Calculation of Weak Protein-Protein Interactions: The pH Dependence of the Second Virial Coefficient

Adrian H. Elcock and J. Andrew McCammon
Department of Chemistry and Biochemistry, Department of Pharmacology, University of California at San Diego,
La Jolla, California 92093-0365 USA

ABSTRACT   Interactions between proteins are often sufficiently weak that their study through the use of conventional structural techniques becomes problematic. Of the few techniques capable of providing experimental measures of weak protein-protein interactions, perhaps the most useful is the second virial coefficient, B_{22}, which quantifies a protein solution's deviations from ideal behavior. It has long been known that B_{22} can in principle be computed, but only very recently has it been demonstrated that such calculations can be performed using protein models of true atomic detail (*Biophys. J.* 1998, 75:2469–2477). The work reported here extends these previous efforts in an attempt to develop a transferable energetic model capable of reproducing the experimental trends obtained for two different proteins over a range of pH and ionic strengths. We describe protein-protein interaction energies by a combination of three separate terms: (i) an electrostatic interaction term based on the use of effective charges, (ii) a term describing the electrostatic desolvation that occurs when charged groups are buried by an approaching protein partner, and (iii) a solvent-accessible surface area term that is used to describe contributions from van der Waals and hydrophobic interactions. The magnitude of the third term is governed by an adjustable, empirical parameter, γ, that is altered to optimize agreement between calculated and experimental values of B_{22}. The model is applied separately to the proteins lysozyme and chymotrypsinogen, yielding optimal values of γ that are almost identical. There are, however, clear difficulties in reproducing B_{22} values at the extremes of pH. Explicit calculation of the protonation states of ionizable amino acids in the 200 most energetically favorable protein-protein structures suggest that these difficulties are due to a neglect of the protonation state changes that can accompany complexation. Proper reproduction of the pH dependence of B_{22} will, therefore, almost certainly require that account be taken of these protonation state changes. Despite this problem, the fact that almost identical γ values are obtained from two different proteins suggests that the basic energetic formulation used here, which can be evaluated very rapidly, might find use in dynamical simulations of weak protein-protein interactions at intermediate pH values.

INTRODUCTION

Macromolecular interactions are often of surprisingly low affinity, even when they are of clear physiological importance. Good examples of low affinity complexes are those formed transiently by proteins involved in electron transfer (Pelletier and Kraut, 1992), and weakly bound multi-enzyme complexes, such as those formed by enzymes of the tricarboxylic acid cycle (Srere et al., 1997). The low stability of these noncovalent associations is such that, in vitro, the complexes usually dissociate, rendering detailed structural study through conventional techniques such as x-ray crystallography almost impossible. Because of these problems, weak interactions have until now received comparatively little attention from researchers, despite the fact that the formation of entities such as multi-enzyme complexes may be extremely important for the regulation of metabolism (see Ovádi, 1991, and references therein).

Given that experimentally based structural studies of weakly interacting macromolecular systems are likely to remain limited, it is worthwhile to consider what other methods may prove useful in providing structural information. One possibility is to develop computational techniques to address the protein docking problem (Janin, 1995; Sternberg et al., 1998). This may be briefly stated as follows: given the structures of two separated proteins, predict the structure of their complex. Many methods have been developed to address this problem, and, although substantial progress has been made, the problem remains essentially unsolved (Sternberg et al., 1998). There are several reasons why a general solution to this problem has not yet been found, but the dominant factor is the extreme computational difficulty associated with including the necessary conformational flexibility in the protein partners (Betts and Sternberg, 1999; Jackson et al., 1998).

For this reason it might be thought that computational methods would be of little use in treating weak interactions, but this may not be the case. It is worth noting, for example, that the development and application of docking methods have so far been aimed more or less exclusively at relatively strong interactions, such as those seen in enzyme-inhibitor and antibody-antigen complexes. Such applications are, of course, important, because crystal structures of the complexes are available, allowing a straightforward determination of the predictive ability of the docking algorithm. However, there is good reason to believe that the protein-protein interfaces involved in such strong interactions may be of a fundamentally different nature from those responsi-

ble for the weak interactions that are of interest here. In particular, the hallmark of many strong complexes, namely a high degree of geometric complementarity and buried hydrophobic surface area (Janin and Chothia, 1990; Jones and Thornton, 1996), is unlikely to be as relevant for weak complexes. Instead, electrostatic interactions are likely to play a more important role. A very good example of the increased importance of electrostatic interactions in more weakly bound complexes has been provided by experimental and computational studies of the cytochrome c–cytochrome peroxidase interaction (Pelletier and Kraut, 1992; Northrup et al., 1988; Roberts and Pique, 1999). Other excellent examples are provided by protein-DNA complexes such as those formed by restriction endonucleases. In particular, several crystallographic studies have shown dramatic differences between the modes of nonspecific and specific protein binding to DNA (Winkler et al., 1993; Albright et al., 1998). Nonspecific binding is largely electrostatic in nature, involving contacts between basic residues on the protein and the sugar and phosphate groups on the DNA. Specific binding (i.e., to the cognate DNA sequence) involves rearrangement of both main-chain and side-chain atoms to form extensive networks of hydrogen bonding and hydrophobic interactions. These observations make the use of computational methods for studying weak interactions more attractive for two reasons: first, meaningful results might be obtained even without including conformational flexibility; and second, proven computational techniques are available for dealing with electrostatic interactions (Honig and Nicholls, 1995).

