

Molecular Origins of Osmotic Second Virial Coefficients of Proteins

B. L. Neal, D. Asthagiri, and A. M. Lenhoff

Center for Molecular and Engineering Thermodynamics, Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716 USA

ABSTRACT The thermodynamic properties of protein solutions are determined by the molecular interactions involving both solvent and solute molecules. A quantitative understanding of the relationship would facilitate more systematic procedures for manipulating the properties in a process environment. In this work the molecular basis for the osmotic second virial coefficient, B_{22} , is studied; osmotic effects are critical in membrane transport, and the value of B_{22} has also been shown to correlate with protein crystallization behavior. The calculations here account for steric, electrostatic, and short-range interactions, with the structural and functional anisotropy of the protein molecules explicitly accounted for. The orientational dependence of the protein interactions is seen to have a pronounced effect on the calculations; in particular, the relatively few protein-protein configurations in which the apposing surfaces display geometric complementarity contribute disproportionately strongly to B_{22} . The importance of electrostatic interactions is also amplified in these high-complementarity configurations. The significance of molecular recognition in determining B_{22} can explain the correlation with crystallization behavior, and it suggests that alteration of local molecular geometry can help in manipulating protein solution behavior. The results also have implications for the role of protein interactions in biological self-organization.

INTRODUCTION

More than half a century ago, Zimm (1946) stated that “it is a task of statistical mechanics to relate the thermodynamics of solutions to the properties of the molecules that compose them.” He used his theoretical results in part to interpret thermodynamic data on protein solutions. With the advent of modern biotechnology, the incentive for such analyses of protein solutions has become much stronger, especially for application to downstream processing operations (Haynes et al., 1993; Watanabe et al., 1994; Coen et al., 1995). Protein thermodynamics in vivo is an even greater challenge, being confounded by the complexity of the intracellular environment. However, it is one that is likely to receive more attention as rapidly emerging genome sequences are used to generate protein structural information and ultimately to develop morphological and functional pictures of intact organisms (Bassingthwaite, 1997; Wilkins et al., 1997). Models describing such systems should faithfully represent the underlying physics and not just fit experimental data. The purpose of the present work is to contribute to such an understanding.

The thermodynamic property of proteins examined by Zimm (1946) was the osmotic second virial coefficient B_{22} , defined by

$$\Pi = RTc_p \left(\frac{1}{M_w} + B_{22} c_p + \cdots \right) \quad (1)$$

where Π is the osmotic pressure, c_p is the protein concentration (in mass units), R is the gas constant, T is the absolute temperature, and M_w is the protein molecular weight. B_{22} reflects the size and direction of deviations from ideality of the solution, with an ideal solution obeying the van't Hoff relation, i.e., Eq. 1 with B_{22} and higher order terms set to zero. At the molecular level, B_{22} reflects the nature of protein-protein interactions: Eq. 1 shows that a positive value of B_{22} indicates an osmotic pressure larger than that for an ideal solution, which can be interpreted intuitively to reflect predominantly repulsive interactions among protein molecules, with the converse true for negative B_{22} .

The insights that B_{22} can provide into molecular interactions have driven ongoing experimental work and parallel efforts to model the interactions. A more complete understanding of the molecular origins of B_{22} can shed light on experimental means of manipulating the thermodynamic properties and phase behavior of proteins and may offer new approaches for extracting partial structural information from solution property measurements. The quantitative relation of B_{22} to molecular interactions is usually presented in terms of the orientationally averaged potential of mean force (PMF), $W(r_{12})$, where r_{12} is the intermolecular center-to-center distance (McMillan and Mayer, 1945; Hill, 1987):

$$B'_{22} = -2\pi \int_0^\infty (e^{-W/kT} - 1) r_{12}^2 dr_{12} \quad (2)$$

Here $B'_{22} = B_{22}M_w^2$, and W is essentially the interaction free energy between two protein molecules in an averaged solvent environment, and k is the Boltzmann constant. (Some care is necessary in comparing calculated B_{22} values with experiment because of differences in the thermodynamic conventions used (Cabezas and O'Connell, 1993), but the

Received for publication 23 March 1998 and in final form 29 June 1998.

Address reprint requests to Dr. Abraham M. Lenhoff, Department of Chemical Engineering, University of Delaware, Newark, DE 19716. Tel.: 302-831-8989; Fax: 302-831-4466; E-mail: lenhoff@che.udel.edu.

Dr. Neal's present address is ARCO Chemical, P.O. Box 38007, So. Charleston, WV 25303.

© 1998 by the Biophysical Society

0006-3495/98/11/2469/09 \$2.00

effects are not large (Coen et al., 1995).) The portion of the integral corresponding to overlap, where $W \rightarrow \infty$, leads to the excluded volume contribution to the virial coefficient, which is always positive. Remaining features of $W(r_{12})$, including the range of r_{12} for which overlap occurs, must be incorporated into models of protein-protein interactions, and they determine the predicted values of B_{22} .

Previous efforts to model W and hence B_{22} have been based on highly idealized descriptions of proteins. The protein molecules are almost always treated as spheres, although Vilker et al. (1981) also modeled bovine serum albumin (BSA) as an ellipsoid instead. For a sphere, the excluded volume is simply four times the sphere volume, but the appropriate sphere radius is uncertain. In addition, a steric barrier in the form of a "hydration layer" is sometimes assumed to surround the molecules and effectively represents an additional excluded volume; the thickness of the hydration layer may then serve as an adjustable parameter. The remaining contributions to B_{22} are energetic, and they are usually modeled using standard colloidal methods (e.g., Hunter, 1987) to account for van der Waals (dispersion) attraction within the Lifshitz-Hamaker framework, and for electrostatic repulsion (Gallagher and Woodward, 1989; Muschol and Rosenberger, 1995). More elaborate treatments (e.g., Vilker et al., 1981; Haynes et al., 1992; Coen et al., 1995; Kuehner et al., 1997) may include such effects as permanent and induced dipole interactions, charge fluctuation terms, and osmotic attraction (depletion flocculation).

