Supporting Information

Fennell et al. 10.1073/pnas.1017130108

SI Text

Nearest-Neighbor Distribution Function Analysis. From the set of precomputation simulations, we generate simplified free energy representations of the local solvent molecules about single spheres. To assemble the solvation behavior around a given arbitrary solute, it is important that these solvent representations are placed on the surface in a physically reasonable manner, basically in a manner that mimics how explicit solvent occupies the surface of the solute. If they are too close to one another, the solvent–solvent interaction term may be too strong, leading to artificially unfavorable solvation. To obtain more physical spatial solvent arrangements, we place the dipoles or quadrupoles in accordance with the surface solvent nearest-neighbor distribution functions.

The idea behind the surface solvent nearest-neighbor distribution function (P_{NN}) is to determine from explicit solvent simulations a solvent arrangement connection to the degree of electrostriction about a solute. Fig. S1 illustrates the observed behavior of water around weakly and strongly charged solutes. As the solute–solvent interactions become stronger (with increasing charge magnitude), the water becomes more strongly pinned to the solute and less able to form optimal hydrogen bonds with neighboring surface waters. Unable to form favorable interactions, surface waters interact in a more repulsive fashion, leading to longer separation distances. To determine these separation distances, we go to the extracted first-shell explicit solvent trajectories discussed in the main text and bin water–water separations between nearest-neighbor waters. This can be determined as

$$P_{\mathrm{NN}}(r_{ij}) = \frac{1}{M \cdot \langle N_{AB} \rangle} \sum^{M} \left(\sum_{i\in A}^{N_A} \sum_{j\in B}^{N_B} \delta(\min r_{ij}) \right). \qquad \textbf{[S1]}$$

Here, M is the number of configuration frames, $\langle N_{AB} \rangle$ is the average number of water molecules in frames where there exist at least two molecules (one termed A and the other B), and $\delta(\min r_{ij})$ is a delta function for each minimum r_{ij} value. Numerically, Eq. **S1** is equal to 1, but the shape of P_{NN} as a function of r_{ij} is what we are interested in. The curve is peaked at the most populated surface solvent separation distance. Thus, for a given solute electric field strength, we know the optimal water spacing about a possible surface site.

In practice, we generate a table (see Table S2) of these P_{NN} peak values as a function of solute charge and size (i.e., local curvature). After placing a water representation dipole on the solute surface, we eliminate all possible subsequent sites that have distances shorter than those composing this peak. This eliminates unphysically close solvent placement.

Dipolar and Quadrupolar Free Energy Representations of Solvent. One of the key goals of semi-explicit assembly (SEA) is to reproduce the surface water polarization response from explicit solvent simulations. This is accomplished through the creation of free energy maps of the surface solvent molecules. These maps then need to be placed about the solute surface accordingly so that we can accumulate the polar solvation free energy due to surface water. In an effort to keep SEA conceptually simple, we decided to start with the most straightforward electrostatic representation, a simple dipole. As the electric field about the simple spherical solutes in our explicit solvent simulations is always normal to the surface, we decided to condense these maps onto a single normal vector to generate our "free energy dipoles."

Fig. S2 illustrates the process used to perform this transformation. Charge density is binned along the normal vector dimension to determine the local charge distribution relative to each spherical solute surface. To generate a dipole representation, the partial charge of the explicit water model's oxygen is assigned to the location of average negative charge density, whereas a neutralizing positive partial charge is assigned to the location of the average positive charge density.

In the main text, we discussed a quadrupolar representation that could be used in place of the simple dipole representation. A quadrupole comes naturally from the free energy maps for water, because the hydrogen atoms easily separate out into two distinct distributions, one near to and the other far from the solute surface. When generating the free energy map, we actually classify these two hydrogens to separate near and far distributions and place water hydrogen partial charges at the average distance of these distributions from the solute. This results in a linear quadrupole as seen in Fig. S2*D*. This is easily represented as two dipoles, and we can apply separate electric field curve fits to each. The benefit to a quadrupole representation is apparent around negatively charged solutes, as one of the solvent partial positive charges can get considerably closer to the solute. Drawbacks of a quadrupole representation include increased complexity and performance degradation in the interaction accumulation loop due the additional solvent site.

