

Continuum Electrostatic Analysis of Preferred Solvation Sites Around Proteins in Solution

Sheldon Dennis, Carlos J. Camacho, and Sandor Vajda*

Department of Biomedical Engineering, Boston University, Boston, Massachusetts

ABSTRACT To understand water-protein interactions in solution, the electrostatic field is calculated by solving the Poisson-Boltzmann equation, and the free energy surface of water is mapped by translating and rotating an explicit water molecule around the protein. The calculation is applied to T4 lysozyme with data available on the conservation of solvent binding sites in 18 crystallographically independent molecules. The free energy maps around the ordered water sites provide information on the relationship between water positions in crystal structure and in solution. Results show that almost all conserved sites and the majority of nonconserved sites are within 1.3 Å of local free energy minima. This finding is in sharp contrast to the behavior of randomly placed water molecules in the boundary layer, which, on the average, must travel more than 3 Å to the nearest free energy minimum. Thus, the solvation sites are at least partially determined by protein-water interactions rather than by crystal packing alone. The characteristic water residence times, obtained from the free energies at the local minima, are in good agreement with nuclear magnetic resonance experiments. Only about half of the potential sites show up as ordered water in the 1.7 Å resolution X-ray structure. Crystal packing interactions can stabilize weak or mobile potential sites (in fact, some ordered water positions are not close to free energy minima) or can prevent water from occupying certain sites. Apart from a few buried water molecules that are strong binders, the free energies are not very different for conserved and nonconserved sites. We show that conservation of a water site between two crystals occurs if the positions of protein atoms, primarily contributing to the free energy at the local minimum, do not substantially change from one structure to the other. This requirement can be correlated with the nature of the side chain contacting the water molecule in the site. *Proteins* 2000;38:176–188. © 2000 Wiley-Liss, Inc.

Key words: solvation free energy; Poisson-Boltzmann electrostatics; T4 lysozyme; water residence time

INTRODUCTION

The structural and functional importance of water associated with proteins is well known.^{1–3} In particular, dehydrated enzymes are not active, but a single water layer restores activity.² Conversely, in receptor-ligand interac-

tions the water shell must be removed, and thus the associating molecules must compete with the bound water.⁴ Thermodynamic and kinetic studies of this competition may elucidate important aspects of molecular recognition. The predominant experimental techniques that provide information on the structure of the solvent around macromolecules are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy.^{1,2} In X-ray crystallography, regions of extra electron density are interpreted as ordered water sites, resulting in about 200 water molecules in a typical protein structure³ with a resolution better than 1.5 Å. NMR data reveal that most of the molecules in the first water layer, including those that appear to be fixed in the crystal, are in rapid motion in solution, with exchange times of less than 500 ps.^{5,6} Exceptions are a few buried waters that may have residence times of up to 0.01 s.

Although X-ray crystallography provides a wealth of structural data, it leaves open many important questions.^{1–3} What is the origin of ordered water sites? Why are some of these sites conserved across different crystal forms of the same protein, whereas others are not? What is the relationship between water positions in crystal structures and in solution? How strongly does water interact with the protein? These questions are not trivial to answer by using structural information alone, as the sites range from highly localized ones in which a water molecule is virtually always present (although possibly rapidly exchanging) to sites that show only a weak preference for water.¹ In addition, the distribution of ordered water can be affected by crystal packing interactions.^{7,8}

A number of attempts have been made to investigate the above questions by using molecular dynamics simulations.^{9–17} These results show that water molecules near the protein surface remain very mobile, with a diffusion coefficient decreased two- to fourfold relative to that of bulk water.¹⁷ However, the simulations differ in their evaluation of the influence of apolar, polar, and charged surface atoms on the mobility of the surrounding solvent.^{1,12} In addition, the residence times obtained are generally shorter than those suggested by NMR, with a

Grant sponsor: National Science Foundation; Grant number: DBI-9630188; Grant sponsor: Department of Energy; Grant number: DE-F602–96ER62263.

*Correspondence to: Sandor Vajda, Department of Biomedical Engineering, Boston University, 44 Cummings Street, Boston, MA 02215. E-mail: vajda@enga.bu.edu

Received 14 April 1999; Accepted 14 September 1999

rather weak correlation between calculated and experimental residence times for individual water molecules.

On the basis of experimental data and simulations, Levitt and Park³ asserted that at an ordered water site the potential of mean force must have a local minimum. Motivated by the observations of Kuhn et al.,¹⁸ they further argued that water molecules only prefer hydrophilic groups that are in crevices on the protein surface, and ordered water sites anywhere else “are essentially an accident of the geometry of the crystal lattice.” This statement was supported by the observation that surface water positions are poorly conserved across different crystal forms of a protein, e.g., three crystal forms of the pancreatic trypsin inhibitor showed only six of the 60 water sites conserved.⁵ We note that this last argument lost some of its power, because recent studies report more extensive conservation of solvation sites.^{19,20}

In this study, we construct free energy landscapes for water molecules interacting with a T4 lysozyme in solution. The calculation is based on the classic continuum electrostatics (CE) model^{21–23} that has been applied to virtually every problem in macromolecular electrostatics. In the CE model the protein is represented by a low dielectric region containing discrete atomic charges at fixed positions, surrounded by a high dielectric medium representing the solvent. We calculate the electrostatic field of the solvated protein by using a finite difference Poisson-Boltzmann (FDPB) method^{21–23} and then map the free energy surface by translating and rotating an explicit water molecule in the precalculated field.^{24,25} The advantage of using continuum electrostatics instead of molecular dynamics is that we can directly calculate the potential of mean force at any water position rather than following and averaging the movement of many individual water molecules. Although this advantage comes at the price of some approximations, we will be able to demonstrate the validity of underlying assumptions and estimate potential errors.

The free energy maps around the ordered water sites provide substantial information on the energetics of solvation. In particular, we will show that almost all conserved sites and a large fraction of nonconserved sites are close to local free energy minima. This is in sharp contrast to the behavior of water molecules randomly placed in the first layer around the protein. Such random water molecules, on the average, must travel more than 3 Å to find a local free energy minimum. The difference between random and ordered water suggests that, in a given crystal, potential solvation sites are at least partially determined by protein-water interactions rather than by crystal packing alone. Apart from a few buried water molecules, these interactions are not very different for conserved and nonconserved sites. As we will show, conservation of a water site between two crystals occurs if the positions of protein atoms primarily contributing to the free energy at the local minimum, do not substantially change from one structure to the other. This requirement can be correlated with the nature of the side chain contacting the water molecule in the site.

MATERIALS AND METHODS

Free Energy Calculation

Consider first a fully solvated rigid protein that serves as the reference state in our calculations. As we mentioned, the system is described by the two-dielectric continuum model.^{21,22} The electrostatic field Φ of the solvated protein is calculated by solving the linear Poisson-Boltzmann equation. We use a finite difference (FD) method as implemented in CONGEN, which features adjustable rectangular grids, a uniform charging scheme that decreases the unfavorable grid energies, and smoothing algorithms that alleviate problems associated with discretization.²³ The calculations were carried out by using a 0.8 Å grid, with uniform charging, anti-aliasing, and 15-point harmonic smoothing. A 8 Å grid margin was maintained around the molecule. The dielectric constants of the protein and the solvent were set to 2 and 78, respectively, and the ionic strength was 0.05 M.

To calculate the potential of the mean force due to the solvated protein at a water site, we place an explicit water represented by the TIP3 model²⁶ at that position. Both the protein and the water molecule are regarded as rigid. The electrostatic free energy of the water molecule in the electrostatic field Φ is calculated by the expression

$$\Delta G_{\text{el}} = \sum_i \Phi_i q_i \quad (1)$$

where q_i is the charge of the i th water atom, and Φ_i is the field at that point. We use the TIP3 model;²⁶ thus, the summation in Equation 1 is over the three atom-centered charges. Because surface waters are quite mobile,⁵ indicating a high dielectric constant, we set $\epsilon = 78$ over the entire solvent region. As we will discuss, our results support this assumption.

