

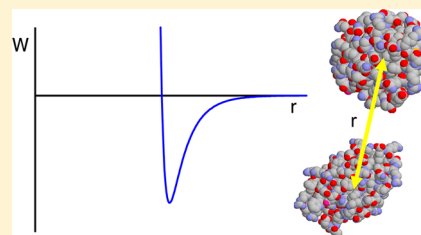
# Anisotropic Contributions to Protein–Protein Interactions

Leigh J. Quang,<sup>†</sup> Stanley I. Sandler, and Abraham M. Lenhoff\*

Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, United States of America

**S** Supporting Information

**ABSTRACT:** The anisotropy of shape and functionality of proteins complicates the prediction of protein–protein interactions. We examine the distribution of electrostatic and nonelectrostatic contributions to these interactions for two globular proteins, lysozyme and chymosin B, which differ in molecular weight by about a factor of 2. The interaction trends for these proteins are computed in terms of contributions to the osmotic second virial coefficient that are evaluated using atomistic models of the proteins. Our emphasis is on identifying the orientational configurations that contribute most strongly to the overall interactions due to high-complementarity interactions, and on calculating the effect of ionic strength on such interactions. The results emphasize the quantitative importance of several features of protein interactions, notably that despite differences in their frequency of occurrence, configurations differing appreciably in interaction energy can contribute meaningfully to overall interactions. However, relatively small effects due to charge anisotropy or specific hydration can affect the overall interaction significantly only if they contribute to strongly attractive configurations. The results emphasize the necessity of accounting for detailed anisotropy to capture actual experimental trends, and the sensitivity of even very detailed atomistic models to subtle solution contributions.



## INTRODUCTION

The interactions between protein molecules play a central role in determining the macroscopic thermodynamic properties and phase behavior of protein solutions and are critical to biological functions. An essential feature of these interactions is their anisotropy, which stems from the nonspherical shape, nonuniform charge distribution, rough local topography, and heterogeneous functionality on the protein surface. Accounting for this anisotropy in modeling protein–protein interactions is important for predictive calculations. While direct molecular simulations using models of proteins represented in full atomistic detail with an explicit solvent are computationally feasible, incorporating them in a systematic exploration of all possible pairwise interactions would be extremely demanding. A more feasible approach therefore entails utilizing a simplified coarse-grained representation of proteins that captures the essential physics of their interactions.

A large body of work has treated protein molecules as spheres within an implicit solvent and interacting via an isotropic pair potential, or more correctly, the potential of mean force (PMF). Several idealized potentials have been used, including the adhesive hard sphere,<sup>1–4</sup> square-well,<sup>5–7</sup> Yukawa,<sup>8–10</sup> modified Lennard-Jones,<sup>11–13</sup> and DLVO<sup>14–20</sup> potentials. While the forms of these potentials are different, each has parameters for the interaction strength, interaction range, and particle diameter. Of particular interest is the range, as it has been recognized for some time<sup>3</sup> that the protein–protein interactions of principal interest are short-ranged relative to protein molecular dimensions.

The effect of the interaction range is especially apparent in predictions of phase behavior,<sup>21,22</sup> which have been widely studied in the colloidal physics literature for a variety of pair

potentials. Such computations can predict the formation of clusters, dense liquid phases, gels, glasses, and crystals such as have been observed experimentally in protein solutions.<sup>23–30</sup> For sufficiently short-ranged potentials, the structures predicted for the phase diagram by these isotropic models, particularly the existence of a metastable liquid–liquid transition within the fluid–solid (crystal) boundary, are qualitatively consistent with those seen experimentally for protein solutions. Nevertheless, these models are clearly simplistic representations of the true interactions, and more elaborate models have also been proposed to include anisotropic contributions. An embedded charge model has been used to account for the charge anisotropy,<sup>31,32</sup> but much more attention has been devoted to models incorporating generically “sticky”, highly attractive patches.<sup>33</sup> Such models include additional parameters and therefore make possible more accurate quantitative representations of protein phase diagrams.<sup>34–36</sup> In addition, they open the possibility of describing competing crystalline phases that can be orientationally ordered or disordered.<sup>37–43</sup>

Although patch models provide a useful conceptual framework for incorporating anisotropy, they are rarely directly associated with details of the molecular structure of proteins and their effects on the interactions. While the excluded volume contribution to the interactions can be captured adequately by a hard-sphere approximation, the dependence of the interactions on the local topography and functionality of the protein surface results in a much more fine-grained orientation dependence than is typically included in patch models. This is primarily a consequence of the specificity of interactions between geo-

Received: July 27, 2013

Published: January 15, 2014

metrically complementary regions, which is intrinsic to biomolecular recognition phenomena.<sup>44</sup>

Structure-based models of protein–protein interactions are much more demanding computationally than patch models, so they have been used mainly to explore behavior of the osmotic second virial coefficient,  $B_{22}$ .  $B_{22}$  is a measure of effective two-body interactions in solution, so bimolecular computations are adequate, compared to the much larger systems needed to study phase behavior. Computations using structure-based models have shown that due to the Boltzmann weighting of the PMF in calculation of  $B_{22}$ , a relatively small number of highly attractive complementary configurations contribute disproportionately.<sup>45–48</sup> That  $B_{22}$  can be dominated by a few highly attractive configurations has been demonstrated experimentally,<sup>49</sup> with computational support,<sup>47,49</sup> by the measurable effects of point mutations on  $B_{22}$  values of T4 lysozyme.

Some of the more recent efforts to calculate  $B_{22}$  have shown the feasibility of atomistic simulations, typically by Brownian dynamics, of multimolecular systems using rigid-body protein representations with an implicit solvent.<sup>46,47</sup> Although such simulations can be used to calculate  $B_{22}$  as well as the growth of dimers or oligomers, they are rarely used for systematic examinations of the nature of the interactions themselves, for instance, for the purpose of developing realistic coarse-grained interaction models that can be applied to more demanding tasks such as calculating phase equilibrium. Such an approach could allow bridging between the colloidal patch models and structure-based biophysical models. Explicit coarse-grained models can be implemented,<sup>50,51</sup> or patch models can be used in which the patch parameters are based on structural details, particularly in accounting for specific interactions between geometrically complementary regions. This has been done for the “patch–antipatch” model,<sup>52</sup> in which the characteristics of high-complementarity interactions were characterized and used in a model based on spheres decorated with patches and complementary “antipatches” in addition to an underlying weakly attractive isotropic potential.

Implementation of such an approach would require knowledge of what density of coverage of patch–antipatch pairs is necessary to represent the pairwise interactions adequately. Local exploration around individual patches using an atomistic model suggests that such patches are no more than a few degrees wide,<sup>52</sup> so that each patch occupies a very small fraction of the overall orientation space for a pairwise system. However, the very large effect of the Boltzmann weighting may make even such small fractions of orientation space exceedingly important in calculating  $B_{22}$ . The objective of the present work is therefore to compute the orientation dependence of the protein–protein PMF for model proteins and to analyze the impact of the PMF distribution, particularly of the most highly attractive patch–antipatch pairs, on the computation of  $B_{22}$ .

The computations are performed for two model proteins, namely hen egg lysozyme and chymosin B. Lysozyme is the most widely used model protein for studies of protein interactions and phase behavior,<sup>4,15,53–59</sup> so a systematic analysis of its anisotropic interactions can enrich the knowledge base for this system. Chymosin B has been less widely studied, although aspects of its phase behavior have recently been reported.<sup>60</sup> However, as a globular protein with a molecular weight more than double that of lysozyme, it is a suitable model protein to include for comparative purposes. The structural data used were from the PDB files that are identified in Table 1, along with other relevant physical properties.

**Table 1. Physical Properties of Proteins Studied**

protein	PDB ID	MW (g/mol)	residues	pI	net charge (pH) <sup>a</sup>
lysozyme	4LYZ	14300	129	11	+8.0 (7.0)
chymosin B	1CMS	35673	323	4.6	−8.5 (5.0)

<sup>a</sup>Calculated using PROPKA.<sup>61,62</sup>

## THEORY AND METHODS

The calculations of interest are of interactions between a pair of identical protein molecules, with the orientation dependence of primary importance. It is appropriate to consider the analysis in terms of the calculation of  $B_{22}$ , since it represents a measure of interactions averaged over all pairwise configurations; nevertheless, explicit evaluation of  $B_{22}$  is not the goal *per se*.  $B_{22}$  is given as<sup>48,63,64</sup>

$$B_{22} = -\frac{1}{16M_W^2\pi^2} \int_0^{2\pi} \int_0^\pi \int_0^{2\pi} \int_0^{2\pi} \int_0^\pi \int_0^\infty (e^{-W/kT} - 1) \times r_{12}^2 dr_{12} \sin\theta d\theta d\phi d\alpha \sin\beta d\beta d\gamma \quad (1)$$

where  $W$  is the PMF,  $r_{12}$  is the center-to-center distance,  $\phi$  and  $\theta$  are the spherical angles representing the translation of the second molecule relative to the first molecule,  $\alpha$ ,  $\beta$ ,  $\gamma$  are the Euler angles denoting the rotation of the second molecule, and  $M_W$  is the molecular weight of the protein. The angles are defined relative to the atomic positions given in the respective PDB files for the two proteins studied. Equation 1 can be rearranged by decomposing the radial integral and consolidating the angular integrals as  $\Omega$  to give

$$B_{22} = \frac{1}{16M_W^2\pi^2} \int_\Omega \left[ \frac{1}{3}r_c^3 - \int_{r_c}^\infty (e^{-W/kT} - 1)r_{12}^2 dr_{12} \right] d\Omega \\ = \frac{1}{16M_W^2\pi^2} \int_\Omega \left[ \frac{1}{3}r_c^3 - I_{in} \right] d\Omega \quad (2)$$

The first term in brackets, which results from integration of  $r_{12}$  between 0 and  $r_c$ , represents the excluded-volume contribution  $B_{22}^{Ex}$  to  $B_{22}$  and is determined directly by  $r_c$ , the center-to-center distance at contact, which is a function of  $\Omega$ . The second term covers the distance-dependent part of the integrand; the result, which is referred to as the inner integral  $I_{in}$ ,<sup>48</sup> is a direct measure of the energetic contribution to  $B_{22}$  due to the intermolecular interactions for orientational configuration  $\Omega$ .

