation of the water molecules (e.g., [29]) rather than by electrostatic energy of the transferred proton. However, now when the empirical valence bond (EVB) and other models demonstrated that the PTR profile follows the electrostatic profile (at least in aquaporin); it is possible to focus on the nature of the electrostatic barrier. This paper will report the results of two complimentary approaches and will quantify the origin of the activation barrier. The paper will also clarify some misunderstandings about the concept of "solvation" by the protein and the ways to examine group contributions.

2. Materials and methods

Accepting, at least for the purpose of this paper, the idea that the PTR profile follows the electrostatic profile of the transferred proton when the barrier is relatively high (for quantitative demonstration see [5]), we can focus on the electrostatic profile. In fact, since we are interested in a proton transfer process we should focus on the electrostatic contributions to the protonation energies (and the corresponding pK_a values) of the different protonation sites, which are nevertheless determined by the electrostatic profile.

Now, in searching for accurate models for evaluation of pK_a values in heterogeneous and sometimes highly charged protein environments, it is important to realize that some formally rigorous methods may not provide converging results, and it is useful to choose methods that have been validated extensively. Thus, we will use here two approaches, the semimacroscopic PDLD/S-linear response approximation (LRA) approach and the microscopic LRA approach [23,32]. These approaches will be briefly described below.

One of the most rigorous ways of evaluating electrostatic free energies is the so-called free-energy perturbation (FEP) method [38,47]. The FEP method evaluates the free energy associated with the change of the potential surface from U_1 to U_2 by gradually changing the potential surface using the relationship (e.g. [44]):

* Corresponding author. Fax: +1-213-740-2701.
E-mail address: warshel@usc.edu (A. Warshel).

0014-5793/$22.00 © 2004 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/j.febslet.2004.06.020

$$U_m(\lambda_m) = U_1(1-\lambda_m) + U_2\lambda_m \quad (1)$$

where λ_m is a parameter that changes between ($0 \leqslant \lambda_m \leqslant 1$). The free-energy increment, associated with the change of U_m, can be obtained by [38]

$$\exp\{-\delta G(\lambda_m \to \lambda_{m'})\beta\} = \langle \exp\{-(U_{m'} - U_m)\beta\}\rangle_m \quad (2)$$

where $\langle\rangle_m$ indicates that the given average is evaluated by propagating trajectories over U_m. The overall free-energy change is now obtained by changing λ_m in n equal increments and evaluating the sum of the corresponding δG:

$$\Delta G(U_1 \to U_2) = \sum_{m=0}^{n-1} \delta G(\lambda_m \to \lambda_{m+1}) \quad (3)$$

The FEP approach was introduced to studies of electrostatic energies in proteins in [44] and has been used extensively in studies of free energies of biological systems (e.g., [21,35]).

In many cases, it is very hard to perform converging FEP calculations (e.g., binding of large ligands). In such cases, it is extremely useful to estimate the free energy of biological processes by an equation derived by [23] and used in studies of ligand binding to proteins [14,32]. This equation expresses the free energy associated with changing the potential of the system from U_1 to U_2 by

$$\Delta G(U_1 \to U_2) = \tfrac{1}{2}\left(\langle U_2 - U_1\rangle_1 + \langle U_2 - U_1\rangle_2\right) \quad (4)$$

The derivation of this equation was based on the assumption that the LRA is valid. Namely, the protein and solvent environments respond linearly to the force associated with the given process. This assumption is, in fact, the basis of macroscopic theory where the free energy of charging an ion of a charge Q_0 is given by the well-known expression

$$\Delta G(Q = 0 \to Q = Q_0) = 1/2\langle U(Q_0)\rangle_{Q_0} \quad (5)$$

where U is the electrostatic potential of the given charge [42]. This expression corresponds to the first term of Eq. (4) and gives an excellent approximation for the solvation energy of ions in water. The second term of Eq. (4) becomes important in the heterogeneous environment of proteins.

Although it is somehow hard to accept that the LRA can provide a reliable way of describing the energetics of macromolecules or of realistic molecular systems, it was found by simulation studies that it is a reasonable approximation, in particular for processes that depend on electrostatic effects [3,15,22].

