PROTEINS: Structure, Function, and Genetics 24:433–438 (1996)

Hydrophilicity of Cavities in Proteins

Li Zhang and Jan Hermans
Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599-7260

ABSTRACT **Water molecules inside cavities in proteins constitute integral parts of the structure. We have sought a quantitative measure of the hydrophilicity of the cavities by calculating energies and free energies of introducing a water molecule into these cavities. A threshold value of the water-protein interaction energy at −12 kcal/mol was found to be able to distinguish hydrated from empty cavities. It follows that buried waters have entropy comparable to that of liquid water or ice. A simple consistent picture of the energetics of the buried waters provided by this study enabled us to address the reliability of buried waters assigned in experiments.** © 1996 Wiley-Liss, Inc.

Key words: structure refinement, buried water, free energy calculation, molecular dynamics simulation

INTRODUCTION

Side-chain packing in proteins is remarkably compact,[1] but not perfect. Cavities of a size that can accommodate one or two waters exist in virtually all large proteins (>100 residues).[2] Some of these cavities are hydrated, as detected by NMR, X-ray, and neutron diffraction methods. Several groups have studied the structural characteristics of the cavities, which include the shape, volume, surface area, polarity, and number of hydrogen bonds made with the buried water.[3–5] It was found that cavity hydration can be predicted with 80% accuracy on the basis of these characteristics.[3,6]

To stabilize a water molecule inside a cavity, the cavity has to provide an energetically favorable environment at least comparable to that provided by liquid water, i.e., the free energy of transfer of a molecule from liquid water to the cavity must be negative. From a thermodynamic analysis based on free energies of transfer of model compounds, Wolfenden and Radzicak[7] have estimated the probability of finding a water in a small hydrophobic cavity as 1/20,000, which is, in practice, unobservable.

Nevertheless, water molecules buried in apolar cavities have been reported[4,8] and it was not clear whether these cases are due to experimental errors or other unknown effect(s) of protein structure. It was speculated that a number of factors, such as polarization, entropy, and water-water hydrogen bonding may stabilize buried water molecules in the absence of water-protein hydrogen bonds.[8] Buried waters found in apolar cavities also led to the estimation that each hydrogen bond contributes little (0.6 kcal/mol) to the stabilization of a buried water molecule,[4] in contrast to 3 kcal/mol per hydrogen bond in liquid water.

We have sought a better quantitative measure of the hydrophilicity of cavities in proteins by calculating energies and free energies of introducing a water molecule into these cavities, using atomic force potentials.[9] The results suggest that these energies represent most of the factors involved in cavity hydration and accurately describe the polarity, and the methods used here may lead to a useful tool in structure determination.

METHODS

Finding Cavities

The molecular surface of the protein was defined as the surface of an object composed of spherical atoms and was calculated by the MS program,[10] with atomic radii 2.67 (tetrahedral carbon), 2.56 (trigonal carbon), 2.40 (nitrogen), 2.20 (oxygen), 0.00 (hydrogen), 2.70 Å (sulfur). The probe radius was set to zero.

The output of the MS program is a set of points evenly distributed on the molecular surface. Any part of the surface that is observable from outside the protein is defined as the outer surface, the remainder as buried. (A water molecule is classified as buried or exposed according to the same criteria.)

Energy Minimization

A program was designed to search for the minimum energy positions of a water molecule in cavities of proteins. Several high resolution structures with better than 1.8 Å resolution were chosen for the survey. Cavities in these proteins are small, which means that the most common size of cavities cannot accommodate a cluster of waters, although some buried waters form hydrogen bonds to neighboring buried waters. The search for the minimum-energy position employed a procedure as follows:

Received August 11, 1995; revision accepted October 27, 1995.

Address reprint requests to Jan Hermans, Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599-7260.

First, the water molecules found in the crystal structure were classified as buried and surface waters. The buried ones were discarded. Hydrogens were added to surface waters and were rotated to the most highly preferred orientation. Since this process dealt with one water molecule at a time, waters set up earlier would not remain in the most preferred orientation after the neighbors were changed. The process was thus repeated several times.

