CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Elsevier - Publisher Connector

Anisotropy of the Coulomb Interaction between Folded Proteins: Consequences for Mesoscopic Aggregation of Lysozyme

Ho Yin Chan,[†] Vladimir Lankevich,[¶] Peter G. Vekilov,[‡§] and Vassiliy Lubchenko[§†*]
Departments of [†]Physics, [‡]Chemical and Biochemical Engineering, and [§]Chemistry, University of Houston, Houston, Texas; and [¶]Department of Physics and Astronomy, University of Rochester, Rochester, New York

ABSTRACT Toward quantitative description of protein aggregation, we develop a computationally efficient method to evaluate the potential of mean force between two folded protein molecules that allows for complete sampling of their mutual orientation. Our model is valid at moderate ionic strengths and accounts for the actual charge distribution on the surface of the molecules, the dielectric discontinuity at the protein-solvent interface, and the possibility of protonation or deprotonation of surface residues induced by the electric field due to the other protein molecule. We apply the model to the protein lysozyme, whose solutions exhibit both mesoscopic clusters of protein-rich liquid and liquid-liquid separation; the former requires that protein form complexes with typical lifetimes of approximately milliseconds. We find the electrostatic repulsion is typically lower than the prediction of the Derjaguin-Landau-Verwey-Overbeek theory. The Coulomb interaction in the lowest-energy docking configuration is nonrepulsive, despite the high positive charge on the molecules. Typical docking configurations barely involve protonation or deprotonation of surface residues. The obtained potential of mean force between folded lysozyme molecules is consistent with the location of the liquid-liquid coexistence, but produces dimers that are too short-lived for clusters to exist, suggesting lysozyme undergoes conformational changes during cluster formation.

MOTIVATION

Protein aggregation is a central problem of biophysics, medicine, and bioengineering (1–6). Thus, better understanding of the thermodynamics of protein phase behavior and the kinetic pathways leading to the formation of protein condensed phases will lead to progress in several fields. Although a protein solution is essentially a two-component—protein plus buffer—mixture, its phase behavior is significantly more complex than the standard textbook picture of binary mixtures (7). For instance, the phase diagram of the protein lysozyme, in addition to two solubility lines corresponding to the rhombohedral and tetragonal crystal phases, contains an additional liquid-liquid coexistence region, which is often accessible before crystallization takes place (8–11).

Even more surprising is that well outside the stability region of the protein-rich phase, protein solutions often host compact inclusions of a protein-rich liquid that are mesoscopic in size, i.e., ~100 times larger than individual proteins (12,13). We have called these inclusions "mesoscopic clusters". Recent analysis by Pan et al. (14) shows indirectly that the clusters result from the formation of transient protein-containing complexes; the complexes are strongly stabilized at protein concentrations typical of the dense protein liquid. This analysis combines the experimentally determined free energy cost of increasing protein concentration–in the homogeneous region of the phase diagram– and classical nucleation theory to show that clusters consisting of a uniform concentrated solution of protein monomers could contain only a few molecules. On the other hand, a diffusion-reaction scheme that includes the possibility of the formation of transient complexes yields that clusters of protein-rich solution should exist with a radius R_{cl} determined by the lifetime τ and diffusivity D of the complexes in the bulk solution:

$$R_{cl} \approx (D\tau)^{1/2}. \quad (1)$$

This microscopic scenario is supported by a number of observations ((14) and Y. Li, V. Lubchenko, and P. G. Vekilov, unpublished). Yet the identity of the complexes and the mechanism of their formation remain unknown. A particularly important aspect of this mechanism is whether the protein molecules undergo partial unfolding or conformational changes during the complex formation. Partial unfolding could result in attraction between solvent-exposed hydrophobic residues or even domain swapping (16). Resulting complexes would have relatively long lifetimes, which, by Eq. 1, could lead to the observed mesoscopic cluster sizes. On the other hand, recent study of γS-crystallin (17) suggests a higher propensity to aggregate when the native structure is more conformationally rigid.

The goal of this work is to establish whether folded protein molecules can form a complex stable enough to give rise to the mesoscopic clusters. This is a necessary first step before addressing the significantly more difficult case of protein binding accompanied by conformational changes. A common, computationally simple model for describing protein-protein interaction is the Derjaguin-Landau-Verwey-Overbeek (DLVO) theory (18,19). In this

Submitted October 14, 2011, and accepted for publication March 2, 2012.

*Correspondence: vas@uh.edu

Editor: Nathan Baker.

© 2012 by the Biophysical Society

0006-3495/12/04/1934/10 $2.00

doi: 10.1016/j.bpj.2012.03.025

model, the total interaction is modeled semiphenomenologically as a sum of screened Coulomb repulsion between two uniformly charged spheres, short-range dispersive attraction, and hard-core repulsion (20,21). This simplified picture is expected to overestimate the electrostatic repulsion, for the charges on the protein surface are of both signs and are distributed nonuniformly. Protein molecules thus can adjust their mutual orientation and substantially mitigate the Coulomb repulsion even if they carry a net charge, a possibility neglected in the DLVO model. Elcock and colleagues (22,23) have developed a microscopically inspired, continuum approach to protein-protein interactions that accounts for the anisotropy of the Coulomb interactions and protein shape, and desolvation penalty for the protein surface.

Even in the unfavorable case of two like charges facing each other during a binding event, deprotonation or protonation of one of the residues in question may occur, thus neutralizing that residue and removing the source of repulsion. Similar acid-base chemistry might result in creating additional pairs of opposite charges that would stabilize the protein complex. Consistent with these notions, various studies have shown water can mediate attractive interactions even between like-charged residues in protein complexes (24–26).

Here, we develop a computationally efficient model that accounts for the anisotropy of the Coulomb component of protein-protein interaction and the possibility of charge regulation with respect to protons. We then apply this methodology to a pair of lysozyme molecules; lysozyme exhibits many of the protein aggregation phenomena, including the mesoscopic clusters and liquid-liquid separation. Our approach allows one to map out the full set of mutual configurations of the proteins and use those to estimate the lower limit on the rate of dissociation of a typical dimer of folded lysozyme molecules. We establish that although two folded molecules can form a relatively stable dimer when they are oriented in a special way, a typical dimer is too short-lived to yield mesoscopic clusters. We then use this result to argue that the majority of protein molecules must undergo conformational changes during cluster formation. In addition, we establish that the DLVO model overestimates the repulsion between the proteins, implying that accounting for the interaction anisotropy is essential for quantitative description of protein aggregation, consistent with conclusions derived from "patch models" of protein-protein interaction (27).