The use of the LRA offers the unique ability to decompose electrostatic free energies of proteins to their individual additive contributions [27]. Such a treatment cannot be accomplished by FEP approaches due to their non-additive nature. The individual LRA contribution of the ith group is given by:

$$\Delta G_i(U_1 \to U_2) = \tfrac{1}{2}\left(\langle U_2^i - U_1^i\rangle_1 + \langle U_2^i - U_1^i\rangle_2\right) \quad (6)$$

While the LRA contribution gives the correct additive components of the total free energy, they do not relate to the corresponding mutational effects. That is, the effect of mutating a given residue should be obtained by performing LRA analysis for the native and mutant form of this residue. Furthermore, the LRA results should be compared to the corresponding results in water (as is done in the PDLD/S-LRA approach discussed below). A reasonable estimate of the effects of mutations can be obtained by using an effective dielectric constant and scaling down the interactions with ionized residues by ~20–40 while scaling down interactions with polar residues by ~2–4 (see [26,28] for discussion).

Despite the formal rigor of the FEP and LRA methods, it was found frequently that such methods are subjected to major convergence problems when one deals with electrostatic effects in protein interiors, and that semimacroscopic models can sometimes give more reliable results. This is true in particular with regards to the PDLD/S-LRA method [24,33] that provides a direct link between the microscopic and macroscopic concepts. This method evaluates the change in solvation free energies upon transfer of a given ligand (H_3O^+ in our case) to the protein by using the effective potential [24]

$$\Delta U_{\mathrm{sol,l}}^{\mathrm{w}\to\mathrm{p}} = \left[-\Delta G_{\mathrm{q,l}}^{\mathrm{w}} + \Delta G_{\mathrm{p}}^{\mathrm{w}}(q = q_{\mathrm{l}}) - \Delta G_{\mathrm{p}}^{\mathrm{w}}(q = 0)\right]\left(\frac{1}{\varepsilon_{\mathrm{p}}} - \frac{1}{\varepsilon_{\mathrm{w}}}\right) + \Delta U_{q\mu}\frac{1}{\varepsilon_{\mathrm{p}}} \quad (7)$$

where $\Delta G_{\mathrm{sol,l}}^{\mathrm{w}}$ is the free energy of solvation of the ligand in water (the self-energy in water), $\Delta G_{\mathrm{p}}^{\mathrm{w}}(q = q_{\mathrm{l}})$ and $\Delta G_{\mathrm{p}}^{\mathrm{w}}(q = 0)$ are the free energies of solvation of the entire protein in water with actual fractional charges on the atoms of the ligand (''the charged state'') and with fractional charges on the atoms of the ligand set to zero (''the uncharged state''), respectively. $\Delta U_{q\mu}$ is the vacuum interaction between the atomic charges on the ligand and the permanent dipoles of the protein (represented by the residual atomic charges), ε_{w} is the dielectric constant of water, and ε_{p} is the dielectric constant of the protein, which is basically a semimacroscopic scaling factor that accounts for the interactions that are not considered explicitly. This factor is quite different than the actual protein dielectric constant (see [30]).

To capture the physics of the reorganization of the protein dipoles in the charging process, it is necessary to relax the protein structure in the relevant charged and uncharged states. Moreover, for accurate free-energy differences, several protein configurations should be averaged. The configurational space can be adequately sampled by utilizing Monte Carlo or molecular dynamics (MD) techniques [2]. In this study, we use a MD approach in the LRA framework described above. This approach approximates the free energy associated with a transformation between the charged and uncharged states by averaging the potential difference between the initial and final states over trajectories propagated on these two states. Using the PDLD/S free energy that corresponds to each protein structure as an effective potential in the PDLD/S-LRA method, the free energy of solvation is given by

$$\Delta\Delta G_{\mathrm{sol},i}^{\mathrm{w}\to\mathrm{p}} = \frac{1}{2}\left[\left\langle \Delta U_{\mathrm{sol},i}^{\mathrm{w}\to\mathrm{p}}\right\rangle_{q=q_i} + \left\langle \Delta U_{\mathrm{sol},i}^{\mathrm{w}\to\mathrm{p}}\right\rangle_{q=0}\right] \quad (8)$$

where the $\Delta U_{\mathrm{sol},i}^{\mathrm{w}\to\mathrm{p}}$ is the PDLD/S effective potential of Eq. (7), the $\langle\rangle_{q=q_i}$ and $\langle\rangle_{q=0}$ terms designate an average over protein configurations generated in the charged and uncharged state of the given group, respectively. Although this approach takes into account the reorganization of the environment explicitly, it may not fully account for some effects such as the complete water penetration and protein reorganization. These factors and the effect of induced dipoles are implicitly included in the model, which lead to the use of ε_{p} in this semimacroscopic treatment (e.g., see [30]).