The differences among the various prior studies lie mainly in the assumptions used to calculate W ; the idealized geometry is retained in all cases. Results are fitted to experimental data by balancing repulsive (mainly electrostatic) against attractive (mainly van der Waals) interactions, with the magnitude and/or the extent of one or more interactions treated as an adjustable parameter. Although these models have been partly successful in capturing some of the principal experimental trends and in obtaining reasonable numerical values with adjustment of parameters, discrepancies remain. For example, the Hamaker constant, which characterizes the magnitude of van der Waals interactions, can be used as an adjustable parameter, but fits to experimental data typically yield unrealistically high values. Among the reasons that have been identified for the discrepancies are the approximations inherent in describing the protein shape, in calculating the intermolecular interaction mechanisms, and in omitting effects that are poorly understood quantitatively, e.g., solvation interactions. A more subtle source of error is that W is usually found by adding independently determined contributions attributed to different mechanisms, each of which is found from the two-molecule pair potential by suitable orientational averaging. A more appropriate approach would be to add all of the contributions, accounting for their orientation dependence, before averaging.

The goal of the present work is to relax some of the assumptions previously made in computing B_{22} . Specifically, we account for details of shape and charge distribu-

tion by using crystallographic data. We have previously shown that a detailed shape representation gives rise to an excluded volume that is higher by $\sim 70\%$ than that for spheres of equivalent volume (Neal and Lenhoff, 1995), and that there is a pronounced orientational dependence for van der Waals interactions within the Hamaker framework (Roth et al., 1996). Here we extend those results by incorporating orientation-dependent interactions in calculating B_{22} . An indication that the orientation dependence may be important comes from an intriguing correlation between B_{22} and protein crystallization, the latter of which is governed by close, specific contacts. Solution conditions conducive to crystallization have been shown to correspond to slightly negative values of B_{22} (George and Wilson, 1994; George et al., 1997). Such conditions denote weak attraction, whereas stronger attraction (more negative B_{22}) was found to correlate with amorphous precipitation. Rosenbaum et al. (Rosenbaum et al., 1996; Rosenbaum and Zukoski, 1996) extended the correlation by showing that the measured virial coefficients could be used to predict solubilities within the framework of the sticky hard sphere model (Baxter, 1968), suggesting that short-range attraction is dominant in governing phase behavior. However, in view of the neglect of orientational dependence, such averaged interpretations may provide an incomplete picture of the underlying physics. This is indeed what emerges from our work, even though our models still rely on some obvious approximations because of both computational limitations and uncertainties regarding physical descriptions.

Two recent sets of virial coefficient measurements obtained in our laboratory by light scattering and small-angle neutron scattering (Velev et al., manuscript submitted for publication) provide a specific focus for our analysis. For hen egg lysozyme, B_{22} decreases monotonically with pH (up to close to the pI) at each of a number of values of ionic strength between 0.005 M and 0.5 M, although the slopes of the curves become rather small at the higher ionic strengths. Within experimental error, the respective curves do not intersect. These results are qualitatively consistent with even the simplest models discussed above: both the decrease in charge with increasing pH and the increased screening with increasing ionic strength reduce repulsive electrostatic interactions—hence the decrease in B_{22} .

For bovine chymotrypsinogen, there is again a monotonic decrease in B_{22} between about pH 3 and 7, consistent with decreasing net charge, but in this case the curves at different ionic strengths all cross at around pH 5. Thus between this region and the pI (> 8), B_{22} increases with increasing ionic strength, indicating that attractive electrostatics are dominant despite the finite positive charge. Similar behavior has been reported for bovine chymotrypsin (Coen et al., 1995), with model results attributing the trend to dipole interactions. However, the calculations described below, which are applied specifically to chymotrypsinogen, link the effect more closely to the geometric complementarity that is essential for specific molecular recognition, such as that manifested in protein crystals.

THEORY AND METHODS

Calculation of B_{22}

Calculations of the second virial coefficient that account for orientational dependence are based on a generalization of Eq. 2 (Zimm, 1946; McQuarrie, 1976):

$$B'_{22} = -\frac{1}{16\pi^2} \int_0^{2\pi} \int_0^\pi \int_0^{2\pi} \int_0^{2\pi} \int_0^\pi \int_0^\infty (e^{-W/kT} - 1) r_{12}^2 dr_{12} \sin \theta d\theta d\phi d\alpha \sin \beta d\beta d\gamma \quad (3)$$

where θ and ϕ denote the angular location of the second molecule relative to the first, and the Euler angles α , β , and γ specify the orientation of the second molecule. The PMF W is now a function of all of the configurational variables, i.e., the angular variables (collectively represented by Ω) as well as the center-to-center distance r_{12} . If the center-to-center distance at contact for a given orientation Ω is defined as r_c , the distance integral in Eq. 3 can be divided into two parts, namely that for r_{12} between 0 and r_c and that between r_c and ∞ . This leads to

$$B'_{22} = \frac{1}{16\pi^2} \int_\Omega \left[\frac{1}{3} r_c^3 - \int_{r_c}^\infty (e^{-W/kT} - 1) r_{12}^2 dr_{12} \right] d\Omega \quad (4)$$

because W is infinite for $r_{12} < r_c$, i.e., configurations corresponding to overlap. The first term in the brackets then represents the excluded volume contribution, which is always positive, and the second term, representing the contribution due to energetic interactions, leads to positive or negative contributions to B'_{22} for repulsive and attractive interactions, respectively. The second term (without the minus sign) is referred to below as the inner integral, I_{in} ; it is a direct measure of the contribution of energetic interactions in each orientation Ω to the second virial coefficient.