THE MODEL

We consider two folded protein molecules. We wish to test for the effects of the mutual orientation of the protein molecules, stemming from the nonuniform charge distribution, while including the possibility of changes in the charge state of the surface residues. To separate these effects from the anisotropy induced by the complicated shape of the protein and with computational efficiency in mind, we assume that the protein molecules are exactly spherical with radius R_p; the ensuing error is quantified below. The radius R_p is chosen so that the volume of the sphere is equal to that of the protein molecule; for lysozyme we adopt $R_p = 1.7$ nm (28). In the model, the total interaction consists of the Coulomb interaction (subject to the Debye screening by the mobile ions in the solution), short-range attraction (due to dispersion and other interactions), and steric repulsion.

Coulomb interactions

Charged residues are represented with point charges located at a depth b beneath the surface of the sphere; this depth b is assumed to be the same for all residues. The charges are located at the same latitude and longitude as in the actual protein molecule; the coordinate center is at the molecule's center of mass. This model harks back to the venerable Tanford-Kirkwood model (29), with the difference that the charge locations are not random but mimic those of the actual protein. If the charge of a residue is distributed over two or three atomic sites, we employ the following procedure: for a protonated arginine, we place charges of $+1/3$ at each of the three nitrogens of the guanidinium group; for a deprotonated aspartate or glutamate, we place $-1/2$ charges at the two oxygens; and for protonated histidine, we place two $+1/2$ charges at the nitrogens of the imidazole group. The resulting charge distribution is displayed in Fig. S1 in the Supporting Material. A finite value of b reflects spatial distribution of charge on a residue and ruggedness of the protein surface. We use two specific values of the depth b, 1.5 and 2 Å, because salt bridges form at separations between the centroids of charged groups ranging between 3 and 4 Å (30).

The charge states of the ionizable residues depend on the pH of the solvent and on the proximity of the other protein, as nearby sources of electrostatic field effectively modify the pK_a value of such residues. To test for the possibility of protonation or deprotonation of ionizable residues, we assume each of them can have exactly two alternative charge states even though more states are possible, in principle; this assumption will turn out to be internally consistent. The charge states of the ionizable residues are determined as follows: First, they are divided into three groups. Group I consists of Asp and Glu, group II of Arg, His, and Lys, and group III of Tyr and Cys, whose reference states, by construction, have charge –1, +1, and 0, respectively. For each ionizable residue i, introduce a variable s_i equal to –1, if the residue belongs to Group I, or +1, if it belongs to Group II or III. In the absence of the other protein, the free energy cost to switch from the protonated (deprotonated) to deprotonated (protonated) state for a Group II and III (Group I) residue, is given by

$$\Delta G_i^{(p)} = s_i \ln(10)(pK_{a,i} - pH)k_B T, \quad (2)$$

where $pK_{a,i}$ is the pK_a value of residue i for an isolated protein molecule. These pK_a values are determined using the package PROPKA3 (31–34). When a source of external field is present—such as resulting from another protein and/or mobile ions from the buffer—the distribution of the variable s_i in Eq. 2 is subject to the energy of the corresponding charge in that field, in addition to the cost $\Delta G_i^{(p)}$ itself. We note that free ions by themselves could affect the pK_a value: According to Wang et al. (35), the resulting pK_a change of a fully solvent accessible ionizable group could change by up to one unit, at the ionic strengths in question. Here, in contrast, we consider a fully folded molecule, implying a relatively low solvent accessibility. In addition, we will see later that the effects of (de)protonation self-consistently turn out to be small. Now, in the presence of another protein molecule, the free energy of a specific protonation configuration is computed according to

$$G^{(p)} = \sum_i \Delta G_i^{(p)} - \sum_i \Delta G_{i,\infty}^{(p)}, \quad (3)$$

where $\Delta G_{i,\infty}^{(p)}$ is the thermally averaged value of $\Delta G_i^{(p)}$ at infinite separation between the proteins and is calculated by Boltzmann averaging Eq. 2 with respect to the two alternative values of s_i with the "energy" parameter equal to $\Delta G_i^{(p)}$ itself.

The Coulomb component of the interaction between the proteins is determined by the instantaneous values of the charges and fluctuates in time. In an "adiabatic" limit of solvent motions being infinitely faster than those of the proteins, the Coulomb interaction is stationary and is equal to its instantaneous value averaged over 2^n configurations, where n is the total number of ionizable residues. Averaging over all 2^n configurations is computationally costly yet usually unnecessary: Only those residues sufficiently close to the interfacial region might appreciably affect each other, owing to the decay of the Coulomb interaction with distance, also enhanced by the Debye screening. Specifically, here we explicitly sample the distinct charge states only on four residues, two residues per each protein molecule. The latter are chosen to be the closest to the midpoint between the proteins, for each orientation. This "interface" subset of residues has $2^4 = 16$ possible protonation states. Any quantity of interest is averaged over these 16 configurations, with corresponding Boltzmann weights. If needed, the subset of the interface residues can be increased (at a higher computational cost).

The pK_a values of those residues not in the interface region are only weakly affected by the other protein. The charge on such residues is assumed to be equal to that on an isolated protein; we will see below this assumption is internally consistent. The average value of this charge is thus equal to

$$z_i = \begin{cases} \dfrac{s_i}{1 + 10^{s_i(pH - pK_{a,i})}} \text{ for Group I and II} \\ \dfrac{-s_i}{1 + 10^{-s_i(pH - pK_{a,i})}} \text{ for Group III,} \end{cases} \quad (4)$$

where $pK_{a,i}$ is the pK_a of residue i on an isolated protein.

The energy of the Coulomb interaction between the two protein molecules in specific orientations is Boltzmann-averaged over the protonation states, also subject to the free energy cost $G^{(p)}$ from Eq. 3,

$$E_{\text{Coulomb}} = \left\langle \frac{1}{2} \sum_i z_i \varphi_i + G^{(p)} \right\rangle_{\text{prot}}, \quad (5)$$

where the summation is over all charges. The angular brackets denote averaging over the 16 distinct protonation states of the interfacial residues with the Boltzmann weight $(1/Z)e^{-E/k_B T}$, where E is the expression inside the angular brackets and Z is the corresponding partition function. The electrostatic potential φ_i on charge i, due to the other protein, is estimated in the Debye-Hückel approximation modified to account for the effects of the dielectric discontinuity,

$$\varphi_i = \sum_j \frac{z_j}{4\pi\varepsilon_0\varepsilon_S} \left(\frac{e^{-\kappa d_{ij}}}{d_{ij}} \right) p_{ij}, \quad (6)$$

where the indices i and j pertain to different protein molecules. In Eqs. 5 and 6, if a residue belongs to the interfacial subset, its charge is set equal to its instantaneous value; otherwise it is computed according to Eq. 4. The distance between residues i and j is denoted with d_{ij}, the dielectric susceptibilities of the vacuum and solvent with ε_0 and ε_s, respectively. The inverse Debye length is computed as

$$\kappa = \sqrt{\frac{2N_A e^2 I}{\varepsilon_0 \varepsilon_S k_B T}},$$

where I stands for the ionic strength of the solution and N_A is Avogadro's number. Here, we take $I = 30$ mM, $T = 298$ K to match the experimental conditions in Pan et al. (14), yielding a Debye length $\kappa^{-1} \approx 0.57$ nm.