The basic PDLD/S-LRA calculations are performed by starting with all the protein groups in their neutral form. The effect of ionizing these groups is then evaluated macroscopically by finding their ionization state in a self-consistent way [33] and then evaluating the effect (in kcal/mol) of these groups using

$$\Delta G_{qq}(r) = 332/(r\varepsilon_{\mathrm{eff}}(r)) \quad (9)$$

where r is the distance between the interacting groups and $\varepsilon_{\mathrm{eff}}$ is an effective dielectric constant whose value is determined by a distance-dependent function [24,43]. The justification of this approximation is discussed in detail elsewhere [24,30,33]. Basically, ε for charge–charge interaction reflects the compensation of the gas phase Coulomb interaction between the charges by the solvation effect of the protein plus solvent system. This compensation has been found to be unexpectedly large even for charge–charge interaction in the protein interior, leading to a large effective $\varepsilon_{\mathrm{eff}}$ (between 20 and 40). This fact has been established repeatedly by both theoretical and experimental studies (e.g., [19,34]). It is also important to realize that $\varepsilon_{\mathrm{eff}}$ is not equal (and typically much larger) than the dielectric constant ε_{p}, that determines $\Delta\Delta G_{\mathrm{sol}}^{\mathrm{w}\to\mathrm{p}}$ (see [30] and discussion below).

The present study was performed on the AQP1 aquaporin water channel (pdb entry 1J4N, see [36]) considering the simulation system shown in Figs. 1 and 2(A). The construction of this system is described in [6] and here we only provide some key details and mention that the PDLD/S-LRA calculations considered two water molecules (on each side of the H_3O^+ ion) in an explicit way.

The PDLD/S-LRA calculations involved as usual two steps (e.g., [33]), first running microscopic MD simulations to generate protein configurations for the charged and uncharged states, and then averaging the PDLD/S results for the generated configurations. The MD runs were performed with the polarizable ENZYMIX force field [24]. All the PDLD/S-LRA calculations were performed by the automated procedure of the MOLARIS program [24], where we generated typically 10 configurations for the charged and uncharged states, using MD simulations of 2 ps, with a 1 fs time step, for each configuration. The calculations were done using the surface constraint all atom solvent (SCAAS) model [20] and the local reaction field (LRF) long range

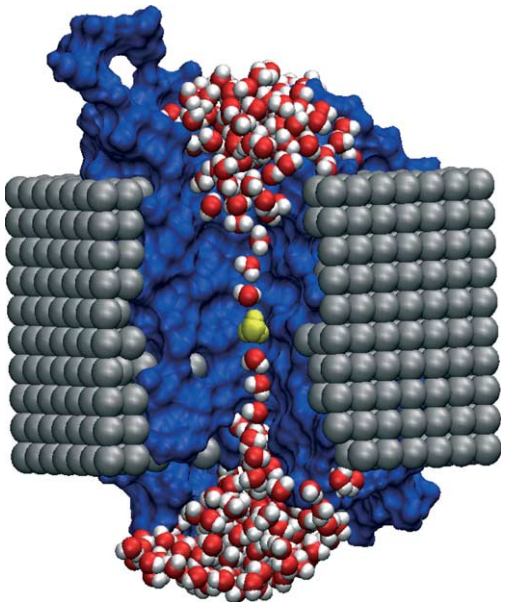

Fig. 1. The all-atom simulation system (spherical boundary conditions). The aquaporin monomer is embedded in a grid of $30 \times 20 \times 20$ Å size and 2.5 Å spacing of carbon-like atoms that represent the low dielectric membrane. An H_3O^+ ion (shown in yellow) is placed in the center of the channel at the NPA region. The water clusters on both sides of the channel represent the SCAAS sphere used in the simulations.

treatment [25].The microscopic LRA calculations were obtained in an automated way from the MD simulations used to generate the PDLD/S configurations.

3. Results

As a starting point, we evaluated the overall PDLD/S-LRA electrostatic barrier for the transfer of an H_3O^+ ion through AQP1. The results of the calculations, which are depicted in Fig. 2(B), indicate, as was found before [6], that there is a high barrier of about 17 kcal/mol at the center of the channel. Our task is now to determine the origin of this high barrier.

The examination of the nature of the barrier was performed by both the PDLD/S-LRA and the microscopic LRA approaches. The results of the PDLD/S-LRA calculations are summarized in Tables 1 and 2 as well as in Fig. 3. The PDLD/S-LRA results of Table 1 are given relative to water. As seen from the table the difference between the energy in the center of the channel and near the entrance is due mainly to the $\Delta G_{\mathrm{p}}^{\mathrm{w}}$ term that reflects the solvation of the charge by the bulk water and by the explicit water molecules, excluding the two water molecules on both sides of the H_3O^+ ion. Obviously an H_3O^+ in the center is not solvated significantly by the bulk solvent while the salvation of an H_3O^+ at the entrance of the channel has a significant contribution from the bulk solvent. Interestingly, the effect of the protein dipoles (the $\Delta U_{q\mu}$ contribution) is not large. This point will be discussed below.

