THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

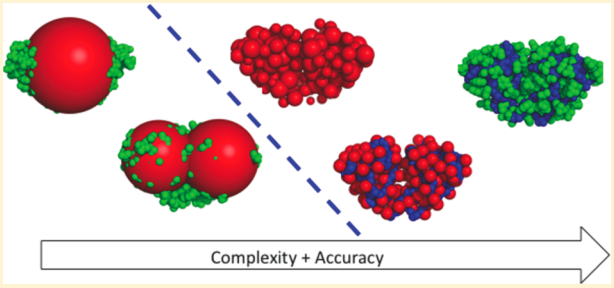
Coarse-Grained Modeling of Protein Second Osmotic Virial Coefficients: Sterics and Short-Ranged Attractions

Alexander Grünberger,[†] Pin-Kuang Lai,[†] Marco A. Blanco, and Christopher J. Roberts*

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

Supporting Information

ABSTRACT: A series of coarse-grained models, with different levels of structural resolution, were tested to calculate the steric contributions to protein osmotic second virial coefficients ($B_{22,S}$) for proteins ranging from small single-domain molecules to large multidomain molecules, using the recently developed Mayer sampling method. $B_{22,S}$ was compared for different levels of coarse-graining: four-beads-per-amino-acid (4bAA), one-bead-per-amino-acid (1bAA), one-sphere-per-domain (1sD), and one-sphere-per-protein (1sP). Values for the 1bAA and 4bAA models were quantitatively indistinguishable for both spherical and nonspherical proteins, and the agreement with values from all-atom models improved with increasing protein size, making the CG approach attractive for large proteins of biotechnological interest. Interestingly, in the absence of detailed structural information, the hydrodynamic radius (R_h) along with a simple 1sP approximation provided reasonably accurate values for $B_{22,S}$ for both globular and highly asymmetric protein structures, while other 1sP approximations gave poorer agreement; this helps to justify the currently empirical practice of estimating $B_{22,S}$ from R_h for large proteins such as antibodies. The results also indicate that either 1bAA or 4bAA CG models may be good starting points for incorporating short-range attractions. Comparison of gD-crystallin B_{22} values including both sterics and short-range attractions shows that 1bAA and 4bAA models give equivalent results when properly scaled to account for differences in the number of surface beads in the two CG descriptions. This provides a basis for future work that will also incorporate long-ranged electrostatic attractions and repulsions.



1. INTRODUCTION

Protein second osmotic virial coefficients (B_{22}) provide a useful link between experimental measures of solvent-averaged protein−protein interactions and molecular scale models for those interactions.[1−5] Fully atomistic models can be prohibitive in terms of computational burden, making coarse-grained (CG) models lucrative if one can determine what level of coarse-graining is sufficient.[6−8] B_{22} is related to the solvent-averaged potential of mean force between two proteins in solution (W_{22}) via the following expression[9]

$$B_{22} = -\frac{1}{2}\int\cdots\int[\exp(-W_{22}(r, \Omega_1, \Omega_2) - 1]\, dr\, d\Omega_1\, d\Omega_2 \quad (1a)$$

$$B_{22} = B_{22,S} - \frac{1}{2}\int\cdots\int[\exp(-W'_{22}(r, \Omega_1, \Omega_2) - 1]\, dr\, d\Omega_1\, d\Omega_2 \quad (1b)$$

In eq 1a, W_{22} is explicitly shown to be a function of the center-to-center distance between two proteins (r), and the orientation vectors of each protein relative to its respective center of mass (Ω_1 and Ω_2, respectively). The expressions above implicitly assume that each protein can be treated as a rigid body: a reasonable assumption for proteins that adopt a stable folded three-dimensional structure in solution.[1,2]

Equation 1b separates B_{22} into the contributions from just steric interactions ($B_{22,S}$) and the remaining contributions, i.e., nonsteric repulsions and all sources of attractions. This is denoted by the prime on W_{22}, indicating that the steric contributions to W_{22} have been integrated into the value of $B_{22,S}$. This is a natural separation, as sterics are the minimum level of interactions that must be included in any molecular model of a real system. It has been argued elsewhere[10−12] that $B_{22,S}$ provides a useful reference point for interpreting experimental B_{22} values when considering whether nonsteric repulsions are a significant factor in stabilizing proteins against unwanted self-association or aggregation. $B_{22,S}$ also provides a natural scaling parameter for B_{22} when attempting to compare B_{22} values for proteins of greatly differing size or shape.[13]

Experimentally, however, it is difficult if not impossible to directly measure $B_{22,S}$ for a given protein. While $B_{22,S}$ is relatively simple to estimate if a protein is well approximated as a single sphere with a known radius of gyration (R_g) or hydrodynamic radius (R_h), when one considers proteins with asymmetric or generally nonspherical structures, it is difficult to know *a priori* how to accurately estimate $B_{22,S}$. This is an increasingly problematic situation when one considers large, multidomain proteins that are of importance in a number of

Received: August 18, 2012
Revised: December 12, 2012
Published: December 17, 2012

biotechnology applications.[14,15] Larger proteins are also more problematic if one seeks to calculate $B_{22,S}$ via all-atom potentials, due to the large computational burden. As such, one goal of the present work is to assess different levels of coarse-graining when calculating $B_{22,S}$ to determine if an optimal level can be identified.

A longer-term objective is to provide accurate coarse grained (CG) models for rapidly calculating B_{22} values for comparison with experimental systems, by including additional interactions, e.g., van der Waals attractions, screened electrostatic attractions/repulsions, etc. In the present report, a first step in this direction is considered by including short-ranged attractions within CG models that have first been shown to accurately capture steric effects for B_{22}. Short-ranged attractions naturally arise in B_{22} as a result of hydrophobic attractions between aliphatic and aromatic protein side chains. They also occur due to van der Waals interactions between side chains and backbone groups, relative to those interactions with surrounding water molecules.[1,2,16] Future work will consider additional interactions such as electrostatics, and ultimately compare the accuracy and speed of CG versus all-atom models. As CG models typically are sought as a means to greatly reduce computational time while retaining some key level of structural detail, these will likely remain implicit-solvent models.