At first sight, it is still not obvious how one might develop and parameterize a computational method with the intention of focusing specifically on weak macromolecular interactions. After all, what is essential for refining any theoretical technique is good experimental data with which to compare, and we have already noted that there is little in the way of structural information. There is however, at least one experimentally accessible quantity that can be readily measured for weakly interacting macromolecules: this is the second virial coefficient, B_{22} (McMillan and Mayer, 1945), formally defined by

$$B_{22} = \frac{-1}{16\pi^2 MW^2} \int (\exp(-\Delta G_{inter}/kT) - 1) d\Omega \quad (1)$$

where ΔG_{inter} represents the interaction energy (potential of mean force) acting between two protein molecules in solution, MW is the molecular weight of the protein(s), and the $d\Omega$ indicates that the integral is carried out over all possible relative orientations of the two. Although B_{22} can be measured for mixed protein solutions, the experimental data studied here are for single-component protein solutions, so that ΔG_{inter} describes the interaction between two identical macromolecules. Estimates of B_{22} can be obtained from a variety of sources such as osmotic pressure, static and dynamic light scattering, and sedimentation equilibrium experiments (reviewed in Velev et al., 1998). Conceptually,

it quantifies the deviations from ideal behavior of a solution due to the presence of two-body interactions between molecules. Of course, the simplest of interactions between proteins is steric: two molecules cannot occupy the same point in space. Such interactions, which give rise to what is typically called an excluded volume contribution, are repulsive, and, if they are the only interactions acting, result in positive values for B_{22}. Other interactions that will clearly make important contributions are electrostatic, van der Waals, hydrophobic, and hydrogen bonding interactions. Favorable interactions between proteins can cause B_{22} to take on negative values.

The aim of the present work is to use experimental B_{22} data as a means of refining and parameterizing a theoretical method for treating weak protein-protein interactions. Many theoretical approaches to the calculation of B_{22} have already been reported, and it is not our purpose here to discuss in detail previous work; an extremely lucid and thoughtful review of the issues involved in such calculations has been provided by Neal et al. (1998). It is worth stating, however, that until recently such calculations involved often severe approximations, perhaps the most blatant of which was that the interacting proteins were typically modeled as spheres. The advantage of such an approximation is that analytical expressions for the second to the seventh virial coefficients can be derived (Ree and Hoover, 1967; Minton, 1983). However, although simplified models can often provide a phenomenological description of experimental results, they are clearly of no use in providing a structural interpretation of B_{22} values. Such insights can only be obtained using more structurally detailed models; only recently, due to increases in computer power, has the use of such models become possible. An important demonstration of the need to consider atomic details came from the work of Neal and Lenhoff (1995), who showed that when such structural models are used to calculate the excluded volume contribution to B_{22}, the result is some 40% larger than would be estimated using spherical models. This increase was attributed to the roughness of the protein surface.

More recent work by the same authors extended their model to incorporate both electrostatic and short-range interactions in the calculations (Neal et al., 1998). Electrostatic interactions were described using a Poisson-Boltzmann (PB) continuum model, an approach that has been extremely successful in accounting for electrostatic effects in macromolecular situations (Honig and Nicholls, 1995). Short-range interactions, encompassing hydrogen bonding, van der Waals interactions, and hydrophobic effects were dealt with using a hybrid Hamaker/Lennard-Jones treatment, with the former model used to describe interactions between atoms separated by more than 6 Å, and the latter used for all closer interactions. This same model has recently been further parameterized (Asthagiri et al., 1999) in an attempt to reproduce the protein-protein binding free energies tabulated by Horton and Lewis (1992). Using their model, Neal et al. calculated values of B_{22} for chymotryp-

sinogen at 0.1 and 0.3 M ionic strength and at pH 3 and 7. Quantitative agreement between the calculated and experimental values was not obtained, but neither was it sought: no attempt was made to adjust the model to optimize the fit. Instead, the focus of the work was on demonstrating the feasibility of performing such calculations and on showing that the energetic model was to some extent at least, realistic. In this respect the work was highly successful, since the qualitative trends were nicely reproduced: especially important was the qualitative reproduction of the experimental result that at pH 3, B_{22} decreases with increasing salt concentration, whereas at pH 7, it increases with increasing salt concentration. A decrease in B_{22} with increasing salt concentration is usually interpreted in terms of repulsive electrostatic interactions between like-charged molecules becoming screened at higher ionic strengths: such effects become greater as the pH is shifted further away from the protein's pI. On the other hand, an increase in B_{22} with increasing salt concentration indicates the presence of favorable dipolar (and possibly higher order) electrostatic interactions that become progressively screened or weakened at higher salt concentrations. The reproduction of both effects in a theoretical model is a highly encouraging result.