Eq. 4 served as the basis for computing B'_{22} using the Monte Carlo integration subroutine D01GBF (NAG Fortran Library Mark 16; Numerical Algorithms Group, Downers Grove, IL) for the five-dimensional orientational integral. Although this routine does not provide extremely high accuracy, it is well suited to the irregular nature of the integrand. For each orientation Ω sampled, r_c was determined iteratively (Neal and Lenhoff, 1995), and I_{in} was computed by an adaptive quadrature routine (D01AJF; NAG Fortran Subroutine Library). Evaluation of this integral at each orientation was accelerated by first calculating interaction energies (see below) at suitably spaced node points and then interpolating using the Akima shape-preserving cubic spline (DSPLEZ; IMSL Fortran Subroutine Library) (De Boor, 1978).

In calculating the PMF for each configuration, the solvent (including electrolyte) is treated as a continuum, and the free energy of interaction is calculated between pairs of

protein molecules. We account for two classes of interactions, which are discussed separately.

Electrostatics

Electrostatic interactions can be quite realistically described and are essential for capturing the pH and ionic strength effects observed experimentally. The model used is a continuum one (Kirkwood, 1934; Warwicker and Watson, 1982; Sharp and Honig, 1990) in which each protein molecule is treated as a dielectric cavity (of dielectric constant 4) containing a set of point charges representing the ionized residues. Thus the charge distribution varies with pH. The molecules are immersed in an unbounded aqueous electrolyte solution. The interaction free energy, representing the electrostatic contribution to the PMF W , can be calculated once the potential field ψ throughout the system is known (Sharp and Honig, 1990). ψ is found by solving Poisson's equation within each protein molecule and the Poisson-Boltzmann equation (Hunter, 1987) in the electrolyte solution. Boundary conditions specifying continuity of the potential field and of the normal gradient of the electric displacement are satisfied at the dielectric boundaries.

The calculations performed here for chymotrypsinogen were based on chain A of the crystal structure in Protein Data Bank file 2CGA (Fig. 1). Charges were assigned to the ionizable residues according to the Henderson-Hasselbalch equation by using intrinsic pK_a values (Zubay, 1984; Stryer, 1988). The resulting charges, $19.2e$ at pH 3 and $4.2e$ at pH 7, are within $\sim 1e$ of the measured titration charge (Marini and Martin, 1971). The potential field was calculated by a boundary element method (BEM) (Yoon and Lenhoff, 1990), with the field in the presence of two molecules found iteratively (Zhou, 1993). For this purpose the molecular surface was discretized by using a cubic mesh (Zauhar, 1995), and the dependent variables (potential and normal

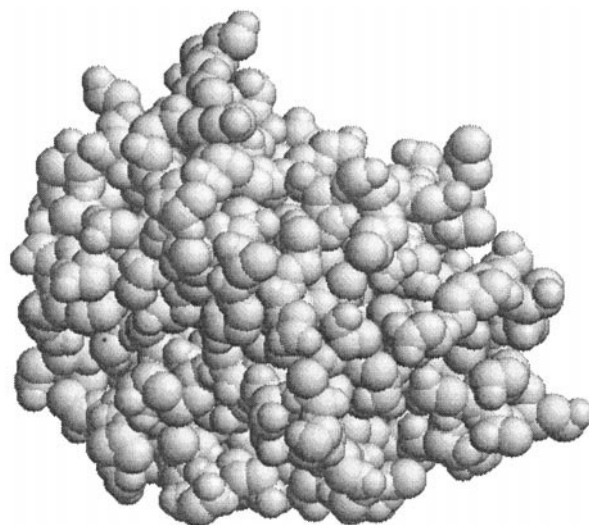


FIGURE 1 Three-dimensional structure of bovine chymotrypsinogen A used as basis for calculations of second virial coefficient.

flux) were approximated as varying linearly over each element (superparametric implementation (Aliabadi and Hall, 1988; Chan et al., 1990, 1992)).

For calculations incorporating the full geometry of chymotrypsinogen, 2553 primary nodes, corresponding to 5102 cubic triangular elements, were used. The electrostatic computations for a system of this size would make the computational time prohibitive for the configurational exploration needed to evaluate the second virial coefficient via Eq. 4. Thus a more efficient approximate description was developed in which the spatial distribution of charges in each protein molecule was retained, but within a spherical dielectric cavity, the radius of which was based on the molecular volume of the protein. In this case, the charges were placed at the same angular positions in the sphere as in the full protein, with each charge at the same distance from the sphere surface as its distance from the nearest BEM mesh point in the full protein case. Because the tessellation used for the spheres comprised 362 primary nodes and 720 elements, the calculations using the sphere approximation were much more economical than those based on the full geometry. The approximation was thus used for the configurational integration in Eq. 4, but an extensive set of calculations using the full geometry was performed to aid in evaluating the accuracy of the approximation.

Short-range interactions

The second class of interactions includes a variety of short-range effects, such as van der Waals (dispersion), solvation, and hydrogen bonding interactions. These are collectively less well understood quantitatively than electrostatics, and because of the short range over which they act, they are very sensitive to the local geometry. These forces are significant in high-affinity protein-protein interactions such as antigen-antibody binding (Davies and Cohen, 1996), as well as in protein folding, because of the dense packing of the protein interior (Creighton, 1993).

Van der Waals interactions are the best understood of the short-range interactions. At the atomic level the interaction potential between atoms i and j (equivalent to the free energy for dispersion interactions) is usually described by a relation of Lennard-Jones form,

$$U_{ij} = 4\epsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (5)$$

that captures the attractive nature of van der Waals interactions and the very short-range Born repulsion due to overlap of electron clouds. Here $-\epsilon$ is the minimum interaction energy and σ_{ij} is the collision diameter. The drawback of relations of this kind for describing protein-protein interactions is that water molecules must be included explicitly, greatly complicating the computational task. The Lifshitz-Hamaker approach (Hunter, 1987) widely used in colloid science overcomes this difficulty via a continuum formulation in which the (usually attractive) interaction free energy

of two bodies, in this case protein molecules, is described by

$$\Delta F = -\frac{A_{12}}{\pi^2} \int_{V_1} \int_{V_2} \frac{1}{r_{12}^6} dV_2 dV_1 \quad (6)$$

The integral is a function purely of the particle geometries, whereas the material properties are represented by A_{12} , the Hamaker constant, which is calculated from the optical properties of the interacting bodies and the intervening medium (water in this case). This approach also has a shortcoming, namely the breakdown of the continuum approximation at very short range. More generally, although Eqs. 5 and 6 are intended to describe the same physical behavior, the results of the two treatments are not easily reconciled.