This report first considers four different levels of coarse-graining to calculate $B_{22,S}$, spanning from the simplest case (1 sphere per protein, 1sP) to a structurally detailed 4-beads-per-amino acid (4bAA) level that is increasingly used to capture the essential physics in protein folding and polypeptide aggregation.[6−8] Proteins spanning from small globular structures to highly asymmetric antibody structures are compared, using Mayer sampling to obtain $B_{22,S}$. Literature values are available for some of the selected proteins and are compared against the different levels of course graining. Overall, the results indicate that CG models can give a quantitatively reasonable approach to determine $B_{22,S}$ for proteins of different size and shape, and also provide theoretical corroboration for some common heuristic approaches to estimate $B_{22,S}$ from indirect experimental measures of the effective hard-sphere diameter of a protein. These results allow one to focus further model refinement on only those CG models that can reasonably reproduce correct steric contributions to B_{22}.

On the basis of those results, one of the proteins—human gamma-D-crystallin (HgD-Crys)—is then used as a test case for how the addition of short-ranged attractions affects the resulting B_{22} value for the 1bAA and 4bAA cases, including the contributions from $B_{22,S}$. The results indicate a simple scaling behavior that makes the two levels of coarse-graining equivalent and allows one to make use of the simpler (and faster) one if desired. As noted above, this also provides a basis for future work to include the effects of screened electrostatics in a more general model of CG protein−protein interactions for calculation of complete B_{22} values for a variety of proteins beyond the HgD-Crys model system tested here.

2. METHODS

Steric Contributions to B_{22}. For $B_{22,S}$ calculations, six model proteins of different sizes and shapes were tested. Figure 1 shows illustrative space-filling representations (only hG-CSF not shown) with side chain atoms (green) and backbone atoms (blue) (hydrogens not shown). The corresponding PDB file names are listed in Table 1. The selected proteins are tuna heart cytochrome C (thCytC, MW ∼ 12 kDa); bovine α-chymotrypsinogen A (aCgn, MW ∼ 25.7 kDa); deoxyhemoglobin (deoxy-HB, MW ∼ 68 kDa); human gamma-D-crystallin (HgD-Crys, MW ∼ 20.6 kDa); human granulocyte colony-stimulating factor (hG-CSF, MW ∼ 18 kDa); and an illustrative immunoglobulin-gamma, type 1 (IgG1) antibody (MW ∼ 147 kDa). The first three in this list have a folded structure that is roughly spherical or globular, while the last three have increasingly asymmetric or extended structures. For the IgG1, a PDB structure produced from the combination of crystal structures of different antibody fragments by Padlan[17] was used.

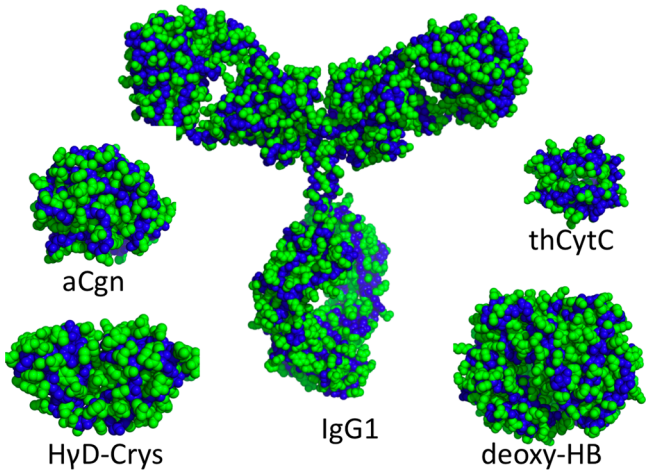


Figure 1. Illustrative all-atom representations of some of the different proteins used as model systems for calculating $B_{22,S}$ values. Structures are semiquantitatively shown to scale, to illustrate the different levels of anisotropy and size of the protein structures. hG-CSF (not shown) is similar in dimensions to HgD-Crys.

Table 1. Summary of Protein Structures

protein	PDB code
thCytC	3CYT
aCgn	1EX3
Deoxy-HB	2HHB
HgD-Crys	1HKO
hG-CSF	1GNC
IgG1	see Padlan et al.[17]

All CG models used in this work were constructed from the coordinates obtained from the Protein Data Bank (PDB) (see also Table 1). Each protein was modeled as either a single sphere per protein (1sP) or a collection of smaller spheres. The different representations are illustrated for HgD-Crys in Figure 2. For the 1sP cases, three different estimates were used to obtain the sphere diameter (σ): experimental R_h values were set equal to $\sigma/2$; values of R_g from experiment or calculated from the PDB structure were used by setting $2R_g = (3/5)^{1/2}\sigma$; alternatively, σ was based on the molecular volume of the protein, estimated from the partial specific volume ν_2 and the molecular weight (MW) as

$$\sigma = 2\left(\frac{3}{4\pi}\nu_2 \mathrm{MW}\right)^{1/3} \tag{2}$$

Table 2 summarizes the σ values for the 1sP calculations based on R_h, R_g, and the product of MW and ν_2.

When using multiple spheres per protein, two different levels of coarse-graining were tested for each protein: a one-bead-per-amino-acid (1bAA) and a four-beads-per-amino-acid (4bAA)

dx.doi.org/10.1021/jp308234j | J. Phys. Chem. B 2013, 117, 763−770

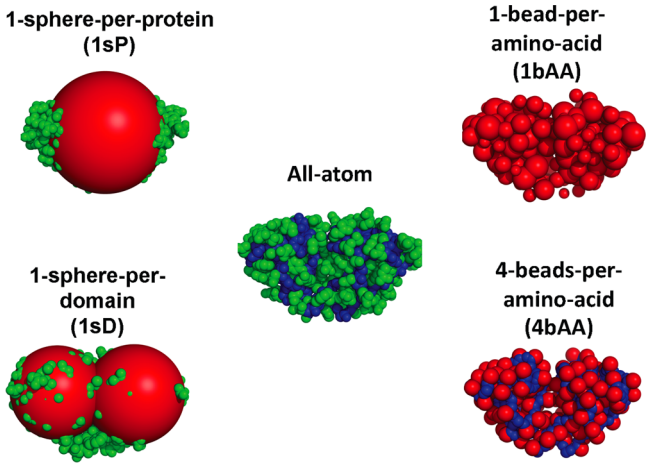


Figure 2. Visual comparison of the all-atom, 4bAA, 1bAA, 1sD, and 1sP (based on R_h) structural representations for HgD-Crys. Large red spheres represent the size and position of the spheres for the 1sP and 1sD CG models, with the green beads showing the regions of the all-atom structure that lie outside the 1sP or 1sD spheres. For the other structures, only the beads corresponding to the atoms or beads included in the structural model of HgD-Crys are shown, with different colors used simply for visual clarity.