By translating and rotating a “probe” water molecule in the electrostatic field Φ we will map the free energy surface in the vicinity of each ordered water site. Because our goal is to find local free energy minima, the electrostatic free energy defined by Equation 1 will also be subjected to minimization in the translational/rotational space of the probe. A natural constraint during minimization is avoiding steric overlaps with the protein. To simplify this constraint, in the excluded volume calculation the water is represented by a sphere of radius 1.4 Å, centered at the oxygen atom, whereas the value and arrangement of charges follow the TIP3 model.²⁶ The constrained minimization problem is solved by a penalty function method,²⁷ i.e., by the unconstrained minimization of the extended target function

$$Q = \Delta G_{\text{el}} + CV_{\text{exc}} \quad (2)$$

where V_{exc} is an excluded volume penalty function defined by $V_{\text{exc}} = \sum_j (d_j - D_j)^2$ for $d_j < D_j$ and $V_{\text{exc}} = 0$ otherwise. In this definition, d_j denotes the distance between the center of the water and the j th water atom, and D_j is the nominal distance between the same atoms, i.e., $D_j = r_j + r_w$, where r_j and r_w are the van der Waals radii of the j th

protein atom and the water, respectively. In penalty function methods,²⁷ the weighting coefficient C is frequently increased during the minimization to remove any potential violation of the constraints. However, in this particular application the constant value of $C = 10^5$ yields so small steric overlaps after convergence that the ΔG_{el} term is almost unaffected, and hence we use this simpler algorithm. The extended potential given by Equation 2 will be used both in the minimization and in the construction of free energy maps; hence, any steric overlap will show up as a sudden increase in the free energy.

As will be discussed, we also want to identify the amino acid residues that contribute the most to the binding of a water molecule at a given location. It is possible to decompose the free energy into residue contributions by using the Poisson-Boltzmann formalism, but the method is either of limited accuracy or computationally expensive. One procedure includes solving the Poisson-Boltzmann equation with the water molecule charged and everything else neutral, reading off the electrostatic field Φ at the residues of interest, and then calculating the free energy by summing over the atoms of each separate residue in Equation 1. The controversial feature of this approach is that we sum the interactions between the charges of the protein and the electrostatic field of a single water molecule, whereas ΔG_{el} has been calculated the other way around, i.e., by summing the interactions between water charges and the electrostatic field Φ of the solvated protein. In principle the two expressions should yield the same free energy, but due to the finite difference approximation deviations may occur.

The accuracy can be improved by a more elaborate procedure. To calculate the contribution of the k th residue to the free energy, we set all atoms of this residue neutral and solve the Poisson-Boltzmann equation to determine the electrostatic field. If ΔG_k denotes the free energy obtained by Equation 1 for this modified field (i.e., with the k th residue neutral), then the difference $\Delta G_{el} - \Delta G_k$ is the contribution of the k th residue to the binding free energy. The problem with this approach is that, for each water position, it requires solving the Poisson-Boltzmann equation $n + 1$ times, where n denotes the number of amino acid residues in the protein. Because the residue contributions are needed only for selecting a “contact” (i.e., most important) residue in each water site, this approach is too expensive, and we use Coulombic calculations with the distance-dependent dielectric $\epsilon = 4r$. The calculated interaction energies are summed separately for side chain and backbone atoms of each residue.

Mapping the Free Energy Surface

Our primary goal is to study whether the observed crystallographic water sites are close to local minima of the free energy calculated for the solvated protein. In principle, this problem can be solved by performing a local minimization of the free energy defined by Equation 2 in the six-dimensional space of water translations and rotations by using each observed water position as a starting point. However, the outcome of local minimization on a

multimodal surface (i.e., with multiple minima) is highly sensitive to small changes in the starting state, even in the orientation of water around its center. Results also depend on the minimization algorithm that may converge to a distant local minimum while ignoring others closer to the starting point.

In view of the above uncertainty, we mapped the free energy surface by moving a “probe” water molecule in the vicinity of each ordered site. The location and orientation of this probe relative to the fixed protein and the selected crystallographic water molecule is specified by six coordinates. The location of the probe is given by the radius r and the Euler angles ϕ and θ in a spherical coordinate system centered at the crystallographic water position. Furthermore, three Euler angles, ϕ_w , θ_w , and ψ_w specify the orientation of the probe.

To find out how far a water molecule can move away from its crystallographic position without an increase in the free energy, we consider the grid points (ϕ_i, θ_i) in the subspace of the Euler angles ϕ and θ , where $\phi_i = 0^\circ, 18^\circ, \dots, 342^\circ$, and $\theta_i = 0^\circ, 18^\circ, \dots, 180^\circ$. Notice that each grid point (ϕ_i, θ_i) defines a line through the crystallographic position, and our goal is to find the local free energy minimum along this line by performing a minimization in terms of the remaining variables r, ϕ_w, θ_w , and ψ_w by using Powell’s conjugate gradient algorithm.²⁸ Because we want to find the minimum closest to the crystallographic position, i.e., the smallest displacement r at which a local minimum occurs, it is obvious to select $r = 0$ as the initial displacement. We also need starting values for the variables ϕ_w, θ_w , and ψ_w .

Although the number of variables has been reduced, the minimization problem is still nonlinear with many potential local minima. For improved robustness of the search, we use a version of the Powell algorithm that includes line minimization based on Brent’s method.²⁸ Nevertheless, for some initial values of $(\phi_w, \theta_w, \psi_w)$ the algorithm may miss the closest local minimum (i.e., the one with the smallest displacement r), and converge to a more distant one. In other cases, the search may terminate early, e.g., due to being restricted to a subspace of the four-dimensional search space. These artifacts can be easily identified by repeating the minimization for each (ϕ_i, θ_i) pair 30 times by using different random orientations of the probe as the starting state. Because early termination or “jumping” over a local minimum along the line defined by (ϕ_i, θ_i) are relatively rare, the majority of the 30 runs yield very similar displacement values, well distinguishable from the few runs that end up in substantially different points. Therefore, for each (ϕ_i, θ_i) we cluster the results of the 30 runs along the line (usually a trivial task), consider the largest cluster, and choose the lowest value of the target function Q within this cluster as the free energy minimum along the line defined by the Euler angles (ϕ_i, θ_i) .

T4 Lysozyme and Conservation of Water-Binding Sites

We study the solvation of T4 lysozyme by using the 1.7 Å resolution X-ray structure (PDB code 4lzm) that includes