The Yukawa-like interaction in the round brackets in Eq. 6 corresponds to the interaction between two pointlike charges in the Debye-Hückel theory. This is an approximation, because the charges are buried inside a sphere with a dielectric constant significantly below that of the solvent. In the Supporting Material, we illustrate the effect of the dielectric discontinuity at the protein-solvent interface and the interface curvature on this interaction. We show that the repulsion between interfacial like-charges can be

significantly enhanced by the dielectric discontinuity, if they face each other, while the corresponding effect on the attraction between opposite charges is modest. The correction factor p_{ij} in Eq. 6 approximately accounts for those effects; it is estimated in the Supporting Material. The correction factor is equated to unity whenever at least one of the residues i or j is at an angular distance from the midpoint between the proteins that exceeds a specific cut-off value θ_c. This value is chosen to be 90°, based on the comparison of the values of the electrostatic potential resulting from Eq. 6 with those produced by numerical solution of the Poisson-Boltzmann equation, which is described at the end of this section. We note that solvation/desolvation of the surface residues is accounted for in the present, continuum approximation, by including the dielectric discontinuity effects. Finally note that there is no explicit contribution of the mobile ions to the interaction energy between the protein molecules at the Debye-Hückel level (36), and so they enter the PMF only through the Debye screening and, possibly, charge regulation (see below).

Dispersion, steric, and other interactions

We model the effective potential stemming from the non-Coulomb interactions by a functional form that smoothly interpolates between two distinct behaviors, as pertinent at long and short separation r_s between the protein surfaces,

$$E_{\mathrm{mol}} = \begin{cases} E_>, r_s > r_2 \\ E_<, r_s < r_1, \end{cases} \tag{7}$$

where $r_1 < r_2$ and a fifth-degree polynomial is used to patch the short- and long-distance behaviors so that the derivatives of order two and below are continuous. At relatively long distances, $r_1 > r_2$, the interaction is designed to match that between two polarizable spheres (37),

$$E_> = -\frac{A_H}{12}\left(\frac{1}{(x+1)^2} + \frac{1}{(x^2+2x)} + 2\ln\left(\frac{x^2+2x}{(x+1)^2}\right)\right), \tag{8}$$

where $x \equiv r_s/2R_p$ is the distance between their surfaces divided by the sphere diameter. The Hamaker constant A_H is set equal to 3 $k_{\mathrm{B}}T$, which is close to its theoretically estimated value for lysozyme (38,39).

At short distances, $r_s < r_1$, the non-Coulomb interactions are dominated by dispersion interaction between surface residues and attraction between hydrophobic patches, if any. Additional sources of attraction may also be present, including: depletion interaction caused by HEPES and charge regulation via transient binding of mobile ions from the buffer (20). The latter appears to be significant and system-dependent (40). This study focuses on cluster formation in HEPES, which is a "mild buffer" (41) and is not expected to bind to or modify the protein surface much. We account for those attractive forces and the steric repulsion phenomenologically, via a Lennard-Jones like functional form (42),

$$E_< = 4\varepsilon\left[\left(\frac{\sigma}{r_s+\delta}\right)^{2\alpha} - \left(\frac{\sigma}{r_s+\delta}\right)^{\alpha}\right], \tag{9}$$

where ε is the depth of the minimum and r_s is the separation between the protein surfaces. The parameters α, δ, σ, and $r_{1,2}$ are chosen to satisfy the following constraints: 1), The surface of the protein is at its van der Waals location, i.e., $r_s = 0$, which yields $\delta = \sqrt[\alpha]{2\sigma}$. 2), The curvature at the minimum of $E_<$ equals 300 nm^{-2}, the latter figure comparable to those observed for ammonium and methane (43,44). Once Constraints 1 and 2 are satisfied, the region in which $E_<$ and $E_>$ intersect is rather insensitive to the precise parameter values; the latter values adopted in the article are given in the Supporting Material.

Potential of mean force

The potential of mean force (PMF) for each orientation of the two proteins is computed by adding together the full Coulomb interaction E_{Coulomb} from Eq. 5 and E_{mol} from Eq. 7,

$$E_{\mathrm{PMF}} = E_{\mathrm{Coulomb}} + E_{\mathrm{mol}}, \tag{10}$$

c.f. the decomposition of the protein-protein PMF by Elcock and McCammon (22).

Given the value of the PMF, the second viral coefficient, B_{22}, is computed in the standard fashion (45,46),

$$B_{22} = -\frac{2\pi N_A}{M^2}\left\langle \int_0^\infty \left(e^{-E_{\mathrm{PMF}}/k_{\mathrm{B}}T} - 1\right) r_c^2 dr_c \right\rangle, \tag{11}$$

where M is the protein mass and

$$r_c \equiv r_s + 2R_p \tag{12}$$

is the distance between the proteins' centers of mass. The angular brackets in Eq. 11 denote averaging with respect to the orientation of the proteins, which is performed here by discrete summation, see details in the Supporting Material.

Tests of the electrostatic potential

In Fig. 1, we compare the potentials determined from Eq. 6 for all charged surface residues, to the corresponding potentials obtained by solving the nonlinear Poisson-Boltzmann (PB) equation with the Adaptive Poisson Boltzmann Solver (APBS) software package (47) for a pair of lysozyme

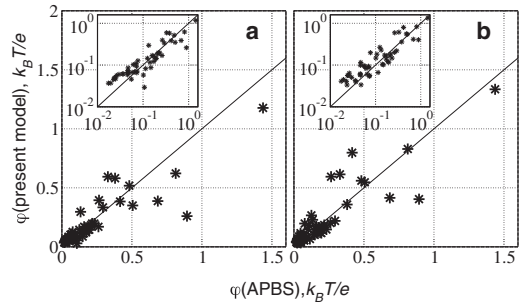

FIGURE 1 Electrostatic potentials on individual residues obtained using our model ($b = 1.5$ Å) and those computed with APBS are plotted against each other (see text and Fig. S1 and Fig. S2 in the Supporting Material for detailed description of the configuration). Only the portions that depend on the interprotein distance are shown. (*Insets*) Replot of the same graphs on the log-log scale to better show the small values of φ. (*a* and *b*) Spherical and actual shape, respectively. Analogous comparisons for $b = 2$ Å are given as Fig. S4, *a* and *b*.

molecules. In the APBS calculation, the size of the grid box is 160.5 × 80.5 × 80.5 Å^3, the grid spacing 0.5 Å, $T =$ 298.15 K; $I = 30$ mM as provided by Na^+ and Cl^- ions; the dielectric susceptibilities of protein and water are 2 (Patargias et al. (48)) and 78.54, respectively. The protein structure, as determined in Wang et al. (49), is obtained from the Protein Data Bank as PDB:2VB1. The residues' protonation states were determined using PROPKA3 (31–34) and imported using PDB2PQR (50) packages (see Fig. S2). Because we are interested in the potential values near charges, where the effect of the finite grid size is most significant, we obtain this potential by averaging its value in nine points: the closest grid point to the charge ("point 1") and the eight grid points closest to point one in directions (± 1, ± 1, ± 1). The position of the charges were determined using the BioMagResBank database (51). Note APBS places charges at the coordinates of actual atoms, not at a fixed depth below the surface, which is a potential source of discrepancy with the predictions of our model.

