ISSN 0027-1349, Moscow University Physics Bulletin, 2015, Vol. 70, No. 5, pp. 362–367. © Allerton Press, Inc., 2015.
Original Russian Text © E.V. Rubtcova, A.B. Solovey, V.I. Lobyshev, 2015, published in Vestnik Moskovskogo Universiteta. Fizika, 2015, No. 5, pp. 33–38.

THEORETICAL AND MATHEMATICAL PHYSICS

The Statistical Characteristics of Protein Hydration Shells: A Computer Simulation

E. V. Rubtcova[a], A. B. Solovey[b], and V. I. Lobyshev[a]

[a] *Department of Physics, Moscow State University, Moscow, 119991 Russia*

[b] *Specialized Educational Scientific Center, Moscow State University, Moscow, 121357 Russia*

e-mail: ev.rubcova@physics.msu.ru; soloveybird@gmail.com; lobyshev@yandex.ru

Received January 15, 2015; in final form, June 1, 2015

Abstract—Water molecules and the hydrogen-bond network are essential structural components of many biological macromolecular systems, such as nucleic acids, as well as fibrillar and globular proteins. The water density in the protein hydration shell is known to differ from that of bulk water. In this paper, spatial density as function of the distance from the protein surface is discussed. The electric potential function is considered in the same way. The resulting profiles of the bound water density are characterized by the presence of several distinct maxima due to the existence of the regular structure of bound water at distances over 7 Å. The minor discrepancies that are observed in both radial functions for different proteins are explained by topological variations that were revealed earlier by studying the valence and dihedral angle-distribution functions for water bound to the same proteins.

Keywords: protein hydration shell, structure modeling, density.

DOI: 10.3103/S0027134915050136

INTRODUCTION

The characteristics of physical properties of protein–solvent surfaces are essential for understanding protein structure, protein folding, and protein interactions with one another and other macromolecules [1]. Protein chains remain virtually immobile under anhydrous conditions; they either exhibit no enzyme activity at all, or this activity is negligibly low. Protein dehydration studies have demonstrated that a protein typically has to be surrounded by at least the first hydration shell to maintain its enzymatic and structural functions [2].

Water molecules in protein solutions can be subdivided into four groups:

• strongly and specifically bound water molecules inside a protein, where water molecules act as a structural component;

• water molecules that comprise the first hydration shell and interact with the protein surface;

• water molecules in the transition state between the first hydration shell and bulk water;

• bulk water.

Strongly bound water molecules occupy internal protein cavities and can be revealed by crystallography and NMR. The properties of the first hydration shell of water mostly depend on the properties of the protein surface, dipole moment, polarizability, and the ability of each water molecule to form hydrogen bonds. The surface-water molecules that form the hydration shell, as well as intermediate water molecules, are more mobile than the strongly bound molecules; their hydrogen bonds can be distorted and our study focuses on this type of water molecule.

The internal parameters of the surface water, i.e., valence and dihedral angles of bound molecules, significantly differ from those of bulk water. It was demonstrated by small-angle X-ray scattering that the average density of the first hydration shell (0–3 Å from the protein surface) is higher than the density of bulk water for different proteins [3]. In some studies this fact was confirmed by X-ray diffraction analysis and neutron scattering, as well as by numerical experiments that employed molecular dynamics simulation. The density of the water shell of a protein up to 5 Å thick was found to be higher than that of bulk water by 15%. Protein hydration is usually discussed using the terms of the hydrophobicity and hydrophilicity of surface protein groups and the number of bound water molecules.

However, disturbance of the hydrogen-bond network is the main reason that the density increases. This disturbance consists of changes in its internal parameters, such as a reduced O...O distance and an increased coordination number of water molecules [4–6]. We have previously shown that the structures of the hydration shell of proteins and bulk water differ topologically. The content of hexacycles and other