Other short-range interactions present additional problems. Hydrogen bonding and solvation effects depend strongly on water and protein structures, and although atomistic simulations using Eq. 5 and pairwise Coulombic electrostatic models are quite successful at capturing local effects (Garde et al., 1997), straightforward means of extracting protein interaction energies are not yet available.

In the face of these difficulties, previous efforts to model the second virial coefficient have generally used Eq. 6, with the protein molecules treated as spheres. More detailed calculations (Roth et al., 1996) in which the idealization of shape is relaxed show that the geometric factor in Eq. 6 is almost always overestimated by the sphere model relative to the value for a more realistic geometric representation. Even here, though, questions remain because of the breakdown of Eq. 6 at short range and the neglect of other interactions. Because of these difficulties, interaction energies for high-affinity contacts are expediently calculated by empirical approaches based on some measure of the complementarity of the apposing surfaces. The buried surface area is frequently used (Novotny et al., 1989; Horton and Lewis, 1992), based on the correlation observed between the accessible surface area of functional groups (usually nonpolar) and the free energy of transfer between water and an organic solvent (Creighton, 1993). An alternative measure of complementarity is the volume of the intervening "gap" or a "gap index," calculated as the ratio of the gap volume to the buried area (Jones and Thornton, 1996).

Because we need to calculate the interaction energy for two molecules not only in a bound state, but also in arbitrary configurations, we use a different semiempirical approach for describing the short-range interactions. First, where the intermolecular gap is large enough to justify a continuum description, we use the Lifshitz-Hamaker approach as applied to realistic representations of the protein shape (Roth et al., 1996), with a protein-water-protein Hamaker constant calculated to be $3.1kT$ (Roth et al., 1996). At shorter range, however, we use Eq. 5 to capture short-range attraction and repulsion. Even in this situation, though, substantial parts of the two protein molecules are widely separated. Thus our method is a hybrid one in which all atom pairs separated by

more than an arbitrary center-to-center distance (here chosen to be 6 Å, the approximate distance that allows a water molecule to fit into the gap) are assumed to interact by Eq. 6, and all other pairs by Eq. 5. Although the transition is taken to be discontinuous, the protein-protein interaction energy curve is fairly smooth as a function of distance because of the large number of atom pairs involved.

This formulation is clearly far from rigorous, but it captures surface complementarity very well and provides a fairly smooth interaction energy profile for two molecules being translated toward each other. A specific advantage over the usual Hamaker approach is that there is no singularity at contact; instead, there is a smooth minimum in the energy trajectory, and for several crystal contacts and high-affinity contacts that we have examined, this minimum lies within a few tenths of an angstrom of the crystallographically observed contact (Neal, 1997).

We have used the OPLS parameter set (Jorgensen and Tirado-Rives, 1988) as the basis for our short-range calculations with Eq. 5, but based on measured binding energies of high-affinity pairs (e.g., Bhat et al., 1994), the resulting interaction energies are too large in magnitude. This discrepancy is not surprising given the important effects, notably those involving solvation discussed above, that are neglected in our short-range calculations. We find empirically that a reduction in the contribution of Eq. 5 by a factor of 2 reduces the short-range energy to a more realistic value; interestingly, this rescaling brings the Lifshitz-Hamaker and OPLS energies calculated in vacuum (Roth et al., 1996) into good agreement.

RESULTS AND DISCUSSION

Because of the emphasis in our calculations on seeking to explain experimental results apparently associated with attractive electrostatic interactions (Coen et al., 1995; Velez et al., manuscript submitted for publication), we first present results of the electrostatics computations and then examine how they contribute to B'_{22} via application of Eq. 4. Electrostatics calculations using the complete description of protein shape and charge distribution were performed for chymotrypsinogen at 0.1 M ionic strength for pH 3 and 7. For each set of solution conditions, five separation distances at each of 36 angular orientations were sampled; the separation distance was based on the minimum distance between the van der Waals surfaces of the nearest atom pairs. The ranges of the electrostatic interaction energies obtained are shown in Fig. 2. At pH 3 all interactions are repulsive, as would be expected from the almost complete protonation of the acidic residues. The interaction energies cover a fairly wide range, however, which is attributable to differences in the proximity of positive charges on each pair of molecules in the different configurations. At pH 7 there is a qualitative difference: in at least one configuration the interaction is attractive, despite the positive net charge. An interesting feature is that the magnitudes of the repulsive interaction

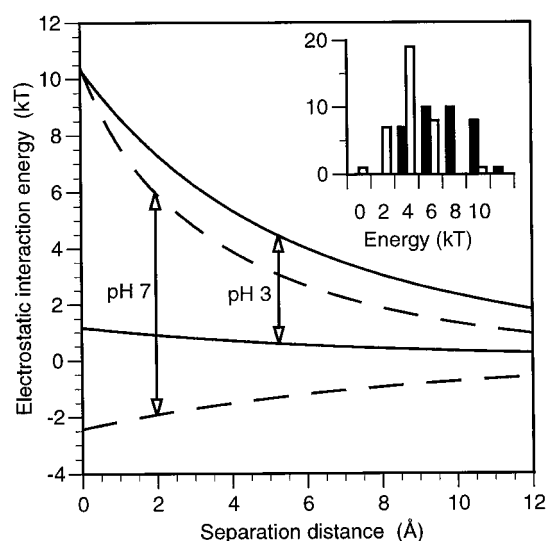


FIGURE 2 Range of calculated electrostatic interaction energies for chymotrypsinogen as a function of separation distance at pH 3 and pH 7, 0.1 M. Full protein structural details were included. Inset shows distribution of energies at 1-Å separation for the 36 configurations studied at each pH value. Filled bars, pH 3; unfilled bars, pH 7.

energies are similar at the two pH values, despite the large differences in net charge. This and the appearance of attractive interactions reflect the importance of the more closely spaced pairs of individual charges as opposed to the global net charge.