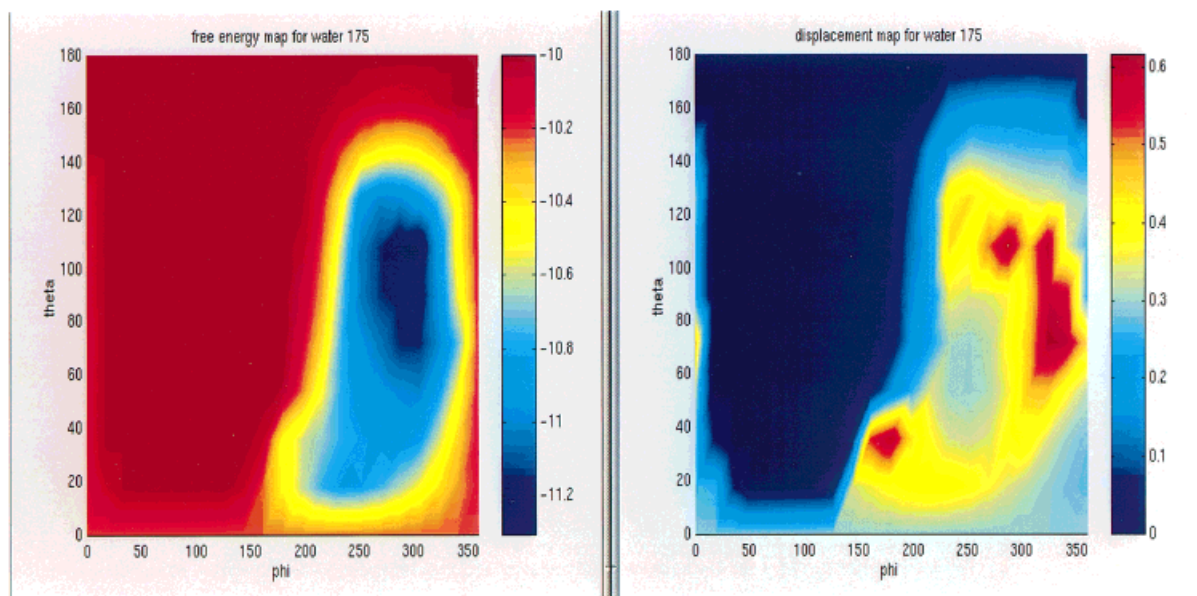


Fig. 1. Characterization of the free energy pocket around crystallographic water site 175 in the T4 lysozyme structure 4lzm. **A:** Map of the local free energy minima (kcal/mol) along the lines defined by the Euler angles ϕ and θ . **B:** Displacement (\AA) from the crystallographic position along the lines defined by the Euler angles ϕ and θ .

137 ordered water sites. The analysis of T4 lysozyme is particularly interesting due to the available information on the conservation of water sites in 18 crystallographically independent molecules, including the wild-type 4lzm and nine mutants.¹⁹ Although the mutations affect the structure in the immediate vicinity of the amino acid substitutions, the rest of the protein remains essentially unchanged. In some cases, some “hinge-bending” motion occurs, but it can be corrected by bringing the rigid body fragments into a reference frame.¹⁹ The overlap of the 18 molecules results in a total of 1,675 water molecules. Zhang and Matthews¹⁹ clustered these waters by taking each water in turn and counting how many other water molecules occurred within 1.2 \AA . The analysis of clusters provided information on the conservation of solvation sites. There were 40 sites occupied in at least 7 structures, with 17 being the largest number of occupancies of a single site. In 4lzm, only 36 of these sites are occupied, because three sites are empty, and one (number 173) contains a chloride anion rather than a water. The conservation of water sites among the 18 molecules is higher if accounting for cases in which the site cannot contain a water molecule due to crystal packing interactions.¹⁹

The X-ray structure of the lysozyme, with the ordered water molecules included, has been refined by 200 steps of energy minimization before free energy calculations. We used version 19 of the Charmm potential²⁹ with polar hydrogens, and 20 kcal/mol/ \AA^2 harmonic constraints on the positions of nonhydrogen atoms. These calculations placed the polar hydrogens and created a plausible hydrogen-bonding network between protein and water. The root-mean-square shift of the water molecules due to the minimization was 0.1 \AA .

RESULTS

Free Energy Maps Around Ordered Water Sites

The primary results of the calculations are the free energy and displacement maps in the vicinity of each ordered water site. Figure 1a and b show these maps for water site No. 175 in 4lzm. The free energy map in Figure 1a shows the minimal free energy as a function of the Euler angles ϕ and θ . The point (ϕ_i, θ_i) is color-coded according to the lowest free energy attained along the line defined by (ϕ_i, θ_i) as we move outward from the center of the local coordinate system, with blue and red representing the low and the high free energies, respectively (see sidebar in Fig. 1a). On the displacement map in Figure 1b, the point (ϕ_i, θ_i) is color-coded according to the value of the displacement r at which the minimal free energy is attained along the line defined by (ϕ_i, θ_i) . We recall that this displacement r is the distance between the starting point and the local free energy minimum. In the color code dark blue represents the smallest (generally 0 \AA) displacement, and the color is changing to red as the displacement increases. Notice that for each (ϕ_i, θ_i) we retain only the lowest free energy along the line and the displacement at which this minimum is attained, irrespective of the corresponding rotational state $(\phi_w, \theta_w, \psi_w)$ of the ligand.

The two maps together clearly show if there is a free energy pocket in the vicinity of the crystallographic water site and indicate the extent of the low free energy region. As represented by the large blue spot on the free energy map in Figure 1a, the free energy pocket around water site 175 is dominated by a deep local minimum at $\phi = 300^\circ$ and $\theta = 110^\circ$, where $Q = -11.31$ kcal/mol. The pocket includes a second, much shallower local minimum at $\phi = 245^\circ$ and

TABLE I. Minimum Free Energy and Displacement Values for the Conserved Water Sites in the T4 Lysozyme Structure 4lzm[†]

Site	Solvent	Interference	Displacement (Å)	Minimum free energy (kcal/mol)	Contact residue	RMSD (Å)
171	17	0	1.05	-8.40	ASP 70	0.59
174	8	2	1.17	-1.60	Backbone	
175	15	0	0.44	-11.31	THR 152	0.43
176	8	1	0.57	-12.83	ASP 47	1.28
177	9	5	0.01	-2.42	Backbone	
179	17	0	0.81	-9.05	Backbone	
180	11	6	1.30	-2.28	Backbone	
182	9	0	1.52	-9.08	ASP 70	0.59
186	7	3	0.48	-4.90	ASN 68	1.42
190	12	0	0.13	-5.20	Backbone	
193	8	2	1.03	-1.39	ASN 144	1.07
195	9	7	0.82	-1.03	CYS 97	1.06
196	8	1	0.28	-4.91	ARG 52	1.37
197	10	7	0.70	-3.76	SER 117	0.73
200	9	2	1.29	-3.09	HIS 31	0.55
201	8	3	0.35	-1.88	ARG 76	1.57
203	9	2	0.63	-7.41	ASP 20	1.65
206	9	2	0.75	-2.46	Backbone	
208	15	0	0.86	-5.05	ASN 101	0.54
210	7	1	1.28	-5.61	ASP 92	1.02
211	9	3	0.65	-5.08	ARG 148	0.47
216	7	0	0.58	-2.53	Backbone	
223	8	8	1.65	-2.64	GLU 5	1.53
231	10	4	0.59	-1.07	Backbone	
232	10	3	1.33	-2.29	ARG 148	0.47
237	8	4	0.66	-0.85	ARG 145	0.63
241	11	5	0.88	-4.03	ASP 72	1.00
245	8	7	1.66	-4.75	ARG 148	0.47
248	9	7	0.00	-0.20	TRP 158	0.70
254	8	1	0.56	-2.09	THR 109	1.28
266	9	2	2.51	-5.92	ASP 70	0.59
274	11	0	0.98	-2.24	ASP 70	0.59
282	8	1	1.00	-1.68	Backbone	
296	10	4	2.17	-1.30	TRP 158	0.70
299	8	1	1.19	-2.13	Backbone	
311	8	7	0.77	-4.36	GLU 128	1.22

[†] Solvent denotes the number of times the site is occupied by a water molecule; interference shows the number of cases in which the site is either occupied by a polar atom from a neighboring protein or a neighboring protein may interfere with the site; displacement is the distance to the closest free energy minimum; contact residue identifies the group of atoms (backbone or side chain) that contribute most to the free energy; RMSD is the root-mean-square deviation of the side chain in 10 T4 lysozyme structures.