Second, a subset of the buried surface points was selected, so that the distance between any pair of selected points is at least 0.9 Å. A simple iterative process was used to search for minimum-energy sites, starting from these points. At each position, the minimum-energy orientation is found, then the position is changed, and the position of lowest energy is retained after each step. Initially, the orientation of minimum energy is found by a global search over a set of more than 2,000 orientations. As the search progresses, the step size for changing the position is gradually decreased, and when this becomes small, the best orientation is found by an incremental rather than a global procedure. (This method is, of course, similar to search procedures developed earlier by Goodford.[11])

Interaction of the probe water with other buried waters was excluded in the calculation of the energy, so that the probe water could only "see" the protein atoms and the surface waters. A 12 Å cutoff was applied to nonbonded interactions.

Third, the set of minimum-energy sites was collected and sent through a filter: a minimum-energy site was discarded if it was not buried, or if the energy was higher than 10 kcal/mol, or if another site with lower energy existed within 1.4 Å.

The error in the calculated energies was estimated to be 2 kcal/mol. The minimum-energy positions were then compared with the water structure determined by crystallography. The survey of all cavities in a protein the size of subtilisin takes several hours of cpu time on a fast workstation; the greater part, by far, of this time is spent on finding the positions of minimum energy. The methods used in this study have been collected in a program package "DOWSER," which is available from the authors (send electronic mail to hermans@med.unc.edu).

Free Energy Calculation

Molecular dynamics simulations were used to calculate free energies of the buried waters in the subtilisin Carlsberg-eglin-C complex. In these calculations, all crystallographic waters, except the weak ones (as defined in the legend to Fig. 1) were represented in the model (263 water molecules). An inelastic restraint with 5 ps relaxation time was applied to all the atoms in the system except the probe water. A 10 Å cutoff of nonbonded interactions and a 1 fs time step were used.

The molecular dynamics simulated a process in which a probe water was gradually diminished to a "ghost" molecule which had zero partial charges and zero van der Waals parameters. The conversion took place in 50 ps, then the system was equilibrated for 10 ps, and then converted back. The temperature of the system was maintained by coupling to a heat bath;[30] the temperature of the probe water was also coupled to a separate heat bath by application of appropriate stochastic and frictional forces. In order to make the process reversible, it was found necessary to impose a harmonic restraint potential on the ghost water so that it would not escape from the system. The force constant of the restraint potential on the ghost water was 5 kcal/(mol·Å^2). The effect of such a restraint can be easily estimated, at 2.3 kcal/mol. The calculation proceeded via non-linear coupling of the energy of the interaction between protein and probe water to the transformation process.[12] The hysteresis was of the order of 1 kcal/mol, and about twice that for Wat^{765} and Wat^{804}, so that the errors in the calculations were estimated to be 1 to 2 kcal/mol.[12] The methods used here were similar to those used in an earlier study which calculated the free energy of xenon binding to myoglobin.[13] (Similar methods have been used also in free energy calculations of buried waters in proteins,[14,15] and the analogous problem of locating water molecules in crystal hydrates according to free energy of transfer has been approached via a different simulation technique.[16])

The free energy of transferring a water from liquid water into the cavity (ΔG°) can be obtained from the free energy of transferring a restrained ghost water into the cavity (ΔG°_s) by subtracting the restraint free energy (−2.3 kcal/mol) and the excess free energy of liquid water (−6.3 kcal/mol), i.e., $\Delta G^\circ = \Delta G^\circ_s + 8.6$ kcal/mol. The calculation of the free energy of transfer at a single site takes over 12 hours of cpu time on a fast workstation (model HP-735, Hewlett Packard, Palo Alto, CA).

RESULTS AND DISCUSSION

Energy Survey

The distance between each minimum-energy position and the nearest experimentally found buried water has been plotted against the value of the energy (Fig. 1a). A clear pattern emerges. The leftmost part of the figure contains the polar cavities, in which water molecules can be placed with energies below −14 kcal/mol, and nearly all of these positions are near a crystallographic water (within 1.0 Å in 90% of the cases). The rightmost part of the figure contains the non-polar cavities, in which water molecules have energies above −10 kcal/mol, and for these no crystallographic water is found within 2.0 Å. As expected, the polar cavities are hydrated and the apolar cavities are empty; the important result is that the calculated energy provides a reliable estimate of the polarity.