The corresponding calculations for the spherical geometry with distributed charge were initially performed for a sphere of volume equal to that of the protein molecule (19.4 Å radius). The approximate model was found to overpredict the electrostatic interaction energy in all but one orientation, by as much as a factor of 2. Imoto (1983) suggested that, in such an approximate model, the spacing between the charges should be kept as close to that in the true protein structure as possible. To achieve this without placing charges outside the low dielectric cavity as Imoto did, we used a larger sphere radius. Calculations of the excluded volume of protein molecules (Neal and Lenhoff, 1995) indicate that from the point of view of protein-protein interactions, the cavity occupied by the protein has a radius ~20% greater than that of the sphere of equivalent volume. Thus the sphere radius was increased to 23.1 Å while keeping the charge depth below the surface and the angular distribution fixed. The results for pH 7 (Fig. 3) show that the sphere model still overestimates the electrostatic energy, but by a smaller factor; comparable results were obtained for pH 3. Although it may be possible to improve agreement further by manipulating the sphere radius, the algorithm used to generate Fig. 3 was used for the full computations because it represents the orientational dependence quite well, and the actual discrepancies are fairly small given the other approximations made, especially in the short-range interaction calculations.

The electrostatic attraction apparent at pH 7 in Figs. 2 and 3 would not be expected to dominate B'_{22} behavior in terms

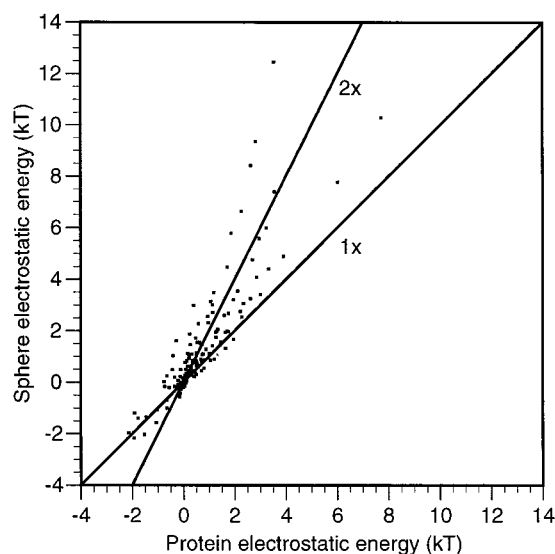


FIGURE 3 Relation of electrostatic interaction energies calculated for full chymotrypsinogen representation (Fig. 1) and sphere approximation (radius 23.1 Å). pH 7, 0.1 M.

of a simple average of the energy, but its importance is boosted by the Boltzmann weighting in Eqs. 3 and 4. However, this weighting applies to the total PMF W , and not to the individual contributions, so it is necessary to examine the short-range interactions as well. The critical attribute of these interactions is that they are attractive, except for the Born repulsion (see Eq. 5) at very short range, so that the integrands in Eqs. 3 and 4 should be positive if only the short-range interactions are considered. The values of I_{in} are then positive for all orientations, corresponding to negative contributions to B'_{22} . The squares in Fig. 4 show the calculated values of I_{in} , based on short-range interactions only, as

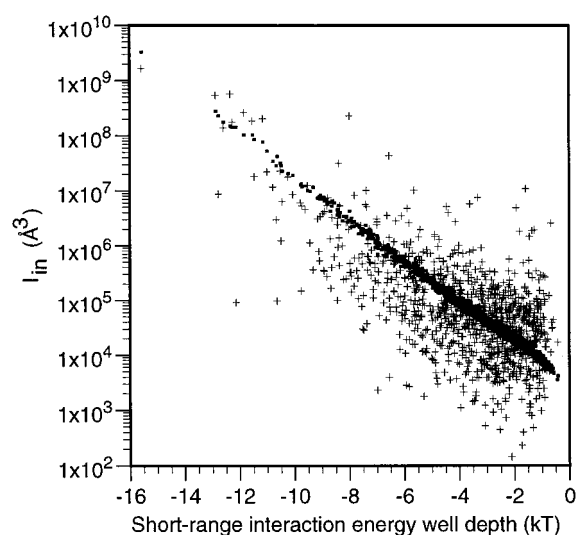


FIGURE 4 Relation of calculated inner integral to well depth of short-range interaction energy profile in 1792 angular configurations of chymotrypsinogen pairs. ■, Short-range interactions only; +, short-range plus electrostatic (pH 7, 0.1 M) interactions.

a function of the well depth of the short-range interaction energy profile with r_{12} ; 1792 configurations, sampled by a Monte Carlo integration routine, are shown. All of the points lie roughly on a straight line on this semilog plot, indicating that the value of I_{in} is controlled mainly by the short-range energy well depth rather than reflecting the full short-range energy profile; this is a consequence of the exponential Boltzmann weighting.

When electrostatic interactions are included in the B'_{22} calculations as well, the I_{in} points in Fig. 4 are perturbed from the values previously examined. For a given configuration, attractive electrostatics increase I_{in} and repulsive electrostatics decrease it, giving rise to the points shown as + in Fig. 4 for electrostatics calculated at pH 7 and an ionic strength of 0.1 M. The extent to which each I_{in} changes depends on the relative magnitudes of the short-range and electrostatic energies. The values in the vicinity of the short-range well are shown in Fig. 5 for the different orientations sampled; there is clearly no correlation between the two energies. Electrostatic interactions are predominantly repulsive, and for some orientations they overcome the short-range attraction, giving rise to negative values of I_{in} ; these are not shown in Fig. 4 because of the logarithmic scale used, but they are relatively small because of the Boltzmann weighting. The attractive electrostatic interactions, on the other hand, are rather weak (Fig. 5), so very large values of I_{in} are due mainly to strong short-range interactions. However, because of the Boltzmann weighting, attractive electrostatics can boost I_{in} substantially above the value due to short-range interactions alone (Fig. 4); as discussed below, these configurations may be central to determining trends in B_{22} .

