ELSEVIER

Journal of Crystal Growth 196 (1999) 377–387

JOURNAL OF CRYSTAL GROWTH

Why is the osmotic second virial coefficient related to protein crystallization?

B.L. Neal[1], D. Asthagiri, O.D. Velev, A.M. Lenhoff*, E.W. Kaler

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

Abstract

A molecular basis is presented for characterizing the osmotic second virial coefficient, B_{22}, of dilute protein solutions, which provides a measure of the nature of protein–protein interactions and has been shown to be correlated with crystallization behavior. Experimental measurements of the second virial coefficient of lysozyme and bovine α-chymotrypsinogen A were performed by static light scattering, as a function of pH and electrolyte concentration. Although some of the trends can be explained qualitatively by simple colloidal models of protein interactions, a more realistic interpretation based on protein crystallographic structures suggests a different explanation of experimental trends. The interactions accounted for are solute–solute excluded volume (steric), electrostatic and short-range (mainly van der Waals) interactions. The interactions depend strongly on orientation, and this profoundly affects calculated second virial coefficients. We find that molecular configurations in which complementary surfaces are apposed contribute disproportionately to the second virial coefficient, mainly through short-range interactions; electrostatic interactions play a secondary role in many of these configurations. Thus molecular recognition events can play a role in determining the solution thermodynamic properties of proteins, and this provides a plausible basis for explaining the observed relationship with crystallization behavior.

PACS: 41.20.Cv; 81.10.Dn; 87.15.Da; 87.15.Kg

Keywords: Protein solution thermodynamics; Protein interactions; Static light scattering; Lysozyme; Chymotrypsinogen

1. Introduction

Determining suitable conditions for growing protein crystals remains largely an empirical art that is usually accomplished in practice by extensive screening experiments. Making protein crystallogenesis more predictable requires understanding of both the thermodynamic conditions controlling the interactions of proteins with the solvent and with other solutes, including ions and other protein molecules, as well as the associated kinetic processes. The complexity of and interactions among these elements suggest that simple measures of the

* Corresponding author. Fax: + 1 302 831 4466; e-mail: lenhoff@che.udel.edu.

[1] Present address: ARCO Chemical, P.O. Box 38007, So. Charleston, WV 25303, USA.

0022-0248/99/$ – see front matter
PII: S0022-0248(98)00855-0

suitability of conditions for crystallization will be difficult to obtain.

Within this context, a particularly intriguing observation [1,2] has been that of a correlation of the value of the osmotic second virial coefficient of protein solutions with solution conditions conducive to crystallization. The osmotic second virial coefficient, B_{22}, is defined by the osmotic virial expansion

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

where Π is the osmotic pressure, c_p the protein concentration (in mass units), R the gas constant, T the absolute temperature, and M_w the protein molecular weight. The recent observations [1,2] have shown that solution conditions conducive to crystallization correspond to slightly negative values of B_{22}. This is not surprising intuitively given the molecular level interpretation of B_{22} – a positive value reflects predominantly repulsive interactions among protein molecules, with the converse true for negative B_{22} – and crystallization depends on a rather delicate approach of monomers to the growing crystal in order to assure that they bind in the correct configurations. Highly negative values of B_{22} typically lead to rapid formation of an amorphous precipitate.

Our objective is to elucidate the mechanistic basis for the relation between B_{22} and crystallization behavior. Specifically, we address an apparent paradox that emerges from the manner in which B_{22} is usually calculated from the protein–protein interactions. The basis is the McMillan–Mayer result [3,4]

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

in terms of the potential of mean force (PMF), $W(r_{12})$, where r_{12} is the intermolecular center-to-center distance. B'_{22}, which has units of volume, is related to the experimental value by $B'_{22} = B_{22}M_w^2$. As the spherically symmetric form of Eq. (2) suggests, most efforts to model W and hence B_{22} have been based on idealized descriptions in which the protein molecules are treated as spheres. The paradox is that this appears inconsistent with the clear dependence of crystallization on highly specific contacts between anisotropic protein molecules.

We revisit here the theoretical basis for calculating B_{22}, in the context of a wider set of studies that we have performed. We have explored experimentally the behavior of B_{22} of two proteins as key solution parameters, namely the pH and the electrolyte concentration, are changed [5]. We have related those results to more detailed information on the structure of these protein solutions determined by small-angle neutron scattering, as well as to crystallization behavior. More germane to the objectives of the present report, though, is that we have extended previous efforts to understand the theoretical basis underlying the trends in virial coefficient behavior [6]. In the present paper we summarize the key experimental results and trends, and discuss the theoretical basis for them in terms of both traditional colloidal models as well as our more realistic calculations.

