Water and proteins: A love–hate relationship

Yaakov Levy and José N. Onuchic*

Center for Theoretical Biological Physics and Department of Physics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093

It is widely appreciated that water molecules play an invaluable role in governing the structure, stability, dynamics, and function of biomolecules. The hydration forces are responsible for packing and stabilization of the protein structure. Particularly, water participates in many hydrogen bond networks and screening electrostatic interactions. However, the exact range of processes mediated by water is far from being understood, and it is only in the recent years that water has been quantitatively treated as an integral component of biomolecular systems. In this issue of PNAS, Papoian *et al.* (1) report a significant improvement in protein structure prediction by adding a water knowledge-based potential to an established Hamiltonian for protein structure prediction. "Wetting" the Hamiltonian improves the predicted structures, especially for large proteins, when long-range interactions between polar or charged groups are mediated by water molecules.

There are a variety of experimental and theoretical studies acknowledging the active role of solvent in protein stability and dynamics. Experimentally, x-ray, neutron diffraction (2), NMR (3, 4), and femtosecond fluorescence (5) measurements reveal the binding sites, structure, and dynamics of water. Theoretically, molecular dynamics simulations offer a detailed atomic description of both the biomolecule and the solvent as well as the time dependency of their dynamics (6–9). These all-atom explicit–solvent simulations of proteins, however, do not appear to provide sufficient conformational coverage to tackle many equilibrium and long-time-scale kinetic properties. Thus, a complementary approach is to adopt simplified models that trade high structural resolution of the water molecule (10, 11) (e.g., generalized born model for the water) or of both the polypeptide and the solvent (e.g., structure-based models) for enhancing conformational sampling (12, 13).

Desolvation during folding processes was studied by using a structure-based (Go) model for the Src homology 3 (SH3) protein assuming that each native contact is formed only after expelling a water molecule that mediates the interaction between any two residues (13). The fully solvated unfolded chain undergoes an initial structural collapse to an overall native topological conformation

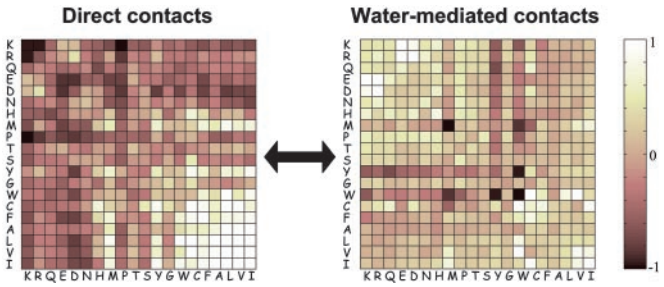

Fig. 1. The knowledge-based potentials incorporated in the structure prediction Hamiltonian by Papoian *et al.* (1). Each pair of residues can interact directly or meditatively by a water molecule (these two types of contacts are also termed first and second wells, respectively). A direct contact was defined as occurring between residues when the distance between their C_β atoms is 4.5–6.5 Å, and, similarly, a water-mediated contact was set to a distance of 6.5–9.5 Å. The second well includes also a protein-mediated contact potential (data not shown), which is highly correlated with the direct-contact potential. Positive values indicate more favorable interactions.

that is followed by a second transition where water molecules are cooperatively squeezed out from the hydrophobic core region, resulting in a dry and packed protein. Gating protein folding by solvent, as implemented in the energetically minimally frustrated model, provides additional microscopic features of the folding events, which have been observed experimentally (13, 14). Atomistic simulation studies of SH3 and proteins A and G support the role of water as a lubricant for the packing of the hydrophobic core after the formation of the transition state (6–8). Moreover, these fully atomic simulations, which are not biased toward direct contacts between the residues, point out that the

Adding water to an established Hamiltonian improves predicted protein structures.

folded state is not completely dry, but a few core water molecules form hydrogen bonds with the protein backbone. These studies argue that water evaporation due to hydrophobically induced drying (as was suggested for the association of two rigid hydrophobic objects) (15, 16) is less realistic for protein folding because water can be gradually expelled from the protein interior due to the chain flexibility and the existence of polar groups in the core.