An additional feature of the I_{in} values in Fig. 4 is the very wide range covered, which is a critical factor in determining

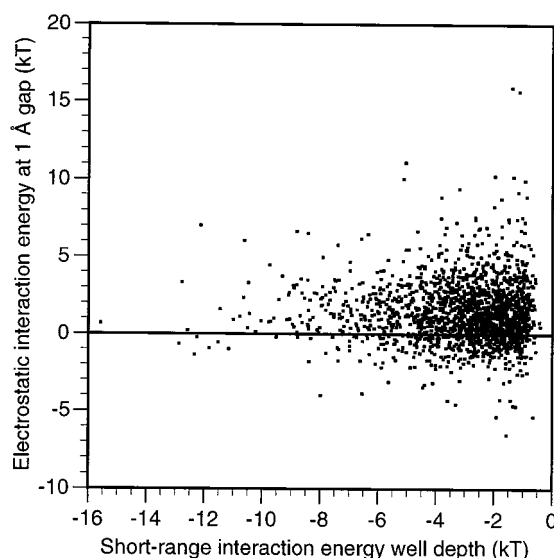


FIGURE 5 Distribution of short-range interaction energy well depths and electrostatic interaction energies (pH 7, 0.1 M) in corresponding configurations.

B'_{22} . Equation 4 shows that B'_{22} comprises the (positive) excluded volume contribution and an unweighted orientational average of I_{in} , and in view of the many orders of magnitude spanned by I_{in} , it is the orientations with the largest values of I_{in} that most strongly affect B'_{22} . This can be seen more clearly from the histogram in Fig. 6, showing the distribution of I_{in} values for the case shown in Fig. 4. Orientations for which I_{in} is smaller in magnitude than $\sim 10^5 \text{ \AA}^3$ contribute negligibly to the energetic term in Eq. 4; this includes those orientations for which repulsive electrostatics lead to negative values of I_{in} . Conversely, the orientations that are most important in determining B'_{22} are those for which I_{in} is large, and as is apparent from Fig. 4, the main effect causing this is strong short-range interactions. Such interactions result from a high degree of complementarity between the apposing surfaces, characteristic of a molecular recognition phenomenon. Qualitatively similar results would be obtained even if the short-range interactions were calculated by some other approach, e.g., one based on buried surface area. What is essential here is to account in detail for the geometry of the interacting molecules, which is overlooked in models treating the protein molecules as spheres.

The observed effect on B'_{22} of electrostatic interactions must then also be related to the high-complementarity configurations: it is the nature of the electrostatic interactions in these orientations that plays the main part in determining experimental electrostatic effects. Attractive electrostatics will increase the I_{in} from those obtained for short-range interactions only, leading to a decrease in the value of B'_{22} , and vice versa for repulsive electrostatics. Because of the nonlinearity of the Boltzmann factor, however, the effects of attraction and repulsion are asymmetrical, and the net result of the orientational average is not easily predicted. As

is apparent from Figs. 4 and 5, however, in this system there is a disproportionate number of high-complementarity configurations for which the electrostatics are attractive at pH 7 and 0.1 M ionic strength. Such attractions are screened at higher ionic strengths, and although repulsive interactions are screened as well, the asymmetry of the Boltzmann dependence, together with the distribution and coupling of electrostatic and short-range energies (see, e.g., Fig. 5), can result in the experimentally observed increase in B_{22} with ionic strength for chymotrypsin and chymotrypsinogen (Coen et al., 1995; Velev et al., manuscript submitted for publication). The histograms of I_{in} values in Fig. 6, all for the same set of orientations, show that there are more configurations with $I_{in} > 10^8 \text{ \AA}^3$ at 0.1 than at 0.3 M.

Beyond the above observations, B_{22} cannot be estimated accurately, because of the uncertainties in the calculated interaction energies and the computational challenge of fully sampling the configurational space. However, the best results obtained using the Monte Carlo integration routine are in reasonable quantitative agreement with experimental results (Table 1), especially in view of the fact that no adjustable parameters were used in the calculations. The counterintuitive trend with electrolyte concentration at pH 7 is also correctly captured. Given the large error estimates on the calculated B_{22} , however, the agreement is significant not so much in the correct prediction of values and trends, but rather in the plausibility of the formulation used. The central role of short-range interactions that emerges from our work was postulated previously (Rosenbaum et al., 1996), but the more specific importance of the high-complementarity configurations represents a new paradigm for interpreting B_{22} data. Another implication is that in view of the physical uncertainties, computational complexity, and dependence on detailed structural information required to calculate B_{22} within this framework, detailed prediction of solution thermodynamic properties is extremely difficult, and one must rely on experimental probes of protein interactions.

The new paradigm also suggests a natural link that may explain the correlation between B_{22} measurements and protein crystallization (George and Wilson, 1994; George et al., 1997): both are governed by molecular recognition in a relatively small number of configurations. To explore this

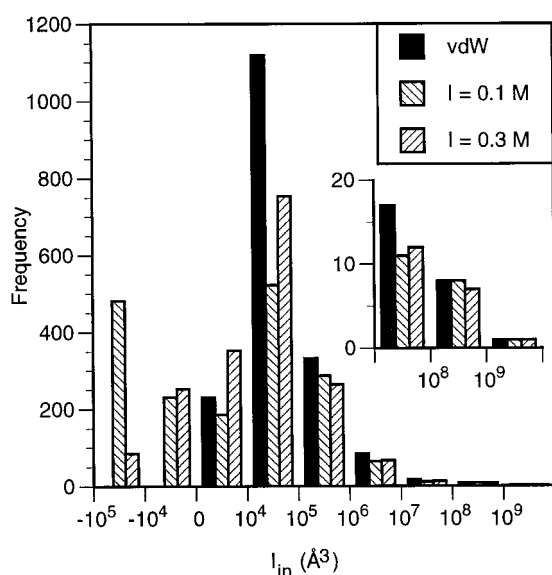


FIGURE 6 Distributions of inner integral values for 1792 angular configurations of chymotrypsinogen pairs, with different contributions included as shown in legend. Inset shows enlarged view of high- I_{in} region.