$\theta = 16^\circ$. According to the displacement map in Figure 1b, along any direction the probe can travel at most 0.6 Å before the free energy starts to increase. In most directions, however, the displacement is much smaller. For example, along the line defined by $\phi = 300^\circ$ and $\theta = 110^\circ$ and leading to the free energy minimum, the displacement is only 0.45 Å. Furthermore, the displacement is close to zero for most directions with $\phi < 150^\circ$ as shown by the blue region on the left half of Figure 1b. Such sudden increase in the free energy generally occurs due to steric overlaps and indicates that the site is close to the protein surface. For other sites, the region of the low free energy, color-coded in blue on the energy map, may be larger or smaller. The pocket at site 175 is somewhat atypical in the sense that, apart from a minor drop in the free energy, the pocket is actually a single minimum, whereas for most sites the

region of low free energy includes three to six distinguishable local minima.

Table I is the summary of results for the 36 conserved water sites, where each site is identified by the water number in the PDB file 4lzm. Based on Table 3 of Zhang and Matthews,¹⁹ under the heading Solvent, we also show the number of T4 lysozyme molecules among the 18 that have a water molecule in the given site. The second integer (Interference) shows the number of cases in which the site is either occupied by a polar atom from a neighboring protein in the crystal lattice, or a neighboring molecule is so close to the empty site that there may be a steric interference preventing its occupation by a water molecule. The next two columns show the minimum free energy and the displacement required to reach it, all extracted from the individual free energy and displacement maps.

The last two columns of Table I provide information on the “contact” residues that contribute most to the binding. As described in the Materials and Methods section, we used Coulombic electrostatics and summed the calculated interaction energies for side chain and backbone atoms of each residue. For more than 95% of the water molecules, the interactions were dominated either by the backbone or by a single side chain shown as the “contact” residue in Table I. Notice that Coulombic expressions were used only to decide which residues contact the water, and the free energy minimum shown in Table I is from the Poisson-Boltzmann calculation. As we will discuss, water molecules contacting a side chain tend to be conserved among various crystal forms if the “contact” side chain conformation does not substantially change across different X-ray structures. The last column of Table I shows the mean root-mean-square deviation (RMSD) of the “contact” side chain, calculated by locally overlapping the backbones in 10 T4 lysozyme structures.¹⁹

Table II shows the minimum free energies and corresponding displacement values for the nonconserved water sites in 4lzm. Although there are 103 nonconserved sites in the structure, we have results only for 96 of them. In seven cases (water sites 205, 209, 220, 233, 234, 255, and 260), the final free energy showed sudden changes at some points of the (ϕ, θ) plain. Repeated minimization starting from further random orientations of the probe were unable to remove the discontinuities in the free energy values. Because these structures are not critical for our analysis, we did not further pursue the construction of their free energy maps. Figure 2 compares the conserved and nonconserved water sites in terms of free energy minima and displacement values.

As we described, the column Solvent in Table I shows how many times a specific site is occupied by a water molecule in the 18 lysozyme structures. This number has been obtained by overlapping the structures and counting how many water molecules occur within a radius of 1.2 Å.¹⁹ Accounting for the 0.1 Å average shift of water molecules during structure refinement, we thus expect that the water in well-defined sites will be within 1.3 Å of the closest local minimum. Table III shows the classification of conserved and nonconserved sites according to this criterion.

Free Energy Maps for Random Water Sites

As shown by Table III, the majority of both conserved and nonconserved water sites are close to local free energy minima. This may indicate that the potential ordered water sites in a given structure are at least partially determined by protein-solvent interactions. However, there is an alternative hypothesis, which maintains that the free energy surface has a large number of densely distributed local minima. To choose between the two hypotheses, we map the free energy surface around 100 water molecules that have been randomly placed in the first water layer around the protein. Figure 3 compares these random sites with the nonconserved ones in terms of free energy minima and displacement values. Table IV shows the means and

standard deviations of minimum free energies and displacement values for conserved, nonconserved, and random water sites.

DISCUSSION

Ordered Water and Free Energy Minima

As mentioned in the Introduction section, Levitt and Park³ suggested that the ordered water sites should be local free energy minima and that apart from hydrophilic sites in crevices, these “are essentially an accident of the geometry of the crystal lattice.” Although the protein conformation we study in this study is from the X-ray structure of the T4 lysozyme, the simulations are for a single solvated protein rather than for the crystalline form. Because the crystal contacts have been removed, we expected that only some of the ordered sites would remain close to local minima and most likely more so for conserved than for nonconserved water.

We found that the majority of both conserved and nonconserved water sites are less than 1.3 Å from a local free energy minimum (Table III), with mean displacements of 0.90 Å and 1.28 Å, respectively. This finding is in sharp contrast to the behavior of randomly placed water molecules, which on the average, have to move 3.22 Å to the nearest local minimum (Table IV). Application of the *t*-test to the means in Table IV shows that the displacement values for nonconserved and random water molecules significantly differ even at the probability level of $P = 0.001$. Thus, the correlation between ordered water sites and free energy minima cannot be explained simply by a large density of minima on the free energy surface. Therefore, we conclude that the potential positions for most ordered water sites are at least partially determined by electrostatic interactions between the protein and the water.

Our second observation is that, with respect to the free energy, there is little difference between conserved and nonconserved solvation sites. In fact, the differences between the mean values shown in Table IV are not significant at $P = 0.2$ for either the free energy or the displacement. The distributions shown in Figure 2 provide a more detailed comparison. According to Figure 2a, the most noticeable difference between the free energies is that, for the conserved sites, the distribution is slightly shifted toward stronger interactions (more negative free energies). In particular, only three of the 96 nonconserved sites have free energies less than -8 kcal/mol (sites 213, 265, and 278), whereas there are five such examples among the 36 conserved sites (171, 175, 176, 179, and 182). As shown by Zhang and Matthews,¹⁹ all five of these sites have little or no exposure to the bulk solvent. However, a number of conserved sites are relatively weak ($\Delta G_{e1} > -4$ kcal/mol), and apart from large shifts in a few nonconserved sites, the distributions of displacement values are very similar (Fig. 2b).

The main conclusion so far is that both conserved and nonconserved site are much closer to local free energy minima than expected for random locations. To further support this statement, we compared the free energy and

TABLE II. Minimum Free Energy and Displacement Values for the Nonconserved Water Sites in the T4 Lysozyme Structure 4lzm[†]