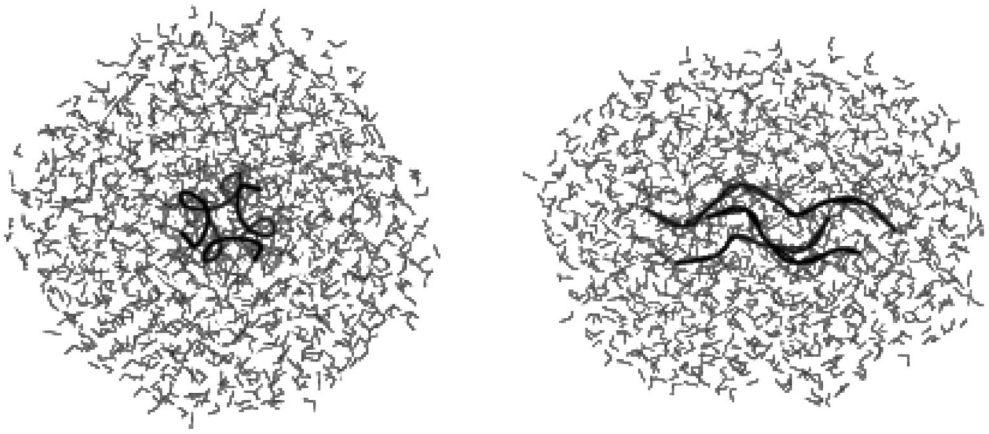

Fig. 1. The hydration shell of the 1ITT.pdb protein: (a) transverse and (b) longitudinal projections.

closed rings that consist of more than six water molecules in the hydration shell of proteins is high [7].

Along with the radial density distribution, the distribution of the electrostatic potential in hydration shell is also of interest, since this distribution is responsible for the type of intermolecular interaction that occurs.

This study aims at simulating the averaged distribution of the density and electrostatic potential as a function of the distance from the protein surface.

With respect to their structures, proteins can be subdivided into two large groups: globular and fibrillar proteins. Fibrillar proteins have periodic helical structures. Collagen is an example of a fibrillar protein. Collagen is among the most common and important structural proteins; it is characterized by a specific sequence of amino-acid residues [–Gly–X–Y–], where Gly is glycine and X, Y are most typically proline or hydroxyproline.

Specific water structuration in procollagen and collagen fibrils, as well as the formation of the native collagen structure that is associated with it, has been repeatedly confirmed using various methods. It has been demonstrated using different experimental methods (e.g., NMR and dielectric relaxation) that water molecules in the structure of collagen fibers are less mobile compared to bulk water [8–11]. It has also been shown that melting of bound water makes the main contribution to the enthalpy of heat denaturation of collagen; in other words, bound water stabilizes the collagen structure [11]. We proposed the hypothesis that the topological patterns of the hydrogen-bond network can be more pronounced in this structure compared to the hydration shell of globular proteins. For this reason, we studied two types of proteins in this work: the globular protein ubiquitin and collagens, as well as collagen fragments.

COMPUTER-SIMULATION EXPERIMENTS

SOLVATE software was used to describe the structure of hydration shells of proteins [13]. This program uses the thickness of a solvent layer as a variable parameter, simulates a convex hydration shell with a desired thickness around a protein complex, which was originally set as a structure from PDB (the high-resolution international Protein Data Bank), approximately calculates the density function, and fills the hydration shell volume with water molecules using the TIP3P water model (a rigid three-point model [14]) based on the resulting data. That is the method by which the spatial arrangement of bound water molecules is determined. The PDB data are presented for proteins in the state of molecular crystals; i.e., proteins surrounded by the minimal amount of water that is required to maintain the native conformation of a protein. In this work, we studied the hydration-shell structures of collagen fragments that are dissolved in water (1CAG.pdb, 1BKV.pdb, and 1ITT.pdb), as well as ubiquitin (1UBQ.pdb).

The first step was to simulate hydration shells for the structures that were mentioned above; let us give the structural characteristics:

• 1CAG.pdb is the collagen helix triple determined at 1.9 Å resolution; each chain is 30 amino acids long; the number of solvent molecules equals 10 062;

• 1BKV.pdb is the collagen helix triple determined at 2 Å resolution; each chain is 30 amino acids long; the number of solvent molecules equals 7164;

• 1ITT.pdb is the collagen helix triple determined at 1.9 Å resolution; each chain is 7 amino acids long; the number of solvent molecules is 2292;

• 1UBQ.pdb is the ubiquitin protein determined at 1.8 Å resolution; the number of solvent molecules is 9630.

The structures of hydration shells for the 1ITT.pdb collagen fragment in two projections ((a) transverse and (b) longitudinal) are shown in Fig. 1 as an example.

Fig. 2. The closed nonconvex surface of the body is approximated by a convex shell (* is the convex shell of the body under study, such as a protein or hydration shell).