APBS yields the full value of the electrostatic potential, which also includes the field from the residues in the same protein molecule and the polarization charge at the two protein-solvent interfaces. To infer the portion φ of the potential that depends exclusively on the interprotein separation, we compute the full potential as a function of the separation and then subtract from it its value at the infinite separation, which is obtained by solving the PB equation for an isolated protein molecule, as illustrated in Fig. S3.

In the configuration corresponding to Fig. 1, Arg^{87} faces Asp^{45}; hereby, actual proteins would face each other with convex regions (see Fig. S2). The distance between the protein surfaces is chosen at 3 Å: On the one hand, this way APBS can place at least one water layer above each protein and so its predictions are representative of the actual values of the potential and are less affected by the continuum nature of the treatment. On the other hand, the 3 Å separation is small enough that the resulting range of the potential values is sufficiently broad to allow for a meaningful comparison. In Fig. 1 *a*, we compare the potentials computed with our model, using the cut-off angle $\theta_c = 90°$, and with APBS. In Fig. 1 *b*, we compare the potentials computed using our model without the sphericity assumption and the APBS produced potentials. In Fig. 2, we show the dependence of the root mean square on the scatter graphs in Fig. 1 on the cut-off angle θ_c. According to Figs. 1 and 2, the sphericity assumption introduces a quantitative, not qualitative error. Fig. 2 justifies using $\theta_c = 90°$ as yielding the smallest overall error in the spherical case and a relatively small error in the actual-shape case. Overall, Fig. 1 indicates that our model is a viable method for evaluating electrostatic interactions between large convex molecules.

RESULTS AND DISCUSSION

The potentials of mean force (PMF) computed using our approximation are presented in Fig. 3, *a* and *b*, for *b* equal to 1.5 and 2 Å, respectively. The thick solid line shows the total PMF from Eq. 10 as a function of the distance between the centers of the protein molecules r_c, averaged over the orientations of the molecules with corresponding Boltzmann weights. The thin solid line shows the minimum energy at each separation, i.e., at the orientation at which the Coulomb term from Eq. 5 has is lowest value, for each value of r_c. The Coulomb and molecular terms, from Eqs. 5 and 7, are shown by the dashed and dash-dotted lines, respectively. The electrostatic components are shown in the insets of Fig. 3, *a* and *b*, where they are compared with the corresponding DLVO values, which were computed using Eq. 15 of Muschol and Rosenberger (52); see also Petsev et al. (21).

The specific values of the depth ε of the molecular term from Eq. 7 were chosen based on the following reasoning: We first compute the model's upper bound on the second

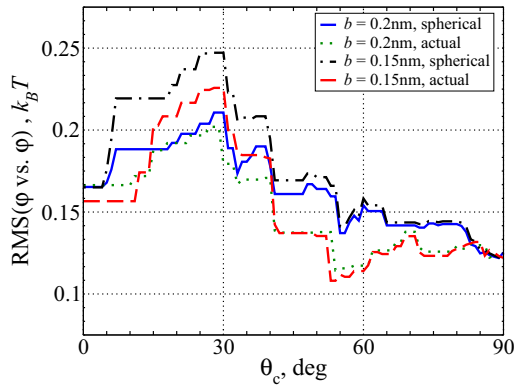

FIGURE 2 Root mean-square deviation from the perfect agreement in the comparison plots for $b = 1.5$ Å, Fig. 1, and $b = 2$ Å, Fig. S4, as a function of the cutoff angle θ_c.

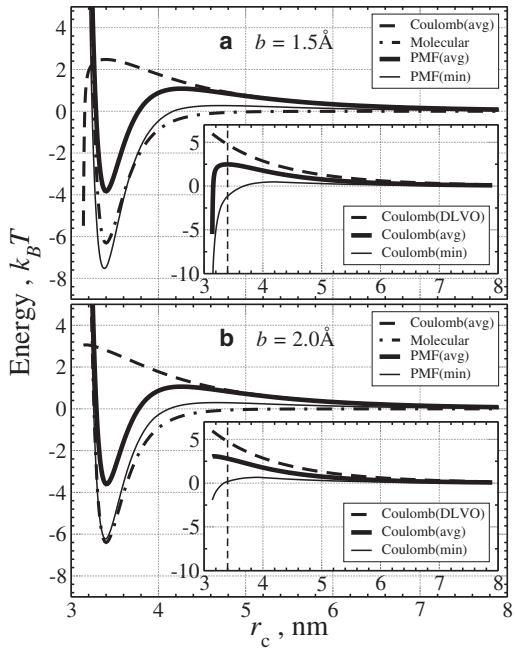

FIGURE 3 Dependence of the interaction energies for a pair of lysozyme molecules on the distance between the protein centers of mass. (*a*) b = 1.5 Å; (*b*) b = 2.0 Å. (*Insets*, *vertical dashed lines*) Separation at which the van der Waals surfaces of the molecules are in contact.

virial coefficient B_{22}, defined in Eq. 11, by including only the electrostatic component E_{Coulomb}, while assuming hard-core repulsion at $r_c = 2R_p$. The resulting value is $B_{22} \approx 1.31 \times 10^{-3}$ and 1.30×10^{-3} ml mol g^{-2} for b =1.5 and 2 Å (the hard sphere portion is 0.24×10^{-3}). This is equal, within the experimental error, to the measured value of 1.4×10^{-3} ml mol g^{-2}, determined from the low-concentration part of the Debye plot in Fig. 2 *a* of Pan et al. (14). According to Neal and Lenhoff (53), the spherical assumption results in underestimating the steric portion of B_{22} by a factor of ~1.7, which would raise our estimate of the repulsive portion of B_{22} to 1.5×10^{-3} or so. This general agreement with the experimental value is reassuring. Still, after including the attractive interactions, B_{22} will likely be below the experimental value, implying our model may somewhat underestimate the repulsive part of the interaction. Note the experimental figure was obtained with HEPES as the buffer. The HEPES anion is significantly larger than, for instance, Cl^-, suggesting a relatively low accessibility to the protein surface and the protein-protein interface, and hence a weaker Debye screening than due to pointlike ions.