Site	Displacement (Å)	Minimum free energy (kcal/mol)	Contact residue	RMSD (Å)
172	0.83	-4.58	THR 142	1.00
181	1.29	-2.26	Backbone	
183	0.90	-2.66	LYS 65	0.97
184	0.80	-3.43	ASP 61	1.22
185	1.72	-3.20	ASP 72	1.00
187	0.89	-3.46	GLU 64	1.52
188	1.44	-1.41	Backbone	
189	0.20	-5.33	Backbone	
191	1.35	-4.18	LYS 19	1.62
192	0.77	-2.56	Backbone	
194	0.56	-5.40	ASP 127	1.45
198	1.23	-5.66	GLU 45	1.35
199	0.95	-4.14	Backbone	
202	1.13	-4.49	ASP 70	0.59
204	1.05	-3.73	GLU 11	0.95
207	0.60	-4.34	GLU 11	0.95
212	0.58	-1.03	Backbone	
213	0.91	-11.97	LYS 124	0.92
214	1.79	-3.91	LYS 124	0.92
215	1.15	-4.37	ASN 140	1.38
217	1.64	-3.48	ASP 127	1.45
218	0.90	-2.34	LYS 135	1.76
219	0.32	-6.80	LYS 135	1.76
221	1.28	-3.20	ARG 8	3.05
222	0.54	-3.71	Backbone	
224	0.29	-2.07	THR 26	0.72
225	0.53	-2.21	Backbone	
226	1.30	-4.53	ASP 89	0.72
227	1.44	-1.00	ARG 8	3.05
228	1.54	-4.10	Backbone	
229	0.89	-6.41	GLU 11	0.95
230	1.00	-2.51	ARG 148	0.47
235	0.86	-3.00	ASP 127	1.45
236	2.11	-2.59	GLU 128	1.21
238	1.20	-1.64	ASP 127	1.45
239	1.57	-2.64	Backbone	
240	1.03	-6.92	GLU 64	1.52
242	0.85	-3.25	ARG 80	1.76
243	0.36	-5.35	ARG 8	3.05
244	0.98	-3.04	MET 1	1.01
246	3.33	-3.50	HIS 31	0.55
247	2.58	-1.90	ARG 14	1.59
249	1.27	-1.88	TYR 139	1.22
250	3.0	-1.98	Backbone	
251	0.91	-2.20	LYS 135	1.76
252	0.56	-1.57	Backbone	
253	0.66	-2.79	Backbone	
256	1.32	-6.07	GLU 11	0.95
257	0.93	-5.14	GLU 64	1.52
258	0.32	-2.62	LYS 16	1.64
259	1.09	-6.14	GLU 62	0.86
261	1.13	-3.17	THR 151	0.54
262	0.14	-1.31	ARG 8	3.04
263	1.23	-3.12	Backbone	
264	0.46	-2.52	GLU 62	0.86
265	1.01	-9.44	ASP 70	0.59
267	0.09	-3.92	GLU 11	0.95
268	0.58	-0.74	GLU 108	1.22

TABLE II. (Continued)

Site	Displacement (Å)	Minimum free energy (kcal/mol)	Contact residue	RMSD (Å)
269	0.40	-2.19	THR 21	2.31
270	1.05	-0.60	Backbone	
271	3.92	-4.67	ASP 127	1.45
272	1.25	-1.70	ASN 116	1.09
273	1.40	-2.43	Backbone	
275	0.82	-2.11	Backbone	
276	0.56	-6.64	LYS 83	1.22
277	1.18	-2.15	ARG 95	0.68
278	0.69	-7.47	ASP 47	1.28
279	1.15	-1.84	THR 115	0.91
280	1.21	-3.40	ASP 70	0.59
281	0.63	-5.67	GLU 45	1.35
283	1.12	-2.40	ARG 8	3.05
284	0.81	-1.17	Backbone	
285	0.25	-5.12	TYR 18	1.04
286	0.95	-3.06	Backbone	
287	1.28	-1.74	Backbone	
288	2.36	-2.97	GLU 5	1.54
289	0.45	-11.04	GLU 108	1.22
290	0.59	-1.52	Backbone	
291	1.18	-1.68	LYS 124	0.92
292	0.66	-2.06	ARG 119	1.59
293	0.95	-2.93	GLU 128	1.21
294	1.21	-3.28	GLU 5	1.54
295	9.79	-6.36	GLU 64	1.52
297	2.55	-2.42	Backbone	
298	0.60	-6.49	TYR 18	1.04
300	3.85	-2.19	ASP 70	0.59
301	1.87	-4.72	GLN 122	0.71
302	1.23	-4.51	ASP 89	0.72
303	1.38	-0.58	ARG 137	3.19
304	4.19	-4.70	GLU 11	0.95
305	0.91	-1.89	Backbone	
306	1.01	-7.68	ASP 20	1.65
307	1.49	-2.84	HIS 31	0.55
308	3.32	-4.74	GLU 62	0.87
309	0.70	-1.88	ARG 137	3.19
310	0.84	-0.59	ASN 55	1.28

[†] For explanation of column headings, see footnote to Table I.

displacement distributions for nonconserved and random sites (Fig. 3). As shown by Figure 3a, the distributions of free energy minima are very similar. This finding indicates that, starting from random locations, the water molecules can shift toward sites that are energetically indistinguishable from nonconserved (and most conserved) sites. However, this requires much larger displacements than in the case of nonconserved sites, and the two displacement distributions substantially differ (Fig. 3b).

Our analysis assumes that the “probe” water interacts only with the solvated protein, and does not interact with any of the ordered water molecules that are, up to some degree, localized in the boundary layer. To test the effects of potential interactions with such high residence time water molecules, we performed a number of test calculations by assuming that the “probe” water in one site can interact with ordered water in all the other sites, including

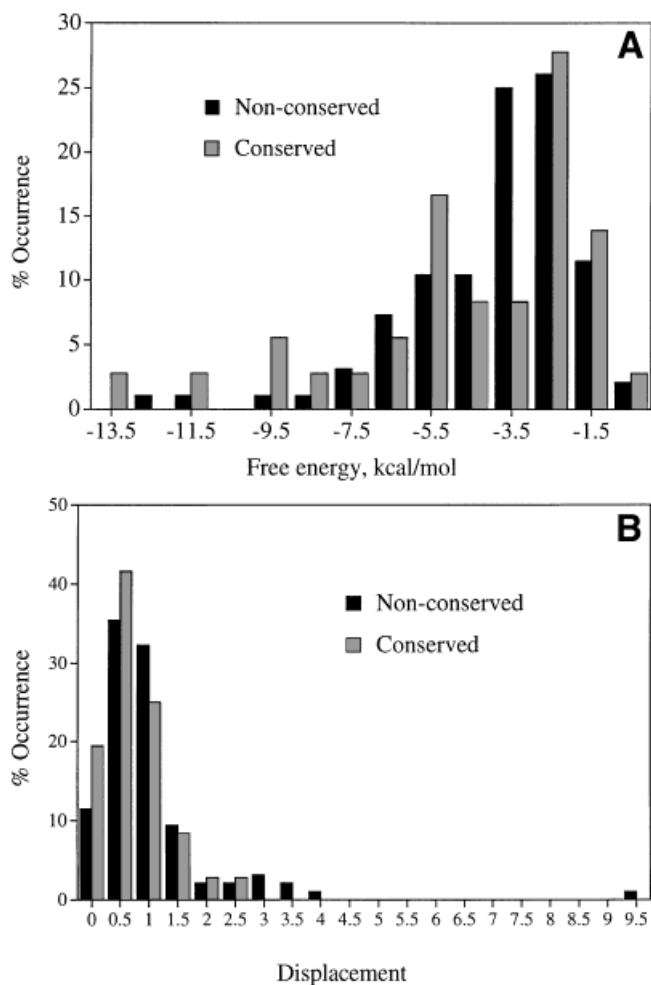


Fig. 2. Comparison of local free energy minima around conserved and nonconserved water sites. **A:** Normalized free energy distribution. **B:** Normalized displacement distribution.

TABLE III. Classification of Conserved and Nonconserved Water Sites in the T4 Lysozyme Structure 4lzm According to the Displacement From the Closest Local Free Energy Minimum[†]

$d < 1.3 \text{ \AA}$		$d > 1.3 \text{ \AA}$
31 (86.1%)	5 (13.9%)	36 conserved sites
72 (75.0%)	24 (25.0%)	96 nonconserved sites

[†] d , displacement from the closest free energy minimum.

electrostatic and steric interactions. Thus, in these tests, we selected an ordered water and regarded all others fixed at their crystallographic positions. We expected that the presence of the immobile water molecules would reduce the displacement of the probe. We have found significant cooperativity in four cases among the 132 water sites. Assuming noncooperativity, four water molecules shift to different crystallographic water sites as follows: 250 to 216, 266 to 265, 297 to 231, and 308 to 259, indicating that these sites are not separated by high-energy barriers.

