Generalized fundamental measure theory for atomistic modeling of macromolecular crowding

Sanbo Qin and Huan-Xiang Zhou*
Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, USA
(Received 10 July 2009; published 26 March 2010)

Macromolecular crowding inside cells affects the thermodynamic and kinetic properties of proteins. The scaled particle theory (SPT) has played an important role toward establishing a qualitative picture for the effects of crowding. However, SPT-based modeling lacks molecular details. Molecular dynamics simulations overcome this limitation, but at great computational cost. Here, we present a theoretical method for modeling crowding at the atomic level. The method makes it possible to achieve exhaustive conformational sampling in modeling crowding effects and to tackle challenges posed by large protein oligomers and by complex mixtures of crowders.

DOI: 10.1103/PhysRevE.81.031919 PACS number(s): 87.15.A−, 87.15.Cc, 87.15.nr

I. INTRODUCTION

The total protein and RNA concentrations inside cells reach 300–400 g/l [1] and the macromolecules together are estimated to occupy over 30% of cellular volume. The crowded environments are expected to have profound effects on the thermodynamics and kinetics of protein folding, binding, aggregation, and other more complex biological events [2]. Numerous *in vitro* experiments now support this expectation [3–20]. Historically the scaled particle theory (SPT) [21] has played an important role toward establishing a qualitative picture for the effects of crowding [2,22]. However, SPT-based modeling of crowding effects lacks molecular details.

This limitation has been overcome by recent molecular dynamics (MD) simulations of crowding, via two approaches [12,23–30]. In the direct simulation approach, a protein molecule is placed inside a box of crowders, and the motions of the protein and the crowders are followed simultaneously. The computational demands presented by such large simulation systems have necessitated the use of coarse-grained models [23–25,27,29].

We have developed a "postprocessing" approach for simulations of crowding [12,26,28,30], whereby the protein is simulated in the absence of crowders and the conformations thus sampled are reweighted according to crowding-induced changes in chemical potential. The postprocessing approach makes it possible to represent the protein at the atomic level, but the calculation of crowding-induced changes in chemical potential, by devising a procedure akin to Widom's insertion method [31], still requires significant computer times. Here, we present a theoretical method for calculating crowding-induced changes in chemical potential for atomistic proteins. This method yields accurate results but at substantially reduced computational cost, opening the possibility of more exhaustive conformational sampling in modeling crowding effects. Moreover, it enables computational studies of challenging problems presented by large protein oligomers, on which effects exerted by crowding are especially dramatic [3,4,15–20], and by complex mixtures of many species of crowders, which are required for realistically mimicking the intracellular milieu.

II. THEORY AND IMPLEMENTATION

Our method is based on the fundamental measure theory (FMT) [32–34], which is a type of density functional theory for fluids but specialized to mixtures of convex hard particles. Unlike the SPT, the FMT does not rely on any assumption about the shapes of the convex particles, and derives the SPT for the special case of spherical particles [33]. When a test protein is randomly placed into a distribution of convex crowders, the FMT predicts the increase in chemical potential of the test protein as [33]

$$\Delta\mu = \Pi_c v_p + \gamma_c s_p + \kappa_c l_p - k_B T \ln(1-\phi), \tag{1}$$

where k_B is Boltzmann's constant and T is the absolute temperature; v_p, s_p, and l_p are the volume, surface area, and linear size of the test protein; Π_c is the osmotic pressure of the crowders, and γ_c and κ_c are the corresponding quantities for surface tension and bending rigidity; and ϕ is the total volume fraction of the crowders. The latter quantities are expressed in terms of the weighted number densities of the different species of crowders:

$$c = \sum_\alpha c_\alpha; \quad c_l = \sum_\alpha l_\alpha c_\alpha; \quad c_s = \sum_\alpha s_\alpha c_\alpha; \quad \phi = \sum_\alpha v_\alpha c_\alpha, \tag{2}$$

where c_α is the number density of species α crowders, and l_α, s_α, and v_α are their linear size, surface area, and volume, respectively. The results are

$$\frac{\Pi_c}{k_B T} = \frac{c}{1-\phi} + \frac{c_l c_s}{(1-\phi)^2} + \frac{{c_s}^3}{12\pi(1-\phi)^3}; \tag{3}$$

$$\frac{\gamma_c}{k_B T} = \frac{c_l}{1-\phi} + \frac{{c_s}^2}{8\pi(1-\phi)^2}; \tag{4}$$

$$\frac{\kappa_c}{k_B T} = \frac{c_s}{1-\phi}. \tag{5}$$

The FMT has been applied to convex particles with relatively simple geometric shapes [33].