Evaluation of eq 2 requires determination of  $r_c$  and  $W(r_{12})$ , both as a function of  $\Omega$ , and then evaluation of the integrals. These are discussed in the remainder of this section, but some general observations are appropriate at this point. For globular proteins the variation of  $r_c$  with  $\Omega$  is fairly limited, but strongly attractive (high-complementarity) configurations can lead to large negative values of  $W$  that, because of the Boltzmann factor, result in extremely large values of  $I_{in}$ <sup>48,65</sup> that add considerable complexity and uncertainty to evaluation of eq 2. These strongly attractive configurations are also implicated in the formation of condensed protein phases, so even outside the context of  $B_{22}$  calculations, they may play a disproportionate role in, for instance, simulations of protein phase behavior.

**Potential of Mean Force,  $W$ .** Molecular mechanics (MM) methods are most widely used for structure-based computations of protein interactions,<sup>66</sup> with the potential of mean force

$W$  usually represented as the sum of electrostatic ( $W_{\text{es}}$ ) and nonelectrostatic ( $W_{\text{ne}}$ ) contributions; in both cases rigid-body calculations based on crystal structures are typical. Different approaches have been adopted for the computation of both the electrostatic and nonelectrostatic contributions, with resulting differences and uncertainties in the values of  $W$  obtained. However, whichever methods are used, typical magnitudes of  $W$  are of the order of several  $kT$ , except for a small fraction of configuration space that is populated by highly attractive configurations with  $W$  of order tens of  $kT$ . Characterizing this small subpopulation is a major source of uncertainty in these calculations, as is discussed again later in considering our results.

MM methods are well suited for calculating the non-electrostatic contribution, but the standard force fields that are used are usually premised on inclusion of explicit solvent, which greatly increases the computational burden. Two alternative approaches have therefore been used. One is a surface-based method in which  $W_{\text{ne}}$  is calculated using the loss in solvent-accessible surface area and an effective interfacial energy.<sup>45,47</sup> The alternative that is used here is a direct adaptation of the MM atomistic approach<sup>48,67</sup> that seeks to account for the presence of solvent in different roles via a hybrid of a direct MM calculation and a Lifshitz-Hamaker continuum colloidal calculation.

In this formulation, interactions between two protein molecules are calculated as a sum of pairwise atomistic interactions, with the pair interaction contribution to the PMF dependent on the separation distance between each atom pair. For atom pairs separated by a center-to-center distance of more than 6 Å, the interactions are computed using the continuum Lifshitz-Hamaker formulation for two spheres,<sup>17,68,69</sup> with the value of the Hamaker constant taken to be that for protein-water-protein interactions.<sup>70,71</sup> For atom pairs separated by a center-to-center distance of less than 6 Å, water molecules are assumed to be excluded and the Lennard-Jones formulation is used to determine the contribution to the PMF. The parameters for the Lennard-Jones equation were taken from the OPLS-AA force field,<sup>72</sup> with scaling by an empirical factor of 0.5<sup>67</sup> allowing consistency between the values predicted by the two limiting models as well as consistency between predicted and experimental values of pairwise protein binding constants.<sup>73</sup>

Electrostatic interactions are widely considered to be described most reliably by solving the Poisson-Boltzmann equation for the configuration of interest and calculating electrostatic free energies from the resulting potential fields.<sup>66,74</sup> Nevertheless, several sources of uncertainty and controversy remain, and the need for solutions for a large number of pairwise configurations has driven the use of approximate methods.<sup>47,48</sup> Here, we use a screened pairwise Coulombic potential

$$W_{\text{es}} = \sum_i \sum_j \frac{q_i q_j}{\epsilon_0 \epsilon_r r} e^{-\kappa r} \quad (3)$$

where  $q_i$  and  $q_j$  are the charges on the two interacting atoms,  $\epsilon_0$  is the permittivity of free space,  $\epsilon_r$  is the dielectric constant of the solvent (taken as 80), and  $\kappa$  is the Debye parameter

$$\kappa = \sqrt{\frac{\frac{1}{2} \sum_i (z_i e)^2 c_i}{\epsilon_0 \epsilon_r kT}} \quad (4)$$

where  $c_i$  is the concentration of the ion  $i$ ,  $e$  is the elementary charge, and  $z_i$  is the valence of the ion  $i$ ; the numerator inside the square root is the ionic strength, in terms of which our electrostatic calculations are reported. The partial charges carried by each atom were taken from the OPLS-AA force field. However, the charges assigned to ionizable residues were determined from the web server PROPKA<sup>61,62</sup> (<http://propka.ki.ku.dk>). Calculations for lysozyme were performed at pH 7 and for chymosin B at pH 5, where the net charges found from PROPKA were +8.0 and −8.5 respectively.

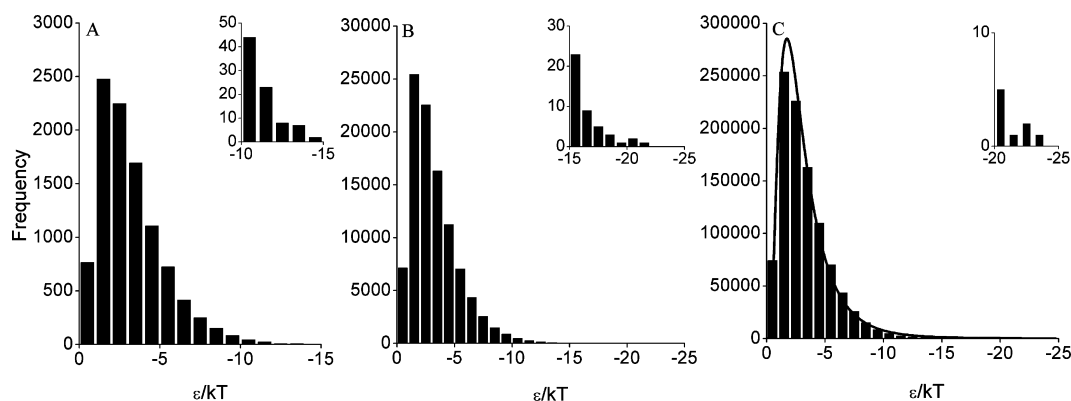
The methods that we have used to calculate  $W$  are, like any others, at best approximate representations that have been adopted for reasons of tractability. It is useful to have some perspective of the implications of these approximations. The most significant contributions to calculations of  $B_{22}$  are from the most attractive pairwise configurations, corresponding to high-complementarity configurations. These should be captured adequately by the nonelectrostatic contribution, via either our MM method, the accessible surface area method or any realistic structure-based alternative; while such methods would be less rigorous than fully atomistic simulations, including solvent, the overall structure of the distributions would still be captured, as would reasonable values of most quantitative details. The electrostatic contributions to  $W$  should be predominantly repulsive for proteins with a finite net charge, and our earlier Poisson-Boltzmann calculations<sup>48</sup> indicated that the attractive electrostatic contributions were generally no more than a few  $kT$  in magnitude. Therefore, despite the approximations our calculations should provide a good indication of the true angular PMF distribution.