Following the appearance of that work, further experimental work has been reported for both chymotrypsinogen and lysozyme by Velev et al. (1998). The attempted reproduction of this data forms the basis of the work reported here. Our approach is similar to that adopted by Neal et al. (1998), but differs in the following ways. First, we aim specifically to reproduce the experimental data quantitatively as a means of empirically parameterizing a general method for treating weak protein-protein interactions. Second, we use a more elaborate electrostatic model that considers the proteins in true atomic detail, avoiding the spherical dielectric approximation made by Neal et al. (1998) (see Methods). The electrostatic treatment that we use has been employed previously in combination with Brownian dynamics (BD) methods to reproduce successfully the association kinetics of several protein-protein systems (Gabdoulline and Wade, 1997; Elcock et al., 1999; Sept et al., 1999). Third, we attempt to account explicitly for electrostatic desolvation effects in protein-protein interactions: the inclusion of this feature was motivated in part by our recent finding that desolvation effects can play an important role in modulating the rates of macromolecular associations (Elcock et al., 1999). Finally, our energetic model is designed for computations sufficiently fast to permit meaningful use in future dynamics simulations.

Our energetic model is described in detail in Methods. In Results, we show that B_{22} is highly sensitive to the details of the energetic model, making quantitative reproduction of experimental values over a range of solution conditions with a single energetic model extremely difficult. Based on further calculations reported in Results, this inability to fit all results with a single model appears to be due to unavoidable

approximations in our treatment of protonation equilibria of the proteins' ionizable residues. Despite this problem, the parameterized model obtained from fitting to the experimental results for lysozyme at intermediate pH values matches well with that obtained from fitting the chymotrypsinogen results. In Discussion, the implications of this result for developing a transferable energetic model suitable for general use in describing protein-protein association energetics are outlined, and the deficiencies of the model are discussed, together with prospects for their improvement.

METHODS

Structures and assignment of ionization states

Structures for both proteins were taken from the protein data bank (PDB; http://www.rcsb.org). For hen egg white lysozyme, PDB file 1hel was used; for chymotrypsinogen A, file 2cga (the same structure employed by Neal et al., 1998) was used. For all structures, hydrogens were added using the molecular simulation program CHARMM (Brooks et al., 1983). The protonation states of the protein's ionizable residues at each pH studied were determined by performing PB continuum electrostatics calculations. The theory and applications of such calculations have been extensively discussed previously (Antosiewicz et al., 1994, 1996). Briefly, PB calculations are used to evaluate the electrostatic energies of the neutral and ionized forms of each ionizable residue. The difference between the two energies gives a measure of the ionization energy of the residue. This ionization energy is calculated in two environments; first, for the residue in the protein and second, for the residue isolated in solution. The difference between the ionization energies in the two environments directly relates to the change in pKa that results from transferring the residue from solution to the protein. If the pKa of the residue in solution (the reference pKa) is known, the pKa of the residue in the protein can be calculated. The complicated business of dealing with interactions with other charged residues that can themselves undergo ionization state changes is dealt with using the multiple site titration method of Gilson (1993).

All pKa calculations were performed using the continuum electrostatics program UHBD (Madura et al., 1994, 1995). Atomic charges and radii for all residues were taken from the CHARMM parameter set (MacKerell et al., 1998). The solvent dielectric in the electrostatics calculations was assumed, as in previous work (Elcock and McCammon, 1998), to be dependent on salt concentration, with its value set to 77.74 for 100 mM ionic strength and 74.71 for 300 mM ionic strength. The protein dielectric value was set to 20.0; the use of such an apparently high value for the protein dielectric has been shown to produce the best results for the special case of pKa calculations (Antosiewicz et al., 1994, 1996). Reference pKa values were taken from model compound values (Antosiewicz et al., 1994). The values assumed are as follows: asp, 4.0; glu, 4.4; his, 6.3; cys, 8.3; tyr, 9.6; lys, 10.4; arg, 12.0. The N- and C-termini are assigned reference values of 7.5 and 3.8, respectively.

For lysozyme, experimental titration curves have been reported by Tanford and Wagner (1954). Since these curves report only on the number of protons released from the protein at each pH and not the net charge of the protein, we compare only the change in net charge between the three pH values studied. Experimentally, the change in net charge in going from pH 3 to pH 6 is approximately −9e, whereas in going from pH 6 to pH 9, the change is approximately −2.5e. From the calculations, the respective values are −5.5e and −2.5e. The change in charge between pH 3 and 6 is underestimated, suggesting that the net charge in our calculations at pH 3 (+13.50e) is actually not as high as it should be. For chymotrypsinogen, experimental values have been reported by Marini and Martin (1971). The experimental change in charge on going from pH 3 to pH 6.8 is approximately −11.5e, whereas in our calculations it is −9.1e. Again, this sug-

gests that the net charge in our calculations at pH 3 (+14.21e) is perhaps not as high as it should be. These underestimates in protein charge at pH 3 are not the cause of the discrepancies reported in Results, since they would be expected to cause errors in the opposite direction to those observed.