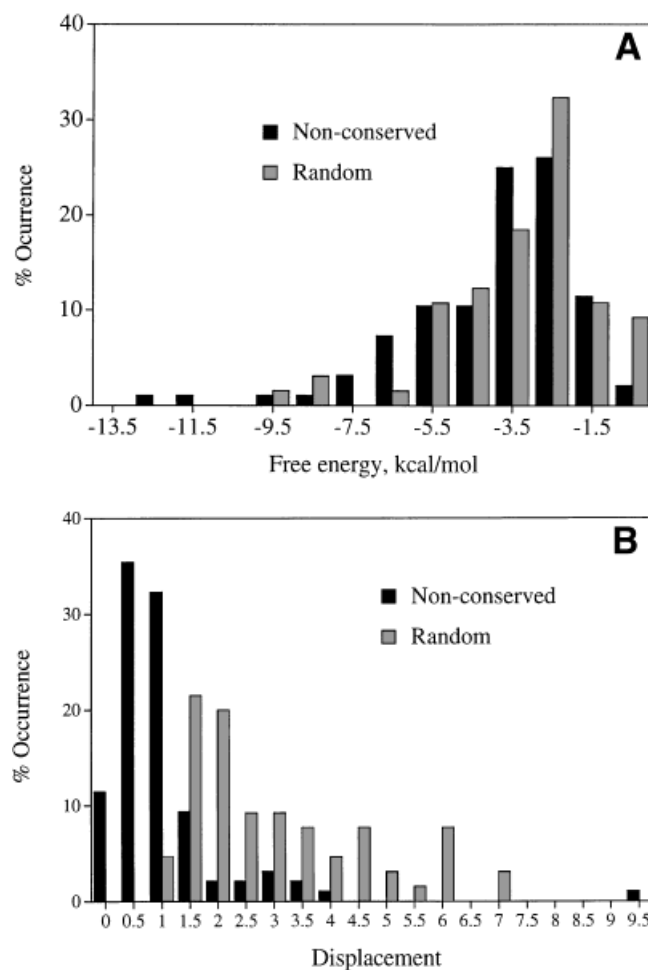


Fig. 3. Comparison of local free energy minima around nonconserved and randomly placed water sites. **A:** Normalized free energy distribution. **B:** Normalized displacement distribution.

TABLE IV. Displacement and Free Energy Statistics of Conserved, Nonconserved, and Random Water Sites[†]

Water sites	Number	Displacement		Free energy	
		Mean	Standard deviation	Mean	Standard deviation
Conserved	36	0.90	0.55	-3.77	3.19
Nonconserved	96	1.28	1.19	-3.38	2.19
Random	100	3.22	1.61	-2.81	1.86

[†] Displacement values in \AA , free energies in kcal/mol.

Notice that the only conserved water for which such shift occurs is 266. In the remaining 128 cases the interactions between the probe in one site and the ordered water in a different site were found to be weak, and the free energy maps were not significantly affected by the assumption of noncooperativity among the ordered water molecules.

Potential and Actual Solvent Binding Sites

We study the local free energy minima found on the free energy maps of in the vicinity of 100 probe positions,

randomly placed in the first water layer. We disregarded eight water sites, because the lowest free energy attained was positive, i.e., no path leading to any free energy pocket was found. In 14 cases, the minimum occurred within 1.3 Å of an ordered water site. In 36 cases, the minimum was farther than 1.3 Å, but the water contacted a side chain that was also in contact with at least one ordered water. In 26 further maps, the water at the free energy minimum contacted a backbone atom. For some of these latter sites, the water-protein interaction was relatively weak, but the mean free energy did not significantly differ from the mean calculated for all sites. In the remaining 16 cases, the waters at the minimal free energy positions contacted side chains that were not seen to contact any ordered water in the X-ray structure. These include four Lys residues, Lys-16, Lys-43 (twice), Lys-60, and Lys-162. Based on the thermal factors, all these side chains are very mobile in the crystal, and water molecules associated with them would not be seen. According to our calculations, water would also contact Asn-2 and Asn-81, both interacting weakly with water, as well as Asp 72 and Asp 92. Because Asp side chains are generally well solvated, we do not fully understand why no ordered water is seen around these two Asp side chains.

It follows from the above analysis that about half of randomly placed water molecules move toward either an ordered water site or toward a side chain that contacts an ordered water in the X-ray structure, and only relatively few (16 of the 92) free energy minima contacting a side chain are far from any of the ordered sites. To confirm these observations, we compared the potential solvation sites obtained from random initial states with the distribution of water contacts in the X-ray structure. According to Tables I and II, the crystallographic water sites tend to contact charged and certain polar side chains. Table V shows the number of these side chains in the T4 lysozyme, as well as the number of those that are “free”, i.e., do not contact any ordered solvation site. The average free energy at the local minima is also shown for each side chain type. In addition, Table V provides information on the total number of water molecules contacting the side chains of each type (some side chains are in contact with up to five water molecules) and the number of water molecules that are in contact with another protein molecule within the crystal lattice.

As shown in the first two columns of Table 5, only about half of the polar/charged side chains (35 of 77) have nearby ordered sites, and this statistic is highly dependent on the type of the side chain. For example, most Arg and Glu side chains contact at least one ordered water, whereas many Asn, Lys, and Tyr side chains are “free.” Although we cannot fully explain this difference, most Asn side chains may be free because they interact with the water relatively weakly, with a free energy of -2.28 kcal/mol (Table V). This assumption is supported by the fact that out of the six water sites contacting an Asn side chain, five are in crystal contacts that are likely to contribute to the stabilization of water at those particular sites. The problem with this observation is that the interaction energies with Arg side

TABLE V. Statistics of Side Chains Contacting Potential and Actual Ordered Water Sites in the T4 Lysozyme Structure 4lzm

Side chain	Side chains			Ordered water sites number	
	Total	Free ^a	Mean free energy (kcal/mol)	Total	Crystal contact
Arg	13	3	-2.86	18	4
Asn	12	8	-2.28	6	5
Asp	10	4	-5.22	23	4
Glu	8	2	-4.60	23	8
Lys	13	7	-4.46	10	3
Tyr	7	5	-5.98	3	0
Thr	11	4	-3.40	7	1
Trp	3	2	-1.33	3	0

^a Number of side chains not contacting any ordered water site.

chains are not much stronger (-2.86 kcal/mol), although these are heavily hydrated. We note that the binding free energies, -2.28 kcal/mol and -2.86 kcal/mol, yield residence times of 226 ps and 594 ps, respectively (see the next section). Although this difference in residence times partially reconciles the apparent contradiction, it will be further discussed at the end of the article. The free energies are larger for Lys and Tyr, comparable to those for Asp and Glu. However, as we will further discuss, Lys and Tyr side chains are frequently unordered in X-ray structures, and the water molecules associated with them may not be clearly visible.

Comparing the potential sites found in the 100 minimizations with the statistics on the actual sites in Table V suggests that about half of the potential solvation sites do not have an ordered water assigned in the 1.7 Å resolution X-ray structure of the T4 lysozyme. First, a charged side chain can be in contact with up to five sites that appear as local minima on the free energy map. In a few cases, all sites are occupied (e.g., Asp 127 is in contact with five water molecules), but most frequently the crystallographer places only one or two water molecules, mainly because the local minima are separated only by moderate free energy barriers and the water molecules may not be strictly localized. Second, some of the potential sites are too weak to affect significantly the electron density map. Crystal packing interactions may affect which of the potential positions is actually visible as an ordered water site. To support this hypothesis, Figure 4 shows an interesting compensation between the free energies of the conserved sites and the number of instances these sites are either occupied by a polar atom from a neighboring protein or are affected by a steric interference (see Table I). The sites with low free energy ($\Delta G_{el} < -5$ kcal/mol) can be conserved without any interference from other proteins in the lattice. However, some of the weakly interacting sites ($\Delta G_{el} > -5$ kcal/mol) are frequently occupied by atoms from a neighboring protein in the crystal lattice. It is possible that such sites may be affected by crystal packing even when occupied by a water molecule. Third, sites contacting a mobile side chain such as Lys frequently do not show up on the X-ray structure.

the desolvation free energy is between 0.09 kcal/mol and 0.6 kcal/mol. Including this term would weaken the binding affinities and would thus reduce the residence times, but the change is very small. We note that Zhang and Hermans³³ used molecular dynamics simulations to calculate free energies of buried waters. For waters in polar cavities, they obtained free energy values between -12.0 kcal/mol and -3.1 kcal/mol, thus, in good agreement with the values found for such sites in the present study.