Table 2. Values of 1sP Diameter for Each Protein, Based on Different Estimates of Molecular Radius or Volume, Using Experimental Values for Hydrodynamic Radii and Radius of Gyration,[34−50] as Well as Measured Partial Specific Volumes[44−46,51−54] and Known MW Values (See Text)

	1sP value of σ ([=] nm)		
	from R_h	from R_g	from v_2MW
thCytC	3.5	4.1	3
aCgn	4.3	5.4	3.9
deoxy-HB	6.4	7.9	5.4
HgD-Crys	3.8	5.6	3.6
hG-CSF	4.7	6.2	3.6
IgG1	11	15.5	7

representation. In the 1bAA model, each amino acid was represented as a single sphere or "bead". The center of each bead was placed at the centroid of the corresponding amino acid.[18] The bead diameter depended on the amino acid of interest, and was based on the average R_g of that type of amino acid when averaged across the folded structures considered here; values are summarized in Table 3, and are in reasonable agreement with those of Levitt.[19]

The 4bAA model was similar to those described previously,[20,21] with each amino acid represented by one bead for the side chain and three beads for the amino acid backbone—one each for the α carbon (σ = 0.37 nm), the carbonyl group (σ = 0.35 nm), and the amine group (σ = 0.29 nm). The parameters used by Bereau and Deserno[20] were used here. Geometric parameters from other models were compared[20−24] and differed only marginally. All side chains were assigned the same bead diameter (0.5 nm), with the exception of glycine, that was not assigned a side chain bead. For proteins that had significantly asymmetric structures (HgD-Crys, hG-CSF, IgG1), calculations were also performed in which the structure was represented by one sphere per domain (1sD). For reasons described in the Results and Discussion section, these calculations are not elaborated upon in detail, and are included only for the sake of completeness.

Table 3. Average R_g Values Calculated from PDB Structures for the Proteins Tested Here (Standard Deviations Less than 4% in All Cases) and the Corresponding 1bAA Bead Diameter Values (σ) for Each Amino Acid

amino acid	R_g (nm)	σ (nm)
Lys	0.259	0.670
Glu	0.233	0.602
Asp	0.200	0.517
Asn	0.196	0.507
Ser	0.163	0.421
Arg	0.296	0.763
Gln	0.233	0.601
Pro	0.169	0.436
Thr	0.174	0.451
Gly	0.134	0.345
His	0.231	0.596
Ala	0.145	0.375
Tyr	0.269	0.696
Cys	0.171	0.443
Trp	0.272	0.703
Val	0.177	0.456
Met	0.231	0.597
Ile	0.198	0.511
Phe	0.245	0.633
Leu	0.206	0.532

For each of the representations described above, the interactions between each bead or sphere in two different proteins were represented by a steeply repulsive soft-sphere potential

$$\phi_{ij} = \left(\frac{\sigma_{ij}}{r_{ij}}\right)^n \tag{3}$$

while spheres within the same protein had zero interaction energy. In eq 3, indices i and j indicate spheres on two different proteins, r_{ij} is the center-to-center distance between spheres i and j, $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, and n = 128 for the calculations performed here. The value of the potential of mean force (W_{22}) for a given configuration of two proteins at a given center-to-center distance and set of orientations was given as a sum of pairwise interactions (ϕ_{ij}) between each bead or sphere (i) of one protein with a counterpart (j) of the other proteins. As the solvent was implicit, W_{22} was therefore formally equivalent to a potential energy function in conventional nomenclature.

Short-Ranged Attractive Contributions to B_{22}. When both steric repulsions and short-ranged attractions were included in the interbead potentials, the following was used

$$\phi_{ij} = 4\varepsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{128} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{4}$$

which is equivalent to a slightly modified version of a standard 12-6 Lennard-Jones potential.

Mayer Sampling. The Mayer sampling algorithm[25,26] is a recently developed computational method for determining virial coefficients based on a biased Monte Carlo integration of eq 1, using a known reference system such as a hard sphere to aid in an umbrella sampling approach. This is a biased sampling method that allows one to avoid sampling configurations that

dx.doi.org/10.1021/jp308234j | J. Phys. Chem. B 2013, 117, 763−770

do not contribute appreciably to the integral in eq 1. For example, when one considers only steric interactions, the value of the integrand in eq 1 is either zero (no overlap of proteins) or −1 (steric overlap). In this case, the Mayer sampling algorithm biases one to avoid sampling configurations where proteins do not overlap at all, and therefore, the integral in eq 1 converges much more rapidly and accurately than if one considers a random search or attempts to integrate the multidimensional integral directly. When attractions are included, configurations without overlap are also sampled, with a bias toward sampling more for those configurations with larger attractive interactions. The bias is corrected for analytically in the final result, by comparing with a reference system with a known value of B_{22}.[25,26]

In principle, the choice of reference is arbitrary, provided one does not select a reference system that causes the Mayer algorithm to artificially over- or under-sample important contributions to the integral in eq 1.[25] A simple HS was used as the reference system here, and the results were confirmed as a function of the location and diameter of the HS within the candidate protein structure (see also Results and Discussion). For multidomain proteins, the HS reference was placed at the center of the largest domain, while for single-domain proteins it was the center of the protein. Preliminary calculations using unbiased MC integration of eq 1 proved untenable to reach converged B_{22} values once one considered proteins larger than aCgn (not shown).

As sampling is conducted in free space in the Mayer algorithm,[25] a maximum displacement (ΔR_{12}) needed to be set for trial translation moves in order to avoid issues with slow convergence or unreasonably small acceptance ratios for trial moves. ΔR_{12} was adjusted during short preliminary runs ($\sim 10^5$–10^6 trials) to yield approximately 50% acceptance. ΔR_{12} was then fixed for the full Mayer sampling runs, which were typically between 10^7 and 10^8 trials per simulation, with translations and rotations sampled randomly.