Second virial coefficient calculations

All virial coefficient calculations were performed with an extensively modified version of the SDA program (Gabdoulline and Wade, 1997) originally developed for performing BD simulations of protein-protein interactions. The integral expressed in Eq. 1 is to be carried out over all possible relative orientations of the two molecules. In configurations for which the free energy of interaction ΔG_{inter} is zero, the exponential term equals one, so that the integrand vanishes. Because of this, the integration only needs to be carried out over the limited number of configurations in which the proteins interact. The integration was performed as follows. One protein molecule was placed at the center of a grid of dimensions $80 \times 80 \times 80$, with a spacing between the grid points of 1.5 Å. The center of a second molecule was then placed in turn at each grid point and ΔG_{inter} was calculated using the energy function defined below. Since the integration is carried out over the various angular orientations that the two proteins can adopt, a large number of different orientations of the second protein were tried at each grid point. This involved stepping through Euler angle increments of 20°, thus corresponding to a total of 2592 different angular orientations. Each orientation's contribution to the integral was scaled by the appropriate volume element: dxdydz for the translational term, $\sin\theta d\theta d\phi d\psi$ for the rotational term. In total, the interaction energies of $80 \times 80 \times 80 \times 2592 = 1.327 \times 10^9$ protein-protein geometries were calculated.

Since the longest dimension of our main grid is 80×1.5 Å, the furthest center-center distance examined in the above calculations is 60 Å. Although this is large enough to include the structures making the largest contributions to the integral, the long-range nature of electrostatic interactions, particularly at low ionic strengths and with large net charges on the proteins, means that appreciable contributions can in principle still be made at even further separations. The contributions made by such protein-protein geometries were accounted for by conducting further calculations using a coarser $70 \times 70 \times 70$ grid of spacing 3 Å. Because the interactions at longer distances are less dependent on the precise orientations of the two molecules, fewer orientations (192) were examined, corresponding to Euler angle increments of 45°. This scheme allows contributions from center-center separations of up to 105 Å to be included. In practice, such contributions make only a modest difference of up to $\sim 0.4 \times 10^4$ ml/mol/g^2 to the calculated B_{22} values.

The energy function

The interaction free energy of each protein-protein geometry is calculated using the following functional form:

$$\Delta G_{inter} = \Delta G_{overlap} + \Delta G_{nonpolar} + \Delta G_{elec,inter} + \Delta G_{elec,desol}$$

where $\Delta G_{overlap}$ is the steric interaction energy, $\Delta G_{nonpolar}$ is the favorable contribution due to burial of solvent-accessible surface area (SASA), $\Delta G_{elec,inter}$ is the electrostatic interaction energy (which may be favorable or unfavorable), and $\Delta G_{elec,desol}$ is the unfavorable contribution due to electrostatic desolvation of both proteins. Each of these terms is dealt with in more detail below.

Steric interactions

Our energetic model contains no provision for repulsive van der Waals interactions (in contrast to the approach of Neal et al., 1998), so a method for dealing with sterically disallowed complex geometries is required. To do this, a three-dimensional grid of spacing 0.5 Å was constructed around the first protein. All grid points lying within 2 Å of the van der Waals surface of the protein were assumed to be inside the protein; all other grid points were considered to be outside. Overlap of the two proteins was then determined based on the positioning of the surface atoms of the second protein: if any surface atom was positioned within a region of the grid occupied by the first protein, overlap was assumed and an infinite interaction energy was assigned. This simple approach saves a considerable amount of time and is possible only because we do not allow for conformational flexibility in either protein. Because assigning a value $\Delta G_{overlap} = \infty$ makes the exponential term vanish in Eq. 1, structures in which there are steric clashes make a positive contribution to the integral, giving rise to the so-called excluded volume contribution noted in the Introduction. For lysozyme, ~ 175 million out of the 1.33 billion structures examined were determined to be overlapping.

Nonpolar interactions

In contrast to the approach adopted by Neal et al. (1998), we assume that short-range interactions, encompassing hydrophobic effects and van der Waals interactions (and also specific hydrogen bonding effects), can be expressed as being proportional to the amount of SASA that is buried when the complex is formed. The use of a SASA term to describe such interactions has been criticized by Asthagiri et al. (1999) on the basis that it provides no repulsive term to prevent steric overlap, which may be a problem during dynamics simulations, and that it neglects contributions from atoms positioned just below the protein surface. Both of these points are undoubtedly true, but the former is actually not a problem if one uses the rather simple expedient, outlined above, of assuming an infinite energy for any configuration in which there is overlap. Such an assumption should be considered only a minor drawback for any method, like our own or Asthagiri's, that also makes the far more drastic approximation of ignoring conformational flexibility. In our opinion, this same point also applies to the second objection raised against a SASA-based approach. Neglect of the van der Waals interaction between atoms that are positioned below the surface is unfortunate, but not likely to be particularly important, given that it is just one of a number of factors that are neglected. Implicit in our approach is that the errors introduced by such assumptions can be compensated by suitable parameterization of the SASA contribution. It should in any case be pointed out that the hybrid Hamaker-Lennard-Jones approach used by Asthagiri et al. (1999) is not free of approximations, exemplified by the fact that a scaling factor α was introduced to connect the Lennard-Jones and Hamaker terms smoothly.