Conserved Versus Nonconserved Sites

If the strength of interactions with the protein is not a major factor in determining whether or not a water site is conserved among different crystal forms, then what is? Although we cannot fully answer this question, we observed that the conserved/nonconserved character of a water site depends on how much its environment changes from one crystal to another. We mentioned that water molecules at local free energy minima predominantly interact either with the backbone or with the side chain of a single residue. We can picture this situation as water molecules “sitting” on the end of charged and polar side chains. If the side chain moves more than 1.2 Å between two X-ray structures, then, by the definition, of Zhang and Matthews¹⁹ the site cannot be conserved.

Figure 5a shows the RMSD values for the 11 side chain types interacting with ordered water sites. As described in the Materials and Methods section, these RMSD values represent the mean of the pairwise side chain RMSDs calculated for 10 T4 lysozyme structures. Figure 5b shows the relative water contact frequencies for each side chain type, normalized to account for the different numbers of conserved and nonconserved sites. It is clear that contacts with side chains whose RMSD is large (say, over 1 Å) work against conservation. From Figure 5a, the residue with the largest RMSD is Arg, but only for the nonconserved sites. It is surprising that for the conserved sites the Arg RMSD is well below 1 Å, suggesting that water sites near freely moving Arg side chains are never conserved. The side chain with the next highest RMSD is Lys. However, Lys never contacts a conserved site (Fig. 5b). In fact, Thanki et al.⁸ also observed that the highly mobile Lys side chains have no preferred orientation for water contacts. The side chain with the next highest RMSD is Glu, which predominantly occurs in contact with nonconserved sites. The only other side chains with more than 1 Å RMSD are Thr, Tyr, and Met. However, the number of these side chains is low, and Tyr and Met occur only in contact with nonconserved water sites. In contrast, conserved water sites most frequently contact Asp side chains that have a relatively small RMSD, particularly at the conserved sites.

CONCLUSIONS

We used the continuum electrostatic model and the FDPB method to map the free energy of water molecules around a protein in solution. In particular, individual maps were created in the vicinity of each ordered water site seen in an X-ray structure of T4 lysozyme. Results can be summarized as follows.

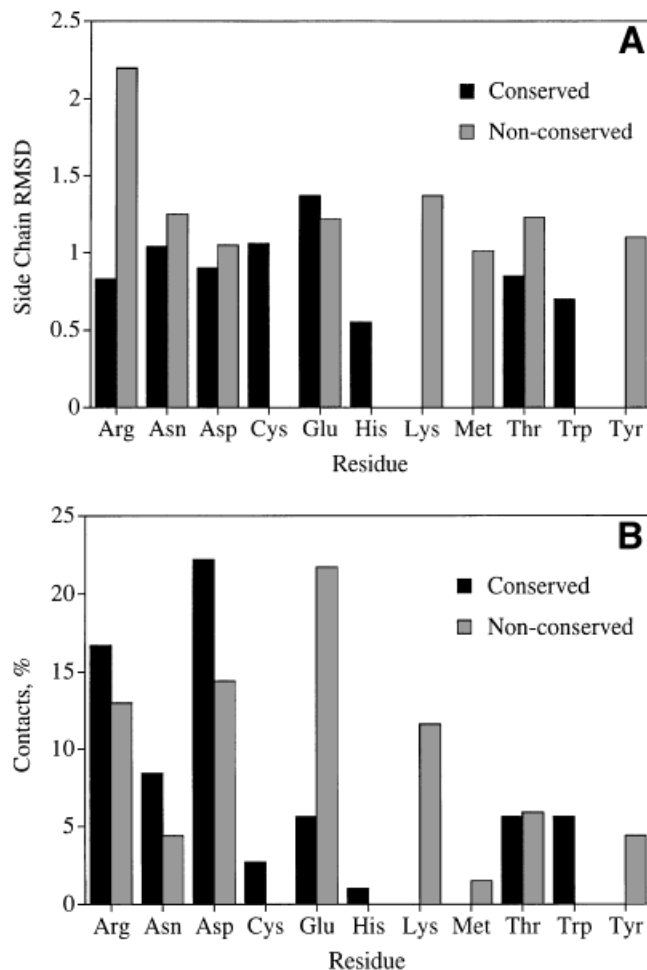


Fig. 5. Statistics of side chain-water interactions. **A:** Root-mean-square deviation (RMSD) values for the 11 side chain types interacting with ordered water sites. **B:** Water contact frequencies for each side chain type, normalized to account for the different numbers of conserved and nonconserved sites.

1. Almost all conserved sites and the majority of nonconserved sites were found within 1.3 Å of local free energy minima. This finding is in sharp contrast to the behavior of randomly placed water molecules that, on the average, have to move 3.2 Å to the nearest local minimum. Thus, we conclude that the potential positions of ordered water sites are at least partially determined by electrostatic interactions between the protein and the water.

2. Based both on the location of the free energy minima reached from random points, and on the analysis of water around charged and polar side chains, only about half of the potential sites are seen in the 1.7 Å resolution structure of T4 lysozyme. Crystal packing interactions can stabilize weak or mobile potential sites (in particular, 5 conserved and 27 nonconserved waters are not close to free energy minima in solution), but can also prevent water from occupying certain sites.

3. By using the calculated electrostatic free energies and accounting for the smaller water entropy in the boundary layer, we estimated water residence times for different

conditions. At ordered water sites the most likely value of electrostatic free energy is $\Delta G_{el} = -2.5$ kcal/mol, resulting in $\tau = 326$ ps residence time. The lowest free energy value, obtained for the almost fully buried water 176, is -12.83 kcal/mol, which yields $\tau = 0.98 \times 10^{-2}$ s. These values are in good agreement with the residence times from NMR experiments.⁵

4. In terms of the free energy, there is very little difference between the nonconserved and most of the conserved solvation sites. A small number of conserved sites with zero or small exposure to the bulk solvent have very low free energies. More generally, the free energy distribution for the conserved sites is slightly shifted toward stronger interactions, but the difference is not significant. Apart from large shifts for a few water molecules in nonconserved sites, the distributions of displacement values are also similar.

5. We observed that the conserved or nonconserved character of a solvation site depends on how much its environment changes from one crystal to another. The ordered water molecules contact either the backbone or a charged/polar side chain. For this latter case, we have found that the side chains contacting conserved sites have substantially smaller RMSDs than the side chains that contact nonconserved sites. Some of these differences are rather strong. For example, the average RMSD of Arg side chains contacting nonconserved sites is 2.2 Å, whereas the average RMSD is only 0.85 Å for Arg side chains in contact with conserved sites. The side chain with the next highest RMSD is Lys, but Lys never contacts a conserved site.