TABLE 1 Comparison of best computational estimates of B_{22} with experimental measurements

pH	Ionic strength (M)	$B_{22} \times 10^4 \text{ (ml mol/g}^2\text{)}$	
		Calculated* (est. % error)	Experiment [#]
3	0.1	5.30 (6)	2.5
3	0.3	1.57 (64)	-1.4
7	0.1	-2.24 (89)	-4.1
7	0.3	-1.66 (46)	-2.1
		-1.50 (47)	
		-2.16 (49)	

*Based on molecular weight of 25,670.

[#]Velev et al. (manuscript submitted for publication).

further, we have compared the energetics for the high-complementarity configurations identified in the Monte Carlo integration (Figs. 4 and 5) with the crystal contacts in the crystallographic coordinate files (Wang et al., 1985). The most attractive crystal contacts yield short-range interaction energy well depths in the range 13–24 kT . The corresponding values in a new crystal form of chymotrypsinogen that we have recently identified (Pjura et al., manuscript in preparation) are also on the order of 20 kT . Thus although relatively few high-complementarity configurations have been found in the configurational exploration, these do indeed have affinities that are comparable to those of crystal contacts. That there are more such configurations than actually observed in a given crystal structure is consistent with previous analyses (Janin and Rodier, 1995) and with the ubiquity of crystal contact configurations observed in the different crystal forms of some proteins (Crosio et al., 1992). Which ones are manifested in a given structure may be determined by such effects as kinetics (not considered here); electrostatic repulsion may play a role.

In summary, our results indicate that B_{22} is determined largely by the contributions of relatively few, highly attractive configurations. The effects of varying electrostatic parameters such as pH and ionic strength are also manifested mainly because of the influence of these configurations. This central role of molecular recognition suggests a plausible link between solution thermodynamic properties and the formation of crystalline phases or other self-organized structures. Our conclusions also suggest that local structural perturbations are an underused route to manipulating protein interactions; these may be effected by, for example, specific adsorption of ions or surfactants. The results also indicate the vast complexity of molecular interactions that may occur in solutions of protein mixtures, not least the intracellular environment.

We gratefully acknowledge the support of the National Science Foundation (grants BCS-9210401 and BES-9510420). We also appreciate informative discussions with Phil Pjura and useful comments on the manuscript by Orlin Velev and Mike Paulaitis.

REFERENCES

- Aliabadi, M. H., and W. S. Hall. 1988. Nonisoparametric formulations for the three dimensional boundary element method. *Eng. Anal.* 5:198–204.
- Bassingthwaighe, J. B. 1997. Design and strategy for the Cardionome Project. *Adv. Exp. Med. Biol.* 430:325–339.
- Baxter, R. J. 1968. Percus-Yevick equation for hard spheres with surface adhesion. *J. Chem. Phys.* 49:2770–2774.
- Bhat, T. N., G. A. Bentley, G. Boulout, M. I. Greene, D. Tello, W. Dall'Acqua, H. Souchon, F. P. Schwarz, R. A. Mariuzza, and R. J. Poljak. 1994. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA.* 91:1089–1093.
- Cabezas, H., Jr., and J. P. O'Connell. 1993. Some uses and misuses of thermodynamic models for dilute liquid solutions. *Ind. Eng. Chem. Res.* 32:2892–2904.
- Chan, C. Y., A. N. Beris, and S. G. Advani. 1990. Simulation of 3-D hydrodynamic interactions around ellipsoidal particles using high order boundary element techniques. In *Computational Engineering with Boundary Elements*, Vol. 1, Fluid and Potential Problems. S. Grilli, C. A. Brebbia, and A. H. D. Cheng, editors. Computational Mechanics Publishers, Boston.
- Chan, C. Y., A. N. Beris, and S. G. Advani. 1992. Second order boundary element method calculations of hydrodynamic interactions between particles in close proximity. *Int. J. Numer. Methods Fluids.* 14:1063–1086.
- Coen, C. J., H. W. Blanch, and J. M. Prausnitz. 1995. Salting out of aqueous proteins: phase equilibria and intermolecular potentials. *AIChE J.* 41:996–1004.
- Creighton, T. E. 1993. *Proteins: Structures and Molecular Properties*. W. H. Freeman and Co., New York.
- Crosio, M.-P., J. Janin, and M. Jullien. 1992. Crystal packing in six crystal forms of pancreatic ribonuclease. *J. Mol. Biol.* 228:243–251.
- Davies, D. R., and G. H. Cohen. 1996. Interactions of protein antigens with antibodies. *Proc. Natl. Acad. Sci. USA.* 93:7–12.
- De Boor, C. 1978. *A Practical Guide to Splines*. Springer-Verlag, New York.
- Gallagher, W. H., and C. K. Woodward. 1989. The concentration dependence of the diffusion coefficient for bovine pancreatic trypsin inhibitor: a dynamic light scattering study of a small protein. *Biopolymers.* 28:2001–2024.
- Garde, S., G. Hummer, M. E. Paulaitis, and A. E. García. 1997. Hydration of biological macromolecules: from small solutes to proteins and nucleic acids. In *Proceedings of the 1996 MRS Fall Meeting Symposium on Statistical Mechanics in Physics and Biology*. D. Wirtz and T. C. Halsey, editors. Materials Research Society, Pittsburgh.
- George, A., Y. Chiang, B. Guo, A. Arabshahi, Z. Cai, and W. W. Wilson. 1997. Second virial coefficient as predictor in protein crystal growth. *Methods Enzymol.* 276:100–110.
- George, A., and W. W. Wilson. 1994. Predicting protein crystallization from a dilute solution property. *Acta Crystallogr. D.* 50:361–365.
- Haynes, C. A., F. J. Benitez, H. W. Blanch, and J. M. Prausnitz. 1993. Application of integral-equation theory to the description of aqueous two-phase partitioning systems. *AIChE J.* 39:215–223.
- Haynes, C. A., K. Tamura, H. R. Korfer, H. W. Blanch, and J. M. Prausnitz. 1992. Thermodynamic properties of aqueous α -chymotrypsin solutions from membrane osmometry. *J. Phys. Chem.* 96:905–912.
- Hill, T. L. 1987. *Statistical Mechanics. Principles and Selected Applications*. Dover, Mineola, NY.
- Horton, N., and M. Lewis. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci.* 1:169–181.
- Hunter, R. J. 1987. *Foundations of Colloid Science*. Oxford University Press, Oxford.
- Imoto, T. 1983. Electrostatic free energy of lysozyme. *Biophys. J.* 44:293–298.
- Janin, J., and F. Rodier. 1995. Protein-protein interactions at crystal contacts. *Proteins Struct. Funct. Genet.* 23:580–587.
- Jones, S., and J. M. Thornton. 1996. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA.* 93:13–20.
- Jorgensen, W. L., and J. Tirado-Rives. 1988. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.
- Kirkwood, J. G. 1934. Theory of solutions of molecules of containing widely separated charges with special application to zwitterions. *J. Chem. Phys.* 2:351–361.
- Kuehner, D. E., C. Heyer, C. Rämisch, U. M. Fornfeldt, H. W. Blanch, and J. M. Prausnitz. 1997. Interactions of lysozyme in concentrated electrolyte solutions from dynamic light-scattering measurements. *Biophys. J.* 73:3211–3224.
- Marini, M. A., and C. J. Martin. 1971. The prototropic behavior of alpha-chymotrypsin and chymotrypsinogen in water and formaldehyde. An examination of the reactivity of the amino groups in formaldehyde. *Eur. J. Biochem.* 19:162–168.
- McMillan, W. G., and J. E. Mayer. 1945. The statistical thermodynamics of multicomponent systems. *J. Chem. Phys.* 13:276–305.
- McQuarrie, D. A. 1976. *Statistical Mechanics*. Harper Collins, New York.