Electric Field Fits and Data Tables. The explicit solvent presimulations provide us with simple measures of surface solvent response. These measures include radial water distances (Table S1), free energy dipole magnitudes in electric fields near a solutes surface (Figs. S3–S5), and water–water separation distances (Table S2). We have also accumulated mean molecular occupancy values (and occupancy distribution variances) for the first solvation shell; however, these rarely get used as the water–water separation distances already help limit the number of "waters" that solvate each solute atom. Although we have already accumulated and continue to accumulate all these quantities for other solvent types, we include below the relevant plots and tables for TIP3P at 300 K and 1 atm, the solvent and state point used for the SEA calculations in the main text.

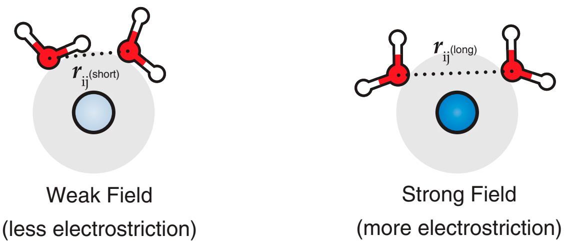

Fig. S1. With weaker solute–solvent interactions (weak field), surface water is more free to form hydrogen bonds, leading to separation distances (r_{ij}) similar to bulk water (2.7 Å). With strong solute–solvent interactions (strong field), a surface water molecule is unable to form hydrogen bonds with neighboring surface waters, leading to longer separation distances.

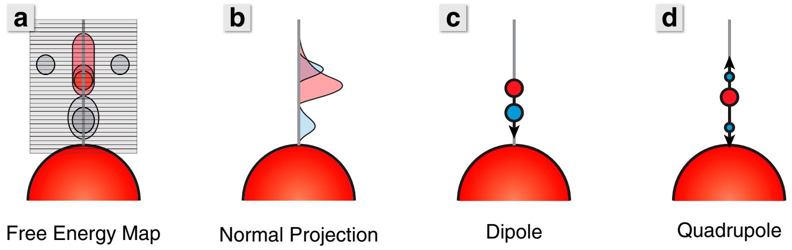

Fig. S2. For each spherical solute simulation, the (*A*) resulting free energy map is binned along the vector normal to the solute to generate a (*B*) projection of the charge density along this vector. This charge density projection can be simplified to a (*C*) dipole or (*D*) quadrupole by calculating the locations of the average density of the positively and negatively charged distributions in the normal projection.

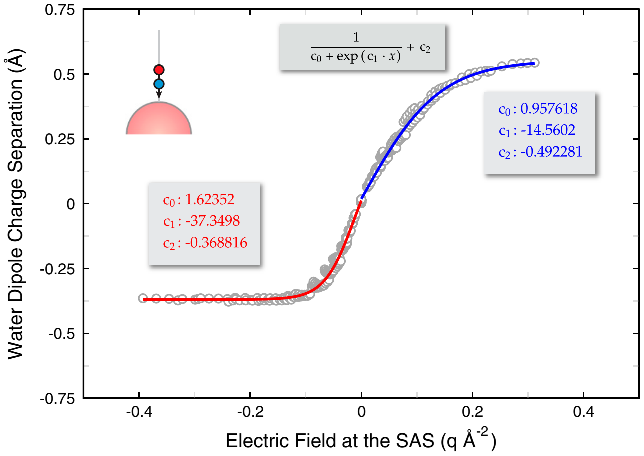

Fig. S3. Sigmoidal curve fits to the positive and negative field regions of the dipole charge separation vs. the electric field at the solvent-accessible surface (SAS) plots. The sign indicates direction of the dipole—with positive, the partial positive charge is farther than the negative from the solute and with negative, it is nearer to the solute. These fits are done in internal units rather than SI units for ease in semi-explicit configuration assembly. One can multiply by 4.006 to convert charge separation to debye and multiply the electric field by $16.02/(4\pi\epsilon_0)$ to convert q Å^{-2} to N/C.