3. RESULTS AND DISCUSSION

To ensure that the values of $B_{22,S}$ obtained from the Mayer sampling algorithm were accurate, converged values were compared as a function of the number of MC trials, initial configurations and seed values, and the size and placement of the single reference hard sphere within the geometry of a given CG protein structure. Generally, convergence was achieved within 10^7–10^8 trials after an initial "pre-equilibration" period, during which the maximum ΔR_{12} was set for each trial move, so as to ensure a reasonable rate of acceptance (see also Methods). The more detailed the structure (e.g., 1bAA and 4bAA), the more trials that were needed for convergence.

Kofke and co-workers[25,26] noted in the original development of Mayer sampling that, while the choice of reference state is arbitrary in principle, there can be subtle but practically important details to consider. As this is based on an umbrella sampling approach, one must ensure that the space that is sampled for the reference system and the system of interest have sufficient overlap. In the present case, this is controlled by choosing the size of the reference HS diameter. In the Mayer sampling algorithm,[25,26] moves are accepted only if the value of the integrand in eq 1a is nonzero for the system of interest (not necessarily the reference system). When calculating steric-only interactions, this means that configurations in which there is no overlap between the proteins are disallowed. However, if one chooses a reference state diameter and position that lies sufficiently outside the molecular volume of the protein of interest, configurations that are relevant for the reference system, i.e., that have HS overlap, will be excluded from the sampling, and the resulting $B_{22,S}$ value may be significantly in error.

In the present case, these factors translate to a question of where to place the center of the reference HS within a given CG protein structure, and what diameter to choose for the reference HS system. As noted in the preceding section, the reference HS was placed with its center either at the geometric center of the protein (for single-domain proteins) or in the center of the largest domain. For HgD-Crys, the domains are of similar size and overlap with each other (see Figure 2). In that case, it is also possible to place the HS reference at the geometric center of the protein and still have the reference system be well embedded within the CG protein structure.

Figure 3 shows the effects of choosing different values of the HS diameter (σ_{HS}) on the converged value of $B_{22,S}$ for one of

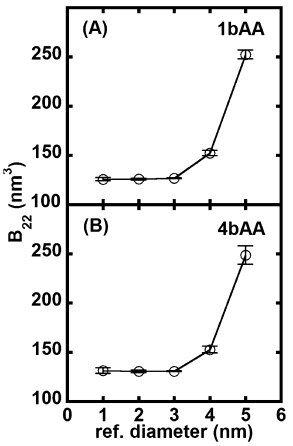


Figure 3. Effect of reference HS diameter on the resulting value of $B_{22,HS}$ from Mayer sampling, using HgD-Crys as an example.

the model proteins: HgD-Crys. The results show that, as long as one chooses a value of σ_{HS} that is sufficiently small to ensure the entire HS reference state is embedded within the protein structure of interest, then the value of $B_{22,S}$ is insensitive to the precise choice of σ_{HS}. Similar results (not shown) were found for each of the other proteins, for a given CG representation. It is worth noting that the direct-sampling version of Mayer sampling was used throughout the work reported here.[25,26] An alternative, overlap-sampling version of Mayer sampling has been developed more recently, that likely would be less sensitive to the effect of the reference choice.[26]

Steric Contributions to B_{22}. To quantitatively compare the results for $B_{22,S}$ as a function of the degree of coarse graining, Figure 4 shows $B_{22,S}$ from different CG methods. The values of $B_{22,S}$ calculated from different CG structures are given by the different bars for each protein. The labeling pattern of the bars for thCytC holds for the other proteins. The 1sP models are labeled as R_h, R_g, and ν_2MW to indicate how the effective radius in the 1sP model was estimated. Values of $B_{22,S}$ are scaled by the all-atom values to allow proteins of significantly different volume to be shown on the same scale. The all-atom values for thCytC, aCgn, and deoxy-HB are available in the literature,[16] while the value for HgD-Crys was calculated using the Mayer sampling as described above but using one bead for each non-hydrogen atom (no explicit

dx.doi.org/10.1021/jp308234j | J. Phys. Chem. B 2013, 117, 763−770

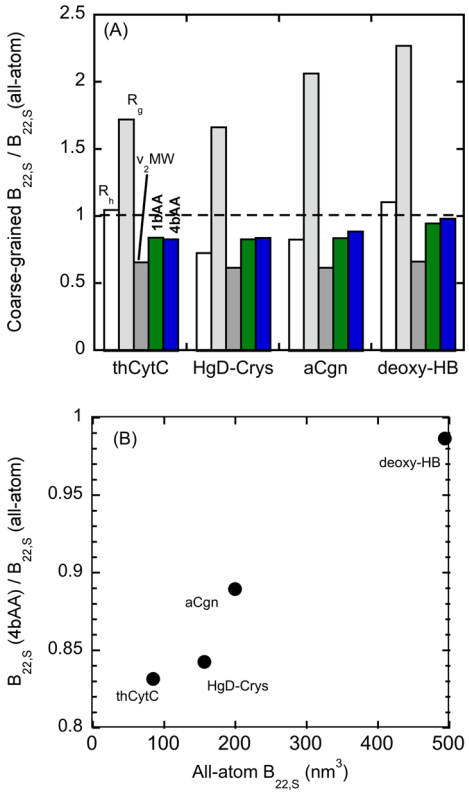


Figure 4. (A) Comparison of $B_{22,S}$ values for the 4bAA, 1bAA, and 1sP models for the proteins with available values from all-atom calculations. The $B_{22,S}$ values are scaled by the corresponding all-atom value for a given protein, so as to place the results on a common scale. (B) Relative value of $B_{22,S}$ from 4bAA to the all-atom value, as a function of the size of the protein in terms of $B_{22,S}$(all-atom).

hydrogens) and using values for σ for each bead in eq 3 from the AMBER force field. As noted in the Methods section, the 1sP values correspond to different choices for σ for the single effective sphere for a given protein, i.e., based on values of R_h, R_g, or the product of MW and v_2 for a given protein.