2. Experimental procedure

Lysozyme and α-chymotrypsinogen were obtained from Sigma. Solutions were prepared using deionized water from a Millipore Milli-Q system. To keep the protein–ion interactions as simple as possible, the electrolyte used comprised 91 mol% NaCl and 9 mol% citrate buffer; acetate buffer was used instead of citrate in a few control experiments to verify that the buffer had no effect on B_{22}. The pH of the aqueous phase and the prepared protein solutions was adjusted by addition of small volumes of 1 M NaOH or HCl. The aqueous phases were filtered through 20 nm cut-off inorganic Anotop filters (Whatman). The protein solutions were filtered through 100 nm filters of the same brand and immediately diluted and sealed in ampoules. All glassware was treated with detergent and Nochromix oxidizing solution. It was abundantly washed with deionized water and dried in a clean environment shortly before each experimental run.

Static light scattering (SLS) data were obtained at an angle of 90° using a Brookhaven light scattering apparatus, equipped with a 488 nm laser and BI9000AT correlator. All measurements were

performed at 25 ± 0.1°C. The absolute Rayleigh ratios, R_θ, of the samples were obtained using pure benzene as a calibrant. The value of $R_\theta(90°)$ used for benzene, 38.6×10^{-6} cm^{-1}, was interpolated from the data of Coumou et al. [7,8]. The dn/dc values of the proteins were set at 0.20 ml/g for lysozyme (our measurements) and 0.192 ml/g for chymotrypsinogen [9]. Disturbances due to dust were minimized by using the built-in statistical dust rejection function and setting a tight rejection ratio (1.05). Each final R_θ data point was based on averaging not fewer than 50 statistically consistent measurements. The results were fitted by linear regression to the well-known equation [10]

$$\frac{Kc_p}{R_\theta} = \frac{1}{M_w} + 2B_{22}c_p \tag{3}$$

where K is an optical constant. Use of this form implicitly assumes that the third and higher virial terms contribute negligibly to the measurements; this assumption should be valid under the low protein concentrations used in our experiments.

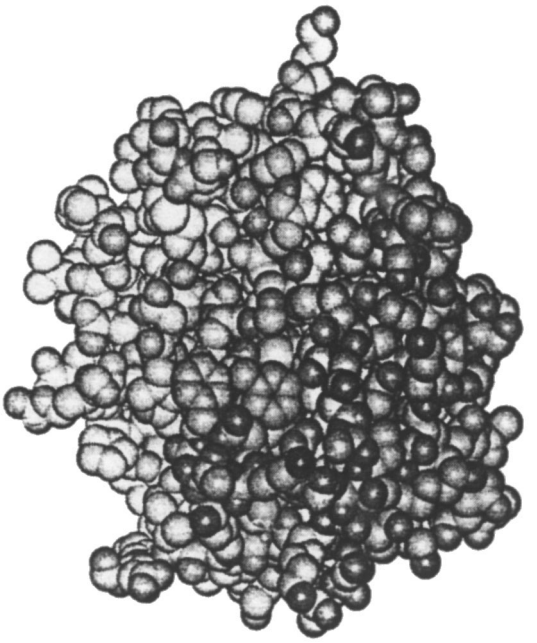

Fig. 1. Three-dimensional structure of bovine chymotrypsinogen A from X-ray crystallographic data. Although the molecule is globular, it is clearly not spherical.

3. Theory

Most previous efforts to model W and hence B_{22} have been based on Eq. (2), using idealized descriptions in which the protein molecules are treated as spheres. The region of overlap (infinite W) leads to an excluded volume contribution to B_{22} that is positive, while the energetic interactions beyond contact are generally modelled using colloidal methods to account for van der Waals (dispersion) attraction and for electrostatic repulsion [11,12]. However, other effects, such as permanent and induced dipole interactions, charge fluctuation terms, and osmotic attraction (depletion flocculation), are sometimes included as well [13–16]. Because of the singularity at contact of the colloidal description of dispersion interactions [17], a "hydration layer" is assumed to prevent the molecules from coming into contact; this layer effectively increases the excluded volume.

Various uncertainties arise from such descriptions, including the approximations in the descriptions of protein shape (Fig. 1) and of intermolecular interaction mechanisms. Nevertheless, the models have proved able to describe experimental data to varying degrees of satisfaction by fitting values of one or more adjustable parameters [5,11–16]. A parameter that is often treated as adjustable is the Hamaker constant, a material property that characterizes the strength of van der Waals interactions. We have calculated values of the Hamaker constant for protein–protein interactions from first principles [18], and found that values of 3.1 and 23.4 kT are appropriate for interactions through water and vacuum, respectively. Fitted values of the Hamaker constant are often considerably higher than the calculated protein–water–protein value.