Another way to look at our results is to examine the microscopic (all atom) LRA calculations of Table 2. This table provides the absolute solvation energy in water and in the protein sites. Here, we see the same trend as in Table 1 but without the scaling by ε_{p}. A part of this compensating effect is

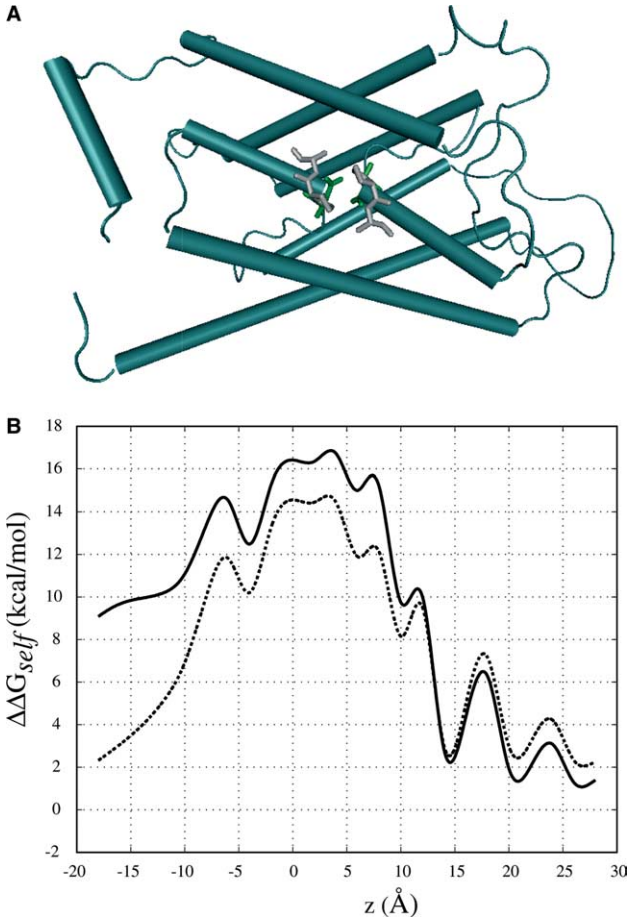

Fig. 2. (A) Simplified representation of AQP1 water channel. The protein and the NPA residues are shown using "cartoon" and "licorice" representations, respectively. (B) The PDLD/S-LRA profile for transferring an H_3O^+ ion along the aquaporin pore both with (solid line) and without (dotted line) the effect of the ionizable residues. The Z axis corresponds to the channel axis (according to Fig. 2(A)) where $Z = 0$ Å corresponds to the center of the channel.

provided by the energy of the induced dipoles (ΔG_{ind}). The difference between the barrier heights obtained by both approaches will be discussed below.

The results of the LRA analysis are also displayed in a graphical way in Fig. 3. Here, again we see the overwhelming effect of the solvation by the surrounding water and the fact that the overall "solvation" in the protein is significantly smaller than in water. This is due to the fact that the sum of the different contributions does not provide as much stabilization as water does. The implication of this desolvation effect is of significant interest, because it is quite different from the perception [45] that the NPA region leads to a large repulsion of the positive charge. That is, the present work has found that the overall effect of the protein permanent dipoles inside the channel (the $\Delta G_{q\mu}^{(ii)}$ of Table 2) does not provide enough solvation to compensate for the lose of solvation by the bulk solvent ($\Delta G_{\mathrm{sol}}^{\mathrm{w}}$). However, it was also found that the $\Delta G_{q\mu}^{(ii)}$ does not lead to a significant electrostatic repulsion (i.e., large positive $\Delta U_{q\mu}$). The finding that the channel does not provide a large positive $\Delta G_{q\mu}^{(ii)}$ is very different from the results obtained from calculations that do not allow the protein dipoles to rearrange themselves upon charge penetration. In such a case,

Table 1
PDLD/S-LRA free-energy contributions (in kcal/mol) at selected points along the AQP1 channel[a]

Site	$\Delta G^{p}_{w}/\varepsilon_p$	$\langle U^{(i)}_{q\mu}\rangle/\varepsilon_p$	$\langle U^{(ii)}_{q\mu}\rangle/\varepsilon_p$	ΔG_{qq}	$-(\Delta G^{w}_{sol}/\varepsilon_p)$	$\Delta\Delta G_{sol}$
Protein, $z = 0$ Å	−5.7	−4.8	−0.5	2.0	26.8	17.8
Protein, $z = 23$ Å	−20.3	−2.0	−0.4	−1.0	26.9	3.2

[a] The different free-energy contributions are designated according to the notation used for the corresponding effective potential of Eq. (7). ΔG_{qq} is evaluated by Eq. (9) for the corresponding contribution to $\Delta U^{w\to p}$. The designated sites are depicted in Fig. 3. The contributions to the $\Delta U_{q\mu}$ are divided into the contribution from the two explicit water molecules in front and after the H_3O^+ ion ($\Delta U^{(i)}_{q\mu}$) and to the contribution from the protein polar groups ($\Delta U^{(ii)}_{q\mu}$).

