Supplemental file

In this supplemental file we report results obtained by alternative solvent dielectric treatments, including calculations with 1–3 explicit water molecules and higher solvent dielectric constants.

Asp26. Analysis of the MD simulation structures showed that the water organization around the Asp26 carboxylate depends strongly on the protonation state. In the neutral and midpoint structures, the Asp26 side chain interacts with one water molecule; in the ionized state, it forms strong interactions with three waters. The details of these interactions are lost when the aqueous solvent is modelled as a structureless continuum. To explore the effect of the continuum approximation we performed calculations in which 1, 2, or 3 water molecules interacting closely with the Asp26 carboxylate were included explicitly as a part of the protein dielectric cavity. The protein dielectric constant ϵ_s^p was set to one. The resulting static free energies are reported in Table S1, together with the MDFE derivatives.

With one explicit water, the static term becomes much more positive in the charged state (116.6 vs. 95.5 kcal/mol without explicit water and $\epsilon_s^p = 1$; compare with Table 1 of main text). The neutral static term has the opposite trend, decreasing to 6.9 kcal/mol (vs. 28.7 previously). The midpoint term is also more positive than before, with a value of 61.8 kcal/mol, in perfect agreement with the average of the neutral and charged state values. All values are in much better agreement with the corresponding MDFE free energy derivatives.

Including two or three explicit waters in the molecular model yields total static terms of 135.7 and 151.6 kcal/mol for the ionized state (see Table S1). These values are significantly higher than the CHARMM MDFE values, indicating that the PB electrostatic potential at the charge insertion sites is overestimated when more explicit waters are employed.

From the static terms with one explicit water we compute the protonation free energy as in the main text. The resulting values are reported in Table S2. The protonation free energy is 61.7 kcal/mol, and does not depend on the pathway, in agreement with linear response.

For the model compound, the static free energies range from 143.7 kcal/mol to -10.1 kcal/mol when a single explicit water is included in the PB calculations (Table S1). The protonation free energy becomes 66.1 kcal/mol, very close to the value without explicit water (66.5 kcal/mol; see Table 1). The PB/LRA protonation

free energy of Asp26 relative to the model compound is -4.4 kcal/mol, close to the experimental value of -4.8 kcal/mol (Table S2).

If the protein dielectric constant is raised to a value of two and one explicit water is retained, the ionized-state static term becomes 122.1 Kcal/mol, in very good agreement with the MDFE derivative (124.2 Kcal/mol). At the protonated state the static term becomes -1.0 Kcal/mol, smaller than before (compare with Table 1 in main text) and in better agreement with the MDFE value (-16.6 Kcal/mol). The protonation free energy (Table S2) varies between 60.6 and 64.6 kcal/mol, depending on the pathway. Using the most reliable, 3-point value (62.6 kcal/mol) and the corresponding ionization free energy of the model compound (67.7 kcal/mol, with $\epsilon_s^p = 2$ and no explicit water; see Table 1), we obtain a relative free energy of -5.1 kcal/mol, in very good agreement with experiment.

Overall, including one explicit water in the definition of the molecular system yields relative protonation free energies (-4.4 – -5.1 kcal/mol) in somewhat better agreement with the experimental value (-4.8 kcal/mol) for thioredoxin Asp26.

The difference between the PB static terms with and without explicit solvent is mainly due to the direct contribution of the water molecules included in the model. To see this, we report a group-decomposition of the static free energy terms for the case of $\epsilon_s^p = 1$ and one explicit water (Table S3). In the charged and midpoint states, the explicit water is oriented by the negative charge on the Asp26 carboxylate, and favors the introduction of a negative charge. At the neutral endpoint, the explicit water's orientation is influenced by the positive charge (+0.44e) on the $H_{\delta 2}$ atom; it opposes the introduction of a negative charge. Interestingly, the sum of the Lys57 and the explicit water contributions obeys linear response to a very good approximation. Also, the effect of displacing the dielectric boundary to accomodate the explicit water is small [compare the Asp26a component with the water (Table S3) and without the water (Table 4 in the main text)].

Asp20. Asp20 is located on the protein surface and interacts with water molecules in the MD simulations of all charge states. For this system we performed additional PB calculations, including a single explicit water interacting closely with Asp20. The resulting static free-energy values vary from 135.0 to -12.3 kcal/mol, depending on the charge state (see Table S1). Thus, the additional explicit water molecule has a much smaller effect on the static free energies of the solvent-exposed Asp20, compared to the thioredoxin internal residue Asp26. This was also observed in the case of the solvent-exposed model compound. The protonation free energy is 60.7 kcal/mol, nearly identical to the value (60.8 kcal/mol; Table 1) without explicit

water. The PB/LRA ionization free energy is -5.4 kcal/mol. Thus, the explicit water does not improve the agreement with the experimental value.