The trends in Figure 4A indicate, as expected, that a higher degree of structural detail (e.g., 1bAA and 4bAA) leads to better agreement with the all-atom calculations. It is clear in Figure 4A that R_g and v_2MW with the 1sP approximation provide a poor estimate of $B_{22,S}$. This is perhaps not surprising, based on earlier arguments[16] that highlight that an approach based on molecular volume (v_2MW) consistently underestimates the steric contributions to B_{22}, and the fact that many of the proteins are not well approximated as spherical. The ability of R_h with the 1sP approximation to estimate $B_{22,S}$ is discussed below.

Figure 4B shows that the agreement between the 4bAA approximation and the all-atom results improves with increasing protein size, and in general is within 10−15% of the all-atom values. This is perhaps not unexpected, as more of the contributions to $B_{22,S}$ arise for smaller proteins from the particular surface geometry and its contribution to excluded volume. In contrast, for larger proteins, the fraction of $B_{22,S}$ that is expected to arise from the overlap of the surface, as opposed to the "core", of two proteins is significantly reduced. From a pragmatic perspective, it is notable that B_{22} values from experiment often have significant statistical uncertainty, once one properly accounts for propagation of errors and the statistical uncertainty inherent in fitting the values; typical 95% confidence intervals on experimental B_{22} values are often 10−15% of the value regressed from measurements such as laser light scattering.[11,27] Thus, even the differences between $B_{22,S}$ from 4bAA and all-atom calculations may be within the uncertainty of the experimental values for B_{22} against which they are compared.

Figure 5A shows the results of $B_{22,S}$ as a function of the different CG approaches for all six test proteins, with all values

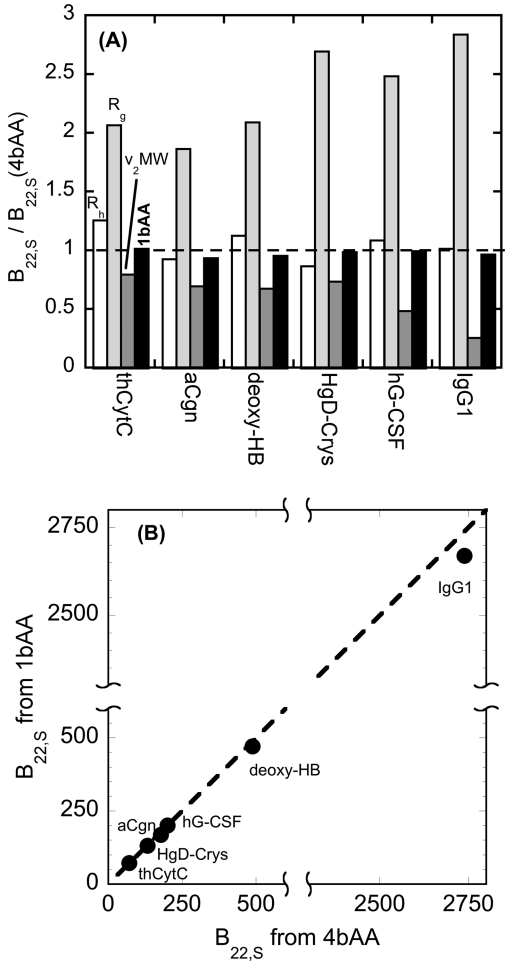


Figure 5. (A) Comparison of $B_{22,S}$ for the 1bAA and each of the 1sP models across all test proteins, scaled by the corresponding 4bAA values to place each on the same scale. (B) Absolute values of $B_{22,S}$ (units = nm^3) for 1bAA versus 4bAA for each model protein.

scaled by the corresponding 4bAA values. Scaling the results for each protein against the 4bAA value for that protein then places all results on the same y-axis scale for easier comparison across proteins with very different sizes. Ideally, one might use the all-atom $B_{22,S}$ values for scaling, but this value is not available for the IgG. As a surrogate, and based on the results above for comparing 4bAA with all-atom results, the $B_{22,S}$ values from the 4bAA model were used as the most structurally detailed and accurate ones available for comparing all proteins here with the other CG models in Figure 5A.

The 1bAA and 4bAA show quantitative agreement across the entire set of test proteins (cf. Figure 5B), while 1sP representations using R_g or v_2MW provide relatively poor results (cf. Figure 5A). In general, using R_g to estimate the effective HS diameter gives values of $B_{22,S}$ that are much greater than the correct value, while that estimated from the product

dx.doi.org/10.1021/jp308234j | J. Phys. Chem. B 2013, 117, 763−770

ν_2MW consistently underestimates $B_{22,S}$ (Figure 5A). Assuming the above reasoning holds regarding the relative contributions from surface residues compared to those in the folded "core" of the protein (or its respective domains) as one increases the size or molecular weight of the protein, one would also expect that the 1bAA and 4bAA CG models will give good quantitative agreement with all-atom values of $B_{22,S}$ when those become available for IgG1 structures. Calculating those values was beyond the scope of this work, for the reasons noted in the introduction regarding the computational cost for B_{22} when dealing with large proteins. Doing so is part of future extensions of this work but may be unneeded based on the discussion below.

Interestingly, $B_{22,S}$ from a 1sP representation with $R_h = \sigma/2$ provides reasonable results across the range of proteins tested here, particularly once one considers that the typical uncertainty in experimental B_{22} values is similar in magnitude to the differences between $B_{22,S}$ from all-atom results and the 1sP $B_{22,S}$ value using R_h to provide σ (see also discussion above). Although only a phenomenological finding, it provides at least a molecular model-based corroboration for the otherwise empirical experimental practice of estimating $B_{22,S}$ for asymmetric proteins based on their R_h values.[10,28] This is potentially useful, in that it suggests one can circumvent detailed or cumbersome calculations of $B_{22,S}$ and simply measure R_h for a given protein. This is particularly attractive when one considers proteins for which no reasonable atomic-resolution structure is available, from which to construct all-atom or CG models for calculating $B_{22,S}$ as a reference state for interpreting experimental B_{22} values.