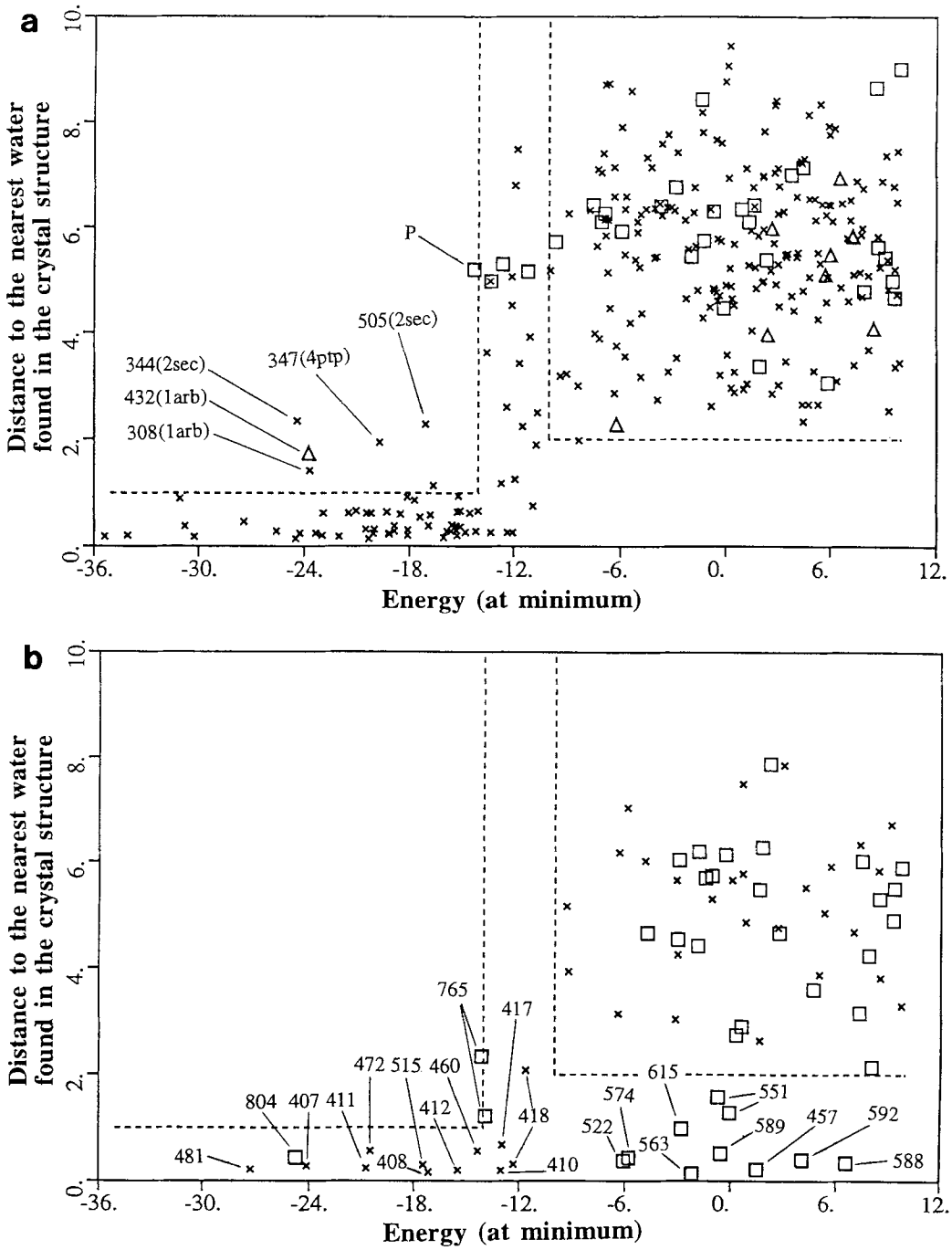

Fig. 1. Plot of minimum energies for introducing a water molecule into interior cavities against the distance to the closest water molecule found in the crystal structure. The protein structures include: **(a)** 1ARB,[25] 2CPL,[17] 5CPV,[26] 4GCR,[25] 135L,[27] 4PTP,[28] 2SEC,[19] 2SGA;[29] **(b)** 1CSE.[18] According to the value of q = 100 × *occupancy* × exp [−B/(4d^2)], where B is the crystallographic B-factor of the water molecule and d is the resolution of the structure, the crystal waters were classified as weak (□: q<1.30), intermediate (△: 1.30<q<1.76), or strong (×: q>1.76). Energy in Kcal/mol; distance in Å.

For five waters the positions of the energy minimum did not coincide precisely with the crystallographic water site, due to the presence of another minimum-energy site in the same cavity, closer to the water site. These have been identified with residue numbers assigned in the PDB file followed by the code of the protein in parentheses (Fig. 1a).