The main 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 distribution using crystallographic data, so our calculations reflect the anisotropic structures and the resulting orientational

dependence of protein–protein interactions. We have previously shown that a detailed shape representation gives rise to an excluded volume that is higher by about 70% than that for spheres of equivalent volume [19], and that there is a pronounced orientational dependence of van der Waals interactions within the continuum colloidal framework [18]. We have extended those results by incorporating orientation-dependent interactions in calculating B_{22} [6]. B_{22} is thus found by integration over r_{12} as well as the angular variables θ and ϕ, denoting the position of the second molecule relative to the first, and the Euler angles α, β and γ, which specify the orientation of the second molecule [20]:

$$B'_{22} = -\frac{1}{16\pi^2}\int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi}\int_0^{\infty}[e^{-W/kT}-1] \times r_{12}^2\,dr_{12}\sin\theta\,d\theta\,d\phi\,d\alpha\sin\beta\,d\beta\,d\gamma. \quad (4)$$

Thus, unlike the case for Eq. (2), all the contributions to W are added, accounting for their orientation dependence, prior to averaging. This eliminates the errors that can result in Eq. (2) from adding independently determined contributions attributed to different mechanisms, each of which is found from the orientation-dependent PMF by orientational averaging.

If the center-to-center distance at contact for a given set of angular variables, denoted for notational convenience by the lumped variable Ω, is defined as r_c, the distance integral in Eq. (4) can be divided into two parts, namely that for r_{12} between 0 and r_c and that between r_c and ∞. Because W is infinite for configurations corresponding to overlap, this leads to

$$B'_{22} = \frac{1}{16\pi^2}\int_\Omega\left[\frac{1}{3}r_c^3 - \int_{r_c}^{\infty}(e^{-W/kT}-1)r_{12}^2\,dr_{12}\right]d\Omega. \quad (5)$$

The first term in the brackets is then the (positive) excluded volume contribution, while the second term, which results from energetic interactions, leads to positive or negative contributions to B_{22} for repulsive and attractive interactions, respectively. This second term (without the minus sign) is referred to below as the inner integral, I_{in}.

B'_{22} was calculated from Eq. (5) using Monte Carlo integration for the five-dimensional orientational integral in view of the irregular nature of the integrand. For each set of orientational variables Ω sampled, r_c was determined iteratively [19], and I_{in} was computed by an adaptive quadrature method. Evaluation of this integral at each orientation was accelerated by interpolation using interaction energies calculated at suitably spaced node points; the energies were calculated as described in the following sections.

3.1. Electrostatic interactions

The electrostatics calculations are based on a continuum description of the electrostatics within the interior of the protein molecules and in the surrounding electrolyte solution (see, e.g., Ref. [21]). Within each molecule we consider a number of point charges determined by the presence of ionizable amino acid residues; the electrostatics within this low-dielectric region are governed by the Poisson equation. Because of the presence of small mobile ions in the electrolyte solution outside the protein molecules, this region is described by the Poisson–Boltzmann equation [17].

Computations based on these equations were performed using a boundary element method [6,22,23] that allows the full protein shape and charge distribution to be accounted for. The results of the computations show a pronounced dependence of interaction energy on both separation distance and orientation. It is essential to capture these in the virial coefficient calculations, but the computational demands of the full electrostatics calculations would make this impossible. We thus used a charge distribution representative of the full one, but placed it in a sphere instead of a cavity with the true protein shape. Although this approach remains approximate, it captures the orientational dependence quite well [6], and to avoid the use of adjustable parameters, we did not refine the approximation further.

3.2. Van der Waals interactions

The physics of the van der Waals interactions is more complicated. The formulation usually used is

a continuum colloidal one, in which the protein molecules are treated as spheres and the Hamaker constant as an adjustable parameter. We extended this framework to account for protein structural details [18], and found that the orientational dependence is quite pronounced, with a particularly notable feature being the presence of a few configurations in which the interactions are strongly attractive; such complementarity underlies molecular recognition processes.

Even with these extensions, however, the continuum formulation breaks down when the intervening gap distance between protein molecules approaches the dimensions of a water molecule. One reason is that although the Hamaker constant used is usually one for protein–protein interactions through water, there may be parts of the two molecules that are too close together to accommodate a water molecule. More important, though, is the fundamental concern that the whole continuum formulation is invalid at such short range. For this region we therefore used an atomistic approach based on the OPLS approach and parameters [24], but because this has the drawback of requiring solvent molecules to be included explicitly, which would increase the computational effort enormously, we have developed a hybrid method [25] in which the computational method used (continuum or atomistic) depends on the intervening gap distance. The parameters in this description have been adjusted so that calculated interaction energies for representative protein–protein complexes are in reasonable agreement with experimentally measured binding energies for these complexes; the parameters were not adjusted further in the B_{22} calculations.