**Computation of  $B_{22}$  Contributions.** The methods described in the previous section were used to calculate  $W(r_{12}, \Omega)$  and, from that dependence, the different elements that contribute to  $B_{22}$  via eq 2. The center-to-center distance at contact  $r_c$  was taken as the value of  $r_{12}$  for which  $W = 0$  at constant  $\Omega$ . Because of the relatively limited variation of  $r_c$  for globular proteins, the configurational integral for the excluded volume contribution in eq 2 can then be readily computed using Monte Carlo integration.<sup>48,75</sup> The value of  $I_{\text{in}}$  for a given set of angles  $\Omega$  was calculated using the cubic spline method to interpolate between discrete points of the PMF and then a one-dimensional Gaussian quadrature from the Fortran subroutine library QUADPACK was used to perform the integration. However, a major challenge is evaluating the configuration integral over  $I_{\text{in}}$  in eq 2. Previous work<sup>48,49,76,77</sup> utilized Monte Carlo integration for this task due to the irregular nature of the integrands, and this method also served as a frame of reference here. In this approach,  $N$  different orientations are randomly sampled from the global configuration space, values of  $r_c$  and  $I_{\text{in}}$  are computed, and  $B_{22}$  is calculated as the arithmetic average of the integrand (brackets in eq 2) for all  $N$  sampled configurations.<sup>78</sup> The error in  $B_{22}$  is estimated from the variance of the integrand  $f$  as

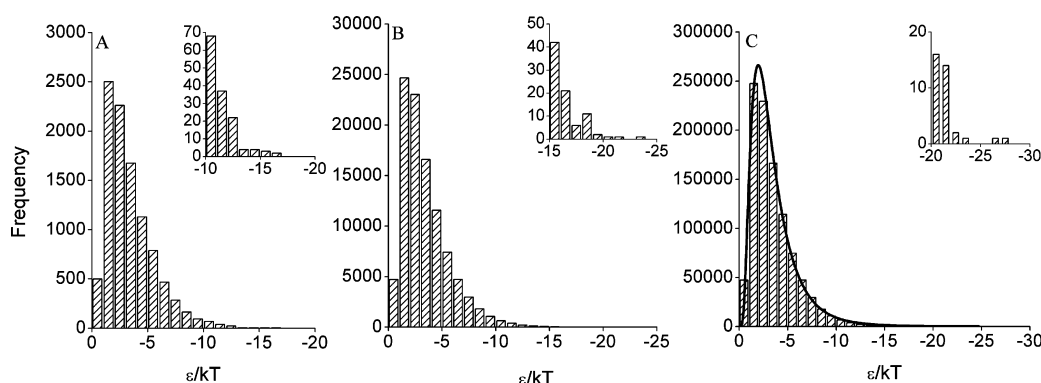
$$\Delta B_{22} = \pm \frac{1}{16M_W^2 \pi^2} V \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}} \quad (5)$$

where  $V$  is the hypervolume of the configuration space explored and the angle brackets denote the arithmetic mean.

The enormous variability in the values of  $I_{\text{in}}$  because of the Boltzmann factor weighting raises questions regarding the accuracy of this approach, and a principal objective of the present work is to provide a more reliable characterization of



**Figure 1.** Histograms of the distribution of the nonelectrostatic interaction well minima for lysozyme from (A)  $10^4$ , (B)  $10^5$ , and (C)  $10^6$  randomly sampled configurations. The inset histograms for each protein are to magnify the tails of the distributions. The line in C shows the best-fit log-normal distribution (eq 8) for  $a = 0.975$ ,  $b = 0.652$ .



**Figure 2.** Histograms of the distribution of the nonelectrostatic interaction well minima for chymosin B from (A)  $10^4$ , (B)  $10^5$ , and (C)  $10^6$  randomly sampled configurations. The inset histograms for each protein are to magnify the tails of the distributions. The line in C shows the best-fit log-normal distribution (eq 8) for  $a = 1.067$ ,  $b = 0.628$ .

the landscape for the integration. To address these numerical concerns,  $B_{22}$  is separated into three contributions

$$B_{22} = B_{22}^{\text{Ex}} + B_{22}^{\text{PA}} + B_{22}^{\text{Background}} \quad (6)$$

where  $B_{22}^{\text{Ex}}$  is the excluded volume contribution,  $B_{22}^{\text{PA}}$  is the contribution from the patch–antipatch energetic interactions, and  $B_{22}^{\text{Background}}$  is the contribution from nonpatch–antipatch energetic interactions. The latter two are distinguished by the depth of the respective energy wells, with configurations with greater well depths considered to be strongly attractive patch–antipatch pairs and included in  $B_{22}^{\text{PA}}$ , and those with less attractive wells incorporated in  $B_{22}^{\text{Background}}$ . Where the division between these two sets occurs is investigated as part of the study.

The very large energetic contribution to  $B_{22}$  from the patch–antipatch interactions requires a careful and detailed integration procedure. For each unique patch–antipatch pair  $i$  centered at  $\Omega_i = \{\phi_i, \theta_i, \alpha_i, \beta_i, \gamma_i\}$  a region of  $\pm\Delta$  for each angle is taken to represent the domain  $\Omega_{\text{PA}}$  of the patch, and the localized patch integration

$$B_{22}^{\text{PA},i} = -\frac{1}{16M_{\text{W}}^2\pi^2} \int_{\Omega_{\text{PA}}} I_{\text{in}} d\Omega_{\text{PA}} \quad (7)$$

is then performed using the globally adaptive multidimensional integration routine DCUHRE.<sup>79,80</sup>  $I_{\text{in}}$  is recalculated at each point in the domain, using the local value of  $r_c$ .

Equation 7 represents the contribution to  $B_{22}$  from an individual patch–antipatch configuration, and the integral in this equation is referred to in what follows as the configurational integral  $I_{\text{config}}$ . If all patch–antipatch pairs occupy distinct, nonoverlapping subregions, then the total contribution to  $B_{22}$  of all  $N_p$  patch–antipatch pairs is obtained simply by summation. To complete the calculation of the overall  $B_{22}$ , the background contributions from the nonpatch–antipatch configurations are accounted for by Monte Carlo integration, but angular configurations within the subregions occupied by the already-accounted-for patch–antipatch pairs are excluded from the sampling. By excluding the patch–antipatch configurations, the  $I_{\text{in}}$  landscape is presumably flatter, and therefore Monte Carlo integration should be suitable for calculating this contribution.

## RESULTS AND DISCUSSION

**PMF Distributions.** Since the largest contributions to the energetic interactions in eq 2 result from high-complementarity configurations, we first seek to characterize the distribution of geometric complementarity by examining the distribution of nonelectrostatic contributions to the PMF. Histograms representing the distributions of the well depths  $\epsilon$  for nonelectrostatic interactions from randomly sampled configurations for lysozyme and chymosin B are shown in Figures 1 and 2, respectively. In each case, results are shown for  $10^4$ ,  $10^5$ , and  $10^6$  configurations; for each orientational configuration the PMF contribution was calculated for a sufficient number of



center-to-center distances (of order tens) to define a smooth functional dependence, with the minimum defining the well depth. For each protein, most of the configurations sampled are weakly to moderately attractive, with the modes of the histograms occurring between  $-1kT$  and  $-2kT$  and the average well depth about  $-3kT$ .

The relative frequencies of the well-depth distribution for both proteins appear to be largely independent of the extent of sampling, with the histograms following approximately a log-normal probability distribution<sup>81</sup>

$$f(x) = \frac{1}{xb\sqrt{2\pi}} \exp\left\{-\frac{(\ln x - a)^2}{2b^2}\right\}; \quad 0 < x < \infty \quad (8)$$

where  $a$  is the location parameter and  $b$  is the scale parameter. The  $a$  and  $b$  parameters for a set of  $N$  configurations are computed from the unbiased estimators

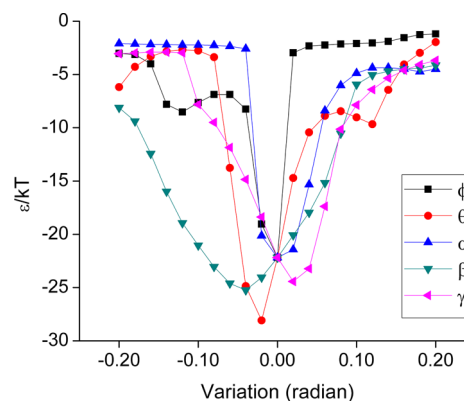
$$a = \frac{\sum_{i=1}^N \ln x_i}{N}$$

$$b = \frac{\sum_{i=1}^N (\ln x_i - a)^2}{N - 1} \quad (9)$$

for  $x$  defined as the absolute value of  $\epsilon/kT$ . The fits of the distributions and the corresponding parameter values are shown in panel C of Figures 1 and 2, which indicates that the log-normal distribution provides a reasonable overall description of the relative frequencies of the well depths for the histograms.

Notwithstanding the relative uniformity of the distributions with sample size, the tails do not necessarily provide a consistent sample of the population of very strongly attractive configurations (insets of Figures 1 and 2). Relatively few strongly interacting configurations were found, with the strongest configuration identified for  $10^4$  configurations of chymosin B having a well minimum of  $-16.6kT$  and that for  $10^6$  configurations having a minimum of  $-27.1kT$ . The configurations with  $\epsilon < -20kT$  are arbitrarily considered in what follows as patch–antipatch pairs, with the specific configurations among the  $10^6$  samples for each protein presented in Tables S1 and S2 in the Supporting Information. These results show that more configurations with  $\epsilon < -20kT$  were identified for chymosin B than for lysozyme even though the number of configurations sampled was the same for both. This observation emphasizes the challenge in the orientation sampling to properly and effectively sample the tails of the well-depth distribution. The importance of the tails increases with the well depth because of the enormous weighting effect of the Boltzmann factor for such deep wells.