*Corresponding author; hzhou4@fsu.edu

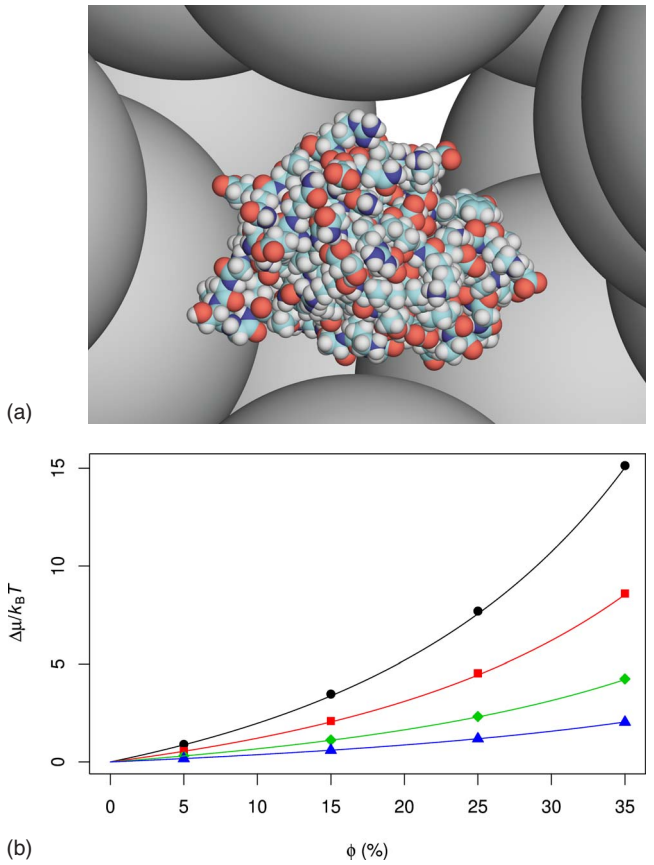

FIG. 1. (Color online) (a) All-atom representation of the barnase-barstar complex surrounded by crowders. (b) Results of $\Delta\mu$ for the protein complex. Symbols are obtained in our previous work [26] by the insertion procedure, with results for crowder radii of 15, 20, 30, and 50 Å shown as circles, squares, diamonds, and triangles, respectively, while curves are predictions of the GFMT. In all GFMT calculations, 10 000 directions were used to define the crowder-exclusion surface; the cubic grid representing the excluded volume had a spacing of 0.5 Å. At each crowder size, the calculation of v_p, s_p, and l_p per protein conformation took ~10 s of CPU time on an AMD Opteron 2382 processor; subsequently using v_p, s_p, and l_p in Eq. (1) to calculate $\Delta\mu$ at different crowder volume fractions required negligible amounts of time. In comparison, the aggregate CPU times for four crowder volume fractions at each crowder size were 80–160 s per snapshot of crowder configurations by the insertion procedure; to reduce statistical errors, the calculations for each protein conformation were repeated on 10–100 snapshots of crowder configurations, bringing the total CPU times per protein conformation at a given crowder size to 800–16 000 s. Overall the GFMT gained a speed up of $\sim 10^2$ to 10^3-fold over the insertion procedure.