This formulation clearly lacks rigor, and may be regarded as a semi-empirical method for describing not only van der Waals interactions, but a more general class of short-range interactions also including solvation effects. Among the advantages that it offers are that it provides a fairly smooth interaction energy profile for two molecules being translated towards each other, that it captures surface complementarity very well, and that there is a well-defined minimum in the energy trajectory rather than a singularity at contact [25].

4. Results and discussion

4.1. Light scattering experiments

Consistent, satisfactorily linear results were obtained from the SLS experiments, although for chymotrypsinogen it was necessary to perform the experiments within 2–3 h after solution preparation to minimize proteolysis and dimerization. The intercepts with the ordinate corresponded to the known molecular weights of the two proteins to within better than $\pm 5\%$. The values of the virial coefficients obtained are in good agreement with the few common points of literature data [1,2,12,15].

The dependence of the second virial coefficient of lysozyme on pH and ionic strength is shown in Fig. 2. At low ionic strengths B_{22} is highly positive at low pH, but decreases with increasing pH; the decrease is steepest above pH 9, and B_{22} is negative at pH 10.5. B_{22} also decreases from positive to negative values with increasing pH at 0.1 and 0.3 M electrolyte, but the curves are much flatter, while at 0.5 M, B_{22} is negative and almost constant throughout. The raw data at pH 10.5–10.6 [5] suggest a molecular weight higher than the known value, which is probably an indication of some aggregation as the isoelectric point is approached (pI > 11).

The SLS data for lysozyme can be explained qualitatively in terms of the simple colloidal models discussed above. The behavior at low ionic strengths is governed by long-range electrostatic interactions. The protein is strongly positively charged at low pH, and intermolecular repulsion sets the positive (repulsive) virial coefficients. The net charge decreases with increasing pH, reducing the strength of the repulsion. The addition of electrolyte suppresses the strength and reduces the range of the electrostatic interactions, with the long-ranged electrostatic contributions attenuated when the ionic strength is raised to 0.3 M and higher.

The virial coefficient values for chymotrypsinogen (Fig. 3) are, in general, smaller in magnitude than those for lysozyme. However, the two sets of data are similar in some respects, most notably the decrease in the value of B_{22} with increasing pH at

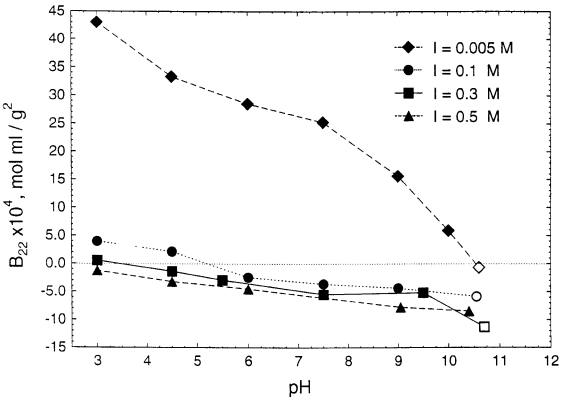

Fig. 2. Osmotic second virial coefficients measured for lysozyme solutions as a function of pH and electrolyte concentration. Open symbols denote conditions at which aggregation was observed.

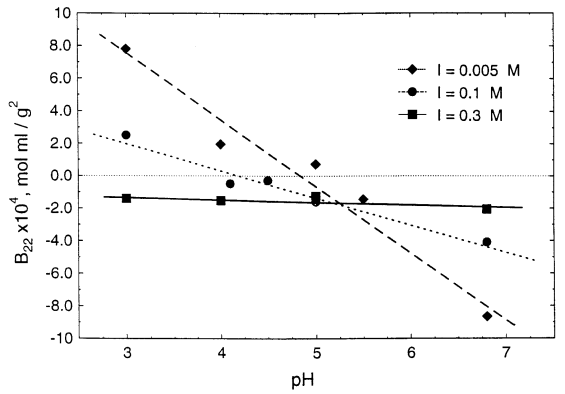

Fig. 3. Osmotic second virial coefficients measured for chymotrypsinogen solutions as a function of pH and electrolyte concentration.

all ionic strengths, and the decreasing slopes of the curves with increasing ionic strength. At low ionic strength, however, B_{22} for chymotrypsinogen appears on the scale shown in the figure to decrease from highly positive at low pH to highly negative at pH 6.8. The data for 0.1 M are intermediate between this trend and that seen in the almost flat curve for 0.5 M. The result is that the three curves appear to intersect around pH 5.2, suggesting that the virial coefficients at this pH are almost unaffected by the value of the ionic strength.