After simulating the dilution of the proteins in water, one needs to make sure the system is unstrained in order to reduce the possible effects of an energetically non-optimal arrangement of the protein–solvent structure. One of the main reasons that artifacts emerge is they attain the local minimum of the potential energy function during minimization using the gradient descent method depending on the number of minimization steps, while the point of the global minimum is not attained. For $|MIN - MIN_{\text{global}}| \gg \epsilon$ (if there is a significant difference between the local and global minima), the system will be strained (in other words, it will have a significant reserve of potential energy) after minimization and the subsequent molecular dynamics will yield different statistical characteristics of the system under different initial conditions. The following sequence of processes was used to eliminate these nonequilibrium states:

(1) minimization of the potential energy of the structure of the primary hydration shell;

(2) the subsequent molecular dynamics procedure at low temperatures (close to the normal conditions; however, the structural integrity is retained), $T \in$ (50–150) K.

The molecular dynamics procedure should not be carried out at high temperatures, since it may yield strongly nonequilibrium structures that have completely different characteristics. In order to rule out a case where the system finds itself in a shallow local minimum, one needs to alternate energy minimization with the molecular dynamics procedure.

The widely used NAMD molecular dynamics software package was employed for the minimization and molecular dynamics simulation of the system [15].

After the quasi-equilibrium state of the protein–hydration shell system was attained, we studied the density of hydration water as a function of the distance from the protein surface.

Let us describe the algorithm for finding the density of this function:

(1) The CGAL software package [16] is used to form the internal and external hydration shells (Fig. 2 shows how the set of points that comprise the convex shell is found on the nonconvex protein shell).

(2) The coordinates of a parallelepiped inside which the structure of hydrated protein could be placed are found.

(3) An arbitrary point, A_{RAND}, in the structure of an isolated parallelepiped is considered (this point does not necessarily have coordinates that coincide with an oxygen atom).

(4) We determine whether this point lies inside the bound water layer. To do so, we ensure that the point lies inside shell W but outside shell P.

(5) If the point is internal and lies inside the hydration shell, A^{RAND} is enclosed in a sphere with radius R (the 1.5 Å value was selected as the parameter R in our case). We first ensured that the sphere entirely belonged to the hydration shell.

(6) The density of the water molecules in the sphere is calculated: $\rho = N_{MOL}/V$, where N_{MOL} is the number of atoms in the water molecules and V is the volume of a sphere with radius R.

(7) Let us assume that the distance between an isolated point and the protein $dist = min(r(A_{\text{RAND}}, C_{protein}, C_{protein}) \in \Omega$, where Ω is the set of points within a protein.

(8) Repeating iterations (1–7) M times ($M \gg 1$) and averaging the ρ values for each $dist$ yields the dependence of the average density of hydration shell on the distance from a protein.

(9) The error is calculated as follows: we assume $g(r)$ to be density-distribution functions.

(a) Let us consider the values of the functions $g_1 = g(r_m)$ and $g_2 = g(r_m+M)$, the functions for steps m and $(m + M)$, respectively, where M is the number of water molecules in the structure. M is set equal to the number of water molecules so that on average each molecule has an effect on function $g(r_m+M)$;

(b) Next, let us consider the normal error of the space $h_{[a,\ b]}$: $\Delta g = 2/(\|g_1\| + \|g_2\|\sqrt{\int (g_1 - g_2)^2 dr}$. The criterion for terminating the iteration process is attaining $|\Delta| < \epsilon$, where $\epsilon \ll 1$, in our case $\epsilon = 0.001$. The physical meaning of this criterion is as follows: the density distribution function reaches a stable level and an increasing number of iterations does not significantly change the distribution.

The following algorithm was used to determine the electric potential as a function of the distance from the protein:

• the procedure for determining the electric potential as a function of the distance from the protein is similar to the one that was used to determine hydration