Table 2
Microscopic LRA free-energy contributions (in kcal/mol) at selected points along the AQP1 channel and in water (see also Fig. 3)[a]

Site	$\Delta G^{(i)}_{q\mu}$	$\Delta G^{(ii)}_{q\mu}$	ΔG_{ind}	ΔG^{w}_{sol}	ΔG_{vdw}	ΔG_{qq}	ΔG_{sol}
Protein, $z = 0$ Å	−21.2	−2.1	−28.7	−11.2	−5.1	2.0	−66.3
Protein, $z = 23$ Å	−10.0	−0.6	−1.7	−85.5	−0.8	−1.0	−99.6
Water	−99.3	0.0	−0.2	−99.3	−1.5	0.0	−101.0

[a] The contributions to the total LRA solvation energy for several points along the channel axis and in water. $\Delta G_{q\mu}$, ΔG_{ind}, ΔG^{w}_{sol} and ΔG_{vdw} are the change in the free energies for, respectively, the ion-permanent dipoles, ion-induced dipoles, ion-water dipoles and van der Waals interaction terms. $\Delta G_{q\mu}$ is divided into the two contributions in the same way as in Table 1.

one gets indeed a large repulsion (see below), but once the protein is allowed (as it should be) to rearrange its polar groups, it tries to stabilize the ion.

A more detailed analysis of the effect of the NPA motif is reported in Table 3, where we examine the effects of the different terms in the LRA equation (the contributions from the simulation with and without the proton charge) and the results obtained when the protein is not allowed to reorganize. We also present the effect of mutating the NPA residues to their non-polar analogues.

Comparing the second and the fifth columns of the table illustrates the crucial effect of the protein reorganization. That is, as seen from the table, calculations that do not allow the NPA to relax upon charge formation drastically overestimate the effects of these residues (34 kcal/mol instead of about 7 kcal/mol). Here, it is important to emphasize here that this

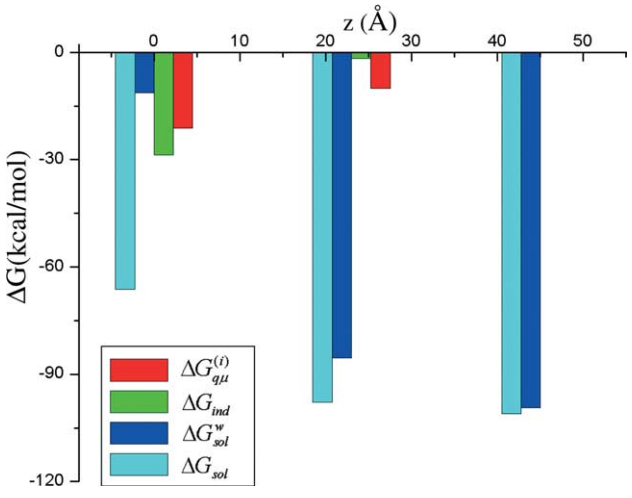

Fig. 3. Microscopic (all-atom) electrostatic contribution for the solvation of an H_3O^+ ion at different sites along the channel pore obtained by the LRA method. The Z axis corresponds to the channel axis in Fig. 2(A). The figure displays the total solvation energy (ΔG_{sol}), and its components (ΔG^{w}_{sol}, $\Delta G^{(i)}_{q\mu}$ and ΔG_{ind}); see also Table 2.

contribution (7 kcal/mol) does not provide half of the total barrier, since all other polar groups compensate the NPA contribution and the combined effect of all the polar groups at the top of the barrier (the $\Delta G^{(ii)}_{q\mu}$ of Table 3) is even negative.

In this respect, it is instructive to see how the ~22 kcal/mol unrelaxed contribution of Asn78 is reduced to around 1 kcal/mol when the protein is allowed to relax. This establishes the fact that the unrelaxed interaction strength cannot be used in estimating group contributions. Further inspection of the table shows that the replacement of the NPA residues by non-polar residues reduces the barrier by only −3.6 kcal/mol (the barrier changes from around 17 kcal/mol to approximately 12 kcal/mol). Similarly mutating Asn78 to its non-polar analogue changes the barrier by only 0.3 kcal/mol.