In crystallographic refinement of protein structures, a common standard for picking up possible water sites is that the electron density peak must be

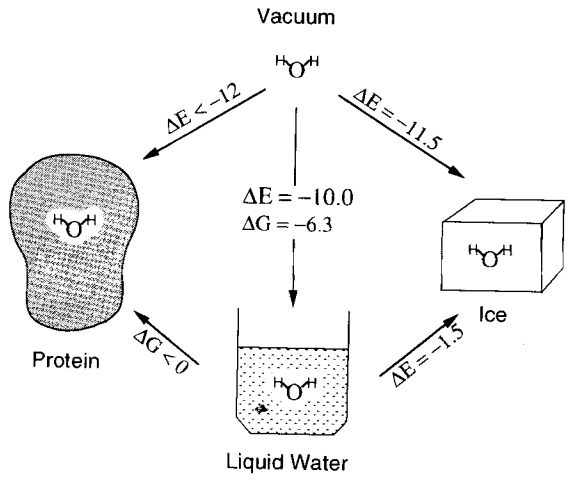

Fig. 2. Energetics of water in different states (energies in kcal/mol).

higher than 3σ or 4σ, where σ is the rmsd of the F_o–F_c difference map, and hydrogen bonding potential and size of the cavity are also considered. Point P in Figure 1a represents a cavity in 2CPL;[17] a water molecule can be located within 0.1 Å of the minimum (the coordinates of the water oxygen position are: 49.70, 20.49, 10.15 Å) by using the 3σ-criterion for identifying water sites, but not with the 4σ-criterion which had been used in the refinement.

Figure 1b shows the match between the minimum-energy positions and crystal waters in 1CSE,[18] a structure of subtilisin Carlsberg-eglin C complex. This protein is of particular interest because its structure in the same crystal packing arrangement has been independently determined at high resolution (1.2 Å in 1CSE[18] and 1.8 Å in 2SEC[19]) by different groups. Comparison of these structures allows an identification of the most reliable water sites.

In the 1CSE structure, the water molecules were assigned by applying an unusually weak criterion (2σ), and this contains nearly twice as many buried waters as the 2SEC structure. The extra water molecules have lower occupancy and higher B-factor ("weak" waters). The energy survey finds minimum-energy sites with energies below −10 kcal/mol within 1 Å of each of the strong waters (in Fig. 1b the numbers correspond to solvent numbers originally assigned in the crystal structure). Only two of the weak waters (Wat^{765} and Wat^{804}) are energetically favorable.

We have also examined buried waters in small apolar cavities in the larger pool of protein structures of an earlier survey,[4] and found that, as a rule, these are "weak" waters judged by their occupancy and B-factor. The statistics strongly suggest that these waters (and the weak waters in 1CSE) are artifacts of the crystallographic refinement process. Any survey that includes these as buried water molecules will produce erroneous conclusions.

Large Cavities

Surface waters and waters in large cavities, in contrast to the small cavities just discussed, can make hydrogen bonds among themselves so that the system is energetically stabilized. However, a few water molecules cannot form a hydrogen bonding network which can compete with that of bulk water, and favorable interactions with the protein are needed in addition. We have found that the sum of water-water and water-protein energies for filling large apolar cavities in two proteins correlates with experimental observations in the following two cases.

The first case: the crystal structure of 1UTG[20] contains an elongated hydrophobic pocket in the (two-fold) symmetrical dimer interface. When the cavity is empty, four sites at the ends of the hydrophobic pocket were found to have favorable energies (−12 and −14 kcal/mol). With these four waters in place, four additional favorable sites become available. The crystal structure here assigns 14 water molecules, some of them rather weak. From the energetic point of view, this is therefore a plausible structure.

The second case: a large nonpolar hydrophobic cavity in human interleukin β is reportedly hydrated with high occupancy in solution,[8] but these water molecules had not been observed in the X-ray structures.[21,22] We have searched for minimum energy positions in this protein and found that the energy for transfer of two water molecules from vacuum into the cavity is −21 kcal/mol in the solution structure. (The oxygen coordinates of the water molecules were 39.43, −14.47, 26.40 Å; 41.79, −14.38, 24.96 Å.). The energy was evaluated to be at least 4 kcal/mol higher in the crystal structures, the main reason being that the cavity is larger in the solution structure.[23] These results would indicate that this cavity's surface is insufficiently polar for it to contain water. A calculation of the free energy for transferring two water molecules from liquid water into the cavity confirms this (unpublished results).