Since SASA calculations for proteins are typically rather slow, incorporation of such calculations into the virial coefficient calculations raises obvious difficulties. To deal with this issue, a fast approximate method was employed. The first protein was placed on a cubic grid large enough to accommodate the entire molecule and of spacing 1.5 Å. A probe atom of radius 1.85 Å, taken as representative of the non-hydrogen atoms in a protein, was then placed in turn at all grid points lying within 1 Å and 6 Å of the protein surface. The surface area of the probe atom buried by positioning it at each point in space was calculated explicitly using UHBD. The amount of SASA buried was then mapped onto the 3-D grid. We then assume that the SASA buried by an approaching protein can be expressed as a sum of contributions from all non-hydrogen atoms lying on the surface of that protein. To calculate the total surface area buried by both proteins, the summation is repeated with the roles of the two proteins reversed: i.e., with the second protein positioned at the center of the SASA grid. The total buried surface area is simply the sum of the two results. To do this summation, the position of each non-hydrogen surface atom on the approaching protein is mapped to the SASA grid, and the value at that position is read from the grid, using a trilinear weighting scheme to interpolate between grid points. In order to prevent unrealistic values from being returned by the interpolation scheme, values were also assigned to the nearest interior grid points. Although these positions are physically

inaccessible because they are within the van der Waals surface of the first protein, they need to be assigned some reasonable SASA value. For simplicity, we assigned all interior grid points a SASA value equal to the mean value found in the first 1 Å shell accessible to the second protein.

In order to evaluate the accuracy of this approximation, a preliminary grid search for chymotrypsinogen was used to identify potential complex structures with substantial amounts of buried SASA. The 200 structures obtained from this initial search were used as a test set and the buried SASA calculated with the grid approximation was compared with that obtained from a rigorous calculation using UHBD. The correlation obtained is quite good ($r^2 = 0.63$), but the approximation consistently overestimates the buried SASA (see filled circles in Fig. 1). Assuming that all surface atoms on the approaching protein contribute equally to the burial of SASA is, therefore, a poor approximation. Instead, much better agreement is obtained when the contribution from each surface atom is scaled by a factor related to its own solvent accessibility. Specifically, the scaling factor was calculated as the accessible surface area of the atom divided by 28.3 Å^2, the maximum area that an atom of radius 1.85 Å can bury through contact with another atom. An upper limit of 1.0 was placed on this scaling factor. The results obtained with this scaling of atom contributions are quite dramatically improved ($r^2 = 0.79$; open circles in Fig. 1).

Electrostatic interactions

Electrostatic interactions between proteins can be calculated rigorously (within a continuum framework) by solving the PB equation (Honig and Nicholls, 1995). To obtain the electrostatic energy of interaction between two proteins requires three calculations: one of the protein-protein complex, and one of each of the two proteins separated from each other. The difference between the electrostatic energy of the complex and the electrostatic energies of the separated proteins gives the electrostatic interaction energy. For the present application, where over 1 billion complex structures are investigated, such an approach is currently out of the ques-

tion, since solving the PB equation using finite difference methods for a modest-sized protein can require several minutes. For the present situation, then, it is clear that a very fast method for treating electrostatic interactions is required. Gabdoulline and Wade (1996) have developed an effective charge method that at least partly fulfills this requirement: the method provides an excellent description of electrostatic interactions in cases where desolvation effects are unimportant. The usefulness of this method has been demonstrated both in the original work (Gabdoulline and Wade, 1996) and through its inclusion in BD simulations of protein-protein association kinetics (Gabdoulline and Wade, 1997; Elcock et al., 1999; Sept et al., 1999). Because of these previous studies, the accuracy of the method is not further demonstrated here.

In order to calculate interaction energetics using this approach, we again use a grid-based method. The electrostatic potential around the first protein is calculated and mapped onto a 3-D grid. The electrostatic interaction energy of a second protein is calculated by summing the contributions from each of its effective charges, where each charge makes a contribution of $\frac{1}{2}q_2\varphi_1$, with q_2 being the effective charge on the second protein and φ_1 being the electrostatic potential at that charge generated by the first protein. Again, to calculate the total interaction energy, the summation is repeated with the roles reversed, i.e., with the second protein at the center of the grid. The total electrostatic interaction energy, $\Delta G_{elec,inter}$, is the sum of the two contributions. This method is extremely fast for two reasons. First, by pre-computing the electrostatic potential around a protein, the interaction with a second protein can be calculated without the need for expensive distance calculations: all that is required is that the potential be read from the grid (again, using a trilinear weighting scheme to interpolate between grid points). Second, the number of effective charges needed to calculate the electrostatic interactions accurately is much less than the total number of atoms in the protein. Accurate results can be obtained, for example, by placing effective charges only at the positions of ionizable residues (Elcock et al., 1999). This reduces the number of charges involved in the calculations from 1971 to 32 for lysozyme and from 3593 to 40 for chymotrypsinogen.