Despite these complications, one may argue that the attractive portion of B_{22} is comparable to or smaller than the Coulomb portion, so that the virial coefficient remains repulsive. This notion yields an approximate lower bound on the depth ε. For the sake of concreteness, we have chosen such values of ε that lower the second virial coefficient, from its Coulomb plus hard-sphere value, by a factor of two. This way, the overall interaction is repulsive, while its numerical value is still comparable to the experimental value. At the same time, the Coulomb and molecular contributions are comparable. Note that simulations of interaction between charged and neutral amino acids indicate the depth of the attractive minimum does not exceed 10 k_BT for oppositely charged residues and is smaller otherwise (54). The values of ε chosen hereby are consistent with this figure. Another consistency check on the numerical values of the parameters of the molecular term from Eq. 7 and the depth b is that at the attractive minimum of the total PMF, the distance between the centroids of the constituents of salt bridges remains close to 3 and 4 Å (see Table 1). This is by no means a given, because the steepness of the Coulomb portion varies rather strongly with interprotein separation, as can be seen in Fig. 3, and could shift appreciably the position of the attractive minimum, which is relatively soft.

Let us now analyze in detail the Coulomb portion of the interprotein interaction. The value of the minimum energy term in Fig. 3 is noteworthy: it shows that despite the rather high net charge on each protein, i.e., ~+7.4, there is always a mutual orientation that barely exhibits repulsion at any separation. This notion is brought home by the histogram in Fig. 4, where we show the distribution of the Coulomb term E_{Coulomb} at two values of the interprotein separation, i.e., those corresponding to the minimum and maximum of the angular-averaged PMF. As expected, the distribution shifts to lower energies with the separation, although the smallest value slightly increases for the b = 1.5 Å case, at small separations. Note the repulsion, though present, is significantly lower than that prescribed by the DLVO theory. This result is of significance for simplified treatments of phase and aggregation behaviors of proteins. Note also the overall change of the Coulomb contribution is small and gradual, resulting in a low, broad barrier.

The actual ensemble of orientations relevant at the ambient temperature is given by the distribution from Fig. 4 multiplied by the corresponding Boltzmann weight. The resulting distributions, corresponding to Fig. 4, *a* and *b*, are shown in Fig. 5, parts *a* and *b*, respectively. We point out that the lowest energy configuration at the minimum of the PMF actually has an attractive Coulomb portion, for b = 1.5 Å. Not surprisingly, the contacting residues in this configuration, i.e., Lys^{116}-Asp^{52} are oppositely charged. Note that the Asp^{52} residue is partially obstructed, and so it is not clear that the hereby predicted most probable contact is actually possible. Although partial accessibility

TABLE 1 Several quantities computed for the two sets of parameters pertaining to Fig. 3, *a* and *b*

b, Å	B_{22}, 10^{-4} ml mol/g^2	d_{min}, Å	e, k_BT	τ, μs	R_c, nm
1.5	6.57	2.86	6.31	0.7–3	5–10
2.0	6.50	3.86	6.40	0.5–2	4–8

In computing the cluster radius, we assume the diffusivity $D = 4.28 \times 10^{-7}$ cm^2/s, based on viscosity 3cPs (54).

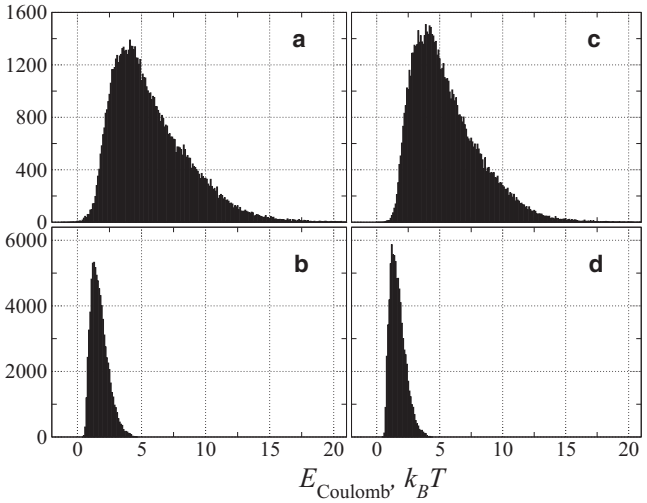

FIGURE 4 Distributions of the Coulomb term E_{Coulomb}. (*a* and *b*) At the interprotein separation corresponding to the minimum and maximum of the PMF, for the $b = 1.5$ Å case from Fig. 3 *a*. (*c* and *d*) Analogous to panels *a* and *b*, respectively, but for the $b = 2.0$ Å case from Fig. 3 *b*.

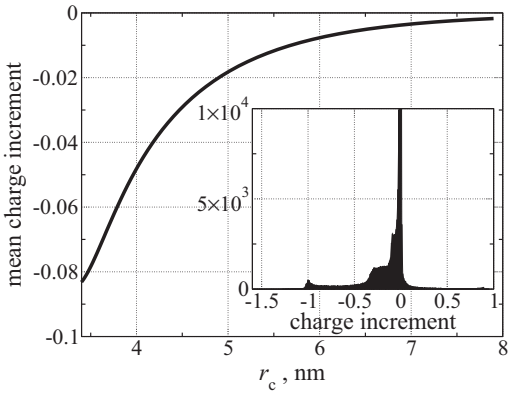

FIGURE 6 Charge increment on the two protein molecules, due to deprotonation or protonation, averaged over the mutual orientation, as a function of the interprotein separation; at $b = 1.5$ Å. The magnitude of this increment is at most four, by construction, because the interface subset contains four residues, two per each protein molecule. For instance, a figure -2 would imply two protons, on average, have detached. (*Inset*) The histogram shows the distribution of the charge increment for all orientations, at the minimum of the PMF in Fig. 3 *a*.

can be modeled by employing a greater depth b for the residue in question, we choose to use a uniform b here to assess strengths and weaknesses of the model in its simplest realization.

In Fig. 6 we display the typical number of residues whose protonation state differs from that of isolated molecules. This number is small, implying that even at close separations, the molecules find configurations that satisfy the Coulomb interactions with almost the same protonation pattern as those on isolated molecules and thus avoid the free energy penalty for protonation/deprotonation from Eq. 2. This observation is of significance for analyses of phase ordering in large assemblies of proteins, because it indicates that, to a good approximation, one can assume the protein charge depends exclusively on the pH of the solution, and not the protein concentration.