These results indicate that the NPA motif is not the primary reason for the barrier. In fact, in order for a given motif to be the reason for the barrier, it is essential that the removal of this motif will eliminate the barrier.

While the general trend of the calculation is qualitatively correct, we should still address several issues concerning the quantitative level of the different methods used. The first issue is the fact that the microscopic LRA gives a higher barrier than the semimacroscopic PDLD/S-LRA approach. This brings us to an interesting dilemma, which has not been widely discussed in the literature. That is, in principle one should obtain the same results for the PDLD/S-LRA and the microscopic LRA with $\varepsilon_p = 2$, but this requires a full convergence of water penetration and related effects [26]. Usually $\varepsilon_p = 4$ gives optimal results in proteins reflecting this incomplete water penetration upon charging. However, in the case of narrow channels where we already have a single file of water molecules, it is reasonable to assume that the water penetration is complete. Thus, we would obtain better agreement between the microscopic and semimacroscopic results with ε_p between 2 and 3. With this in mind we believe that the actual barrier should be close to its microscopic estimate of 30 ± 6 kcal/mol than to the semimacroscopic estimate obtained with $\varepsilon_p = 4$. It is important to note that we also perform full free-energy perturbation (FEP) calculations for the solvation of the H_3O^+ ion in the NPA region. This calculation gave a barrier of ~26 kcal/mol.

Table 3
The energetics (in kcal/mol) of the different residues in the NPA region with different approximations[a]

Group	$(\Delta U_{q=q_0})_{\text{unrelaxed}}$	$\langle \Delta U \rangle_{q=0}$	$\langle \Delta U \rangle_{q=q_0}$	$(\Delta G_{q\mu})_{\text{relaxed}}$	$\Delta G_{\text{polar}\rightarrow\text{non-polar}}$
Asn78	21.9	1.7	0.2	1.0	−0.3
Pro89	0.8	0.4	0.4	0.4	−1.28
Ala80	1.8	0.8	0.5	0.7	−0.11
Asn194	7.5	4.9	2.4	3.7	−1.02
Pro195	0.6	0.3	0.2	0.3	−0.55
Ala196	1.3	0.7	0.6	0.7	−0.94
Total NPA	33.9	8.8	4.3	6.6	−3.6
All other polar groups	−2.8	−1.1	−11.5	−7.7	–
Total polar groups	31.1	7.7	−7.2	−1.1	–

[a] $(\Delta U_{q=q_0})_{\text{unrelaxed}}$ corresponds to the group contributions obtained using the unrelaxed protein structure. $\langle \Delta U \rangle_{q=q_0}$, $\langle \Delta U \rangle_{q=0}$ and $\Delta G_{q\mu}$ are, respectively, the contributions of the two terms in Eq. (6) and $\Delta G_{q\mu}$ is the corresponding free-energy contribution of the indicated group. $\Delta G_{\text{polar}\rightarrow\text{non-polar}}$ designates the overall free-energy obtained by mutating the given group to an identical non-polar group where all the residual charges were set to zero.

Table 4
Approximate group contributions (in kcal/mol) to the electrostatic free energy of an H_3O^+ at the NPA site[a]

Residue	ΔG	Residue	ΔG	Residue	ΔG	Residue	ΔG
GLU17	−1.8	LEU65	0.5	ALA110	0.3	PRO195	0.3
PHE18	−0.4	HIS76	−0.5	GLU144	−1.2	ALA196	0.7
ALA20	−0.5	LEU77	−4.7	ARG161	0.4	ARG197	1.6
MET21	−3.1	ASN78	1.0	ARG163	0.4	SER198	0.5
PHE24	−0.7	PRO79	0.4	ILE174	−0.4	PHE199	0.4
SER28	0.5	ALA80	0.7	VAL178	0.3	GLY200	0.4
ASP50	−0.5	VAL81	0.8	HIS182	0.8	ASP210	−0.5
SER55	−0.5	LEU83	0.4	GLY192	−0.8	HIS211	−0.5
PHE58	−0.7	ARG95	0.3	ILE193	−0.6	ASP230	−0.4
THR64	0.5	GLN103	−0.6	ASN194	3.7	ARG236	0.4

[a] The table lists the largest contributions. The contributions for polar residues were obtained from the corresponding $\langle \Delta U_{q\mu} \rangle / \varepsilon_p$. The contributions for ionized residues were obtained using Eq. (9).