In terms of attempting to justify the above result, fundamentally the R_h value reflects the size of an "effective" sphere with the same value of the tracer diffusion coefficient as that for the real molecule. For a nonspherical protein, to some extent this involves averaging how the excluded volume of the protein perturbs the solvent spatially as the protein moves through the solvent—with different regions of the protein contributing larger or smaller "effective sizes" to the net average R_h value. Thus, the R_h value for a nonspherical object can be much less than its R_g value. For $B_{22,S}$, a somewhat conceptually similar averaging process is involved, except that now it is a question of the excluded volume of one protein with respect to another protein. The excluded volume of a highly nonspherical protein with maximum dimension R_{max} is much lower than what one would obtain for a spherical object with radius R_{max}. This argument is, however, at best only semiquantitative, as the excluded volume of one protein with respect to another is significantly different from the excluded volume of a protein with respect to a water molecule. Therefore, while it is perhaps not surprising that R_g is not a useful predictor for $B_{22,S}$ with a 1sP model, this argument cannot fully justify the almost quantitative prediction of $B_{22,S}$ values from R_h with the 1sP model.

1sD models for the asymmetric proteins HgD-Crys, hG-CSF, and IgG1 were also tested, but the results are not shown for pragmatic reasons. First, they did not show significant improvement over the 1sP models. Second, and more importantly, when one considers a 1sD model, the choice of where to place the sphere for each domain and the diameter of that domain becomes essentially arbitrary. The results above for the 1sP model showed that using a simple and easily accessible estimate for σ from the MW and ν_2 values, or from the R_g value (if such a value is even known), is not reasonably accurate. While R_h provided a good estimate of σ for $B_{22,S}$ values, such R_h values are typically obtained experimentally only for the entire protein, not for a given domain. While algorithms exist to estimate R_h for globular proteins or subdomains,[29] this requires one to know *a priori* the crystal or NMR structure of the protein of interest. Thus, if one is interested solely in the value of $B_{22,S}$ for a given protein—independent of whether a detailed crystal or NMR structure is available—the present results indicate that using $\sigma = 2R_h$ may be sufficient, with R_h from simple light scattering or analytical ultracentrifugation measurements.

Short-Ranged Attractive Contributions to B_{22}. From the perspective of building more detailed CG models for folded proteins, including both sterics and attractions or nonsteric repulsions, the results here suggest that either a 1bAA or a 4bAA representation is a reasonable starting point. In addition, the Mayer sampling algorithm proved to be a robust approach to determining B_{22} values, and is easily extended to consider additional interactions between proteins. As a first test, B_{22} values for HgD-Crys were calculated using Mayer sampling, employing either a 1bAA or a 4bAA representation. The steric contributions were handled as described above, while the attractions were described by a simple short-ranged attraction akin to the attractive portion of a Lennard-Jones potential. The scaling parameters for interbead distance (σ_{ij}) remained the same as used above, leaving only the well depth (ε) as an additional parameter. Equivalently, this leaves the effective dimensionless temperature ($T^* = k_BT/\varepsilon$, with k_B denoting Boltzmann's constant) as an adjustable parameter.

Figure 6A shows the results of B_{22} vs $1/T^*$ for the 1bAA and 4bAA cases of HgD-Crys based on the calculations described in the preceding paragraph and the Methods section. The trends are semiquantitatively similar, with B_{22} approaching its expected limit of $B_{22,S}$ at high T^* (low inverse temperature), and with increasingly negative values of B_{22} as T^* is reduced. The values of B_{22} are more sensitive to T^* for the 4bAA case, as expected

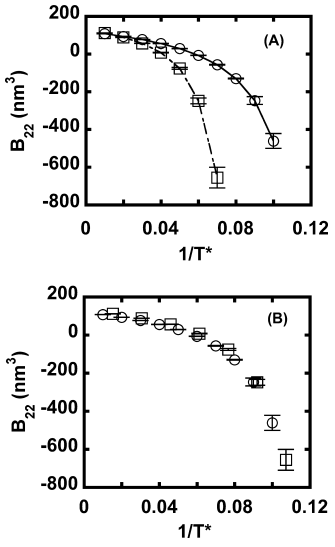


Figure 6. (A) Comparison of B_{22} for the 1bAA and 4bAA models of HgD-Crys when both steric repulsions and short-ranged attractions are included. (B) Comparison of the same B_{22} values from panel A but with the temperature values of the 4bAA results rescaled by a factor of 1.533 to empirically account for the different effective ε value for the 1bAA model compared to a 4bAA representation.

dx.doi.org/10.1021/jp308234j | J. Phys. Chem. B 2013, 117, 763−770

on the basis of the following argument. The 4bAA model has more beads accessible on the surface of a given protein molecule when compared to the 1bAA model. As such, there are more pairwise interactions that will contribute appreciably to eq 1 for any given configuration of a pair of proteins if the proteins are not overlapping. Thus, a 1bAA representation underestimates the number of bead−bead interactions between two proteins if one assumes ε is the same for 1bAA and 4bAA representations.

That is, W is a sum of pairwise interactions between all beads in the two proteins, but the bead−bead interaction is short ranged (∼ a bead diameter). As such, most of the beads that are not on the surface will experience weak or negligible attractive interactions with beads of the other protein. A naïve expectation therefore might be that W for the 4bAA model will be 4 times larger than that for the 1bAA model if one considers the same protein−protein distance and relative orientation. However, many of the beads that represent the backbone of surface amino acids are buried behind side-chain beads, and one would then expect a scaling factor that is between 1 and 4.

In addition, the range of the attractions is somewhat longer for the 1bAA model than for the 4bAA model, as the range scales as the bead diameter, and the 1bAA diameters are typically larger those in the 4bAA model. Both of these effects are expected to make the scaling parameter significantly less than 4. The value of W always appears as W/kT in the evaluation of B_{22}, and therefore, it is natural to consider rescaling temperature when considering whether the 1bAA and 4bAA models give similar results. Equivalently, one may consider that the 4bAA representation should employ a smaller value of ε, or a higher T^*, in order for the 1bAA and 4bAA models to give the same results.

Figure 6B shows the same values of B_{22} from Figure 6A but now with the T^* values for the 4bAA results rescaled by a factor of 1.533. The values of B_{22} match quantitatively if the temperature is rescaled for the 4bAA relative to the 1bAA calculations. The Supporting Information (Figure S1) also illustrates that the distribution of configurations (or, more accurately, the values of the integrand in eq 1 sampled in the Mayer algorithm) are the same for the two CG models if this temperature rescaling is employed.