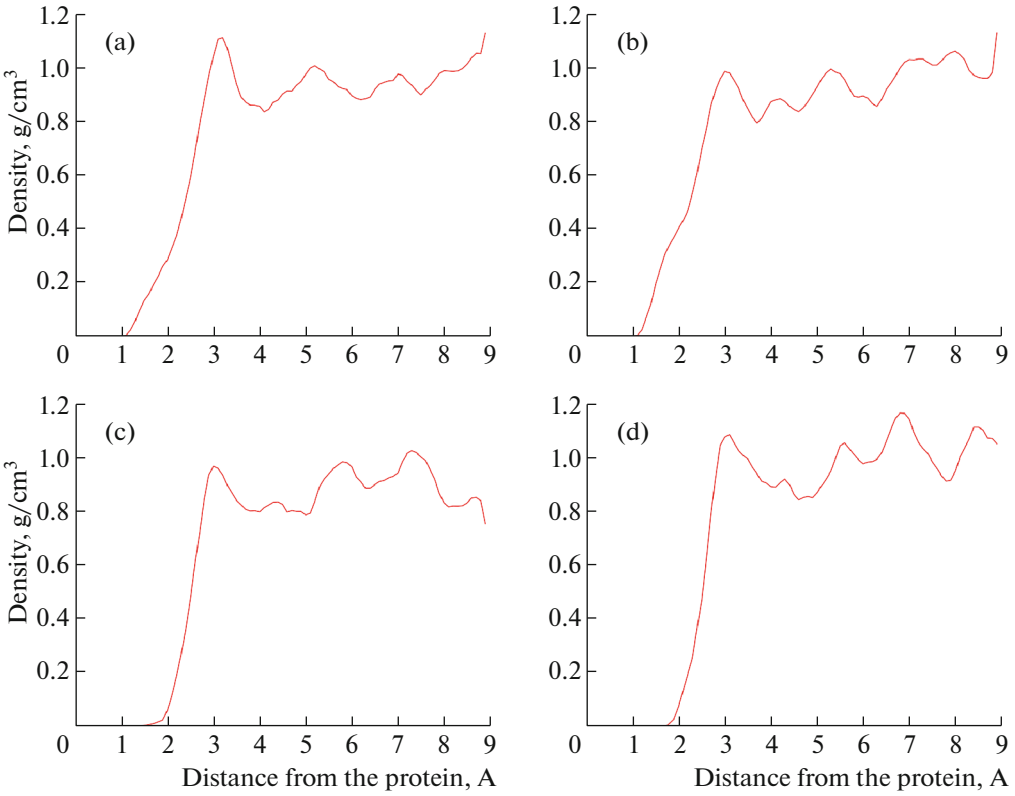

Fig. 3. The density-distribution functions in hydration shells, N is the number of alterations of the minimization and heating processes: (a) 1UBQ.pdb, $N = 3$; (b) 1UBQ.pdb, $N = 2$; (c) 1CAG.pdb, $N = 3$; and (d) 1Cag.pdb, $N = 2$.

shell density; however, we calculate the potential of the electric field, $\sum_i \frac{q_i}{\rho_i}$ (where summation is made for all water atoms for a given point), at each point, A_{RAND}, instead of calculating the density $\rho = N_{MOL}/V$ using the aforementioned algorithm. The charge values for our model are as follows: $q_{\text{oxygen}} = -0.64$ and $q_{\text{hydrogen}} = 0.32$.

RESULTS AND DISCUSSION

Figure 3 shows the averaged radial distributions of the hydration shell density for two proteins with different numbers of alterations of the minimization and heating procedures ($N = 2, 3$). There is a qualitative difference in the radial distribution of bound water for globular and fibrillar proteins. The first peak for collagen is observed at a distance of 2.9Å, while the position of the maximum for a globular protein is the same but the peak is broadened and water molecules are detected at a closer distance. This is most likely to be caused by the heterogeneity of the surface of the globular protein. For all the radial distributions of the mass density, we observed clearly pronounced maxima that belong to the first, second, third, and fourth hydration shells (which correspond to the following distances from the protein surface: 2.9, 4.2, 5.5, and 6.8–7.1 Å). The subsequent peaks probably arise from boundary effects; thus, they were not taken into consideration. Let us note that the second, third, and fourth maxima in the hydration shell of collagen are slightly shifted to larger distances than those for a globular protein. The third and fourth maxima are pronounced more clearly for collagen. This could be caused by topological variations of the structure of hydration shell, which were revealed by comparing the valence and dihedral angles [7].

Figures 4 and 5 show the functions of the electrostatic potentials, both for the structure of the hydration shell that includes oxygen atoms and protons and for the shells that take only oxygen atoms into account.

The results demonstrate that the functions of the electrostatic potential in hydration shells of proteins, as well as the density functions, have an ordered quasi-periodic structure at distances that include more than three molecular layers of water. The functions that were obtained for the structures with oxygen atoms and protons are more detailed compared to those that took only oxygen atoms into account. The potential profiles for the 1ITT.pdb protein were qualitatively different from those of the other proteins. This difference is due to topological variations, which can be seen from the specific distributions of valence and dihedral angles compared to the other proteins [7].