Another issue that arises in the random orientation sampling is whether a given patch–antipatch pair at the sampled orientation represents the true minimum. To explore this, the first highly attractive configuration for lysozyme in Table S1 (entry 1; Supporting Information) was selected as a test case. Each individual orientation angle for this configuration was varied within  $\pm 0.20$  rad around the central orientation while holding the other angles fixed at their respective central values (Figure 3). Although there is certainly variability among the different angles, the well appears to be roughly 0.1 radians wide for each angle, similar to the example shown previously.<sup>52</sup> Accurate determination of the true well depth therefore



**Figure 3.** Well depth as a function of the angles for the orientation listed in entry 1 of Table S1, Supporting Information. The largest change occurs when  $\theta$  is decreased by  $-0.02$  rad, which indicates that the originally sampled orientation is not the optimum alignment.

requires refinement of the initial locations identified for the patch–antipatch pairs.

To improve the accuracy of the energy minima estimates,  $10^5$  random orientations were sampled within  $\pm 0.10$  rad around the central orientation for each of several of the patch–antipatch configurations for each protein. The refined minima are shown in Tables S3 and S4, Supporting Information, along with the adjusted values for each of the five angles. Although the differences in the angles are small, these small variations can lead to substantial changes in the well depths. The most significant case is for the first patch–antipatch pair for lysozyme, in which a configuration with a well depth on the order of  $-40kT$  was identified. This configuration corresponds approximately to a dimer with rotational symmetry and two contact regions. Similar strong patch–antipatch pairs are seen for chymosin B, in which the refinement led to a well depth as high as  $-39kT$ . The high sensitivity of  $\epsilon$  to the orientation angles indicates that the region representing a given patch–antipatch interaction is only a very small fraction of the global angular space, an important issue involved in the integration required for determining  $B_{22}$ .

**Initial Estimation of  $B_{22}$ .** The excluded-volume contribution to  $B_{22}$  was calculated using Monte Carlo integration. The calculated values are compared with the theoretical hard-sphere  $B_{22}$  values based on the diameter of a sphere of equal volume for each protein (Table 2). The values are comparable in magnitude, but the value is higher for the atomistic representation of the protein. This result is consistent with previous findings<sup>50,75</sup> that the excluded volume contribution for the atomistic representation can be more than 50% greater than that for a sphere because of the roughness of the protein surface.

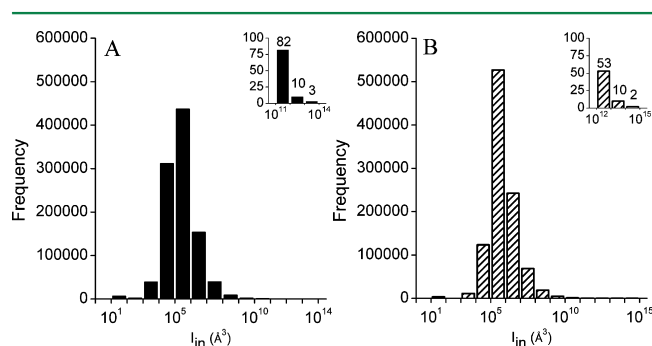
The contributions to  $B_{22}$  from energetic interactions result largely from the nonelectrostatic interactions, which are dominant for the patch–antipatch configurations. The  $B_{22}$  values based on the excluded volume and nonelectrostatic interactions calculated using Monte Carlo integration are presented in Table 2. There are two symptoms of the computational challenges evident in these results. First, the  $B_{22}$  values computed for each protein show significant uncertainties even for  $10^6$  configurations, and second, the final computed values are orders of magnitude greater than typical experimentally measured values, which are closer to  $10^{-4}$ – $10^{-3}$  ml mol/g<sup>2</sup> in magnitude.<sup>53,59,82,83</sup> The computed  $I_{in}$

**Table 2.**  $B_{22}$  Calculated from  $10^6$  Randomly Sampled Configurations Based on Excluded Volume Contribution and Both Excluded Volume and Nonelectrostatic Attraction<sup>a</sup>

protein	$B_{22}^{\text{Ex}} (\times 10^{-4} \text{ mL mol/g}^2)$	$\sigma (\text{\AA})$	$B_{22}^{\text{HS}} (\times 10^{-4} \text{ mL mol/g}^2)$	total $B_{22} (\text{mL mol/g}^2)$
LYZ	$2.997 \pm 0.001$	32.0	2.02	$-0.24 \pm 0.07$
CMS	$1.3990 \pm 0.0005$	42.9	0.783	$-0.95 \pm 0.59$

<sup>a</sup>The  $\sigma$  value is the sphere equivalent diameter determined from the empirical correlation of Neal and Lenhoff.<sup>48</sup> The error in the Monte Carlo estimate is calculated from eq 5.

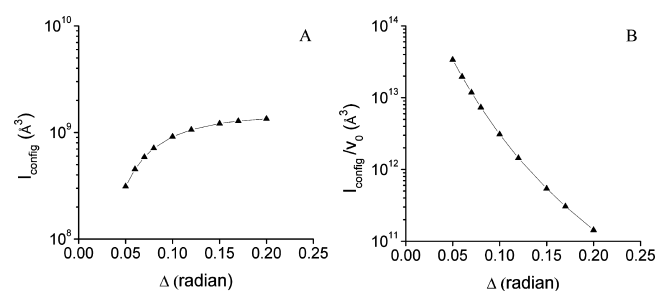
values for the  $10^6$  configurations sampled for lysozyme and chymosin B (Figure 4) span more than 10 orders of magnitude,



**Figure 4.** Histogram of the computed  $I_{\text{in}}$  for the  $10^6$  randomly sampled configurations for (A) lysozyme and (B) chymosin B. The inset shows an enlarged view of the high- $I_{\text{in}}$  tail of the distribution.

reflecting the Boltzmann weighting of the PMF. Because the Monte Carlo integration method takes the arithmetic average of all  $I_{\text{in}}$  values, it implicitly assumes that the same  $I_{\text{in}}$  value applies over the whole subdomain of the integration hypervolume that the local value is intended to represent. For the few configurations that have extremely large  $I_{\text{in}}$  values, the local variation shown in Figure 3 is neglected, and those configurations contribute disproportionately to the mean value of  $I_{\text{in}}$  and consequently to the value of the integral overall, making this a poor method for obtaining an accurate numerical estimate of  $B_{22}$ .

**Contributions to  $B_{22}$ : Patch–Antipatch Pairs.** To improve the accuracy of the integration of  $I_{\text{config}}$  (eq 7) around the patch–antipatch configurations, integration over a domain of  $\pm\Delta$  radians for each angle around the central orientation was performed using adaptive quadrature.<sup>79,80</sup>  $\Delta$  directly determines the size of the subregion over which the interaction energy is strongly attractive before it decays, and therefore represents the patch size. However, increasing  $\Delta$  enlarges the hypervolume of the integration domain and consequently  $I_{\text{config}}$  increases monotonically. This can be seen in Figure 5A, in which  $I_{\text{config}}$  is computed using the integration routine for a representative patch–antipatch pair, the fourth lysozyme pair listed in Table S3, Supporting Information. Thus, the angular size of a patch–antipatch pair cannot be determined directly from the dependence of  $I_{\text{config}}$  on  $\Delta$ . Instead, therefore,  $I_{\text{config}}$  was normalized by the hypervolume of integration  $v_0$  and the result was plotted against  $\Delta$  (Figure 5B). At approximately  $\Delta = 0.10$  rad the normalized integral is about a factor of 10 below its highest value, suggesting that  $\Delta = 0.10$  rad comfortably incorporates essentially all of the main contribution of the patch and so reasonably characterizes the patch size for this particular orientation. Although the patch size for each patch–antipatch pair may be somewhat different,  $\Delta = 0.10$  rad appeared from the cases examined to be a reasonable universal value for integration of all patch–antipatch pairs, that is, one that

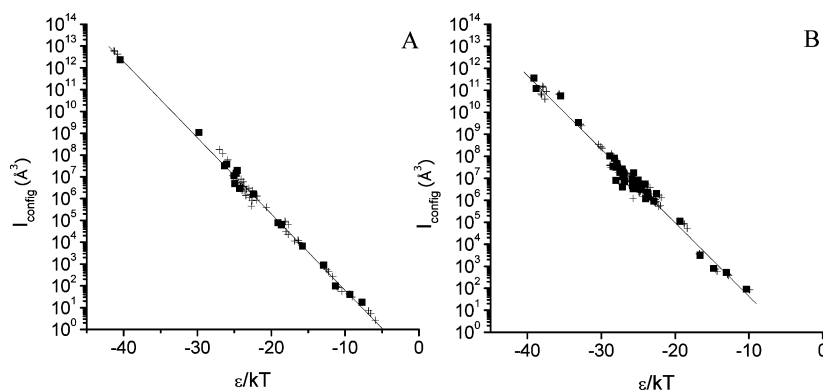


**Figure 5.** (A)  $I_{\text{config}}$  computed from the DCUHRE routine as a function of  $\Delta$  for lysozyme patch–antipatch pair 4 in Table S3, Supporting Information.  $I_{\text{config}}$  increases monotonically as  $\Delta$  increases due to the increase in the hypervolume of the integration. (B)  $I_{\text{config}}$  normalized by the volume of integration  $v_0$  as a function of  $\Delta$  for lysozyme patch–antipatch pair 4 in Table S3. The normalized integral decreases as  $\Delta$  is increased.