The smaller magnitudes of B_{22} for chymotrypsinogen than for lysozyme at low pH can be explained qualitatively by the larger net charge at pH 7 for the smaller lysozyme molecule. More interesting, though, is the behavior around pH 7, where B_{22} for chymotrypsinogen *increases* with increasing ionic strength; in view of the screening effect of the electrolyte, this behavior can be interpreted as being governed by attractive electrostatic interactions. Since pI > 8 for chymotrypsinogen, the net charge on each molecule is still finite and positive around pH 7, which rules out the applicability of straightforward colloidal models in describing the data. Data for chymotrypsin, which is almost identical to chymotrypsinogen in structure, show the same trend as that seen in Fig. 3 [15]. The key to explaining the observations appears to lie with the charge distribution, which is very uniform for lysozyme but much less so for chymotrypsinogen [26], resulting in a larger dipole moment for the latter. Colloidal models that account for dipole interactions are indeed capable of describing the high pH trend [5,15], but reasonable agreement with experiment is possible only by treating the Hamaker constant as an adjustable parameter.

It is the high pH trend in B_{22} for chymotrypsinogen that is the focus of our detailed virial coefficient calculations. Specifically, we explore some of the uncertainties that may arise in the simpler colloidal models, and we then compute the interactions with inclusion of the anisotropic structural features of the molecule.

4.2. Calculations of second virial coefficients

Although we have fitted our data to a colloidal model in which the molecules are treated as spheres [5], the uncertainties in the assumptions and parameters on which such models are based are cause for concern. We briefly illustrate these uncertainties here by calculating B_{22} for chymotrypsinogen for simple limiting cases of the colloidal model, and then demonstrate that a quite different picture emerges when anisotropy is accounted for. The limiting cases examined are guided by the observation that B_{22} is negative at high ionic strengths (Fig. 3), where electrostatic contributions to B_{22} are negligible. We thus examine the conditions under which the excluded volume and van der Waals interactions alone can lead to prediction of

negative B_{22}. A sphere of volume equal to that of chymotrypsinogen (radius 19.3 Å) was used, within the framework of Eq. (2). The excluded volume contribution is positive and given by 4 times the sphere volume, while the van der Waals contribution is found using the standard expression [17] that is linear in the Hamaker constant A. Because the van der Waals interactions are attractive, Eq. (2) indicates that they lead to a negative contribution to B_{22}.

Results of these calculations are shown as a function of the hydration layer thickness in Fig. 4, for the two limiting values of the Hamaker constant for proteins interacting through water and vacuum. The excluded volume contribution is a weak function of the hydration layer thickness because that layer is included in the calculation of the excluded volume. For large values of the hydration layer thickness, van der Waals interactions are weak and B_{22} drops quite slowly as the thickness is reduced. However, below a thickness of approximately 1 Å, B_{22} for the protein–water–protein Hamaker constant (3.1 kT) drops sharply due to the approach to the singularity at short range. The corresponding curve when the protein–vacuum–protein Hamaker constant (23.4 kT) is used decreases visibly over a much wider range of values of the hydration layer thickness because of the stronger van der Waals interactions that it denotes.

These results show the great sensitivity of the calculated B_{22} to the assumed hydration layer thickness. A fixed value of this parameter (e.g., 1 or 3 Å) is often assumed, and it is likely that the effect of this assumption is compensated by adjustment of the Hamaker constant in fitting experimental data. In view of the sensitivity shown in Fig. 4, however, the physical significance of the fitted Hamaker constants appears questionable. That only isolated water molecules appear in the contact region between molecular pairs in crystallographic structures further places in question the assumption of an intervening hydration layer.

Eq. (4) was applied in using the full shape of the chymotrypsinogen molecules to calculate the excluded volume [19] and continuum van der Waals [18] contributions to B_{22}. The Hamaker constant was set to the protein–water–protein value of 3.1 kT. The results are shown in Fig. 5; the error

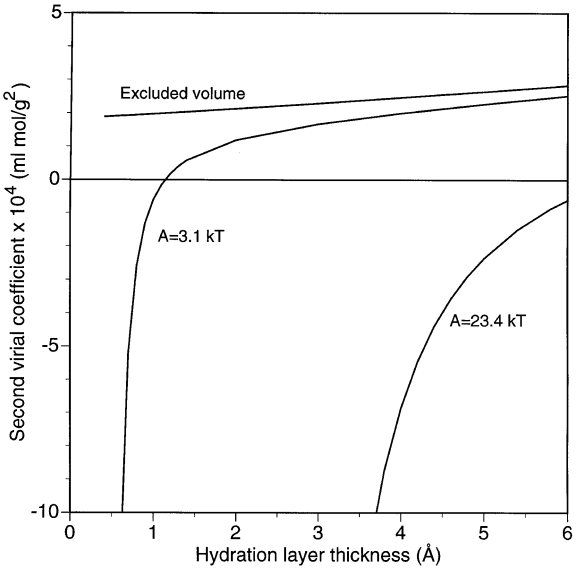

Fig. 4. Calculated values of second virial coefficient for sphere model, showing dependence on value of hydration layer thickness and Hamaker constant. Only excluded volume and van der Waals interactions are included.