Water is fundamental in protein folding mainly because of its role in defining hydrophobic attractions (17, 18) that are responsible for the rapid gluing of hydrophobic residues. The hydration water in the proximity of the protein surface, which exhibits dynamical properties markedly deviating from those of bulk (19), are crucial for stabilizing folded proteins. Fluctuations of the hydration water can slave the protein dynamics and thus affect its function (20, 21), yet the interplay between the protein and solvent complexity is an intriguing open question. Furthermore, water not only interacts with the protein surface, but it can directly interact with the protein backbone and side chains in the protein interior or even form clusters of two or more water molecules in hydrophobic cavities (3). Buried water molecules have much longer mean residence time than water in the first hydration shell, and thus they constitute an integral part of the protein structure (4, 22, 23). However, interior water molecules can escape to the bulk and be replaced by water from the hydration shell (23). Accordingly, mutations can affect the number of structural water molecules within the core and disrupt essential main-chain interaction network mediated by ordered water contacts (24), resulting in destabilization.

See companion article on page 3352.

*To whom correspondence should be addressed. E-mail: jonuchic@ucsd.edu.

Papoian *et al.* (1) uniquely incorporate water into a protein structure prediction Hamiltonian by allowing any tertiary pairwise interaction in the protein (including non-native contacts) to be either a direct contact or a long-range contact mediated by either a water molecule or by the protein itself, resulting in a highly nonadditive potential (25). The potentials of direct, as well as protein- and water-mediated, contacts for any pair of residues were obtained by a bioinformatic approach based on data set of 156 monomeric proteins (Fig. 1). This indirect definition of specific water-mediated contacts overcomes the likely underestimated number of interactions with water as found by high-resolution x-ray structures. The "wet" potential shows a remarkable improvement of predicting the structure of α-helical proteins, especially for those with >115 residues, through long-range water-mediated interactions. These interactions are important in the early stage to guide the structural search by the formation of long-range contacts. Late events include the formation of short-range contacts, the exclusion of water from the protein interior, and stabilizing the folded state by bridging hydrophilic groups at the protein surface. Water molecules can guide folding and facilitate packing of supersecondary structural elements by mediating long-range interactions between polar and charged amino acids, pointing out its important role for folding and stabilization of large and multidomain proteins.

The aqueous environment, thus, has a more active role in protein dynamics and stability than what is traditionally imagined and may have many applications. For instance, incorporating water-mediated contacts between hydrophilic residues in protein design may result in more stable proteins. Moreover, the common hydrophilic nature of the interfaces of protein–protein (26, 27) and protein–DNA (28) complexes together with the limited success of coarse-grained folding potentials for binding problems (25) and the abundance of water molecules at the interfaces suggest that water is indispensable for biomolecular recognition and self-assembly (29) (see Fig. 2). For some complexes, it was concluded that water can contribute to exquisite specificity (30), whereas for others, water was found to allow promiscuous binding by acting as a buffer that weakens unfavorable polar interactions (31). Thus, the enthalpy gain from water-mediated contacts is greater than the entropic cost that must be paid for immobilizing interfacial water. Kinetically, water molecules can guide a fully solvated protein to recognize a fully solvated DNA by a gradual expulsion of water layers, presumably via the fly casting mechanism (32), resulting with a wet interface complex. We must mention, however, that the gross features of the association mechanisms of various homodimers (33), trimers, and tetramers were obtained based on their topology alone. The binding transition state of several protein complexes obtained by using a Go model is in agreement with their experimental Φ value (unpublished data). This finding suggests that the protein topology is a dominant factor governing protein folding and binding, yet water has to be taken into account for studying desolvation effects on binding.

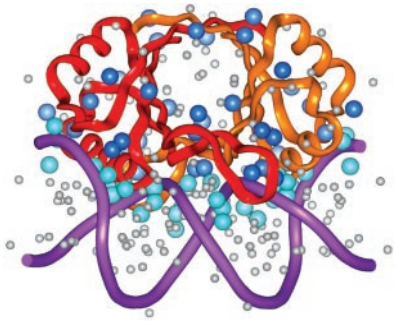

Fig. 2. Water at the protein–protein and protein–DNA interfaces. The complex of bovine papillomavirus-1 E2 with DNA (PDB ID code 2bop) contains 241 water molecules (represented by spheres), of which 40 water molecules mediate protein–protein interactions (blue spheres) and another 42 water molecules mediate protein–DNA interactions (light blue spheres).