The electrostatic potentials around proteins were obtained by solving the PB equation on a 3-D grid of $149 \times 149 \times 149$ grid points spaced 1.5 Å apart. We note that this grid spacing is considerably larger than is typically used in energetic calculations using the PB equation; spacings of <1.0 Å are more usual for such applications. Use of a larger grid spacing is justified because the present application requires that we calculate interaction energies over a wide range of protein-protein separations: a large grid is required so that all important contributions to the integral in Eq. 1 are included. In any case, concerns about the precision of the potentials should be weighed against other weaknesses of the model, in particular, its neglect of conformational flexibility: there is no guarantee that a more precise potential grid would give a more accurate representation of binding energetics, given that several important factors contributing to the binding affinity are missing.

As with the pKa calculations, charges and radii for all atoms of non-ionizable residues were obtained from the CHARMM22 parameter set (MacKerell et al., 1998). Charges for atoms of ionizable residues were set to values intermediate between their charged and neutral forms to an extent depending on the calculated degree of ionization of the residue at the pH and ionic strength of interest. The solvent dielectric was set to the same value used in the pKa calculations, but the protein dielectric constant for these calculations was set to 2.0. Before solving the PB equation for each system, the electrostatic potential at the boundary of the grid was set by summing contributions from every atom in the protein, assuming that all such contributions are subject to Debye-Hückel screening. The linearized PB equation was then solved straightforwardly using the finite difference method implemented in UHBD (Madura et al., 1994, 1995). These same electrostatic potentials were also used to derive effective charges for each protein using the method developed and implemented in the program ECM by Gabdoulline and Wade (1996).

- ● contribution of surface atoms not scaled (r^2=0.63)
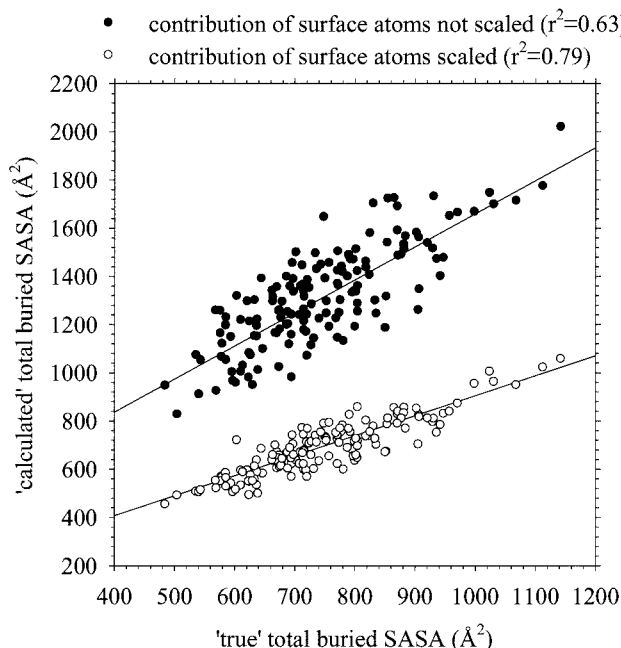- ○ contribution of surface atoms scaled (r^2=0.79)

FIGURE 1 Comparison of rigorous and approximate values of the amount of SASA buried in a set of 200 complex structures calculated for chymotrypsinogen A.

Electrostatic desolvation effects

Although an excellent approximation at intermediate to long separations, the use of effective charges becomes increasingly poor as the proteins approach contact (Gabdoulline and Wade, 1996). The primary reason for this is that electrostatic desolvation effects are neglected. Physically, such effects result from the displacement of high dielectric solvent around a protein by the low dielectric of an approaching second protein. This energetically unfavorable phenomenon is thermodynamically important, e.g., it is the reason why protein salt bridges are of only limited stability at room temperature (Hendsch and Tidor, 1994). However, in certain cases electrostatic desolvation can also have quite dramatic structural consequences; we have previously shown, for example, that the approach of a low dielectric protein can cause the large deformations in DNA structure observed experimentally (Elcock and McCammon, 1996).

Such effects are not straightforward to include in a computationally efficient way. The rigorous approach of re-solving the PB equation for each protein-protein geometry would incorporate electrostatic desolvation effects implicitly, but these are beyond current computational resources. In the work of Neal et al. (1998), attempts were made to circumvent this problem by using simplified PB calculations: the shapes of the proteins were approximated as spheres, thus avoiding the very time-consuming process of defining internal and external dielectric regions. Although this means that the PB equation is rigorously solved, it is questionable to what extent the spherical approximation is likely to be accurate, particularly since the strength of electrostatic interactions appears to be consistently overestimated by the approximation (Neal et al., 1998).