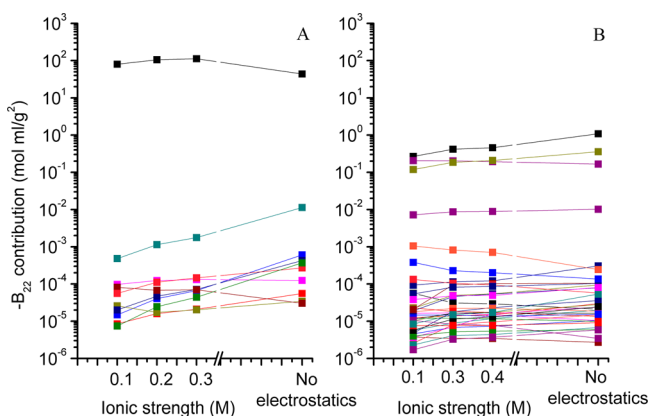
accounts for the dominant contribution of the patch to the interaction calculations. Contributions further from the well minimum and therefore omitted for the finite values of  $\Delta$  used are accounted for in the background contribution to  $B_{22}$ .

The localized integration was performed for configurations spanning a range of well depths for both proteins, including those both above and below  $-20kT$ , and for cases where only nonelectrostatic interactions were considered as well as where both nonelectrostatic and electrostatic interactions were included. The results represent the contribution of each individual region, in units of  $\text{\AA}^3$ , and they are plotted as a function of the total well depth of the configuration in Figure 6. There is some scatter in the results, which is to be expected since there is variability in the shapes of the energy profiles and uncertainty in the energy minima. Nevertheless, the data can be well-fitted by exponential correlations that are similar for both proteins, apart from a less negative exponent for chymosin B. Although electrostatics can affect the contribution to  $B_{22}$  of a configuration relative to the nonelectrostatic value, in a direction that depends on whether the electrostatics are attractive or repulsive, the same overall correlation for each protein appears applicable despite the different distance dependence for nonelectrostatic and electrostatic contributions. Therefore, within the range of well depths covered by Figure 6, the correlations can be used to estimate the contribution to  $B_{22}$  of a strongly attractive configuration simply from knowledge of the well depth.

The  $B_{22}$  contributions, in units of  $\text{mL mol/g}^2$ , for the individual lysozyme and chymosin B patch–antipatch pairs with  $\epsilon < -20kT$  are presented in Figure 7; these results include electrostatic interactions over a range of ionic strengths. The configuration of entry 1 in Table S3, Supporting Information, which is the  $-40kT$  patch identified for lysozyme after local sampling refinement, has a  $B_{22}$  contribution that is orders of magnitude greater than those of the other pairs, but even some of the weaker patch–antipatch pairs have contributions of



**Figure 6.** Plot of the localized configuration integral for (A) lysozyme and (B) chymosin B as a function of the total well depth.  $I_{\text{config}}$  was computed based on nonelectrostatic interactions alone (■) and nonelectrostatic and electrostatic interactions combined (+). Electrostatics were incorporated at pH 7 for lysozyme and pH 5 for chymosin B, at different ionic strengths. Integration was performed within the limits of  $\pm 0.10$  rad around the central orientation of the configurations. The regressed curve for lysozyme is  $F(x) = 0.0200 \exp(-0.805x)$ ,  $R^2 = 0.9941$ , and that for chymosin B is  $F(x) = 0.0195 \exp(-0.770x)$ ,  $R^2 = 0.9806$ , where  $x = \varepsilon/kT$ .



**Figure 7.** Contributions to  $B_{22}$  from individual patch-antipatch pairs with  $\varepsilon < -20kT$  for (A) lysozyme and (B) chymosin B based on nonelectrostatic interaction energies alone and with electrostatics. The contributions were determined by integrating within  $\pm 0.10$  rad around the central orientation using the DCUHRE integration routine.

order approaching  $10^{-4}$  ml mol/g<sup>2</sup>, a typical experimental magnitude for  $B_{22}$ . These magnitudes are discussed further below, but what is of interest here is that the dependence on ionic strength is strong enough that  $B_{22}$  can change appreciably with added salt; this is, of course, consistent with experimental observations. The basis for this sensitivity lies in the nature of the Boltzmann factor, specifically the decomposition of nonelectrostatic and electrostatic contributions as  $\exp[-W/kT] = \exp[-(W_{\text{ne}} + W_{\text{es}})/kT] = \exp[-W_{\text{ne}}/kT] \exp[-W_{\text{es}}/kT]$ . For high-complementarity configurations ( $W_{\text{ne}}/kT \ll -1$ ), even relatively small changes in  $W_{\text{es}}$  can be amplified by the  $\exp[-W_{\text{ne}}/kT]$  factor into large changes in the contribution to  $B_{22}$ .

The sign of the change in each  $B_{22}$  contribution is also significant. The inclusion of electrostatics significantly reduces the calculated attraction for most of the patch-antipatch pairs, as indicated by the increased contribution to  $B_{22}$  as the ionic strength increases. This reflects the predominantly repulsive nature of the electrostatic interactions in the configurations represented by the patch-antipatch pairs. However, two of the 10 patch-antipatch pairs with  $\varepsilon < -20 kT$  for lysozyme and about one-third of those for chymosin B appear to have an increased  $B_{22}$  contribution when electrostatics are incorporated,

suggesting the presence of attractive electrostatics in these configurations. These results indicate that in general there are weaker repulsive electrostatic effects for chymosin B than for lysozyme, which is not surprising in view of the proximity to the pI for the conditions chosen for chymosin B. It is such local electrostatic contributions that are thought to lead to salting-in behavior,<sup>84</sup> but their explicit computational exploration has not previously been reported.

**Contributions to  $B_{22}$ : Background.** The decomposition of the contributions to  $B_{22}$  (eq 6) allows the contributions of the most attractive configurations to be calculated by direct integration and the weaker background contribution by Monte Carlo integration. While the latter set of configurations contribute much less individually than the former, they occupy a much larger fraction of the overall angular integration hypervolume, so they may still contribute measurably to  $B_{22}$ . However, the threshold of  $-20kT$  was chosen arbitrarily, so some care is needed in evaluating the respective contributions. Computational exploration indicated that while Monte Carlo integration was indeed adequate for weakly attractive and repulsive configurations, unacceptably high errors resulted from using this method for configurations closer to the  $-20kT$  threshold because of the lack of accounting for the energy profile (cf. Figure 3), so a further subdivision into weak and moderately attractive configurations was made.

To estimate a reasonable cutoff value  $\varepsilon_{\text{cutoff}}$ , a useful benchmark is the average well depth from all the sampled configurations, i.e.  $\varepsilon \approx -3kT$ , which corresponds to  $I_{\text{in}} \approx 3.9 \times 10^4 \text{ Å}^3$ . The effect of the Boltzmann factor is seen mainly for the highly attractive configurations, as captured in Figure 6 and the associated correlations for  $I_{\text{config}}$ .  $\langle I_{\text{in}} \rangle$  for a configuration is determined by  $I_{\text{config}}$  scaled by the subdomain of integration  $v_0$ . For  $\Delta = 0.10$  rad, for which  $v_0 = 1.97 \times 10^{-4}$ , the correlations give  $\langle I_{\text{in}} \rangle \leq 3.9 \times 10^4 \text{ Å}^3$  for approximately  $\varepsilon > -7kT$ . Therefore  $\varepsilon_{\text{cutoff}}$  was taken as  $-7kT$ , and configurations reflecting weaker interactions were included in the Monte Carlo integration, which was applied to the angular hypervolume that excluded the patch-antipatch and moderate contributions.

The contribution from configurations with stronger interaction ( $\varepsilon < \varepsilon_{\text{cutoff}}$ ) should be determined by explicit integration, but the very large number of such configurations would be computationally expensive. This explicit approach was therefore reserved for the strongest patch-antipatch pairs with  $\varepsilon <$

**Table 3.** Background  $B_{22}$  for Lysozyme and Chymosin B Based on Nonelectrostatic Interactions Alone and with Addition of Electrostatics<sup>a</sup>

	$B_{22}^{\text{Background}} (\times 10^4 \text{ mol mL/g}^2)$					
	lysozyme			chymosin B		
	weak	moderate	total	weak	moderate	total
nonelectrostatics	−10.3	−10.4	−20.7	−3.66	−1.73	−5.39
0.10 M	−5.62	−5.45	−11.1	−3.58	−2.04	−5.62
0.20 M	−7.76	−6.99	−14.8			
0.30 M	−8.66	−8.58	−17.2	−3.65	−1.98	−5.63
0.40 M				−3.65	−1.76	−5.41

<sup>a</sup>Electrostatics were incorporated at pH 7 for lysozyme and pH 5 for chymosin B. Configurations with  $\epsilon < -7kT$  were excluded in the Monte Carlo integration.