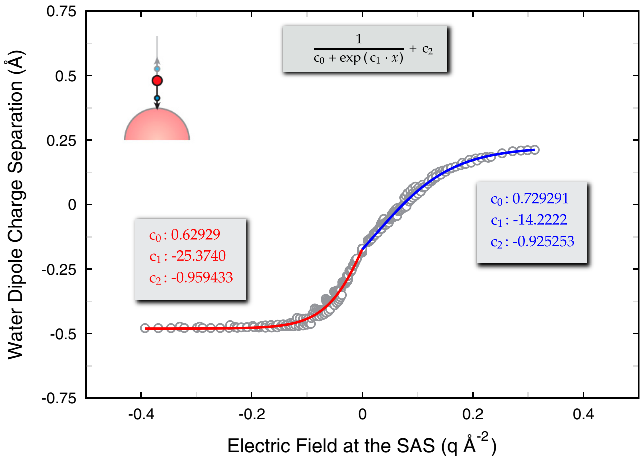

Fig. S4. Sigmoidal curve fits to the positive and negative field regions of the inner dipole charge separation of the quadrupole vs. the electric field at the SAS plots.

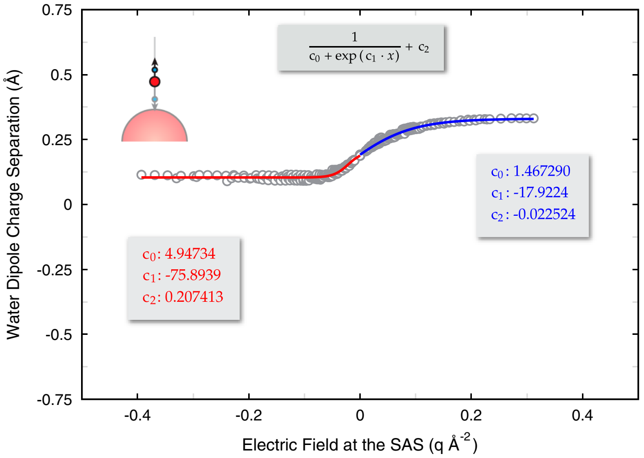

Fig. S5. Sigmoidal curve fits to the positive and negative field regions of the inner dipole charge separation of the quadrupole vs. the electric field at the SAS plots.

Table S1. Extracted r_w values from Lennard–Jones (LJ) sphere presimulations

ϵ_{LJ}, kcal/mol	σ_{LJ}, Å					
	1.4	2.2	3.0	3.8	4.6	5.4
−1						
0.015625	1.74	2.04	2.39	2.91	3.41	3.88
0.03125	1.60	2.08	2.46	2.98	3.47	3.94
0.0625	1.65	2.14	2.56	3.05	3.54	4.01
0.125	1.70	2.19	2.65	3.13	3.61	4.08
0.25	1.76	2.26	2.73	3.21	3.68	4.14
0.5	1.83	2.33	2.82	3.29	3.76	4.21
−0.8						
0.015625	1.65	2.01	2.49	3.00	3.50	3.97
0.03125	1.66	2.10	2.56	3.07	3.57	4.03
0.0625	1.71	2.18	2.64	3.15	3.63	4.09
0.125	1.76	2.25	2.72	3.22	3.70	4.15
0.25	1.82	2.33	2.81	3.30	3.77	4.21
0.5	1.88	2.40	2.88	3.37	3.83	4.27
−0.6						
0.015625	1.62	2.07	2.62	3.15	3.63	4.07
0.03125	1.69	2.15	2.69	3.21	3.68	4.12
0.0625	1.79	2.23	2.77	3.28	3.74	4.18
0.125	1.85	2.32	2.84	3.34	3.80	4.23