Fig. 5 shows that during dissociation, the vast majority of the angular orientations are close in energy to the average energy at the respective protein-protein separation. Thus, a much simplified picture for binding/dissociation is adequate, in which the orientations of the protein molecules are averaged out and so the only reaction coordinate left is the interprotein separation. Furthermore, the protein motions in water are strongly overdamped, implying we are in the Kramer's limit of the transition state theory. Neglecting hydrodynamic effects, the typical decay time of the complex can be then expressed as (55)

$$\tau = \frac{2\pi\zeta}{M\omega^{\ddagger}\omega_B} e^{G^{\ddagger}/k_BT}, \tag{13}$$

where the friction coefficient, ζ, is related to the viscosity of solvent (56), η, by the Stokes Law $\zeta \approx 6\pi\eta R_p$. $G^{\ddagger}$ is the altitude of the maximum of the PMF relative to its minimum, while ω_B and $\omega^{\ddagger}$ are the vibrational frequencies at the minimum and the inverted maximum at the transition state. By Eq. 1, the cluster radius can be presented as

$$R_{\text{cl}} = \left(\frac{2\pi k_{\text{B}}TD_2/D_1}{M\omega^{*}\omega_B} e^{G^{\ddagger}/k_{\text{B}}T}\right)^{1/2}, \tag{14}$$

where $D_2/D_1 \approx 0.76$ is the diffusivity of a dumbbell composed of two identical spheres relative to the diffusivity of one of those spheres (57). Note R_{cl} does not explicitly depend on the solution's viscosity.

The expressions in Eqs. 13 and 14 are underestimates because: a), the Coulomb component in the initial state of

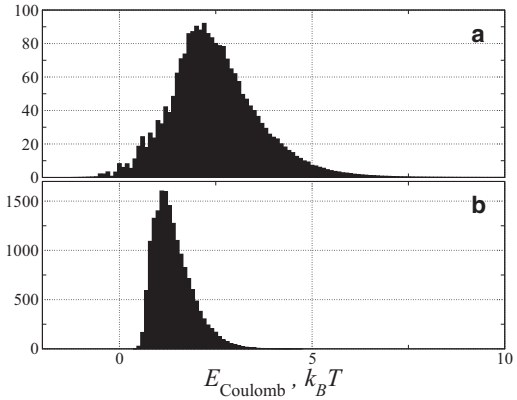

FIGURE 5 Boltzmann-weighted distribution of the Coulomb term at two values of the interprotein separation: at the minimum (*a*) and maximum (*b*) of the PMF; $b = 1.5$ Å. (*a* and *b*) Distributions from Fig. 4, parts *a* and *b*, respectively.

dissociation is distributed (see Figs. 4 and 5); b), the average lifetime is determined by averaging Eq. 13 with respect to the initial state, with the corresponding Boltzmann weight; and c), for a distributed quantity x, $\langle e^x \rangle \geq e^{\langle x \rangle}$. Thus the lifetime in Eq. 13 is underestimated, approximately, by a factor $\langle e^{-E_{\text{Coulomb}}/k_{\text{B}}T} \rangle / e^{-\langle E_{\text{Coulomb}}/k_{\text{B}}T \rangle}$, where the averaging is with respect to the distribution from Fig. 5. In Table 1, we list the lifetime following from Eq. 13 as the lower bound and the latter figure multiplied by the factor $\langle e^{-E_{\text{Coulomb}}/k_{\text{B}}T} \rangle / e^{-\langle E_{\text{Coulomb}}/k_{\text{B}}T \rangle}$, as the upper bound. These lifetimes yield cluster sizes that are at least one order-of-magnitude lower than the observed value (see Table 1).

Finally, we discuss the potential error stemming from the assumption that the attractive term from Eq. 7 is isotropic and our neglecting additional modulations to the PMF, due to the discreteness of water, which are likely present and appear to be ~2 $k_{\text{B}}T$ in magnitude (54). These errors, if any, would not lead to a longer-lived complex, because such a longer-lived complex necessarily implies a lower yet B_{22}. Indeed, the integral in the attractive portion of the second virial coefficient, in Eq. 11, is dominated by values of the argument that minimize the exponent (which is the basis of what is called the steepest-descent approximation). Thus, this attractive portion is proportional to $e^{V/k_{\text{B}}T}$, averaged of the orientation, where V is the depth of the PMF at a given mutual orientation. On the other hand, the average lifetime of complex, which enters Eq. 1, is proportional to the orientational average of $e^{G^{\ddagger}/k_{\text{B}}T} = e^{(V+\delta V)/k_{\text{B}}T}$, where δV is the altitude of the typical dissociation barrier of the PMF. Hence, the second virial coefficient is approximately a linear function of the lifetime, with a negative slope. As a result, the presence of deep attractive minima, requisite for longer times, would imply a second virial coefficient lower than its observed value. Note conformational changes that enable density-stabilized complexes would not lead to a decrease in the second virial coefficient because the latter is pertinent to low densities.

CONSEQUENCES FOR MESOSCOPIC AGGREGATION

The goal of this work was to test whether a pair of folded and conformationally rigid lysozyme molecules could form transient complexes that live long enough to give rise to mesoscopic clusters observed in lysozyme solutions. Toward this goal, we have developed a computationally efficient model for protein-protein interaction in solutions with moderate ionic strength, in which the Debye screening length is comparable or larger than the protein size. This model originates in the classic Tanford-Kirkwood model, but applies it to interaction between the residues on distinct protein molecules. The model partially accounts for hydration interactions, in a continuum approximation, by including effects of the dielectric discontinuity at the protein-solvent interfaces. We explicitly consider the possibility of changes in the charge state of the ionizable residues stemming from the changes in their pK_a values caused by the proximity of another protein. The dispersive and excluded-volume parts of the interaction, as well as attraction between hydrophobic patches and other ion-induced effects, are included phenomenologically. The resulting uncertainty is mitigated by testing the resulting PMF against the measured second virial coefficient.

Our model accounts rather fully for the anisotropic nature of the Coulomb interactions between biomolecules that stems from inhomogeneous charge distribution on their surfaces. We have shown that by neglecting this anisotropy, the DLVO approximation overestimates the repulsion between the proteins. This observation is consistent with the known general trend that to recover experimental second virial coefficients using the DLVO theory, one must use significantly greater values of the Hamaker constant than those calculated from first principles (38). Also, we have seen that changes of the protonation state of the surface residues are relatively unimportant, which greatly simplifies the task of modeling large assemblies of proteins.

The model is sufficiently simple to allow for complete sampling of the protein orientations with modest computational effort. The resulting PMF is smooth enough to allow for accurate estimates of its second derivatives at the initial and transition states for the dissociation of a transient protein dimer. We have thus established that typical dimers of folded lysozyme molecules are too short-lived to explain the mesoscopic size of the clusters found in solutions of lysozyme and other proteins. Still, according to Pan et al. (14), the radius of the cluster is determined by the longest living complex, which a priori could consist of more than two proteins. Such complexes would have to be relatively compact, because the lifetime of a chainlike structure is determined by the lifetime of the weakest link, i.e., a dimer, which we have already shown to be too short to explain the clusters. Now, the large N limit of such compact complexes has been already analyzed by Hutchens and Wang (58), who have shown that for isotropically interacting protein molecules, a purely electrostatic scenario in combination with short-range attraction is inconsistent with the presence of the mesoscopic clusters. The isotropic assumption is appropriate in this context even though the interaction within a pair of molecules is rather anisotropic: it is unlikely that in large collections of such molecules, the "favorable" contacts will be typically satisfied. To demonstrate this notion explicitly, we analyze the energetics of a lysozyme trimer in the Supporting Material and confirm that those conform to the large N trend established by Hutchens and Wang (58). We thus conclude that, at least in the case of lysozyme, the protein molecules do not remain fully rigid but undergo conformational changes during cluster formation. These conformational changes may be relatively small, so as to increase the contact area, or could include partial unfolding, in which case attraction between solvent-exposed

hydrophobic residues or even domain swapping (16,59) could take place.