−20kT (Tables S3 and S4, Supporting Information) and the remaining ones with moderate interactions ( $-20kT < \epsilon < -7kT$ ) were evaluated using the empirical correlations for  $I_{\text{config}}$  (Figure 6). The number of such moderate configurations is, in fact, large enough that their relative frequency as a function of  $\epsilon$  was estimated from statistical distributions (Figures 1 and 2 for the nonelectrostatic contribution alone).

The contributions to  $B_{22}^{\text{Background}}$  calculated for the weak and moderate configurations for lysozyme at pH 7 and chymosin B at pH 5 are shown in Table 3. The background  $B_{22}$  contributions from the weak and moderate configurations are similar for all the cases shown. For lysozyme,  $B_{22}^{\text{Background}}$  decreases as the ionic strength increases due to decreased contributions from both components. For chymosin B, however, the inclusion of electrostatics increases the overall attraction for the moderate contribution as well as for the overall background, albeit only slightly. Results such as this show the counterintuitive trends that may emerge despite the very large number of configurations accounted for in the ensemble. Not only can such trends differ from what is expected based on average properties such as net charge, but their origins are much more complex than suggested by the idealized models that are sometimes used, for example using dipole moments.<sup>85,86</sup>

**Sources of Uncertainty in Estimation of  $B_{22}$ .** The overall values of  $B_{22}$  are obtained by addition to the background values in Table 3 of those for the excluded volume (Table 2) and the patch–antipatch configurations (Figure 7). The former two contributions are of magnitude  $10^{-4}$  to  $10^{-3}$  ml mol/g<sup>2</sup>, which is consistent with most experimental data.<sup>53,59,82,83</sup> In addition, most of the individual patch–antipatch pairs contribute of order  $10^{-6}$  to  $10^{-4}$  ml mol/g<sup>2</sup>, again in a range consistent with experimental data. However, the contributions of the small number of most strongly attractive patch–antipatch pairs (below about −30kT) for each protein would completely overwhelm the other contributions and increase the overall  $B_{22}$  value orders of magnitude beyond the usual measured range.

There are several possible sources of uncertainty that may have contributed to this discrepancy, as well as to other less apparent errors. An obvious one is the inherently approximate nature of the MM calculations, and the approach used for the electrostatic interactions is one that accounts more clearly for the Coulombic interactions and less so for desolvation contributions to the free energy. It is the patch–antipatch configurations that are most sensitive to the electrostatic contribution, and it may be reasonable to perform more detailed numerical calculations to evaluate the electrostatic contribution for this much more limited set of configurations.

While the electrostatic calculations in the present work did not account explicitly for desolvation effects as in our earlier calculations, the electrostatic contribution to the interaction remains relatively small by either method in the patch–antipatch configurations.

It is the nonelectrostatic contribution on which the validity of the MM calculations depends mainly, so the accuracy of the method used should be considered. The parameters in this method were calibrated<sup>67</sup> against experimental binding data<sup>73</sup> in the range −10kT to −31kT, so well depths of the magnitude shown are reasonable. This method yields interaction free energies in the range from −0.010kT Å<sup>−2</sup> to −0.029kT Å<sup>−2</sup>, which are smaller in magnitude than the coefficients used in calculations based on buried surface area.<sup>45,47</sup>

Nevertheless, different causes of overestimation of the well depth are possible. The patch–antipatch configurations are especially sensitive to the atomistic structures used, but since conformational adaptation may occur during, and promote, bimolecular association, it seems unlikely to suppress the unreasonably high predicted affinity. A more important factor that may reduce affinity is the disruption to the local topography and hence to complementarity due to binding of small molecules, with water,<sup>76,77,87</sup> small ions,<sup>88</sup> and additives such as arginine<sup>89</sup> the most likely contenders. That a single such bound molecule can disrupt the binding in a patch–antipatch configuration indicates the challenge of predictive calculations, and this is aggravated by the presumed increasing likelihood of strong binding with increasing area of the contact surface. Molecular dynamics simulations<sup>90</sup> that find a 432.5 Å<sup>2</sup> lysozyme–lysozyme contact yielding a well depth of just a few kT, similar to that for a 51.2 Å<sup>2</sup> contact, suggest the possible magnitude of such effects.

Another possibility is that the magnitudes of our calculated interaction free energies are uniformly overestimated. In this case, all the well depths should be reduced, which would reduce the weakly and moderately attractive contributions and presumably make an especially small number of patch–antipatch configurations dominant.

Finally, the very high PMFs in individual configurations raise the possibility of formation of dimers that are irreversible on the time scale of a typical virial coefficient measurement. The rate of formation of such a complex would be very small because of the very small fraction of configuration space that it occupies, but the very slow dissociation kinetics may preclude effective sampling of all possible configurations. If such dimers are present, the contribution of the monomer–monomer PMF would not be fully reflected in the experimental  $B_{22}$  measurement. The presence of small quantities of dimers has



been reported in lysozyme solutions in the course of the careful identification of impurities.<sup>91</sup>

Another possible source of error is the density of orientational sampling. As Figures 1 and 2 show, the general PMF distribution is adequately captured even with only  $10^4$  configurations, and denser sampling is necessary only to identify the patch–antipatch pairs. For patches with  $\Delta = 0.10$  rad, the full configuration space would be covered by a few million evenly spaced samples, so the  $10^6$  random samples used here would be likely to have found most of the significant patch–antipatch pairs. Figure 7 suggests that these would have to be unusually strongly attractive in order to alter the results of our calculations appreciably. An additional computational parameter to which there may be some sensitivity is the patch size, with  $\Delta = 0.10$  rad used here. Based on analyses of the kind shown in Figure 7 that we have performed for numerous patch–antipatch pairs, this appears unlikely to be a major factor; although the contribution to  $B_{22}$  changes with  $\Delta$ , the changes for larger values of  $\Delta$  are due to relatively low interaction energies that would be adequately accounted for in the background contribution.

## CONCLUSIONS

The issues raised in this work illustrate the difficulty and uncertainties of computing  $B_{22}$  at the atomistic scale, but at the same time, the results presented provide considerable new insight into the structure of the computational problem and hence the character of the biophysical system. At their most basic level, the results re-emphasize the fine-grained influence of structure on the anisotropy of protein–protein interactions and therefore the solution properties of proteins; idealized model systems may provide some conceptual insight but they are totally inadequate for predicting experimental behavior based on structure. The heterogeneity of the  $I_{in}$  function makes calculation of the  $B_{22}$  integral particularly challenging, but our analysis shows the delicate interplay among the orientational distribution of the PMF, the coupling to electrostatic contributions, and the dominant role of the Boltzmann factor. A measured solution property represents an average over an extraordinarily complex ensemble in which individual contributions can be extremely sensitive to small changes in solution conditions. While we do not claim to have a definitive method for computing such properties accurately, this work elucidates the nature of the problem and the unlikelihood of reliability of any less rigorous approach.

A key challenge remains the identification and treatment of anomalously attractive configurations with well depths  $\sim -30kT$  to  $-40kT$ ; their potentially overwhelming contribution to  $B_{22}$  makes the outcome extremely sensitive to the assumed underlying physics, including very subtle effects such as strongly bound water molecules. An aspect that warrants further investigation is the possible ameliorating effect of kinetic limitations to the exploration of these dominant configurations under normal experimental conditions, for example, in determining the presence of stable dimers.

## ASSOCIATED CONTENT

### Supporting Information

Tables listing individual configurations and energies for patch–antipatch pairs (i.e., highly attractive configurations) as well as the methods for calculating the atomic positions. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*Phone +1 302 831 8989. E-mail [lenhoff@udel.edu](mailto:lenhoff@udel.edu).

### Present Address

<sup>†</sup>Eastman Chemical Company, 10380 Worton Rd., Chestertown, MD 21620, U.S.A.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was made possible by support from the National Science Foundation (BES-0519191 and DMR-1209424).