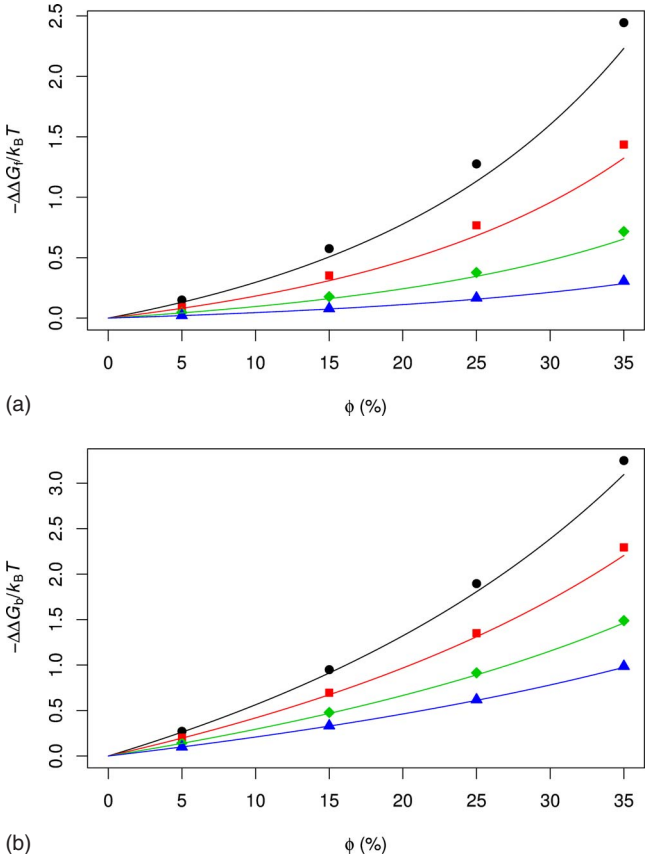

FIG. 2. (Color online) Effects of crowding on the (a) folding free energy of cytochrome b_{562} and (b) binding free energy of barnase and barstar. Symbols are obtained in our previous work [26] by the insertion procedure, while curves are predictions of the GFMT. The different curves represent the different crowder radii listed in Fig. 1(b).

We generalize the FMT to test proteins represented at the atomic level and refer to our method the generalized fundamental measure theory (GFMT). A hint of the GFMT appeared when we found that $\Delta\mu$ obtained in our previous work [12,26] by the insertion procedure could be fitted to Eq. (1). However, the fitting did not produce a predictive method, because it was not clear how v_p, s_p, and l_p could be calculated. An atomistic protein is not a convex particle. But, we realized that, regardless of whether the test particle is convex, what matters to $\Delta\mu$ is the spatial region excluded to the crowders (the magnitude of $\Delta\mu$ depends on both the size and the shape of this region). For a convex test particle, this region is the same as the physical space occupied by the particle. However, for a nonconvex test particle, the particular region excluded to a crowder depends on the latter's size and shape.

In this work, we model the test protein at the atomic level but the crowders as spheres [Fig. 1(a)], like in previous studies using MD simulations [12,23–27]. For spherical crowders at a given radius, the test protein's region of exclusion is enclosed by what we refer to as the crowder-exclusion surface. This surface is the same as the solvent-exclusion surface, except here the solvent probe is a crowder. (The solvent-exclusion surface when calculated with a 1.4-Å solvent probe is better known as the molecular surface [35]; it consists of contact parts, which are convex, and reentrant parts, which are concave.) It is obvious to identify v_p as the volume enclosed by the crowder-exclusion surface and s_p as the area of this surface. This identification is similar in spirit to the morphometric approach of Roth *et al.* [36]. For the linear size l_p, we looked for a quantity that would measure how far the different parts of the crowder-exclusion surface are dispersed in space. Such a quantity seemed to be pro-

vided by the radius of gyration, r_g, of points uniformly distributed on the surface. We therefore assumed that $l_p \propto r_g$. To determine the proportionality constant, we noted that, for a spherical test protein, l_p is known to be its radius, and r_g is also calculated to be the radius; therefore, we set the proportionality constant to be 1, i.e., $l_p = r_g$. (Further details on the calculation of r_g, including a comparison of r_g and the FMT result for l_p can be found in the supplementary information [37].) In contrast to the fact that the morphometric approach has no predictive power, we show below that Eq. (1), with v_p, s_p, and l_p calculated according to the crowder-exclusion surface, gives accurate predictions of $\Delta\mu$.

Existing codes for generating the molecular surface are either too slow or do not work at all for the range of probe radius of interest here. We therefore devised a simple and fast procedure to generate the crowder-exclusion surface. The basic idea is to place a crowder probe at many positions (totaling N) around the test protein, where the crowder comes into contact with the test protein. These positions are along rays with uniformly separated directions emanating from the geometric center of the test protein. Along each ray the crowder is positioned at the farthest point where it is tangent to at least one protein atom. From this position, the ray is traced backward for a length equal to the crowder radius, arriving at the crowder surface. That final location is taken to be a sample point on the crowder-exclusion surface. The solid angle around each of the N directions is $4\pi/N$. If the radial distance of sample point i is r_i, then the surface area, Δs_i, accorded to the sample point is $(4\pi/N)r_i^2$. Because the r_i and hence Δs_i values are not uniform, the sample points are not uniformly distributed on the crowder-exclusion surface. When averaging over the surface, Δs_i must be introduced as a weight.