The value of the scaling parameter for T^* was determined empirically by regressing different scaling values for T^* (or for ε) until the B_{22} curves vs inverse temperature overlapped for the 1bAA and 4bAA cases. The value of the scaling parameter is numerically close to 3/2; however, we have not yet identified whether a simple geometric argument can produce this scaling value, or whether it will depend on the protein size, shape, and/or degree of curvature of the protein surface. As noted above, while there are 4 times as many beads per amino acid for the 4bAA versus the 1bAA model, clearly the scaling parameter is much less than 4. Structurally, this is not surprising because many of the backbone beads are not as surface accessible as the bead corresponding to the side-chain in the 4bAA model. However, a precise argument is not yet apparent as to why the particular value near 3/2 was obtained. As such, it may be highly dependent on the protein in question, and future work will be needed to test this for other proteins and/or provide a more fundamental basis for the value of such a scaling parameter.

The observation that a single scaling parameter can reduce the 1bAA and 4bAA results to a common curve supports the argument above that one may choose either CG representation so long as one can tune the effective temperature. For example, experimentally B_{22} often reaches a plateau value at high enough salt concentrations, i.e., in the limit of small screening lengths for electrostatics, where sterics and short-ranged (non-electrostatic) attractions may be dominant contributions to B_{22}.[2] For typical protein dimensions (overall diameter of the order of 1−20 nm), salt concentrations on the order of a few hundred millimolar can produce such conditions, where electrostatic interactions are greatly screened, but before higher-order solvation effects become important at high (∼ molar) salt concentrations.[10,28,30−32] The present results suggest that either a 1bAA or a 4bAA level of coarse-graining may be sufficient for building more sophisticated models that add interactions such as screened electrostatics between each titratable amino acid on the surface, provided one is not focused on very high salt concentrations where some type of explicit treatment of the solvent is arguably unavoidable.[1,33] Testing such an approach is part of ongoing work that will be provided in a future report.

In addition, the observation that either 1bAA or 4bAA models provide equivalent results for B_{22} (Figure 6) and the distribution of states contributing to B_{22} (Figure S1, Supporting Information) suggests that the structural "complementarity" of the surfaces is not the overarching contribution to B_{22}, as long as a reasonable resolution CG model is employed, i.e., 1bAA and 4bAA were very similar, but lower-resolution CG models may not work as well, and clearly do not accurately capture the steric contributions. Future work will seek to address, in more detail, the question of whether just a small number of highly specific interactions dominate the net value of B_{22} at experimentally relevant conditions for crystallization and protein stability studies.[1−4,10−13]

4. SUMMARY

Comparison of steric contributions to B_{22} for a series of proteins of increasing size and non-globular morphology showed that 1bAA and 4bAA coarse-grained modeling approaches are reasonably accurate, particularly as one considers larger proteins. Phenomenologically, reasonable values are also obtained if one employs a single hard sphere with radius equal to the experimental hydrodynamic radius. Using gamma-D-crystallin as a test case for also incorporating short-ranged attractions with either 1bAA or 4bAA descriptions, a simple scaling relationship was obtained that makes either approach quantitatively equivalent if one rescales the temperature or the well depth of the attractions (on a per bead basis) to account empirically for the different number of interacting spheres that are exposed at the protein surface for a 1bAA versus 4bAA coarse-grained description. Future work will extend the present results to consider electrostatic interactions that provide longer-range attractions and repulsions within a more general CG model for protein B_{22} calculations.

■ ASSOCIATED CONTENT

Supporting Information

Figure S1 shows the distribution of energy states sampled with a 1bAA and 4bAA representation under conditions with both steric and short-ranged attractive interactions for gD-Crys. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Address: 150 Academy St., Newark, DE 19716. Phone: +1.302.831.0838. Fax: +1.302.831.1048. E-mail: cjr@udel.edu.

Author Contributions

†These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

A. M. Lenhoff is thanked for helpful discussions, and the National Science Foundation (CBET - 0853639) and Friedrich-Naumann Foundation are gratefully acknowledged for financial support.