bars show estimates of the error in the results as determined by the Monte Carlo integration routine. Because of the high dimensionality of the integral and the complexity of the individual integrand evaluations, some of the estimated errors are quite large, but they do not detract significantly from the conclusions that may be drawn from the results. Furthermore, our experience in using the Monte Carlo routine suggests that the error estimates are generally quite conservative.

The most striking aspect of the results is the very gradual decline in B_{22} with decreasing hydration layer thickness even down to 0.1 Å (cf. Fig. 4 for spheres). Although the slope increases as the thickness approaches zero, the obvious conclusion of these calculations is that the second virial coefficient is dominated by the excluded volume. This is not entirely unexpected as the detailed treatment of the protein shape leads to an increased excluded volume contribution [19] and weaker van der Waals interactions [18] compared to the sphere representation. Therefore B_{22} for the continuum van der Waals description remains positive, and thus incapable of representing the data adequately.

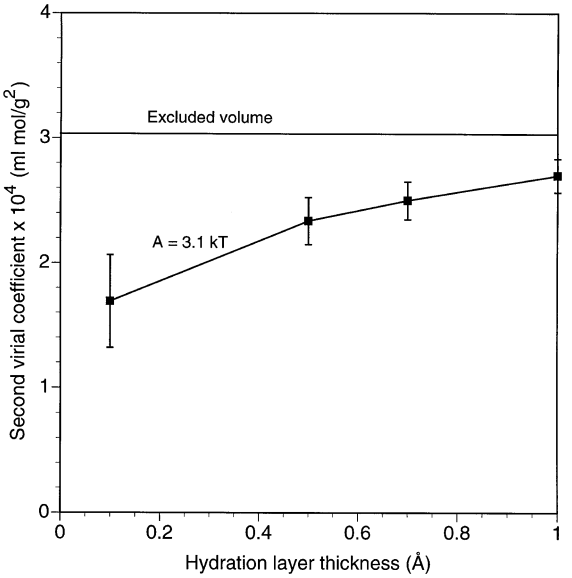

Fig. 5. Calculated values of second virial coefficient based on crystallographic structure of chymotrypsinogen, shown as a function of hydration layer thickness. Only excluded volume and van der Waals interactions are included.

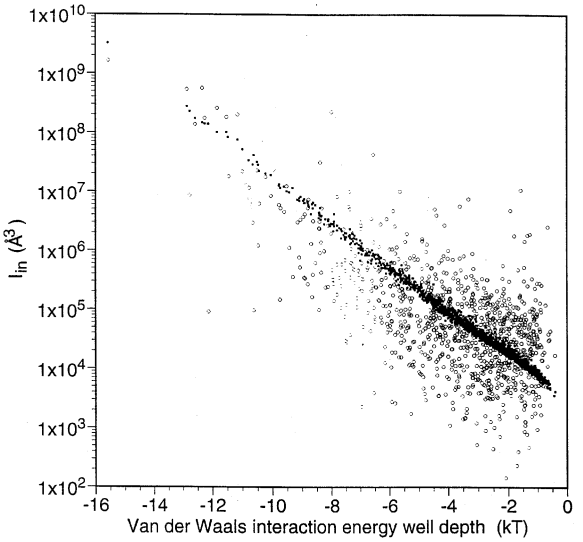

Fig. 6. Relation of calculated inner integral to well depth of short-range interaction energy profile for angular configurations of chymotrypsinogen pairs sampled by Monte Carlo integration routine. Short-range interactions only, filled squares; short-range plus electrostatic (pH 7, 0.1 M) interactions, open circles.

The reason for this is presumably the unsuitability of the continuum description of van der Waals interactions using the protein–water–protein Hamaker constant for capturing interactions close to contact. On the other hand, use of the protein–vacuum–water Hamaker constant is also inappropriate, as water is excluded from only a small region between the two molecules.