- Muschol, M., and F. Rosenberger. 1995. Interactions in undersaturated and supersaturated lysozyme solutions: static and dynamic light scattering results. *J. Chem. Phys.* 103:10424–10432.
- Neal, B. L. 1997. Molecular thermodynamic properties of aqueous protein solutions. Ph.D. thesis. University of Delaware, Newark, DE.
- Neal, B. L., and A. M. Lenhoff. 1995. Excluded volume contribution to the osmotic second virial coefficient for proteins. *AIChE J.* 41:1010–1014.
- Novotny, J., R. E. Bruccoleri, and F. A. Saul. 1989. On the attribution of binding energy in antigen-antibody complexes McPC 603, D1.3, and HyHEL-5. *Biochemistry.* 28:4735–4749.
- Rosenbaum, D., P. C. Zamora, and C. F. Zukoski. 1996. Phase behavior of small attractive colloidal particles. *Phys. Rev. Lett.* 76:150–153.
- Rosenbaum, D. F., and C. F. Zukoski. 1996. Protein interactions and crystallization. *J. Cryst. Growth.* 169:752–758.
- Roth, C. M., B. L. Neal, and A. M. Lenhoff. 1996. Van der Waals interactions involving proteins. *Biophys. J.* 70:977–987.
- Sharp, K. A., and B. Honig. 1990. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* 19:301–332.
- Stryer, L. 1988. *Biochemistry*. W. H. Freeman and Co., New York.
- Vilker, V. L., C. K. Colton, and K. A. Smith. 1981. The osmotic pressure of concentrated protein solutions: effect of concentration and pH in saline solutions of bovine serum albumin. *J. Colloid Interface Sci.* 79:548–566.
- Wang, D., W. Bode, and R. Huber. 1985. Bovine chymotrypsinogen A x-ray crystal structure analysis and refinement of a new crystal form at 1.8 Å resolution. *J. Mol. Biol.* 185:595–624.
- Warwicker, J., and H. C. Watson. 1982. Calculation of the electric potential in the active site cleft due to α -helix dipoles. *J. Mol. Biol.* 157:671–679.
- Watanabe, E., S. Tsoka, and J. A. Asenjo. 1994. Selection of chromatographic protein purification operations based on physicochemical properties. *Ann. N.Y. Acad. Sci.* 721:348–365.
- Wilkins, M. R., K. L. Williams, R. D. Appel, and D. F. Hochstrasser. 1997. *Proteome Research: New Frontiers in Functional Genomics*. Springer-Verlag, Berlin.
- Yoon, B. J., and A. M. Lenhoff. 1990. A boundary element method for molecular electrostatics with electrolyte effects. *J. Comp. Chem.* 11: 1080–1086.
- Zauhar, R. J. 1995. SMART: a solvent-accessible triangulated surface generator for molecular graphics and boundary element applications. *J. Computer-Aided Mol. Design.* 9:149–159.
- Zhou, H.-X. 1993. Boundary element solution of macromolecular electrostatics: interaction energy between two proteins. *Biophys. J.* 65: 955–963.
- Zimm, B. H. 1946. Applications of the methods of molecular distribution to solutions of large molecules. *J. Chem. Phys.* 14:164–179.
- Zubay, G. 1984. *Biochemistry*. Addison-Wesley, New York.