## REFERENCES

- (1) Miller, M. A.; Frenkel, D. Phase diagram of the adhesive hard sphere fluid. *J. Chem. Phys.* **2004**, *121* (1), 535–545.
- (2) Miller, M. A.; Frenkel, D. Simulating colloids with Baxter's adhesive hard sphere model. *J. Phys.: Condens. Matter* **2004**, *16* (42), S4901.
- (3) Rosenbaum, D.; Zamora, P. C.; Zukoski, C. F. Phase behavior of small attractive colloidal particles. *Phys. Rev. Lett.* **1996**, *76* (1), 150.
- (4) Rosenbaum, D. F.; Zukoski, C. F. Protein interactions and crystallization. *J. Cryst. Growth* **1996**, *169* (4), 752–758.
- (5) Duda, Y. Square-well fluid modelling of protein liquid–vapor coexistence. *J. Chem. Phys.* **2009**, *130* (11), 116101–2.
- (6) Liu, H.; Garde, S.; Kumar, S. Direct determination of phase behavior of square-well fluids. *J. Chem. Phys.* **2005**, *123* (17), 174505–4.
- (7) Pagan, D. L.; Gunton, J. D. Phase behavior of short-range square-well model. *J. Chem. Phys.* **2005**, *122* (18), 184515–6.
- (8) Foffi, G.; McCullagh, G. D.; Lawlor, A.; Zaccarelli, E.; Dawson, K. A.; Sciortino, F.; Tartaglia, P.; Pini, D.; Stell, G. Phase equilibria and glass transition in colloidal systems with short-ranged attractive interactions: Application to protein crystallization. *Phys. Rev. E* **2002**, *65* (3), 031407.
- (9) Hagen, M. H. J.; Frenkel, D. Determination of phase diagrams for the hard-core attractive Yukawa system. *J. Chem. Phys.* **1994**, *101* (5), 4093–4097.
- (10) Lomba, E.; Almaraz, N. G. Role of the interaction range in the shaping of phase diagrams in simple fluids. The hard sphere Yukawa fluid as a case study. *J. Chem. Phys.* **1994**, *100* (11), 8367–8372.
- (11) Lutsko, J. F.; Nicolis, G. The effect of the range of interaction on the phase diagram of a globular protein. *J. Chem. Phys.* **2005**, *122* (24), 244907–8.
- (12) ten Wolde, P. R.; Frenkel, D. Enhancement of protein crystal nucleation by critical density fluctuations. *Science* **1997**, *277*, 1975–1978.
- (13) ten Wolde, P. R.; Frenkel, D. Enhanced protein crystallization around the metastable critical point. *Theor. Chem. Acc.: Theory, Comput., Model. (Theor. Chim. Acta)* **1999**, *101* (1), 205–208.
- (14) Curtis, R. A.; Lue, L. A molecular approach to bioseparations: Protein–protein and protein–salt interactions. *Chem. Eng. Sci.* **2006**, *61* (3), 907–923.
- (15) Curtis, R. A.; Prausnitz, J. M.; Blanch, H. W. Protein–protein and protein–salt interactions in aqueous protein solutions containing concentrated electrolytes. *Biotechnol. Bioeng.* **1998**, *57* (1), 11–21.
- (16) Hiemenz, P. C.; Rajagopalan, R. *Principles of Colloid and Surface Chemistry*, 3rd ed.; Marcel Dekker, Inc.: New York, 1997.
- (17) Israelachvili, J. *Intermolecular and Surface Forces*, 2nd ed.; Academic Press: New York, 1992.
- (18) Pellicane, G.; Costa, D.; Caccamo, C. Microscopic determination of the phase diagrams of lysozyme and  $\gamma$ -crystallin solutions. *J. Phys. Chem. B* **2004**, *108* (23), 7538–7541.
- (19) Pellicane, G.; Costa, D.; Caccamo, C. Phase coexistence in a DLVO model of globular protein solutions. *J. Phys.: Condens. Matter* **2003**, *15* (3), 375.

- (20) Pellicane, G.; Costa, D.; Caccamo, C. Theory and simulation of short-range models of globular protein solutions. *J. Phys.: Condens. Matter* **2004**, *16* (42), S4923.
- (21) Gast, A. P.; Hall, C. K.; Russel, W. B. Polymer-induced phase separations in nonaqueous colloidal suspensions. *J. Colloid Interface Sci.* **1983**, *96* (1), 251–267.
- (22) Ilett, S. M.; Orrock, A.; Poon, W. C. K.; Pusey, P. N. Phase behavior of a model colloid-polymer mixture. *Phys. Rev. E* **1995**, *51* (2), 1344.
- (23) Berland, C. R.; Thurston, G. M.; Kondo, M.; Broide, M. L.; Pande, J.; Ogun, O.; Benedek, G. B. Solid-liquid phase boundaries of lens protein solutions. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89* (4), 1214–1218.
- (24) Broide, M. L.; Berland, C. R.; Pande, J.; Ogun, O. O.; Benedek, G. B. Binary-liquid phase separation of lens protein solutions. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88* (13), S660–S664.
- (25) Broide, M. L.; Tominc, T. M.; Saxowsky, M. D. Using phase transitions to investigate the effect of salts on protein interactions. *Phys. Rev. E* **1996**, *53* (6), 6325.
- (26) Dumetz, A. C.; Chockla, A. M.; Kaler, E. W.; Lenhoff, A. M. Protein phase behavior in aqueous solutions: Crystallization, liquid–liquid phase separation, gels, and aggregates. *Biophys. J.* **2008**, *94* (2), 570–583.
- (27) Piazza, R. Interactions and phase transitions in protein solutions. *Curr. Opin. Colloid Interface Sci.* **2000**, *5* (1–2), 38–43.
- (28) Schurtenberger, P.; Chamberlin, R. A.; Thurston, G. M.; Thomson, J. A.; Benedek, G. B. Observation of critical phenomena in a protein–water solution. *Phys. Rev. Lett.* **1989**, *63* (19), 2064.
- (29) Taratuta, V. G.; Holschbach, A.; Thurston, G. M.; Blankschtein, D.; Benedek, G. B. Liquid–liquid phase separation of aqueous lysozyme solutions: Effects of pH and salt identity. *J. Phys. Chem.* **1990**, *94* (5), 2140–2144.
- (30) Thomson, J. A.; Schurtenberger, P.; Thurston, G. M.; Benedek, G. B. Binary liquid phase separation and critical phenomena in a protein/water solution. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84* (20), 7079–7083.
- (31) Carlsson, F.; Malmsten, M.; Linse, P. Monte Carlo simulations of lysozyme self-association in aqueous solution. *J. Phys. Chem. B* **2001**, *105* (48), 12189–12195.
- (32) Rosch, T. W.; Errington, J. R. Investigation of the phase behavior of an embedded charge protein model through molecular simulation. *J. Phys. Chem. B* **2007**, *111* (43), 12591–12598.
- (33) Bianchi, E.; Blaak, R.; Likos, C. N. Patchy colloids: state of the art and perspectives. *Phys. Chem. Chem. Phys.* **2011**, *13* (14), 6397–6410.
- (34) Kern, N.; Frenkel, D. Fluid–fluid coexistence in colloidal systems with short-ranged strongly directional attraction. *J. Chem. Phys.* **2003**, *118* (21), 9882–9889.
- (35) Liu, H.; Kumar, S. K.; Sciortino, F. Vapor–liquid coexistence of patchy models: Relevance to protein phase behavior. *J. Chem. Phys.* **2007**, *127* (8), 084902–5.
- (36) Lomakin, A.; Asherie, N.; Benedek, G. B. Aeolotopic interactions of globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (17), 9465–9468.
- (37) Chang, J.; Lenhoff, A. M.; Sandler, S. I. Determination of fluid–solid transitions in model protein solutions using the histogram reweighting method and expanded ensemble simulations. *J. Chem. Phys.* **2004**, *120* (6), 3003–3014.
- (38) Li, X.; Gunton, J. D.; Chakrabarti, A. A simple model of directional interactions for proteins. *J. Chem. Phys.* **2009**, *131* (11), 115101–8.
- (39) Noya, E. G.; Vega, C.; Doye, J. P. K.; Louis, A. A. Phase diagram of model anisotropic particles with octahedral symmetry. *J. Chem. Phys.* **2007**, *127* (5), 054501–11.
- (40) Noya, E. G.; Vega, C.; Doye, J. P. K.; Louis, A. A. The stability of a crystal with diamond structure for patchy particles with tetrahedral symmetry. *J. Chem. Phys.* **2010**, *132* (23), 234511–13.
- (41) Romano, F.; Sanz, E.; Sciortino, F. Role of the range in the fluid–crystal coexistence for a patchy particle model. *J. Phys. Chem. B* **2009**, *113* (46), 15133–15136.
- (42) Romano, F.; Sanz, E.; Sciortino, F. Phase diagram of a tetrahedral patchy particle model for different interaction ranges. *J. Chem. Phys.* **2010**, *132* (18), 184501–9.
- (43) Song, X. Role of anisotropic interactions in protein crystallization. *Phys. Rev. E* **2002**, *66* (1), 011909.
- (44) Jones, S.; Thornton, J. M. Principles of protein–protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93* (1), 13–20.
- (45) Elcock, A. H.; McCammon, J. A. Calculation of weak protein–protein interactions: The pH dependence of the second virial coefficient. *Biophys. J.* **2001**, *80*, 613–625.
- (46) McGuffee, S. R.; Elcock, A. H. 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.* **2006**, *128* (37), 12098–12110.
- (47) Mereghetti, P.; Gabdouliline, R. R.; Wade, R. C. Brownian dynamics simulation of protein solutions: Structural and dynamical properties. *Biophys. J.* **2010**, *99* (11), 3782–3791.
- (48) Neal, B. L.; Asthagiri, D.; Lenhoff, A. M. Molecular origins of osmotic second virial coefficients of proteins. *Biophys. J.* **1998**, *75* (5), 2469–2477.
- (49) Chang, R. C.; Asthagiri, D.; Lenhoff, A. M. Measured and calculated effects of mutations in bacteriophage T4 lysozyme on interactions in solution. *Proteins: Struct., Funct., Bioinf.* **2000**, *41* (1), 123–132.
- (50) Grunberger, A.; Lai, P. K.; Blanco, M. A.; Roberts, C. J. Coarse-grained modeling of protein second osmotic virial coefficients: Sterics and short-ranged attractions. *J. Phys. Chem. B* **2013**, *117* (3), 763–770.
- (51) Lund, M.; Jönsson, B. A mesoscopic model for protein–protein interactions in solution. *Biophys. J.* **2003**, *85* (5), 2940–2947.
- (52) Hloucha, M.; Lodge, J. F. M.; Lenhoff, A. M.; Sandler, S. I. A patch–antipatch representation of specific protein interactions. *J. Cryst. Growth* **2001**, *232* (1–4), 195–203.
- (53) Bonneté, F.; Finet, S.; Tardieu, A. Second virial coefficient: Variations with lysozyme crystallization conditions. *J. Cryst. Growth* **1999**, *196* (2–4), 403–414.
- (54) Gripon, C.; Legrand, L.; Rosenman, I.; Vidal, O.; Robert, M. C.; Boué, F. Lysozyme–lysozyme interactions in under- and super-saturated solutions: A simple relation between the second virial coefficients in H<sub>2</sub>O and D<sub>2</sub>O. *J. Cryst. Growth* **1997**, *178* (4), 575–584.
- (55) Guo, B.; Kao, S.; McDonald, H.; Asanov, A.; Combs, L. L.; Wilson, W. W. Correlation of second virial coefficients and solubilities useful in protein crystal growth. *J. Cryst. Growth* **1999**, *196* (2–4), 424–433.
- (56) Moon, Y. U.; Curtis, R. A.; Anderson, C. O.; Blanch, H. W.; Prausnitz, J. M. Protein–protein interactions in aqueous ammonium sulfate solutions. Lysozyme and bovine serum albumin (BSA). *J. Solution Chem.* **2000**, *29* (8), 699–717.
- (57) Piazza, R.; Pierno, M. Protein interactions near crystallization: A microscopic approach to the Hofmeister series. *J. Phys.: Condens. Matter* **2000**, *12* (8A), A443–A449.
- (58) Ries-Kautt, M. M.; Ducruix, A. F. Relative effectiveness of various ions on the solubility and crystal growth of lysozyme. *J. Biol. Chem.* **1989**, *264* (2), 745–748.
- (59) Velev, O. D.; Kaler, E. W.; Lenhoff, A. M. Protein interactions in solution characterized by light and neutron scattering: Comparison of lysozyme and chymotrypsinogen. *Biophys. J.* **1998**, *75* (6), 2682–2697.
- (60) Maurer, R. W.; Sandler, S. I.; Lenhoff, A. M. Salting-in characteristics of globular proteins. *Biophys. Chem.* **2011**, *156* (1), 72–78.
- (61) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pK<sub>a</sub> values. *Proteins* **2005**, *61* (4), 704–721.
- (62) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pK<sub>a</sub> predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.