In the present work we have employed an alternative approach that maintains the atomic resolution of the protein surface but avoids re-solving the PB equation for each protein-protein geometry. By analogy to the SASA grid, we calculated an electrostatic desolvation grid around each protein: this grid maps the energetic consequences of having a probe atom of internal dielectric 2.0 occupying each grid position around the protein. Again, these calculations were performed for all grid positions lying within 1 and 6 Å of the van der Waals surface. At each grid point, the desolvation energy caused by the presence of an uncharged probe atom of radius 1.85 Å was calculated by subtracting the electrostatic energy of the protein with the probe present from the electrostatic energy of the protein without the probe. These calculations were performed using small grids ($20 \times 20 \times 20$ by 0.5 Å) centered on the probe atom, and the potentials at the edges of each grid were obtained by focusing from a larger grid encompassing the entire protein. The calculated values of the electrostatic desolvation at each point were then mapped to a grid of identical spacing to the electrostatic potential grid. In line with the nonpolar calculations, the total desolvation contribution due to the approach of a second protein was then obtained by summing the values of the desolvation term at each heavy atom on the surface of the second protein. As before, better results were obtained when the contribution due to each atom on the second protein was scaled by its solvent accessibility (Fig. 2). To find the desolvation energy caused by approach of the first protein to the second protein, the roles of the two proteins were reversed. Our approach to calculating electrostatic desolvation effects is clearly only approximate (Fig. 2), and obviously more work remains to be done before a rapid but truly accurate method is developed. Despite worries over the accuracy of the approximation, it is worth noting that we also performed B_{22} calculations without the electrostatic desolvation term, obtaining results that were, if anything, slightly worse (see Results for an exception). Because neglecting the desolvation term in effect means omitting an unfavorable contribution to ΔG_{inter}, the values of γ required to fit experimental results with this model were much smaller ($0.007–0.010$ kcal/mol/Å²).

Calculation of changes in ΔG_{inter} due to protonation state changes

In order to calculate the change in ΔG_{inter} that occurs when the proteins are allowed to undergo protonation state changes in the complex, the structures were subjected to explicit pKa calculations as described above for the
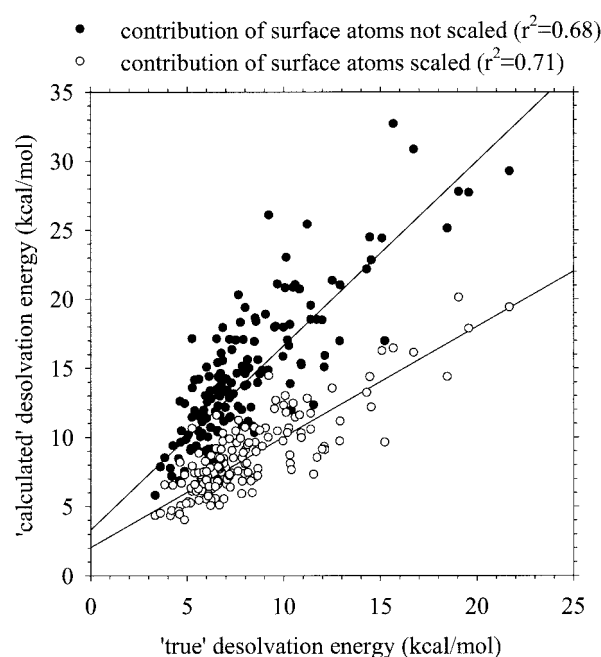

FIGURE 2   Comparison of rigorous and approximate values of the electrostatic desolvation energies for a set of 200 complex structures calculated for chymotrypsinogen A.

isolated proteins. In addition to returning the mean charge of each ionizable residue at the pH value of interest, these calculations give the value of the electrostatic potential at each ionizable residue that would be generated by a unit charge placed at the location of every other ionizable residue, and the self potential generated by placing a unit charge at the ionizable residue itself (Antosiewicz et al., 1994). In line with the usual uses of the linear PB equation (Honig and Nicholls, 1995), the electrostatic energy of the complex after protonation state changes can be calculated as:

$$\Delta G_{elec} = 1/2 \sum_i \sum_j q_i q_j \phi_{ij}$$

where q_i and q_j are the new charges on all ionizable residues and ϕ_{ij} is the above-mentioned unit charge potential (or the self potential if $i = j$). The double summation runs over all ionizable residues in the complex. The change in electrostatic energy due to the change in protonation states can then be estimated by repeating the above calculation with the old charges, i.e., the charges obtained from pKa calculations on the isolated proteins.

RESULTS

Lysozyme

The model we use to describe the energetics of protein-protein interactions contains two components that together account for the electrostatic interactions and an additional single component that describes contributions from nonpolar interactions (see Methods). For the moment we will assume that our description of the electrostatic interactions is accurate and nonadjustable; we will assume, on the other hand, that the magnitude of the nonpolar term, which is proportional to the buried SASA, can be freely varied by incorporating a scaling factor γ that is optimized to fit the

experimental values. Fig. 3 shows a plot of the calculated B_{22} for lysozyme at [100 mM, pH 6] as a function of γ. At low values of γ, the calculated B_{22} is insensitive to γ, but at intermediate values (>0.010 kcal/mol/Å^2), it becomes progressively more sensitive, so that when B_{22} approaches the experimental value (-2.46×10^4 ml.mol/g^2), it begins to decrease drastically. As has been alluded to by Neal et al. (1998), it is this extreme sensitivity to the details of the energetic model that makes accurate calculation of B_{22} especially difficult.