Free Energy Survey

To include entropic effects of thermal motion of protein atoms and water molecules, molecular dynamics simulations were used to calculate free energies of the buried waters in the subtilisin Carlsberg-eglin-C complex.

As shown in Table I, the calculated free energies for the hydrated cavities are dramatically different from those for the empty cavities and the cavities with weak waters, and correlate well with the results of the simple energy calculations. The free energies for Wat^{804} and Wat^{765} are near zero, and at

TABLE I. Calculated Free Energies for Transfer of Water Into Various Cavities in the 1CSE Structure*

Cavity site	Occupancy	B-factor	ΔG°_s	ΔG°	E_{min}
Wat^{410}	1.03	7.17	−11.7	−3.1	−13
Wat^{412}	1.13	10.22	−11.0	−4.4	−15
Wat^{417}	1.10	7.41	−16.9	−8.3	−13
Wat^{418}	1.16	14.06	−13.2	−3.6	−12
Wat^{488}	1.04	17.73	−20.6	−12.0	−28
Wat^{588}	0.66	31.98	6.0	14.6	7
Wat^{592}	0.63	35.14	4.5	12.1	4
Wat^{599}	0.61	33.03	1.8	10.4	−2
Wat^{765}	0.70	27.33	−6.7	1.9	−14
Wat^{804}	0.43	30.30	−10.5	−1.5	−15
Cavity 1†	0		1.6	10.1	−9
Cavity 2†	0		6.2	14.8	−9

*ΔG°_s: Free energy of transferring a restrained ghost water into the cavity. ΔG°: Free energy of transferring a water molecule from liquid water into the cavity. Crystallographic occupancy and B-factor are also given. E_{min} is the energy at the energy minimum in the energy survey.

†Oxygen coordinates of the probe waters in cavities 1 and 2 were 34.69, −31.36, −1.68 and 50.96, −23.47, 11.88 (Å), respectively. These empty cavities were selected because of their low energy in the energy survey.

the present accuracy, are not decisive indicators of the state of these two sites.

CONCLUSIONS

The calculated energies and free energies provide a physically consistent picture of the energetics of buried waters in proteins, as follows. Consider a water molecule in four states: liquid water, protein cavity, vacuum, and ice (see Fig. 2). The free energy of transfer from liquid water to a cavity must be below zero. The energy of liquid water is low, therefore, either the energy of the buried water must also be low, or the entropy must be high. As the size of the cavity severely limits the entropy of a buried water molecule, the protein must provide several hydrogen bonds to the buried water in order to attain the required low energy of interaction. The threshold energy of −12 kcal/mol is below the energy of liquid water (−10 kcal/mol) and near the energy of ice (−11.5 kcal/mol). This indicates that buried waters have entropy as low as that of liquid water or ice, in agreement with the limits calculated by Dunitz from measured entropies of water, ice, and crystal hydrates.[24]

ACKNOWLEDGMENTS

We thank our colleagues Charles Carter, Hengming Ke, Anna Radzicka, and Richard Wolfenden for helpful comments. This research was supported by the National Science Foundation (Biophysics Program grant MCB-9314854 and Biological Instrumentation Program grant BIR-9123574) and the NIH Division of Research Resources (grant RR08102).