To calculate v_p, the volume inside the sphere with radius equal to the largest value of r_i is discretized into a cubic grid. A voxel in the grid is eliminated when either it is separated from every atom by more than the sum of the atomic radius and crowder radius or it is inside the crowder probe while the latter is placed at each of the N positions. The remaining voxels make up v_p. The calculation of s_p follows an algorithm proposed by Windreich *et al.* [38]. Basically, each voxel comprising v_p is assigned a weight according to how its six faces are buried by its neighbors. (Our v_p and s_p results are virtually identical to those obtained from an existing method [39] but require substantially less CPU times; see supplementary information [37].) Finally l_p is calculated as the radius of gyration of sample points on the crowder-exclusion surface:

$$l_p^2 = \frac{\sum_{i,j} \Delta s_i \Delta s_j r_{ij}^2}{2\sum_{i,j} \Delta s_i \Delta s_j}, \tag{6}$$

where r_{ij} are the distances between sample points and the weight factors Δs_i account for the non-uniform distribution of the sample points.

In our postprocessing approach to crowding simulations [12,26], the motions of the test protein and those of the crowders are followed in two separate simulations. For each representative conformation of the protein in a given state, $\Delta\mu$ is then calculated by the insertion procedure over different snapshots of crowder configurations. Finally the results for $\exp(-\Delta\mu/k_BT)$ are averaged over protein conformations and over crowder configurations. With the GFMT, the crowder simulation is no longer needed. Instead, for each protein conformation, we calculate v_p, s_p, and l_p for the given crowder size and then use Eq. (1) to calculate $\Delta\mu$.

III. APPLICATIONS

We illustrate the predictive power of the GMFT for crowding effects on two central biophysical problems: the folding equilibria of individual proteins and the binding

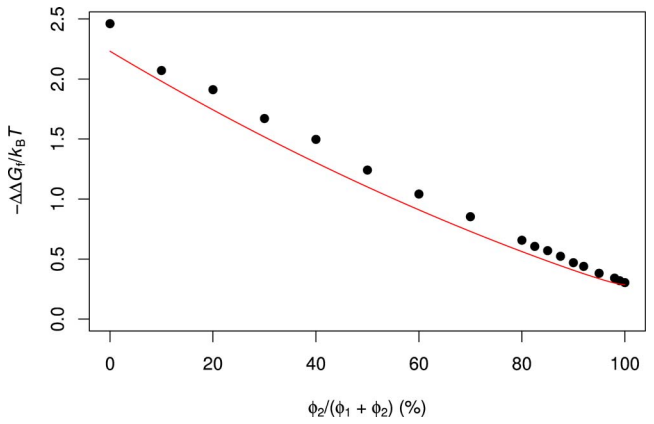

FIG. 3. (Color online) Effects of mixed crowding on the folding free energy of cytochrome b_{562}. Two species of crowders, with 15-Å and 50-Å radii, are mixed at different ratios. Symbols are obtained in our previous work [26] by the insertion procedure, while curves are predictions of the GFMT.

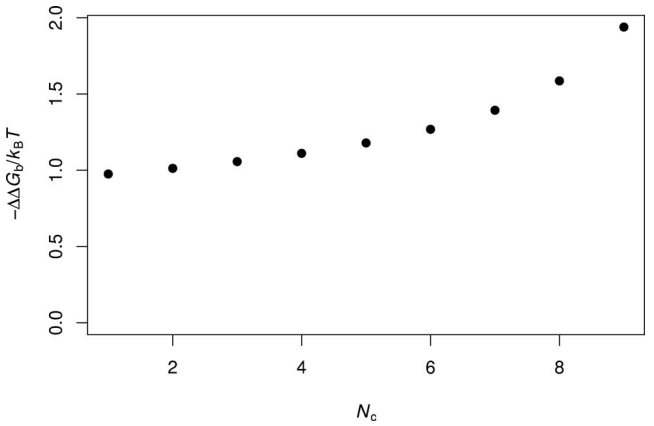

FIG. 4. GMFT predictions for the effects on the binding free energy of barnase and barstar by increasingly complex mixtures of crowders. The total volume fraction of crowders is fixed at 35%. N_c denotes the number of crowder species. At $N_c=1$, all crowders have a 50-Å radius. At $N_c=2$, crowders with 50-Å and 45-Å radii are mixed. Each successive increase in N_c introduces a new species that is 5 Å smaller than the previous smallest crowders. At $N_c=9$, the crowder radii range from 10 to 50 Å. At each N_c, all the species have the same volume fraction.

TABLE I. Average values of v_p, s_p, and l_p for the test systems.