Finally, we provide in Table 4 the contributions from different protein residues to the barrier at the NPA region. This qualitative result is obtained from the PDLD/S-LRA contributions for the polar groups and from Eq. (9) for ionized residues. This table can be used as a rough guide for mutating different residues to non-polar residues. Predicting the effect of mutation to a polar or ionized residue requires similar calculations.

4. Concluding remarks

This work examined the origin of the barrier for the transport of protons through the aquaporin channel. Considering our finding that the barrier for PTR follows the electrostatic profile we focus on the origin of the electrostatic barrier. It was found that the barrier is due to the fact that the absolute value of the "solvation" of the charge is smaller in the channel than in water. It should be clarified in this respect that the solvation of the charge reflects all the electrostatic components including the effect of the protein polar groups and ionized residues [4,41,42].

Although early structural studies (e.g., [36]) have assumed that the Grotthuss mechanism will overcome the electrostatic barrier, they provided a reasonable qualitative discussion of the factors that control the barrier for ion transport. However, such structure-based analysis is not sufficiently quantitative to provide a consistent analysis. Here, it is important to use electrostatic calculations that convert the structural information to energetics, while considering the protein reorganization effects.

As to the effect of specific residues, we emphasize that such effects cannot be determined computationally by just evaluating the protein–ion electrostatic interaction at a fixed protein structure. Such effects can be evaluated qualitatively by the LRA formulation, which takes the protein reorganization into account and in a more quantitative way by "mutating" the given residue. Both approaches indicated that the NPA region destabilizes the H_3O^+ ion, but this effect is much smaller than the corresponding effect of water dipoles in the reference bulk system. It is also demonstrated that the NPA effect is quite small since the replacement of this residue by other residues does not change the barrier in a drastic way. Thus we predict that mutations of the NPA region will not lead to proton penetration through the aquaporin channel (unless the folding of the channel is completely destroyed). We also would like to clarify once more what must be meant by the argument that the barrier is "due" to the NPA motif. Such a proposal means that a removal of the NPA will reduce the barrier in a drastic way. If the removal of the NPA does not remove the barrier, the barrier is not due to this motif.

The assumption that the NPA motif plays a major role in the water/protein selectivity seems to overlook several important points about what is required from an ion (or proton) channel. In general, it is rather trivial to create a channel that will not conduct protons. All that is needed is a relatively hydrophobic interior. Even a mildly polar interior is not sufficient for allowing proton transfer since the protons will be less stable than

in water. Similarly, it is not difficult to design a channel that will allow a moderate transport of water, since the solvation energy of water is not large. The true biological challenge is the construction of a channel that will have the correct polarity to allow proton transport.

Although this paper seems to send a "negative" message with regards to the prediction that simple mutation experiments will not be successful (see related prediction in [16]), it does not mean that the situation is hopeless. In principle, if it was possible to keep the channel folded while performing arbitrary mutations, it should be possible to increase the proton conductance. For example, lining the NPA region with serines would drastically reduce the barrier at this point. It would be challenging to provide conclusive predictions for such mutations, since they would require simultaneous calculations of protein stability and proton binding for different mutants, which would inevitably help the experimental progress in the field.

Acknowledgements: This work was supported by the National Institute of Health Grant GM40283. We thank the University of Southern California's High-Performance Computing and Communications (HPCC) Center for computer time.