Finally, according to calculations in Asherie et al. (9), ten Wolde and Frenkel (60), and Brandon et al. (61), the width of the attractive minimum and the height of the repulsive hump are consistent with the observed liquid-liquid separation in lysozyme. Combined with the above result that cluster formation should be accompanied by protein-conformational changes, this notion suggests the intriguing possibility that the formation mechanism of the macroscopic dense-liquid phase and the mesoscopic clusters in lysozyme are distinct.

SUPPORTING MATERIAL

Seven figures, one table, and references (47,62–68) are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00336-0.

This work is supported by the National Science Foundation, grant No. MCB 0843726 (to V.L. and P.G.V.), and the Norman Hackerman Advanced Research Program, grant No. 003652-0078-2009 (to P.G.V.). V.L. is supported in part by the Arnold and Mabel Beckman Foundation Beckman Young Investigator Award, the Alfred P. Sloan Research Fellowship, and The Welch Foundation.

REFERENCES

1. Eaton, W. A. 2003. Linus Pauling and sickle cell disease. *Biophys. Chem.* 100:109–116.
2. Benedek, G. B., J. Pande, …, J. I. Clark. 1999. Theoretical and experimental basis for the inhibition of cataract. *Prog. Retin. Eye Res.* 18:391–402.
3. McPherson, A. 2009. Introduction to Macromolecular Crystallography. John Wiley, Hoboken, New Jersey.
4. Long, M. L., J. B. Bishop, …, L. J. DeLucas. 1996. Protein crystal growth in microgravity review of large scale temperature induction method. *J. Cryst. Growth.* 168:233–243.
5. Peseta, S., J. A. Langer, …, C. E. Samuel. 1989. Interferons and their actions. *In* Annual Review of Biochemistry. C. C. Richardson, P. D. Boyer, I. B. Dawid, and A. Meister, editors. Annual Reviews, Palo Alto, CA. 727–778.
6. Bromberg, L., J. Rashba-Step, and T. Scott. 2005. Insulin particle formation in supersaturated aqueous solutions of poly(ethylene glycol). *Biophys. J.* 89:3424–3433.
7. Berry, P. S., S. A. Rice, and J. Ross. 2000. Physical Chemistry. Oxford University Press, New York.
8. Thomson, J. A., P. Schurtenberger, …, G. B. Benedek. 1987. Binary liquid phase separation and critical phenomena in a protein/water solution. *Proc. Natl. Acad. Sci. USA.* 84:7079–7083.
9. Asherie, N., A. Lomakin, and G. B. Benedek. 1996. Phase diagram of colloidal solutions. *Phys. Rev. Lett.* 77:4832–4835.
10. Muschol, M., and F. Rosenberger. 1997. Liquid-liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization. *J. Chem. Phys.* 107:1953–1962.
11. Rosenbaum, D. F., A. Kulkarni, …, C. F. Zukoski. 1999. Protein interactions and phase behavior: Sensitivity to the form of the pair potential. *J. Chem. Phys.* 111:9882–9890.
12. Pan, W., O. Galkin, …, P. G. Vekilov. 2007. Metastable mesoscopic clusters in solutions of sickle-cell hemoglobin. *Biophys. J.* 92:267–277.
13. Gliko, O., N. Neumaier, …, P. G. Vekilov. 2005. A metastable prerequisite for the growth of lumazine synthase crystals. *J. Am. Chem. Soc.* 127:3433–3438.
14. Pan, W., P. G. Vekilov, and V. Lubchenko. 2010. Origin of anomalous mesoscopic phases in protein solutions. *J. Phys. Chem. B.* 114:7620–7630.
15. Reference deleted in proof.
16. Bennett, M. J., M. R. Sawaya, and D. Eisenberg. 2006. Deposition diseases and 3D domain swapping. *Structure.* 14:811–824.
17. Brubaker, W. D., J. A. Freites, …, R. W. Martin. 2011. Separating instability from aggregation propensity in γS-crystallin variants. *Biophys. J.* 100:498–506.
18. Derjaguin, B. V. 1989. Theory of Stability of Colloids and Thin Films. Plenum Press, New York.
19. Verwey, E. J. W., and J. T. G. Overbeek. 1948. Theory of Stability of Lyophobic Colloids. Elsevier, Amsterdam.
20. Israelachvili, J. N. 1995. Intermolecular and Surface Forces. Academic Press, New York.
21. Petsev, D. N., N. D. Denkov, and K. Nagayama. 1993. Diffusion and light scattering in dispersions of charged particles with thin electrical double layers. *Chem. Phys.* 175:265–270.
22. Elcock, A. H., and J. A. McCammon. 2001. Calculation of weak protein-protein interactions: the pH dependence of the second virial coefficient. *Biophys. J.* 80:613–625.
23. McGuffee, S. R., and A. H. Elcock. 2006. Atomically detailed simulations of concentrated protein solutions: the effects of salt, pH, point mutations, and protein concentration in simulations of 1000-molecule systems. *J. Am. Chem. Soc.* 128:12098–12110.
24. Vazdar, M., J. Vymětal, …, P. Jungwirth. 2011. Like-charge guanidinium pairing from molecular dynamics and ab initio calculations. *J. Phys. Chem. A.* 115:11193–11201.
25. Pednekar, D., A. Tendulkar, and S. Durani. 2009. Electrostatics-defying interaction between arginine termini as a thermodynamic driving force in protein-protein interaction. *Proteins.* 74:155–163.
26. Papoian, G. A., J. Ulander, …, P. G. Wolynes. 2004. Water in protein structure prediction. *Proc. Natl. Acad. Sci. USA.* 101:3352–3357.
27. Lomakin, A., N. Asherie, and G. B. Benedek. 1999. Aeolotopic interactions of globular proteins. *Proc. Natl. Acad. Sci. USA.* 96:9465–9468.
28. Rosenbaum, D. F., and C. F. Zukoski. 1996. Protein interactions and crystallization. *J. Cryst. Growth.* 169:752–758.
29. Tanford, C., and J. G. Kirkwood. 1957. Theory of protein titration curves. 1. General equations for impenetrable spheres. *J. Am. Chem. Soc.* 79:5333–5339.
30. Kumar, S., and R. Nussinov. 2002. Close-range electrostatic interactions in proteins. *Chem. Bio. Chem.* 3:604–617.
31. Bas, D. C., D. M. Rogers, and J. H. Jensen. 2008. Very fast prediction and rationalization of pK_a values for protein-ligand complexes. *Proteins.* 73:765–783.
32. Li, H., A. D. Robertson, and J. H. Jensen. 2005. Very fast empirical prediction and rationalization of protein pK_a values. *Proteins.* 61: 704–721.
33. Olsson, M. H. M., C. R. Sondergaard, …, J. H. Jensen. 2011. PROPKA3: consistent treatment of internal and surface residues in empirical pK_a predictions. *J. Chem. Theory Comput.* 7:525–537.
34. Sondergaard, C. R., M. H. M. Olsson, …, J. H. Jensen. 2011. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK_a values. *J. Chem. Theory Comput.* 7:2284–2295.
35. Wang, D., R. J. Nap, …, I. Szleifer. 2011. How and why nanoparticle's curvature regulates the apparent pK_a of the coating ligands. *J. Am. Chem. Soc.* 133:2192–2197.
36. Overbeek, J. T. G. 1990. The role of energy and entropy in the electrical double layer. *Colloids Surf.* 51:61–75.
37. Hamaker, H. C. 1937. The London-van der Waals attraction between spherical particles. *Physica.* 4:1058–1072.