- (63) McQuarrie, D. *Statistical Mechanics*; Harper and Row: New York, 1976.
- (64) Zimm, B. H. Application of the methods of molecular distribution to solutions of large molecules. *J. Chem. Phys.* **1946**, *14* (3), 164–179.
- (65) Neal, B. L.; Asthagiri, D.; Velez, O. D.; Lenhoff, A. M.; Kaler, E. W. Why is the osmotic second virial coefficient related to protein crystallization? *J. Cryst. Growth* **1999**, *196* (2–4), 377–387.
- (66) Elcock, A. H.; Sept, D.; McCammon, J. A. Computer simulation of protein–protein interactions. *J. Phys. Chem. B* **2001**, *105* (8), 1504–1518.
- (67) Asthagiri, D.; Neal, B. L.; Lenhoff, A. M. Calculation of short-range interactions between proteins. *Biophys. Chem.* **1999**, *78* (3), 219–231.
- (68) Hamaker, H. C. The London–van der Waals attraction between spherical particles. *Physica* **1937**, *4* (10), 1058–1072.
- (69) Hunter, R. J. *Foundations of Colloid Science*, 2nd ed.; Oxford University Press: Oxford, 2004.
- (70) Roth, C. M.; Lenhoff, A. M. Improved parametric representation of water dielectric data for Lifshitz theory calculations. *J. Colloid Interface Sci.* **1996**, *179* (2), 637–639.
- (71) Roth, C. M.; Neal, B. L.; Lenhoff, A. M. van der Waals interactions involving proteins. *Biophys. J.* **1996**, *70* (2), 977–987.
- (72) Jorgensen, W. L.; Tirado-Rives, J. The OPLS potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657–1666.
- (73) Horton, N.; Lewis, M. Calculation of the free energy of association for protein complexes. *Protein Sci.* **1992**, *1* (1), 169–181.
- (74) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268* (5214), 1144–1149.
- (75) Neal, B. L.; Lenhoff, A. M. Excluded volume contribution to the osmotic second virial coefficient for proteins. *AIChE J.* **1995**, *41* (4), 1010–1014.
- (76) Asthagiri, D.; Paliwal, A.; Abras, D.; Lenhoff, A. M.; Paulaitis, M. E. A consistent experimental and modeling approach to light-scattering studies of protein-protein interactions in solution. *Biophys. J.* **2005**, *88* (5), 3300–3309.
- (77) Paliwal, A.; Asthagiri, D.; Abras, D.; Lenhoff, A. M.; Paulaitis, M. E. Light-scattering studies of protein solutions: Role of hydration in weak protein–protein interactions. *Biophys. J.* **2005**, *89* (3), 1564–1573.
- (78) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes: The Art of Scientific Computing*; Cambridge University Press: Cambridge, 1986.
- (79) Berntsen, J.; Espelid, T. O.; Genz, A. Algorithm 698: DCUHRE: An adaptive multidimensional integration routine for a vector of integrals. *ACM Trans. Math. Softw.* **1991**, *17* (4), 452–456.
- (80) Berntsen, J.; Espelid, T. O.; Genz, A. An adaptive algorithm for the approximate calculation of multiple integrals. *ACM Trans. Math. Softw.* **1991**, *17* (4), 437–451.
- (81) Ogunnaike, B. A. *Random Phenomena: Fundamentals of Probability and Statistics for Engineers*; CRC Press: Boca Raton, FL, 2010.
- (82) George, A.; Wilson, W. W. Predicting protein crystallization from a dilute solution property. *Acta Crystallogr. Sect. D* **1994**, *50* (4), 361–365.
- (83) Muschol, M.; Rosenberger, F. Interactions in undersaturated and supersaturated lysozyme solutions: Static and dynamic light scattering results. *J. Chem. Phys.* **1995**, *103* (24), 10424–10432.
- (84) Tanford, C. *Physical Chemistry of Macromolecules*; John Wiley & Sons: New York, 1961.
- (85) Coen, C. J.; Blanch, H. W.; Prausnitz, J. M. Salting out of aqueous proteins: Phase equilibria and intermolecular potentials. *AIChE J.* **1995**, *41* (4), 996–1004.
- (86) Vilker, V. L.; Colton, C. K.; Smith, K. A. The osmotic pressure of concentrated protein solutions: Effect of concentration and pH in saline solutions of bovine serum albumin. *J. Colloid Interface Sci.* **1981**, *79* (2), 548–566.
- (87) Baron, R.; Setny, P.; McCammon, J. A. Water in cavity–ligand recognition. *J. Am. Chem. Soc.* **2010**, *132* (34), 12091–12097.
- (88) Grigsby, J. J.; Blanch, H. W.; Prausnitz, J. M. Cloud-point temperatures for lysozyme in electrolyte solutions: effect of salt type, salt concentration and pH. *Biophys. Chem.* **2001**, *91* (3), 231–243.
- (89) Vagenende, V.; Han, A. X.; Mueller, M.; Trout, B. L. Protein-associated cation clusters in aqueous arginine solutions and their effects on protein stability and size. *ACS Chem. Biol.* **2013**, *8* (2), 416–422.
- (90) Pellicane, G.; Smith, G.; Sarkisov, L. Molecular dynamics characterization of protein crystal contacts in aqueous solutions. *Phys. Rev. Lett.* **2008**, *101* (24).
- (91) Thomas, B. R.; Vekilov, P. G.; Rosenberger, F. Heterogeneity determination and purification of commercial hen egg-white lysozyme. *Acta Crystallogr. Sect. D* **1996**, *52*, 776–784.