Proteins	v_p [a] ($Å^3$)	s_p [a] ($Å^2$)	l_p [a] (Å)
Folded cytochrome b_{562}	21022.4; 22861.1; 24013.8	4462.2; 4525.9;4586.0	21.1; 21.4; 21.8
Unfolded cytochrome b_{562}	25674.3; 29931.6; 32897.8	5711.0; 5875.4; 6026.7	24.5; 25.4; 25.9
Barnase	21393.7; 23301.2; 24377.7	4326.7; 4409.4; 4473.2	19.7; 20.2; 20.5
Barstar	17132.1; 18203.7; 18809.9	3569.2; 3617.4; 3654.4	17.6; 17.9; 18.1
Barnase-barstar complex	39011.1; 42311.1; 44266.2	6577.8; 6673.7; 6752.6	24.5; 25.2; 25.6

[a]Averaged values for v_p, s_p, and l_p are listed for three crowder sizes: 15 Å, 30 Å, and 50 Å.

equilibria of proteins pairs. The test systems, cytochrome b_{562} in the folded and unfolded states and barnase and barstar in the unbound and bound states, are the same as those in our previous work presenting the postprocessing approach [26]; the same protein conformations generated in the absence of crowders are used here (additional information is found in supplementary information [37]). In Table I, we list the averaged values of v_p, s_p, and l_p (over conformations) for the test systems, calculated for three crowder sizes ranging from 15 to 50 Å. The averaged v_p, s_p, and l_p increase with increasing crowder size, as expected. Note that we do not use these averaged values in Eq. (1) to calculate the final $\Delta\mu$; rather, we use the v_p, s_p, and l_p values from the different protein conformations to obtain individual results for $\Delta\mu$ and then average the results for $\exp(-\Delta\mu/k_BT)$. The agreement between the GFMT predictions on $\Delta\mu$ and those obtained in our previous work [26] by the insertion procedure is illustrated in Fig. 1(b).

In vitro experiments [5–9] and MD simulations [23,25,26,29] have found modest enhancements of protein folding stability by crowding; similar effects have been found on the binding stability of small proteins [12]. Figure 2 shows that, when the difference in $\Delta\mu$ is taken between the folded and unfolded states of cytochrome b_{562} or between the bound and unbound states of barnase and barstar, the modest increases in folding stability and binding stability found by the postprocessing approach in our previous work [26] are still predicted well by the GFMT.

The concentrated intracellular milieu is comprised of many species of macromolecules, each at a low concentration. As a step toward realistically mimicking *in vivo* environments, we have studied in *in vitro* experiments the effect of a mixture of two crowding agents on folding stability [9]. By using the postprocessing approach for simulations of crowding, we also studied the effect of mixing crowders of two different sizes [26]. As shown by Table I, the v_p, s_p, and l_p parameters for atomistic proteins depend on crowder size. To signify this dependence, we denote by $v_{p\alpha}$ the volume of the test protein excluded to a species α crowder, and similarly by $s_{p\alpha}$ and $l_{p\alpha}$ the crowder-species specific area and linear size. When a mixture of crowders is present, we propose the following mixing rule:

$$q_p = \sum_\alpha (c_\alpha/c) q_{p\alpha}, \tag{7}$$

where $q=v$, s, or l. Figure 3 shows that the resulting predictions for the effects of a mixture of 15-Å and 50-Å crowders on folding stability are in good agreement with results obtained by the postprocessing approach in our previous work [26]. As we emphasized previously [26], the stabilization effect of the mixture is greater than the sum of the two constituents. This nonadditive effect of mixed crowding is captured by the GFMT.

We have demonstrated that the GFMT gains significant computational speed over the insertion procedure without loss of accuracy. The speed up may be particularly useful for exhaustive sampling of states (such as the unfolded state of a protein; Qin *et al.*, to be published) comprised of vastly different conformations. More importantly, the GFMT applies equally well to situations where the insertion procedure faces tremendous challenges. Examples of such situations are large protein oligomers and mixtures of many crowder species. A particularly large oligomeric complex is the ribosome, which carries out the essential function of protein synthesis. There is experimental evidence that macromolecular crowding may significantly shift the association/dissociation equilibrium of the large and small subunits of the ribosome [4]; our preliminary calculation with the GFMT indicates that macromolecular crowding may stabilize the ribosome against dissociation by as much as 40 k_BT. The GFMT can similarly be used to model at the atomic level protein aggregation and polymerization under crowding; there the effects of crowding have been found to be especially dramatic [3,15–20]. We will also be able to model mixtures of many crowder species, as illustrated in Fig. 4. Future work will be directed at further development of the GFMT when crowders are also represented at the atomic level.