ϵ_{LJ}, kcal/mol	σ_{LJ}, Å					
	1.4	2.2	3.0	3.8	4.6	5.4
0.25	1.91	2.40	2.92	3.40	3.85	4.28
0.5	1.97	2.48	2.99	3.46	3.91	4.33
−0.4						
0.015625	1.74	2.26	2.82	3.31	3.75	4.16
0.03125	1.80	2.34	2.89	3.37	3.79	4.20
0.0625	1.86	2.42	2.95	3.42	3.84	4.25
0.125	1.93	2.49	3.01	3.46	3.89	4.29
0.25	2.00	2.56	3.06	3.51	3.93	4.34
0.5	2.08	2.63	3.12	3.56	3.97	4.38
−0.2						
0.015625	2.00	2.58	3.04	3.44	3.83	4.22
0.03125	2.08	2.63	3.08	3.48	3.87	4.26
0.0625	2.15	2.68	3.11	3.51	3.91	4.30
0.125	2.21	2.72	3.15	3.55	3.94	4.34
0.25	2.27	2.76	3.18	3.58	3.98	4.37
0.5	2.32	2.79	3.21	3.61	4.01	4.41
0						
0.015625	2.37	2.74	3.11	3.49	3.86	4.24
0.03125	2.39	2.76	3.14	3.52	3.90	4.28
0.0625	2.41	2.79	3.17	3.55	3.93	4.31
0.125	2.43	2.81	3.19	3.58	3.96	4.35
0.25	2.44	2.83	3.22	3.61	3.99	4.38
0.5	2.46	2.85	3.24	3.63	4.02	4.42
+0.2						
0.015625	2.25	2.67	3.08	3.46	3.84	4.22
0.03125	2.29	2.71	3.10	3.50	3.88	4.26
0.0625	2.32	2.74	3.14	3.53	3.92	4.30
0.125	2.35	2.77	3.17	3.56	3.95	4.34
0.25	2.38	2.80	3.20	3.59	3.99	4.38
0.5	2.41	2.83	3.22	3.62	4.02	4.41
+0.4						
0.015625	2.04	2.52	2.98	3.40	3.80	4.19
0.03125	2.09	2.57	3.02	3.44	3.84	4.23
0.0625	2.15	2.63	3.06	3.48	3.88	4.27
0.125	2.20	2.67	3.11	3.52	3.92	4.32
0.25	2.26	2.72	3.15	3.56	3.96	4.36
0.5	2.30	2.76	3.18	3.59	4.00	4.39
+0.6						
0.015625	1.93	2.35	2.84	3.30	3.72	4.13
0.03125	1.97	2.42	2.90	3.35	3.78	4.19
0.0625	2.02	2.49	2.96	3.40	3.83	4.23
0.125	2.07	2.55	3.02	3.46	3.88	4.28
0.25	2.13	2.61	3.07	3.51	3.92	4.33
0.5	2.19	2.67	3.12	3.55	3.96	4.37
+0.8						
0.015625	1.85	2.23	2.70	3.18	3.64	4.06
0.03125	1.89	2.31	2.77	3.25	3.70	4.12
0.0625	1.93	2.38	2.85	3.32	3.76	4.18
0.125	1.98	2.45	2.92	3.38	3.82	4.24
0.25	2.03	2.51	2.99	3.44	3.87	4.29
0.5	2.10	2.58	3.05	3.50	3.92	4.34
+1						
0.015625	1.79	2.19	2.57	3.06	3.54	3.99
0.03125	1.83	2.25	2.65	3.14	3.61	4.06
0.0625	1.87	2.31	2.74	3.22	3.68	4.12
0.125	1.92	2.37	2.82	3.29	3.75	4.18
0.25	1.99	2.43	2.90	3.37	3.82	4.24
0.5	2.09	2.50	2.97	3.44	3.88	4.30

Table S2. Extracted solvent–solvent spacing distances from the LJ sphere presimulations