The study by Papoian *et al.* (1) indicates that water not only induces protein folding and binding but also actively participates via long-range water-mediated contacts. Both hydrophobic and hydrophilic effects are dominant driving forces for biochemical processes. Adding water may improve protein docking as well as protein and drug design strategies and results with higher specificity and affinity. "Wet" Hamiltonians are expected to predict more accurate structures. Moving on wet funneled energy landscapes has provided insight on the coupled slaving between the protein and water dynamics and the role played by water in biological functionality.

We thank the National Science Foundation (Grants PHY-0216576, PHY-0225630, and MCB-0084797) for supporting our research.

1. Papoian, G. A., Ulander, J., Eastwood, M. P., Luthey-Schulten, Z. & Wolynes, P. G. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 3352–3357.
2. Savage, H. & Wlodawer, E. (1986) *Methods Enzymol.* **127,** 162–183.
3. Ernst, J. A., Clubb, R. T., Zhou, H.-X., Gronenborn, A. M. & Clore, G. M. (1995) *Science* **267,** 1813–1817.
4. Otting, G., Liepinsh, E. & Wuthrich, K. (1991) *Science* **254,** 974–980.
5. Pal, S. K., Peon, J. & Zewail, A. H. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 1763–1768.
6. Sheinerman, F. B. & Brooks, C. L., III (1998) *J. Mol. Biol.* **278,** 439–456.
7. Shea, J.-E., Onuchic, J. N. & Brooks, C. L., III (2002) *Proc. Natl. Acad. Sci. USA* **99,** 16064–16068.
8. Garcia, A. E. & Onuchic, J. N. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 13898–13903.
9. Tobi, D., Elber, R. & Thirumalai, D. (2003) *Biopolymers* **68,** 359–369.
10. Bursulaya, B. D. & Brooks, C. L. (2000) *J. Phys. Chem. B* **104,** 12378–12383.
11. Zagrovic, B., Sorin, E. J. & Pande, V. (2001) *J. Mol. Biol.* **313,** 151–161.
12. Sorenson, J. & Head-Gordon, T. (1998) *Fold. Des.* **3,** 523–534.
13. Cheung, M. S., Garcia, A. E. & Onuchic, J. N. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 685–690.
14. Fernandez-Escamilla, A. M., Cheung, M. S., Vega, M. C., Wilmanns, M., Onuchic, J. N. & Serrano, L. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 2834–2839.
15. Lum, K., Chandler, D. & Weeks, J. D. (1999) *J. Chem. Phys. B* **103,** 4570–4577.
16. Huang, X., Margulis, C. J. & Berne, B. J. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 11953–11958.
17. Kauzmann, W. (1959) *Adv. Prot. Chem.* **14,** 1–59.
18. Dill, K. A. (1990) *Biochemistry* **29,** 7133–7141.
19. Bizzarri, A. R. & Cannistraro, S. (2002) *J. Phys. Chem. B* **106,** 6617–6633.
20. Ansari, A., Jones, C. M., Henry, E. R., Hofrichter, J. & Eaton, W. A. (1992) *Science* **256,** 1796–1798.
21. Fenimore, P. W., Frauenfelder, H., McMahon, B. H. & Parak, F. G. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 16047–16051.
22. Denisov, V. P. & Halle, B. (1995) *J. Mol. Biol.* **245,** 682–697.
23. Garcia, A. E. & Hummer, G. (2000) *Proteins Struct. Funct. Genet.* **38,** 261–272.
24. Covalt, J. C., Roy, M. & Jennings, P. A. (2001) *J. Mol. Biol.* **307,** 657–669.
25. Papoian, G. A., Ulander, J. & Wolynes, P. G. (2003) *J. Am. Chem. Soc.* **125,** 9170–9178.
26. Xu, D., Tsai, C.-J. & Nussinov, R. (1997) *Protein Eng.* **10,** 999–1012.
27. Sheinerman, F. B., Norel, R. & Honig, B. (2000) *Curr. Opin. Struct. Biol.* **10,** 153–159.
28. Janin, J. (1999) *Structure (London)* **7,** R277–R279.
29. Ladbury, J. E. (1996) *Chem. Biol.* **3,** 973–980.
30. Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I. & Tello, D. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 1089–1093.
31. Sleigh, S. H., Seavers, P. R., Wilkinson, A. J., Ladbury, J. E. & Tame, J. R. H. (1999) *J. Mol. Biol.* **291,** 393–415.
32. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 8868–8873.
33. Levy, Y., Wolynes, P. G. & Onuchic, J. N. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 511–516.