In the remainder of the calculations we therefore use the hybrid model described earlier (continuum at long range, atomistic at short range) in calculating van der Waals interactions. In order to show the effects of the van der Waals interactions alone, separate from the excluded volume, the results are shown in terms of the inner integral introduced via Eq. (5). Since the interactions are strictly attractive, the values of I_{in} are positive for all orientations, corresponding to negative contributions to B_{22}. The filled squares in Fig. 6 show the calculated values of I_{in}, based on van der Waals interactions only, as a function of the well depth of the van der Waals interaction energy profile with r_{12}; 1792 configurations, sampled by the Monte Carlo integration routine, are shown. All the points lie roughly on a straight line on this semilog plot, indicating that the value of I_{in} is controlled mainly by the van der Waals energy well depth rather than reflecting the full energy profile; this is a consequence of the exponential Boltzmann weighting.

The form of the plot shown in Fig. 6 allows the role of electrostatics to be elucidated as well. The I_{in} points in Fig. 6 examined above are shifted when electrostatics are included, with the I_{in} result for each configuration increased or decreased when electrostatics are attractive or repulsive, respectively; the points shown as open circles in Fig. 6 are obtained for calculations at pH 7 and an ionic strength of 0.1 M. Electrostatic interactions are predominantly repulsive, so most of the points move down. Indeed, for some orientations they cause I_{in} to become negative; these are not shown because of the logarithmic scale used, but the values of I_{in} represented are relatively small because the magnitude of the Mayer function (in brackets in Eq. (5)) can be no greater than unity. The attractive electrostatic interactions, on the other hand, are rather weak, so very large values of I_{in} are due mainly to strong van der Waals interactions.

However, because of the Boltzmann weighting, attractive electrostatics can boost I_{in} substantially above the value due to van der Waals interactions alone (Fig. 6); as discussed below, these configurations may be central to determining trends in B_{22}.

An additional feature of the plotted I_{in} values is that they cover a very wide range, which is a critical factor in determining B_{22}. Apart from simple multiplicative constants, B_{22} is calculated by adding the (positive) excluded volume contribution and an unweighted orientational average of I_{in} (Eq. (5)), and given 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. 7, showing the distribution of I_{in} values for the case shown in Fig. 6. Orientations for which I_{in} is smaller in magnitude than about 10^5 Å^3 contribute negligibly to the energetic term in Eq. (5); 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. 6, the main effect causing this is strong van der Waals interactions. These strong interactions result from a high degree of complementarity between the apposing surfaces, characteristic of a molecular recognition phenomenon [27,28]; such behavior can obviously not be captured by models that treat the protein molecules as spheres.

The observed trend seen for B_{22} of chymotrypsinogen around pH 7 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 van der Waals interactions only, leading to a decrease in the value of B_{22}, and vice versa for repulsive electrostatics. Fig. 6 shows that 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, can result in the experimentally observed increase in B_{22} (Fig. 3) [5,15]. The histograms of I_{in} values in Fig. 7, all for the same set of orientations, show there to be more configurations with $I_{in} > 10^8$ Å^3 at 0.1 than at 0.3 M.

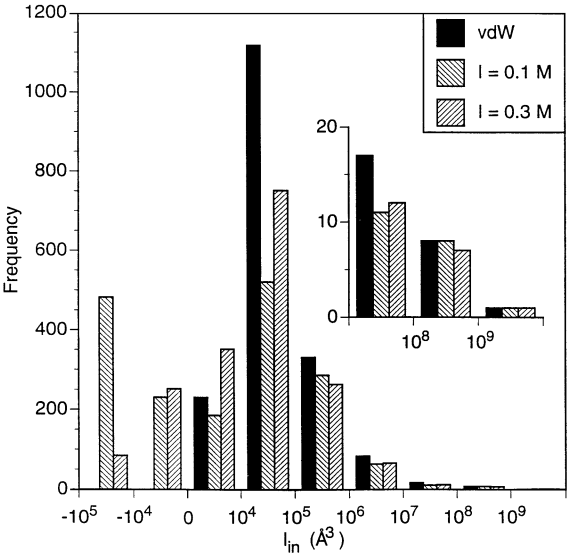

Fig. 7. Distributions of inner integral values for angular configurations of chymotrypsinogen pairs, with I_{in} calculated for different conditions as shown in legend. Inset: enlarged view of distribution for large values of I_{in}.

Although considerable uncertainty remains in the actual calculated values of B_{22} for chymotrypsinogen under different conditions, 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 trend with electrolyte concentration at pH 7 is also correctly captured. Given the large error estimates on B_{22}, however, the agreement is significant less in the correct prediction of values and trends than in the plausibility of the formulation used. The central role of short-range interactions that emerges from our work was postulated previously [29], but the more specific importance of the high-complementarity configurations represents a new paradigm for interpreting B_{22} data. Furthermore, it is precisely this central role of molecular recognition that suggests a mechanistic explanation for the link between

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

pH	Ionic strength (M)	$B_{22} \times 10^4$ (ml mol/g^2)	
		Calculated[a] (est. % error)	Experiment [5]
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)	

[a]Based on molecular weight of 25 670.