ϵ_{LJ}, kcal/mol	σ_{LJ}, Å					
	1.4	2.2	3.0	3.8	4.6	5.4
−1						
0.015625	2.50	2.83	3.01	3.07	2.78	2.76
0.03125	2.55	2.78	3.04	3.10	2.77	2.77
0.0625	2.62	2.84	3.11	3.10	2.77	2.78
0.125	2.70	2.90	3.07	3.19	2.79	2.78
0.25	2.77	2.96	3.09	3.22	2.79	2.77
0.5	2.85	3.01	3.07	3.23	2.79	2.79
−0.8						
0.015625	2.59	2.98	3.08	2.78	2.77	2.76
0.03125	2.64	2.94	3.10	2.81	2.78	2.76
0.0625	2.70	2.91	3.18	2.81	2.79	2.77
0.125	2.77	2.98	3.20	2.83	2.79	2.77
0.25	2.85	3.03	3.28	2.82	2.78	2.78
0.5	2.92	3.07	3.24	2.80	2.79	2.80
−0.6						
0.015625	2.71	3.09	3.17	2.78	2.77	2.76
0.03125	2.80	3.16	3.30	2.78	2.79	2.76
0.0625	2.83	3.19	3.27	2.79	2.78	2.77
0.125	2.89	3.21	3.30	2.79	2.78	2.78
0.25	2.95	3.24	3.25	2.80	2.77	2.77
0.5	3.01	3.17	3.34	2.79	2.77	2.77
−0.4						
0.015625	2.88	3.34	2.80	2.79	2.76	2.75
0.03125	2.96	3.30	2.80	2.78	2.77	2.76
0.0625	3.04	3.40	2.80	2.79	2.77	2.76
0.125	3.13	3.38	2.80	2.78	2.77	2.76
0.25	3.16	3.37	2.82	2.79	2.76	2.76
0.5	3.21	3.36	2.80	2.77	2.79	2.77
−0.2						
0.015625	3.42	2.78	2.77	2.75	2.75	2.75
0.03125	3.37	2.80	2.78	2.76	2.76	2.74
0.0625	3.21	2.77	2.77	2.77	2.76	2.75
0.125	3.10	2.79	2.79	2.76	2.75	2.75
0.25	3.03	2.77	2.77	2.76	2.75	2.76
0.5	2.86	2.79	2.76	2.76	2.76	2.76
0						
0.015625	2.76	2.75	2.74	2.75	2.73	2.75
0.03125	2.75	2.75	2.75	2.75	2.75	2.75
0.0625	2.78	2.75	2.75	2.76	2.75	2.75
0.125	2.78	2.76	2.76	2.75	2.75	2.75
0.25	2.76	2.75	2.76	2.75	2.75	2.76
0.5	2.77	2.76	2.77	2.76	2.76	2.75
+0.2						
0.015625	2.74	2.74	2.74	2.73	2.74	2.74
0.03125	2.74	2.73	2.73	2.74	2.74	2.73
0.0625	2.75	2.74	2.74	2.74	2.75	2.75
0.125	2.76	2.74	2.74	2.74	2.74	2.75
0.25	2.73	2.76	2.74	2.75	2.74	2.74
0.5	2.72	2.76	2.73	2.74	2.74	2.76
+0.4						
0.015625	3.13	2.77	2.74	2.73	2.73	2.74
0.03125	3.16	2.74	2.74	2.75	2.73	2.73
0.0625	3.16	2.74	2.74	2.74	2.73	2.73
0.125	3.21	2.75	2.73	2.73	2.73	2.74
0.25	3.30	2.73	2.76	2.74	2.74	2.74
0.5	3.24	2.76	2.74	2.73	2.74	2.74
+0.6						
0.015625	2.97	3.17	2.74	2.73	2.73	2.73
0.03125	3.00	3.16	2.74	2.73	2.73	2.73
0.0625	3.04	3.18	2.74	2.74	2.73	2.73
0.125	3.08	3.05	2.74	2.74	2.73	2.73
0.25	3.13	2.76	2.74	2.73	2.74	2.73
0.5	3.10	2.79	2.74	2.74	2.74	2.73

ϵ_{LJ}, kcal/mol	σ_{LJ}, Å					
	1.4	2.2	3.0	3.8	4.6	5.4
+0.8						
0.015625	2.87	3.04	3.15	2.74	2.74	2.73
0.03125	2.91	3.06	2.80	2.74	2.73	2.73
0.0625	2.96	3.09	2.77	2.73	2.73	2.73
0.125	3.01	3.10	2.77	2.74	2.73	2.73
0.25	3.04	3.13	2.74	2.74	2.73	2.73
0.5	3.04	3.14	2.74	2.74	2.73	2.74
+1						
0.015625	2.80	2.91	3.10	2.76	2.74	2.72
0.03125	2.85	2.95	3.12	2.76	2.73	2.74
0.0625	2.90	3.00	3.11	2.74	2.73	2.73
0.125	2.94	3.04	3.14	2.76	2.73	2.74
0.25	2.97	3.07	3.09	2.74	2.74	2.73
0.5	2.87	3.09	3.10	2.73	2.73	2.73