REFERENCES

1. Richards, F.M. The interpretation of protein structures: Total volume, group volume distributions and packing density. J. Mol. Biol. 82:1–14, 1974.
2. Hubbard, S.J., Argos, P. Cavities and packing at protein interfaces. Protein Sci. 3:2194–2206, 1994.
3. Rashin, A.A., Iofin, M., Honig, B. Internal cavities and buried waters in globular proteins. Biochemistry 25:3619–3625, 1986.
4. Williams, M.A., Goodfellow, J.M., Thornton, J.M. Buried waters and internal cavities in monomeric proteins. Protein Sci. 3:1224–1235, 1994.
5. Hubbard, S.J., Gross, K.-H., Argos, P. Intramolecular cavities in globular proteins. Protein Eng. 7:613–626, 1994.
6. Vedani, A. An algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds. J. Am. Chem. Soc. 113:5860–5862, 1991.
7. Wolfenden, R., Radzicka, A. On the probability of finding a water molecule in a hydrophobic environment. Science 265:936–937, 1994.
8. Ernst, J.A., Clubb, R.T., Zhou, H.-X., Gronenborn, A.M., Clore, G.M. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. Science 267:1813–1815, 1995.
9. Hermans, J., Berendsen, H.J.C., van Gunsteren, W.F., Postma, J.P.M. A consistent empirical potential for water protein interactions. Biopolymers 23:1513–1518, 1985.
10. Connolly, M.L. Solvent accessible surfaces of proteins and nucleic acids. Science 221:709–713, 1983.
11. Goodford, P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J. Med. Chem. 28:849–857, 1985.
12. Hermans, J. A simple analysis of noise and hysteresis in free energy simulations. J. Phys. Chem. 95:9029–9032, 1991.
13. Hermans, J., Shankar, S. The free energy of xenon binding to myoglobin from molecular dynamics simulation. Israel J. Chem. 27:225–227, 1986.
14. Wade, R.C., Mazor, M.H., McCammon, J.A., Quiocho, F.A. A molecular dynamics study of thermodynamics and structural aspects of the hydration of cavities in proteins. Biopolymers 31:919–931, 1991.
15. Helms, V., Wade, R.C. Thermodynamics of water mediating protein-ligand interactions in cytochrome P45cam: A molecular dynamics study. Biophys. J. 69:810–824, 1995.
16. Resat, H., Mezei, M. Grand canonical Monte Carlo simulation of water positions in crystal hydrates. J. Am. Chem. Soc. 116:7451–7452, 1994.
17. Ke, H. Similarities and differences between human cyclophilin A and other beta-barrel structures. Structural refinement at 1.63 angstroms resolution. J. Mol. Biol. 228: 539–550, 1992.
18. Bode, W., Papamokos, E., Musil, D. The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech *Hirudo medicinalis*. Eur. J. Biochem. 166:673–692, 1987.
19. McPhalen, C.A., James, M.N.G. Structural comparison of two serine proteinase-protein inhibitor complexes: Eglin-c-subtilisin Carlsberg and CI-2-subtilisin novo. Biochemistry 27:6582–6598, 1988.
20. Morize, I., Surcouf, E., Vaney, M.C., Epelboin, Y., Buehner, M., Fridlansky, F., Milgrom, E., Mornon, J.P. Refinement of the $c222_1$ crystal form of oxidized uteroglobin at 1.34 Å resolution. J. Mol. Biol. 194:725–739, 1987.
21. Finzel, B.C., Clancy, L.L., Holland, D.R., Muchmore, S.W., Watenpaugh, K.D., Einspahr, H.M. Crystal structure of recombinant human interleukin-1β at 2.0 angstroms resolution. J. Mol. Biol. 209:779–791, 1989.
22. Camacho, N.P., Smith, B.R., Goldman, A., Schneider, B., Green, B., Young, P.R., Berman, H.M. The structure of an interleukin-1β mutant with reduced bioactivity shows multiple subtle changes in conformation that affect protein-protein recognition. Biochemistry 32:8749–8757, 1993.
23. Clore, G.M., Wingfield, P.T., Gronenborn, A.M. High-res-

olution three-dimensional structure of interleukin 1β in solution by three- and four-dimensional nuclear magnetic resonance spectroscopy. Biochemistry 30:2315–2323, 1991.
24. Dunitz, J.D. The entropic cost of bound water in crystals and biomolecules. Science 264:670, 1994.
25. Tsunasawa, S., Masaki, T., Hirose, M., Soejima, M., Sakiyama, F. The primary structure and structural characteristics of *Achromobacter lyticus* protease I, a lysine-specific serine protease. J. Biol. Chem. 264:3832–3839, 1989.
26. Swain, A.L., Kretsinger, R.H., Amma, E.L. Restrained least squares refinement of native (calcium) and cadmium-substituted carp parvalbumin using X-ray crystallographic data at 1.6-Å resolution. J. Biol. Chem. 264: 16620–16628, 1989.
27. Harata, K. X-ray structure of monoclinic turkey egg lysozyme at 1.3 angstroms resolution. Published by deposition in Protein Data Bank (entry: 135L).
28. Chambers, J.L., Stroud, R.M. The accuracy of refined protein structures, comparison of two independently refined models of bovine trypsin. Acta Crystallogr. B 35:1861–1874, 1979.
29. Moult, J., Sussman, F., James, M.N.G. Electron density calculations as an extension of protein structure refinement. *Streptomyces griseus* protease at 1.5 angstroms resolution. J. Mol. Biol. 182:555–566, 1985.
30. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684–3690, 1984.