References

[1] Agre, P. and Kozono, D. (2003) FEBS Lett. 555, 72–78.
[2] Allen, M.P. and Tildesley, D.J. (1987) Computer Simulation of Liquids. Oxford University Press, Oxford.
[3] Aqvist, J. and Hansson, T. (1996) J. Phys. Chem. 100, 9512–9521.
[4] Aqvist, J. and Warshel, A. (1989) Biophys. J. 56, 171–182.
[5] Braun-Sand, S., Strabjl, M. and Warshel, A. (2004) submitted.
[6] Burykin, A. and Warshel, A. (2003) Biophys. J. 85, 3696–3706.
[7] Chakrabarti, N., Tajkhorshid, E., Roux, B. and Pomes, R. (2004) Structure 12, 65–74.
[8] Day, T.J.F., Schmitt, U.W. and Voth, G.A. (2000) J. Am. Chem. Soc. 122, 12027–12028.
[9] de Groot, B. and Grubmuller, H. (2001) Science 294, 2353–2357.
[10] de Groot, B.L., Frigato, T., Helms, V. and Grubmuller, H. (2003) J. Mol. Biol. 333 (2), 279–293.
[11] Decoursey, T.E. (2003) Physiol. Rev. 83, 475–579.
[12] Denker, B.M., Smith, B.L., Kuhajda, F.P. and Agre, P. (1988) J. Biol. Chem. 263, 15634–15642.
[13] Eigen, M. (1963) Angew. Chem. Int. Ed. 75 (12), 489–508.
[14] Florian, J., Goodman, M.F. and Warshel, A. (2002) J. Phys. Chem. B 106, 5739–5753.
[15] Hwang, J.-K. and Warshel, A. (1987) J. Am. Chem. Soc. 109, 715–720.
[16] Hwang, J.-K. and Warshel, A. (1988) Nature 334, 270.
[17] Ilan, B., Tajkhorshid, E., Schulten, K. and Voth, G.A. (2004) Proteins: Struct. Funct. Bioinform. 55, 223–228.
[18] Jensen, M.O., Tajkhorshid, E. and Schulten, K. (2003) Biophys. J. 85 (5), 2884–2899.
[19] Johnson, E.T. and Parson, W.W. (2002) Biochemistry 41, 6483–6494.
[20] King, G. and Warshel, A. (1989) J. Chem. Phys. 91 (6), 3647–3661.
[21] Kollman, P. (1993) Chem. Rev. 93, 2395–2417.
[22] Kuharski, R.A., Bader, J.S., Chandler, D., Sprik, M., Klein, M.L. and Impey, R.W. (1988) J. Chem. Phys. 89, 3248–3257.
[23] Lee, F.S., Chu, Z.T., Bolger, M.B. and Warshel, A. (1992) Prot. Eng. 5, 215–228.
[24] Lee, F.S., Chu, Z.T. and Warshel, A. (1993) J. Comp. Chem. 14, 161–185.
[25] Lee, F.S. and Warshel, A. (1992) J. Chem. Phys. 97, 3100–3107.
[26] Muegge, I., Qi, P.X., Wand, A.J., Chu, Z.T. and Warshel, A. (1997) J. Phys. Chem. B 101, 825–836.
[27] Muegge, I., Schweins, T. and Warshel, A. (1998) Proteins: Struct. Funct. Genet. 30, 407–423.
[28] Muegge, I., Tao, H. and Warshel, A. (1998) Prot. Eng. 10 (12), 1363–1372.
[29] Nagle, J.F. and Morowitz, H.J. (1978) Proc. Natl. Acad. Sci. USA 75, 298–302.
[30] Schutz, C.N. and Warshel, A. (2001) Proteins: Struc. Funct. Genet. 44, 400–417.
[31] Sham, Y., Muegge, I. and Warshel, A. (1999) Proteins 36, 484–500.
[32] Sham, Y.Y., Chu, Z.T., Tao, H. and Warshel, A. (2000) Proteins: Struct. Funct. Genet. 39, 393–407.
[33] Sham, Y.Y., Chu, Z.T. and Warshel, A. (1997) J. Phys. Chem. B 101, 4458–4472.
[34] Sham, Y.Y., Muegge, I. and Warshel, A. (1998) Biophys. J. 74, 1744–1753.
[35] Shurki, A. and Warshel, A. (2003) Adv. Prot. Chem. 66, 249–312.
[36] Sui, H., Han, B.-G., Lee, J.K., Walian, P. and Jap, B.K. (2001) Nature 414, 872–878.
[37] Tajkhorshid, E., Nollert, P., Jensen, M., Miercke, L., Stroud, R.M. and Schulten, K. (2002) Science 296, 525–530.
[38] Valleau, J.P. and Torrie, G.M. (1977) Modern Theoretical Chemistry. in: A Guide to Monte Carlo for Statistical Mechanics. 2. Byways (Berne, B.J., Ed.), vol. 5, pp. 169–194, Plenum Press, New York.
[39] Vuilleumier, R. and Borgis, D. (1997) J. Mol. Graph. 436–437, 555–565.
[40] Warshel, A. (1979) Photochem. Photobiol. 30, 285–290.
[41] Warshel, A., Aqvist, J. and Creighton, S. (1989) Proc. Natl. Acad. Sci. USA 86, 5820–5824.
[42] Warshel, A. and Russell, S.T. (1984) Q. Rev. Biophys. 17, 283–421.
[43] Warshel, A., Russell, S.T. and Churg, A.K. (1984) Proc. Natl. Acad. Sci. USA 81, 4785–4789.
[44] Warshel, A., Sussman, F. and King, G. (1986) Biochemistry 25, 8368–8372.
[45] Yarnell, A. (2004) Chem. Eng. News 82 (4), 42–44.
[46] Zundel, G. and Fritcsh, J. (1986) In: The Chemical Physics of Solvation, vol. 2, Elsevier, Amsterdam.
[47] Zwanzig, R.W. (1954) J. Chem. Phys. 22, 1420.