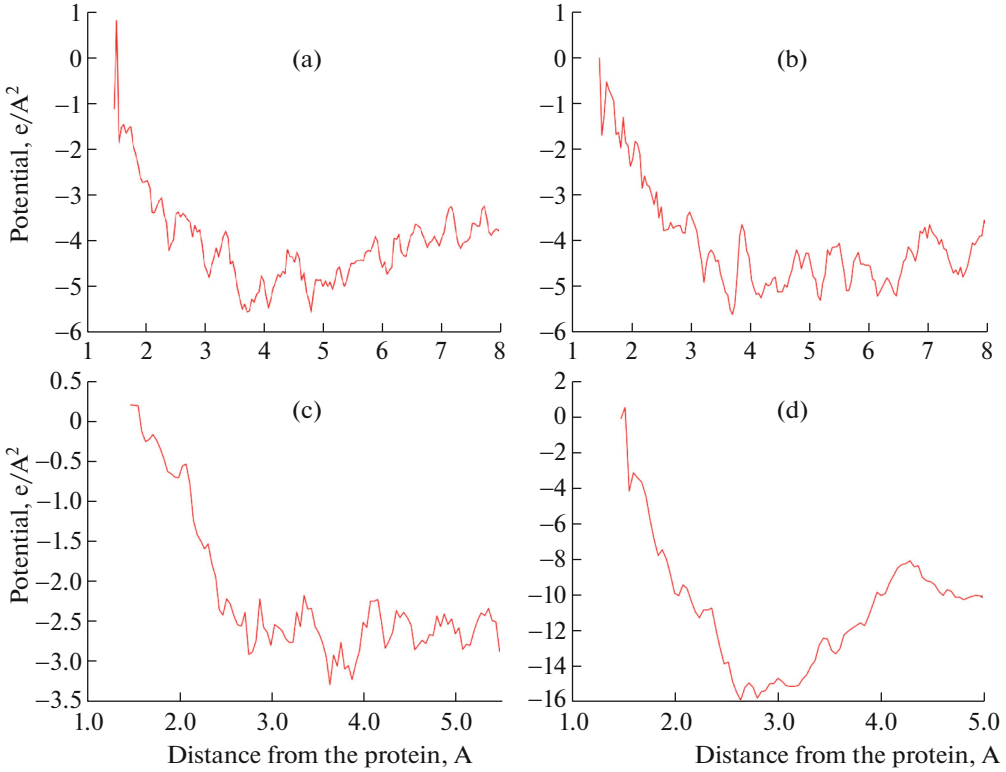

Fig. 4. The potential distribution function, including both oxygen atoms and protons for the shells of (a) 1UBQ.pdb; (b) 1CAG.pdb; (c) 1BKV.pdb; and (d) 1ITT.pdb.