It is important to keep in mind that the analysis presented here has been carried out on a relatively coarse-grain grid (with the step size of 18°) over the subspace (ϕ, θ). We have recently repeated the analysis of solvation sites of the T4 lysozyme by using an off-grid procedure, based on multistart local minimizations in the space of all the six variables.³⁸ Robustness has been attained by the classic nonlinear simplex method.³⁹ In the vicinity of each solvation site, 30 minimizations have been performed by using random initial simplexes around the starting point (i.e., an ordered water position or a random point in the first water layer), and the 30 solutions were clustered. The new method confirmed all conclusions of the present study, and the distributions of displacement values were virtually identical in the two calculations. However, some of the free energy values have been reduced when going off the grid. The low free energy values in Tables I and II (below -6 kcal/mol) changed very little, but many higher values have been reduced by as much as 4 kcal/mol, indicating that the corresponding local minima are very narrow and are somewhat smoothed out when restricting consideration to the grid. Some of these changes may be relevant for the interpretation of ordered water positions. In particular, the mean free energy of water molecules contacting an Asn side chain changed from -2.28 kcal/mol (see Table V) to -2.77 kcal/mol, whereas the mean free energy for water contacting Arg has been reduced to -6.2 kcal/mol, thereby explaining why Arg side chains are substantially better solvated than Asn. However, the residence time data do

not support the existence of a large number of such strong solvation sites. This contradiction may suggest that deep but very narrow minima of the electrostatic free energy are entropically so unfavorable that they have little effect on the residence times in solution.

Finally, we note that the method developed here, i.e., translating and rotating molecular probes (a water molecule in this case) along a protein for the construction of the free energy maps can be extended to study the weakly specific binding of organic solvent molecules to proteins, either in crystal⁴⁰ or in solution.⁴¹ In particular, one can calculate the free energy of replacing water (mobile or partially bound) in the reference state by a given molecule. A limited number of preliminary runs that used continuum electrostatics and a structure-based desolvation potential gave results that were in good agreement with experimental data.⁴¹

ACKNOWLEDGMENTS

The authors thank Xing Hung for providing data on crystal contacts in the lysozyme structures. We also thank Martha Teeter and Bruce Tidor for helpful discussions.

REFERENCES

1. Karplus PA, Faerman CH. Ordered water in macromolecular structure. *Curr Opin Struct Biol* 1994;4:770–776.
2. Teeter MM. Water-protein interactions: theory and experiment. *Annu Rev Biophys Chem* 1991;20:577–600.
3. Levitt M, Park B. Water: now you see it, now you don't. *Structure* 1993;1:223–226.
4. Mattos C, Ringe D. Locating and characterizing binding sites on proteins. *Nat Biotechnol* 1996;14:595–599.
5. Otting G, Liepinsh E, Wuthrich K. Protein hydration in aqueous solution. *Science* 1991;254:974–980.
6. Belton PS. NMR studies of protein hydration. *Prog Biophys Mol Biol* 1994;61:61–79.
7. Finer-Moore JS, Kosiakoff AA, Hurley JH, Earnest T, Stroud RM. Solvent structure in crystals of trypsin determined by X-ray and neutron diffraction. *Proteins* 1992;12:203–222.
8. Thanki N, Thornton JM, Goodfellow JM. Distribution of water around amino acid residues in proteins. *J Mol Biol* 1988;202:637–657.
9. Levitt M, Sharon R. Accurate simulation of protein dynamics in solution. *Proc Natl Acad Sci USA* 1988;85:7557–7561.
10. Brooks CL, Karplus M. Solvent effects on protein motion, and protein effects on solvent motion. Dynamics of the active site region of lysozyme. *J Mol Biol* 1989;208:159–181.
11. Komeiji Y, Uebayasi M, Someya J, Yamato I. A molecular dynamics study of solvent behavior around a protein. *Proteins* 1993;16:268–277.
12. Brunne RM, Liepinsh E, Otting G, Wuthrich K, van Gunsteren WF. Hydration of proteins. A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. *J Mol Biol* 1993;231:1040–1048.
13. Lounnas V, Pettitt BM. A connected-cluster of hydration around myoglobin: correlation between molecular dynamics simulations and experiments. *Proteins* 1994;18:133–147.
14. Lounnas V, Pettitt BM. Distribution function implied dynamics versus residence times and correlations: solvation shells of myoglobin. *Proteins* 1994;18:148–160.
15. Wang CX, Chen WZ, Tran V, Douillard R. Analysis of interfacial water structure, and dynamics in an α -maltose solution by molecular dynamics simulation. *Chem Phys Lett* 1996;251:268–274.
16. Rocchi C, Bizarri AR, Cannistraro S. Water residence times around copper plastocyanin: a molecular dynamics simulation approach. *J Chem Phys* 1997;214:261–276.
17. Makarov VA, Feig M, Andrews BK, Pettitt BM. Diffusion of solvent around biomolecular solutes: a molecular dynamics simulation study. *Biophys J* 1998;75:150–158.

18. Kuhn LA, Siani MA, Pique ME, Fisher CL, Getzoff ED, Tainer JA. The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. *J Mol Biol* 1992;228:13–22.
19. Zhang XJ, Matthews JW. Conservation of solvent-binding sites in 10 crystals of T4 lysozyme. *Protein Sci* 1994;3:1031–1039.
20. Sanschagrín PC, Kuhn LA. Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteases and contributions to ligand specificity. *Protein Sci* 1998;7:2054–2064.
21. Gilson MK, Honig B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins* 1988;4:7–18.
22. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. *Science* 1995;268:1144–1149.
23. Bruccoleri RE. Grid positioning independence and the reduction of self-energy in the solution of the Poisson–Boltzmann equation. *J Comp Chem* 1993;14:1417–1422.
24. Sezerman OU, Vajda S, DeLisi C. Free energy mapping of class I MHC molecules and structural determination of bound peptides. *Protein Sci* 1996;5:1272–1281.
25. Camacho CJ, Weng Z, Vajda S, DeLisi C. Free energy landscapes of encounter complexes in protein-protein association. *Biophys J* 1999;76:1176–1178.
26. Jorgensen WL, Chandrasekhar J, Madura JD. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–935.
27. Rao SS. Optimization. Theory and applications. New York: John Wiley and Sons; 1978.
28. Press W, Flannery BP, Teukolsky SA, Vetterling WT. Numerical recipes. Cambridge: Cambridge University Press; 1990.
29. Brooks BR, Bruccoleri RE, Olafson B, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* 1983;4:197–214.
30. Vajda S, Weng Z, Rosenfeld R, DeLisi C. Effect of conformational flexibility and solvation on receptor–ligand binding free energies. *Biochemistry* 1994;33:13977–13988.
31. Dunitz JD. The entropic cost of bound water in crystals and biomolecules. *Science* 1994;264:670–670.
32. Amzel LM. Loss of translational entropy in binding, folding, and catalysis. *Proteins* 1997;28:144–149.
33. Zhang L, Hermans J. Hydrophilicity of cavities in proteins. *Proteins* 1996;24:433–438.
34. Roux B, Nina M, Pomes R, Smith JC. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. *Biophys J* 1996;71:670–681.
35. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. *Nature* 1986;319:199–203.
36. Vajda S, Weng Z., DeLisi C. Extracting hydrophobicity parameters from solute partition and protein mutation/unfolding experiments. *Protein Eng* 1995;8:1081–1092.
37. Noyes RM. Effects of diffusion rates on chemical kinetics. *Prog React Kinet* 1961;1:129–160.
38. Dennis S, Camacho CJ, Vajda S. Exploring potential solvation sites of proteins by multistart local minimization. In: Floudas CA, Pardalos PM, editors. Optimization in Computational Chemistry and Molecular Biology. Norwell, MA: Kluwer Academic; 2000.
39. Nelder JA, Mead R. A simplex method for function minimization. *Computer J* 1964;7:308–313.
40. Mattos C, Ringe D. Locating and characterizing binding sites on proteins. *Nat Biotech* 1996;14:595–599.
41. Liepinsh E, Otting G. Organic solvents identify specific ligand binding sites of protein surfaces. *Nat Biotech* 1997;15:264–268.