REFERENCES

(1) Asthagiri, D.; Paliwal, A.; Abras, D.; Lenhoff, A.; Paulaitis, M. *Biophys. J.* **2005**, *88*, 3300−3309.
(2) Neal, B. L.; Asthagiri, D.; Lenhoff, A. M. *Biophys. J.* **1998**, *75*, 2469−2477.
(3) Curtis, R. A.; Lue, L. *Chem. Eng. Sci.* **2005**, *61*, 907−923.
(4) Curtis, R. A.; Blanch, H. W.; Prausnitz, J. M. *J. Phys. Chem. B* **2001**, *105*, 2445−2452.
(5) Young, T. M.; Roberts, C. J. *J. Chem. Phys.* **2009**, *131*, 125104/1−125104/9.
(6) Ding, F.; Buldyrev, S. V.; Dokholyan, N. V. *Biophys. J.* **2006**, *90*, 4574−4584.
(7) Marchut, A. J.; Hall, C. K. *Biophys. J.* **2006**, *90*, 4574−4584.
(8) Yun, S.; Urbanc, B.; Cruz, L.; Bitan, G.; Teplow, D. B.; Stanley, H. E. *Biophys. J.* **2007**, *92*, 4064−4077.
(9) Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Springer-Verlag: New York, 1991.
(10) Sahin, E.; Grillo, A. O.; Perkins, M. D.; Roberts, C. J. *J. Pharm. Sci.* **2010**, *99*, 4830−4848.
(11) Weiss, W. F.; Young, T. M.; Roberts, C. J. *J. Pharm. Sci.* **2009**, *98*, 1246−1277.
(12) Li, Y.; Ogunnaike, B. A.; Roberts, C. J. *J. Pharm. Sci.* **2010**, *99*, 645−662.
(13) Kern, N.; Frenkel, D. *J. Chem. Phys.* **2003**, *118*, 9882−9889.
(14) Holliger, P.; Hudson, P. J. *Nat. Biotechnol.* **2005**, *23*, 1126−1136.
(15) Brekke, O. H.; Sandlie, I. *Nat. Rev. Drug Discovery* **2003**, *2*, 52−62.
(16) Neal, B. L.; Lenhoff, A. M. *AIChE J.* **1995**, *41*, 1010−1014.
(17) Padlan, E. *Mol. Immunol.* **1994**, *31*, 169−217.
(18) Cherfils, J.; Duquerroy, S.; Janin, J. *Proteins* **1991**, *11*, 271−280.
(19) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59−107.
(20) Bereau, T.; Deserno, M. *J. Chem. Phys.* **2009**, *130*, 235106/1−235106/15.
(21) Smith, A. V.; Hall, C. K. *J. Mol. Biol.* **2001**, *312*, 187−202.
(22) Ding, F.; Borreguero, J. M.; Buldyrey, S. V.; Stanley, H. E.; Dokholyan, N. V. *Proteins* **2003**, *53*, 220−228.
(23) Irbäck, A.; Sjunnesson, F.; Wallin, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13614−13618.
(24) Takada, S.; Luthey-Schulten, Z.; Wolynes, P. G. *J. Chem. Phys.* **1999**, *110*, 11616.
(25) Singh, J. K.; Kofke, D. A. *Phys. Rev. Lett.* **2004**, *92*, 220601.
(26) Benjamin, K. M.; Singh, J. K.; Schultz, A. J.; Kofke, D. A. *J. Phys. Chem. B* **2012**, *111*, 11463−11473.
(27) Blanco, M. A.; Sahin, E.; Li, Y.; Roberts, C. J. *J. Chem. Phys.* **2011**, *134*, 225103/1−225103/12.
(28) Brummitt, R. K.; Nesta, D. P.; Chang, L.; Chase, S. F.; Laue, T. M.; Roberts, C. J. *J. Pharm. Sci.* **2011**, *100*, 2087−2103.
(29) de la Torre, J. G.; Huertas, M. L.; Carrasco, B. *Biophys. J.* **2000**, *78*, 719−730.
(30) Moon, Y. U.; Curtis, R. A.; Anderson, C. O.; Blanch, H. W.; Prausnitz, J. M. *J. Solution Chem.* **2000**, *29*, 699−717.
(31) Velev, O. D.; Kaler, E. W.; Lenhoff, A. M. *Biophys. J.* **1998**, *75*, 2682−2697.
(32) Dumetz, A. C.; Chockla, A. M.; Kaler, E. W.; Lenhoff, A. M. *Biochim. Biophys. Acta, Proteins Proteomics* **2008**, *1784*, 600−610.
(33) Paliwal, A.; Asthagiri, D.; Abras, D.; Lenhoff, A. M.; Paulaitis, M. E. *Biophys. J.* **2005**, *89*, 1564−1573.
(34) Kolvenbach, C. G.; Narhi, L. O.; Philo, J. S.; Li, T.; Zhang, M.; Arakawa, T. *J. Peptide Res.* **2009**, *50*, 310−318.
(35) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* **2012**, *38*, 16424−16431.
(36) Mirkin, N.; Jaconcic, J.; Stojanoff, V.; Moreno, A. *Proteins* **2007**, *70*, 83−92.
(37) Bonincontro, A.; Bultrini, E.; Calandrini, V.; Cinelli, S.; Onori, G. *J. Phys. Chem. B* **2012**, *104*, 6889−6893 (2012)..
(38) Tayyab, S.; Haq, S. K.; Sabeeha; Aziz, M. A.; Khan, M. M.; Muzammil, S. *Int. J. Biol. Macromol.* **1999**, *26*, 173−180.
(39) Alazard, R.; Mourey, L.; Ebel, C.; Konarev, P. V.; Petoukhov, M. V.; Svergun, D. I.; Erard, M. *Nucleic Acids Res.* **2007**, *35*, 4420−4432.
(40) Lister, I.; Schmitz, S.; Walker, M.; Trinick, J.; Buss, F.; Veigel, C.; Kendrick-Jones, J. *EMBO J.* **2004**, *23*, 1729−1738.
(41) Doster, W.; Longeville, S. *Biophys. J.* **2007**, *93*, 1360−1368.
(42) Valette, I.; Waks, M.; Wejman, J. C.; Arcoleo, J. P.; Greer, J. *J. Biol. Chem.* **1981**, *256*, 672−679.
(43) Jones, C. R.; Jhonson, C. S.; Penniston, J. T. *Biopolymers* **1978**, *17*, 1581−1593.
(44) McManus, J. J.; Lomakin, A.; Ogun, O.; Pande, A.; Basan, M.; Pande, J.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16856−16861.
(45) Bonneté, F.; Vivarès, D. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 1571−1575.
(46) Jossang, T.; Feder, J.; Rosenqvist, E. *J. Protein Chem.* **1988**, *7*, 165−171.
(47) Saito, H.; Takahashi, S.; Nagata, M.; Tsuchiya, T.; Mugishima, H.; Yan, K.; Kondo, Y.; Matsuyama, T.; Sekine, T.; Igarashi, T. *Pediatr. Nephrol.* **2008**, *24*, 609−612.
(48) Sukumar, M.; Doyle, B. L.; Combs, J. L.; Pekar, A. H. *Pharm. Res.* **2004**, *21*, 1087−1093.
(49) Pilz, I.; Schwartz, E.; Durchschein, W.; Licht, A.; Sela, M. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 117−121.
(50) Kilar, F.; Simon, I.; Lakatos, S.; Vonder Viszt, S.; Medgyesi, G.; Zavodszky, P. *Eur. J. Biochem.* **1985**, *147*, 17−25.
(51) Pjura, P. E.; Paulaitis, M. E.; Lenhoff, A. M. *AIChE J.* **1995**, *41*, 1005−1009.
(52) Chalikian, T.; Gindikin, V.; Breslauer, K. *FASEB J.* **1996**, *10*, 164−170.
(53) Bernhardt, J.; Pauly, H. *J. Phys. Chem.* **1975**, *79*, 584−590.
(54) Horan, T.; Wen, J.; Narhi, L.; Parker, V.; Garcia, A.; Arakawa, T.; Philo, J. *Biochemistry* **2012**, *35*, 4886−4896.