ACKNOWLEDGMENT

This work was supported in part by Grant No. GM058187 from the National Institutes of Health.

[1] S. B. Zimmerman and S. O. Trach, J. Mol. Biol. **222**, 599 (1991).
[2] H.-X. Zhou, G. Rivas, and A. P. Minton, Annu Rev Biophys **37**, 375 (2008).
[3] R. S. Fuller, J. M. Kaguni, and A. Kornberg, Proc. Natl. Acad. Sci. U.S.A. **78**, 7370 (1981).
[4] S. B. Zimmerman and S. O. Trach, Nucleic Acids Res. **16**, 6309 (1988).
[5] Y. Qu and D. W. Bolen, Biophys. Chem. **101-102**, 155 (2002).
[6] D. S. Spencer *et al.*, J. Mol. Biol. **351**, 219 (2005).
[7] X. Ai *et al.*, J. Am. Chem. Soc. **128**, 3916 (2006).
[8] A. Roberts and S. E. Jackson, Biophys. Chem. **128**, 140 (2007).
[9] J. Batra, K. Xu, and H. X. Zhou, Proteins **77**, 133 (2009).
[10] J.-M. Yuan *et al.*, Protein Sci. **17**, 2156 (2008).
[11] D. Homouz *et al.*, Proc. Natl. Acad. Sci. U.S.A. **105**, 11754 (2008).
[12] J. Batra *et al.*, Biophys. J. **97**, 906 (2009).
[13] D. Drenckhahn and T. D. Pollard, J. Biol. Chem. **261**, 12754 (1986).
[14] Y. Y. Kuttner *et al.*, J. Am. Chem. Soc. **127**, 15138 (2005).
[15] M. D. Shtilerman, T. T. Ding, and P. T. Lansbury, Jr., Biochemistry **41**, 3855 (2002).
[16] V. N. Uversky *et al.*, FEBS Lett. **515**, 99 (2002).
[17] D. M. Hatters, A. P. Minton, and G. J. Howlett, J. Biol. Chem. **277**, 7824 (2002).
[18] M. del Alamo, G. Rivas, and M. G. Mateu, J. Virol. **79**, 14271 (2005).
[19] M. Rotter *et al.*, J. Mol. Biol. **347**, 1015 (2005).
[20] K. Snoussi and B. Halle, Biophys. J. **88**, 2855 (2005).
[21] J. L. Lebowitz, E. Helfand, and E. Praestgaard, J. Chem. Phys. **43**, 774 (1965).
[22] A. P. Minton, Methods Enzymol. **295**, 127 (1998).
[23] M. S. Cheung, D. Klimov, and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **102**, 4753 (2005).
[24] D. D. Minh *et al.*, J. Am. Chem. Soc. **128**, 6006 (2006).
[25] L. Stagg *et al.*, Proc. Natl. Acad. Sci. U.S.A. **104**, 18976 (2007).
[26] S. Qin and H.-X. Zhou, Biophys. J. **97**, 12 (2009).
[27] D. L. Pincus and D. Thirumalai, J. Phys. Chem. B **113**, 359 (2009).
[28] S. Qin *et al.*, J Phys. Chem. Lett. **1**, 107 (2010).
[29] J. Mittal and R. B. Best, Biophys. J. **98**, 315 (2010).
[30] H. Tjong and H.-X. Zhou, Biophys. J. (to be published).
[31] B. Widom, J. Chem. Phys. **39**, 2808 (1963).
[32] Y. Rosenfeld, Phys. Rev. Lett. **63**, 980 (1989).
[33] S. M. Oversteegen and R. Roth, J. Chem. Phys. **122**, 214502 (2005).
[34] J. Wu and Z. Li, Annu. Rev. Phys. Chem. **58**, 85 (2007).
[35] F. M. Richards, Annu. Rev. Biophys. Bioeng. **6**, 151 (1977).
[36] R. Roth, Y. Harano, and M. Kinoshita, Phys. Rev. Lett. **97**, 078101 (2006).
[37] See supplementary material at http://link.aps.org/supplemental/10.1103/PhysRevE.81.031919 for further details on the calculation of the geometric parameters v_p, s_p, and r_g and on the proteins studied here.
[38] G. Windreich, N. Kiryati, and G. Lohmann, Pattern Recognit. **36**, 2531 (2003).
[39] N. R. Voss *et al.*, J. Mol. Biol. **360**, 893 (2006).