As discussed in the main text, one possible source of the disagreement between the theoretical and experimental relative ionization free energies could be that solvent close to Asp20 is more strongly ordered than in the bulk. To explore this possibility, we performed calculations with a solvent dielectric ϵ^{w}=120 and a protein dielectric $\epsilon_{\mathrm{s}}^{\mathrm{p}}$=1. The resulting static free energies vary between 137.7 kcal/mol and -8.9 kcal/mol, and the protonation free energy increases by 1.1 kcal/mol, to 61.8 kcal/mol. The relative free energy is -4.7 kcal/mol, still far from the experimental value (0.0).

Table S1. Static free energies from PB/LRA and MDFE, with explicit water*.

		$\epsilon_s^p = 1$			$\epsilon_s^p = 2$	$\epsilon_s^p = 1$ ϵ^w=120	
residue	state	1†	2†	3†	1†	0	MDFE$^{\sharp}$
	ASPH	6.9(1.0)			-1.0(0.5)		-16.6(4.6)
Asp26	Midpoint	61.8(0.6)			64.6(0.3)		44.8(2.0)
	ASP	116.6(0.7)	135.7(1.0)	151.6(0.4)	122.1(0.7)		124.2(2.2)
	ASPH	-12.3(0.5)				-8.9(0.3)	-19.0(0.4)
Asp20	Midpoint	60.1(0.7)				61.3(0.4)	57.0(1.2)
	ASP	135.0(0.7)				137.7(0.5)	143.5(0.8)
	ASPH	-10.1(0.5)					-19.3(1.4)
model	Midpoint	65.3(0.6)					58.4(0.8)
	ASP	143.7(0.4)					144.6(1.4)

*In kcal/mol. The signs correspond to the ASP $\longrightarrow$ ASPH direction (protonation). The protein dielectric constant is $\epsilon_s^p = 1$ or 2. Mean values over 100-200 MD structures are reported. The statistical uncertainty is in parentheses.

†Number of explicit water molecules retained in the PB calculation. $^{\sharp}$Free energy derivatives from a molecular dynamics free energy simulation (MDFE) starting from the ionized state ASP ('backward' run); see Table 1.

Table S2. Protonation free energies from PB/LRA and MDFE*

residue	pathway†	$\epsilon_s^p = 1$ $1^{\sharp}$	$\epsilon_s^p = 2$ $1^{\sharp}$	$\epsilon_s^p = 1$, ϵ^w=120	MDFE$^{\|}$	Exp$^{\sharp\sharp}$
	2-point	61.7(0.6)	60.6(0.4)			
Asp26	1-point	61.8(0.6)	64.6(0.3)			
	3-point	61.7(0.4)	62.6(0.3)		49.3(1.6)	
Asp20	3-point	60.7(0.4)		61.8(0.2)	59.6(0.6)	
model	3-point	66.1(0.3)			60.5(0.6)	
		$\Delta\Delta G^{\ddagger}$				
Asp26		-4.4(0.5)	-5.1(0.3)		-10.9(1.8)	-4.8
Asp20		-5.4(0.5)		-4.7(0.2)	-0.9(0.8)	0.0

*In kcal/mol. The signs correspond to the ASP $\longrightarrow$ ASPH direction (protonation). The protein dielectric constant is $\epsilon_s^p = 1$ or 2. Mean values over 100-200 MD structures are reported. Statistical uncertainty in parentheses.

†The pathways have been explained in Table 2. $^{\sharp}$Number of explicit water molecules retained in the PB calculation. $^{\|}$See Table 2, main text. $^{\sharp\sharp}$See Table 3, main text. ‡Relative to the model compound.

Table S3. Group contributions to the PB static free energy. Thioredoxin Asp26*

Group	ASPH	midpoint	ASP
	1†	1†	1†
Asp9	-2.0(0.2)	-2.1(0.3)	-0.9(0.6)
Asp43	-1.6(1.0)	-2.4(1.0)	-1.2(0.2)
Lys57	12.0(3.9)	15.2(3.8)	43.6(3.5)
Asp26a‡	4.9(1.3)	16.8(0.4)	31.2(1.3)
Asp26b‖	3.7(1.5)	6.9(1.4)	6.6(1.5)
Wat‡‡	-22.6(4.2)	11.1(2.0)	20.4(7.8)
Lys57+Wat	-10.6	26.3	64.0
Total	6.9	61.8	116.6

*In kcal/mol. The signs correspond to the ASP $\longrightarrow$ ASPH direction (protonation). Standard deviation along the MD trajectories in parentheses.

†Results with 1 explicit water molecules are reported. In the charged end state, the water molecule closest to atom $O_{\delta 2}$ was retained. ‡Contribution due to the reaction field induced by source charges at the insertion sites (see Sec. 4.2, main text). ‖Contribution from Asp26, excluding the source charges. ‡‡Contribution from the explicit water molecule.