38. Beretta, S., G. Chirico, and G. Baldini. 2000. Short-range interactions of globular proteins at high ionic strengths. *Macromolecules.* 33:8663–8670.

39. Nir, S. 1977. van der Waals interactions between surfaces of biological interest. *Prog. Surf. Sci.* 8:1–58.

40. Materese, C. K., A. Savelyev, and G. A. Papoian. 2009. Counterion atmosphere and hydration patterns near a nucleosome core particle. *J. Am. Chem. Soc.* 131:15005–15013.

41. Perutz, M. F., A. M. Gronenborn, …, D. T. Shih. 1985. The pK_a values of two histidine residues in human hemoglobin, the Bohr effect, and the dipole moments of α-helices. *J. Mol. Biol.* 183:491–498.

42. Vliegenthart, G. A., J. F. M. Lodge, and H. N. W. Lekkerkerker. 1999. Strong weak and metastable liquids structural and dynamical aspects of the liquid state. *Physica A.* 263:378–388.

43. Kristof, T., J. Vorholz, …, G. Maurer. 1999. A simple effective pair potential for the molecular simulation of the thermodynamic properties of ammonia. *Mol. Phys.* 97:1129–1137.

44. Tabisz, G. C. 1977. An intermolecular potential for CH4-CH4 calculated within the electron gas approximation. *Chem. Phys. Lett.* 52:125–128.

45. Allahyarov, E., H. Löwen, …, A. A. Louis. 2003. Nonmonotonic variation with salt concentration of the second virial coefficient in protein solutions. *Phys. Rev. E.* 67:051404.

46. Neal, B. L., D. Asthagiri, …, E. W. Kaler. 1999. Why is the osmotic second virial coefficient related to protein crystallization? *J. Cryst. Growth.* 196:377–387.

47. Baker, N. A., D. Sept, …, J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA.* 98:10037–10041.

48. Patargias, G. N., S. A. Harris, and J. H. Harding. 2010. A demonstration of the inhomogeneity of the local dielectric response of proteins by molecular dynamics simulations. *J. Chem. Phys.* 132:235103.

49. Wang, J., M. Dauter, …, Z. Dauter. 2007. Triclinic lysozyme at 0.65 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* 63:1254–1268.

50. Dolinsky, T. J., P. Czodrowski, …, N. A. Baker. 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 35 (Web Server issue):W522–W525.

51. Ulrich, E. L., H. Akutsu, …, J. L. Markley. 2008. BioMagResBank. *Nucleic Acids Res.* 36 (Database issue):D402–D408.

52. Muschol, M., and F. Rosenberger. 1995. Interactions in undersaturated and supersaturated lysozyme solutions: static and dynamic light scattering results. *J. Chem. Phys.* 103:10424–10432.

53. Neal, B. L., and A. M. Lenhoff. 1995. Excluded volume contribution to the osmotic second virial coefficient for proteins. *AIChE J.* 41:1010–1014.

54. Masunov, A., and T. Lazaridis. 2003. Potentials of mean force between ionizable amino acid side chains in water. *J. Am. Chem. Soc.* 125:1722–1730.

55. Frauenfelder, H., and P. G. Wolynes. 1985. Rate theories and puzzles of hemeprotein kinetics. *Science.* 229:337–345.

56. Fredericks, W. J., M. C. Hammonds, …, F. Rosenberger. 1994. Density, thermal expansivity, viscosity and refractive index of lysozyme solutions at crystal growth concentrations. *J. Cryst. Growth.* 141:183–192.

57. Petsev, D. N., B. R. Thomas, …, P. G. Vekilov. 2000. Interactions and aggregation of apoferritin molecules in solution: effects of added electrolytes. *Biophys. J.* 78:2060–2069.

58. Hutchens, S. B., and Z.-G. Wang. 2007. Metastable cluster intermediates in the condensation of charged macromolecule solutions. *J. Chem. Phys.* 127:084912.

59. Bennett, M. J., S. Choe, and D. Eisenberg. 1994. Domain swapping: entangling alliances between proteins. *Proc. Natl. Acad. Sci. USA.* 91:3127–3131.

60. ten Wolde, P. R., and D. Frenkel. 1997. Enhancement of protein crystal nucleation by critical density fluctuations. *Science.* 277:1975–1978.

61. Brandon, S., P. Katsonis, and P. G. Vekilov. 2006. Multiple extrema in the intermolecular potential and the phase diagram of protein solutions. *Phys. Rev. E.* 73:061917.

62. Allen, R., and J.-P. Hansen. 2002. Density functional approach to the effective interaction between charges within dielectric cavities. *J. Phys. Condens. Matter.* 14:11981.

63. Nakajima, Y., and T. Sato. 1999. Calculation of electrostatic force between two charged dielectric spheres by the re-expansion method. *J. Electrostatics.* 45:213–226.

64. Stratton, J. A. 1941. Electromagnetic Theory. McGraw-Hill, Englewood Cliffs, NJ.

65. Finkelshtein, A. V. 1977. Electrostatic interactions of charged groups in an aqueous-medium and their effect on formation of polypeptide-chain secondary structure. *Mol. Biol.* 11:627–634.

66. Lindell, I. V. 1992. Electrostatic image theory for the dielectric sphere. *Radio Sci.* 27:1–8.

67. Rakhmanov, E. A., E. B. Saff, and Y. M. Zhou. 1994. Minimal discrete energy on the sphere. *Math. Res. Lett.* 1:647–662.

68. Saff, E., and A. Kuijlaars. 1997. Distributing many points on a sphere. *The Mathematical Intelligencer.* 19:5–11.