B_{22} and crystallization, since protein–protein interactions at the crystal contacts are clearly manifestations of molecular recognition as well. 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 at present, and one must rely on experimental probes of protein interactions.

In summary, our calculations, based on a realistic representation of proteins, 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 due to the influence of these configurations. The physical picture presented by this approach is quite different from that suggested by more idealized colloidal models, and in addition does not depend on fitting values of adjustable parameters. More generally, the results show the vast complexity of molecular interactions that may occur in solutions of protein mixtures, which are likely to make detailed quantitative predictions extremely difficult. On the other hand, the central role of complementarity suggests an explanation for the utility of B_{22} in exploring the suitability of different solvent conditions for crystallization.

Acknowledgements

We are grateful for support from the National Science Foundation under grant number BES-9510420.

References

[1] A. George, W.W. Wilson, Acta Crystallogr. D 50 (1994) 361.
[2] A. George, Y. Chiang, B. Guo, A. Arabshahi, Z. Cai, W.W. Wilson, Meth. Enzym. 276 (1997) 100.
[3] W.G. McMillan, J.E. Mayer, J. Chem. Phys. 13 (1945) 276.
[4] T.L. Hill, Statistical Mechanics, Principles and Selected Applications, Dover, Mineola, NY, 1987.
[5] O.D. Velev, E.W. Kaler, A.M. Lenhoff, Biophys. J. 75 (1998), in press.
[6] B.L. Neal, D. Asthagiri, A.M. Lenhoff, Biophys. J. 75 (1998) 2469.
[7] D.J. Coumou, J. Colloid Sci. 15 (1960) 408.
[8] D.J. Coumou, E.L. Mackor, J. Hijmans, Trans. Faraday Soc. 60 (1964) 1539.
[9] M.H. Smith, in: H.A. Sober (Ed.), Handbook of Biochemistry, Chemical Rubber Co., Cleveland, 1970.
[10] B.H. Zimm, J. Chem. Phys. 16 (1948) 1093.
[11] W.H. Gallagher, C.K. Woodward, Biopolymers 28 (1989) 2001.
[12] M. Muschol, F. Rosenberger, J. Chem. Phys. 103 (1995) 10424.
[13] V.L. Vilker, C.K. Colton, K.A. Smith, J. Coll. Interf. Sci. 79 (1981) 548.
[14] C.A. Haynes, K. Tamura, H.R. Korfer, H.W. Blanch, J.M. Prausnitz, J. Phys. Chem. 96 (1992) 905.
[15] C.J. Coen, H.W. Blanch, J.M. Prausnitz, A.I.Ch.E. J. 41 (1995) 996.
[16] D.E. Kuehner, C. Heyer, C. Rämsch, U.M. Fornefeld, H.W. Blanch, J.M. Prausnitz, Biophys. J. 73 (1997) 3211.
[17] R.J. Hunter, Foundations of Colloid Science, Oxford University Press, Oxford, 1987.
[18] C.M. Roth, B.L. Neal, A.M. Lenhoff, Biophys. J. 70 (1996) 977.
[19] B.L. Neal, A.M. Lenhoff, A.I.Ch.E. J. 41 (1995) 1010.
[20] B.H. Zimm, J. Chem. Phys. 14 (1946) 164.
[21] K.A. Sharp, B. Honig, Ann. Rev. Biophys. Biophys. Chem. 19 (1990) 301.
[22] B.J. Yoon, A.M. Lenhoff, J. Comp. Chem. 11 (1990) 1080.
[23] H.-X. Zhou, Biophys. J. 65 (1993) 955.

[24] W.L. Jorgensen, J. Tirado-Rives, J. Am. Chem. Soc. 110 (1988) 1657.
[25] D. Asthagiri, B.L. Neal, A. M. Lenhoff, (1998), submitted.
[26] D.J. Barlow, J.M. Thornton, Biopolymers 25 (1986) 1717.
[27] W.A. Hendrickson, J.L. Smith, W.E. Royer Jr., in: R.M. Burnett, H.J. Vogel (Eds.), Biological Organization: Macromolecular Interactions at High Resolution, Academic Press, New York, 1987, p. 235.
[28] S. Jones, J.M. Thornton, Proc. Natl. Acad. Sci. 93 (1996) 13.
[29] D. Rosenbaum, P.C. Zamora, C.F. Zukoski, Phys. Rev. Lett. 76 (1996) 150.