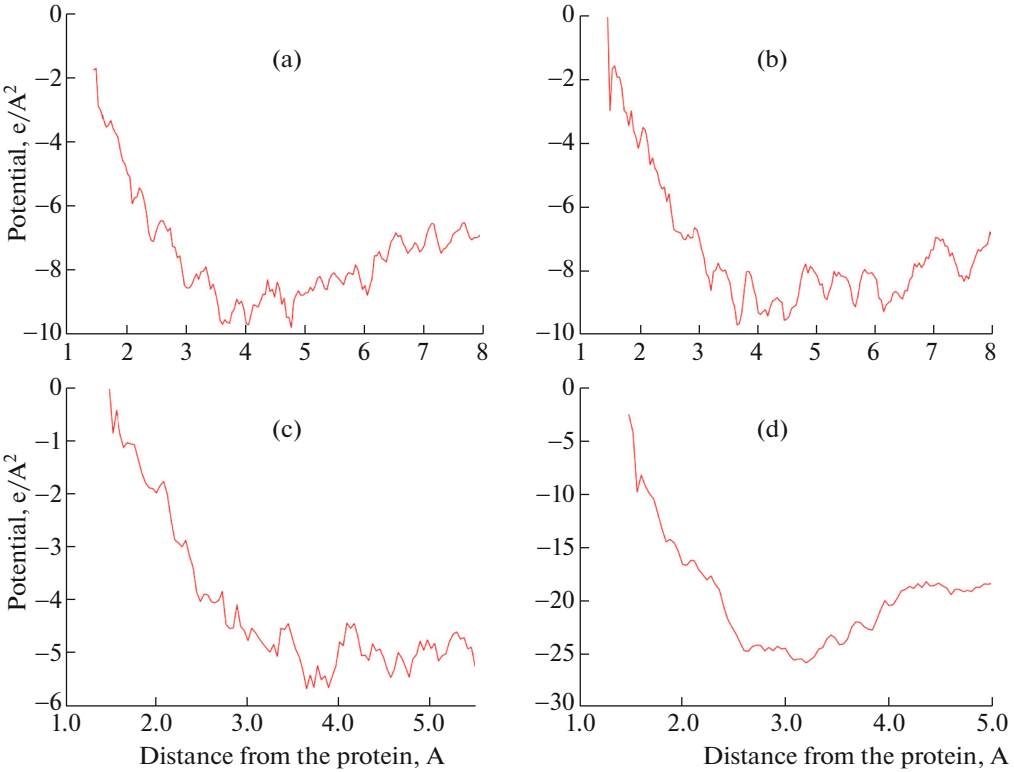

Fig. 5. The potential distribution function, including only oxygen atoms for the shells of (a) 1UBQ.pdb; (b) 1CAG.pdb; (c) 1BKV.pdb; and (d) 1ITT.pdb.

CONCLUSIONS

It was demonstrated that the density distribution in the bulk layer of hydration water indicates that multilayer organization is present in the hydration shell of the proteins that were under study. Taking the pronounced periodicity into account, it is fair to say that structural ordering occurs at distances from the protein surface of up to 7Å. The density-distribution functions of the hydration shell of globular and fibrillar proteins are qualitatively similar; however, small variations in density functions for the hydration shells of different proteins are attributed to topological changes that correlate with our previous results [7]. Similar conclusions were made by analyzing the distribution functions of the electrostatic potential.

More complete research into the topological features of the hydration shell can be performed to continue this research.

REFERENCES

1. W. Kauzmann, Adv. Protein Chem. **14**, 1 (1959).
2. C. Mattos and D. Ringe, Curr. Opin. Struct. Biol. **11**, 761 (2001).
3. D. I. Svergun, S. Richard, M. H. J. Koch, et al., Proc. Natl. Acad. Sci. USA **95**, 2267 (1998).
4. S. H. Lee and P. J. Rossky, J. Chem. Phys. **100**, 3334 (1994).
5. A. D. Friesen and D. V. Matyushov, J. Chem. Phys. **135**, 104501 (2011).
6. S. Sarupria and S. Garde, Phys. Rev. Lett. **103**, 037803 (2009).
7. E. V. Rubtcova, A. B. Solovey, and V. I. Lobyshev, Biophysics **59**, 869 (2014).
8. Y. K. Cheng, W. S. Sheu, and P. J. Rossky, Biophys. J. **76**, 1734 (1999).
9. G. E. Chapman, S. S. Danyluk, and K. A. McLauchlan, Proc. R. Soc. London, Ser. B **178**, 465 (1971).
10. J. Bella, B. Brodsky, and H. Berman, Structure **3**, 893 (1995).
11. V. I. Lobyshev and L. P. Kalinichenko, *Isotopic Effects D_2O in Biological Systems* (Moscow, 1978) [in Russian].
12. N. G. Esipova, N. S. Andreeva, and T. V. Gatovskaya, Biofizika **3**, 529 (1958).
13. http://www.mpibpc.mpg.de/grubmueller/solvate
14. P. Mark and L. Nilsson, J. Chem. Phys. **105**, 9954 (2001).
15. NAMD – Scalable Molecular Dynamics. http://www.ks.uiuc.edu/Research/namd/
16. CGAL – Computational Geometry Algorithms Library . http://www.cgal.org
17. F. Burling, W. I. Weis, K. M. Flaherty, and A. T. Brunger, Science **271**, 72 (1996).
18. M. Levitt and R. Sharon, Proc. Natl. Acad. Sci. USA **85**, 7557 (1988).
19. RCSB Protein Data Bank . 1CAG Structure Summary. http://www.rcsb.org/pdb/explore.do?structureId=lCAG
20. RCSB Protein Data Bank. 1UBQ Structure Summary. http://www.rcsb.org/pdb/explore/explore.do?structureId=lUBQ
21. B. Trumbore, J. Graphics Tools **2**, 21 (1997).
22. Y. Levy and J. N. Onuchic, Annu. Rev. Biophys. Biomol. Struct. **35**, 389 (2006).
23. L. Vitagliano, R. Berisio, and A. de Simone, Biophys. J. **100**, 2253 (2011).

Translated by D. Terpilovskaya