Fig. 4 shows the difference between the calculated and experimental values of B_{22} for lysozyme in six different solution conditions as a function of γ. For perfect agreement with experiment we look for all six curves to cross the *x*-axis at the same value of γ: this would indicate that a single value of γ can be found that, in combination with our electrostatic model, correctly describes B_{22} in different conditions. This is not the case: for [100 mM, pH 3] the best value of γ is 0.029 kcal/mol/Å^2, whereas for [300 mM, pH 9] the best value is 0.023 kcal/mol/Å^2. These values are not drastically different: in percentage terms the difference between the extreme estimates of γ amount to only ~25%. Nevertheless, because of the high sensitivity of the results to the exact value of γ used, the errors in calculated B_{22} values using either of the two extreme values are large.

For fitting all experimental results simultaneously, the best value for γ is 0.0237 ± 0.003 kcal/mol/Å^2. The calculated values of B_{22} obtained with this value are compared to the experimental values in Table 1. From a quantitative perspective, the results are in reasonable agreement, although the calculated values at pH 3 appear much too high. Unfortunately, however, the calculations do not reproduce
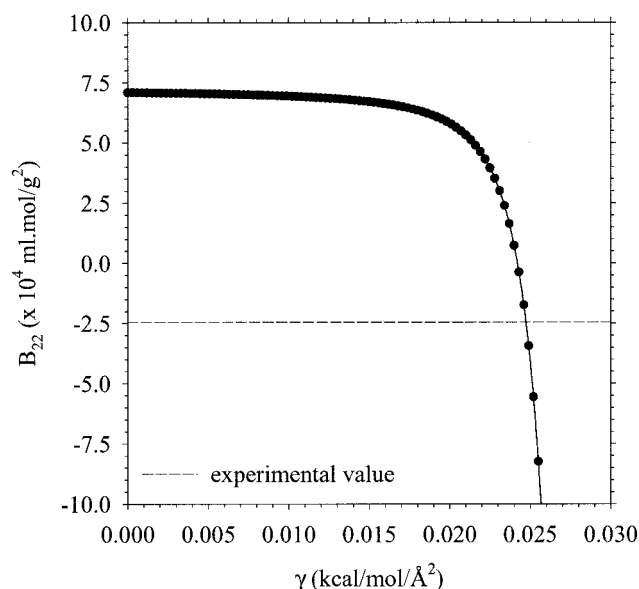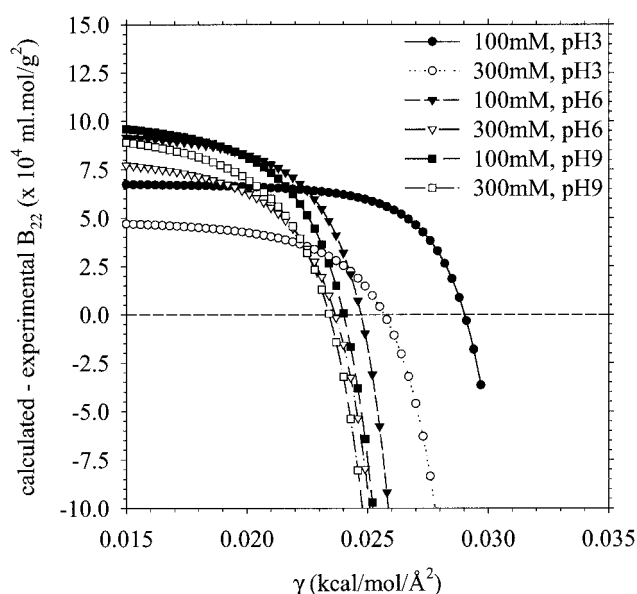

FIGURE 4   Dependence of the difference between calculated and experimental B_{22} values for lysozyme in different solution conditions as a function of γ.

the correct ordering: the calculated B_{22} at [300 mM, pH 6] is lower than that obtained at [100 mM, pH 9], whereas the reverse is true experimentally. In fact, the results we obtain using our more elaborate energetic model are actually not much better than the simple Derjaguin-Landau-Verwey-Overbeek (DLVO) model used by Velev et al. (1998) to interpret their experimental results. It should be noted, however, that the incorrect ordering of the [300 mM, pH 6] and [100 mM, pH 9] results is even worse with the DLVO model. In a subsequent section, we attempt to establish the reason why our model provides only a modest improvement in accuracy.

In order to provide a more detailed basis for understanding the calculated B_{22} values, the 200 structures with lowest ΔG_{inter} in each solution condition were retained for further analysis. The geometries of the 200 best structures found at 100 mM ionic strength are shown in Fig. 5 as a function of pH. Most of the most favorable structures found at pH 3 are identical with those found at pH 9, reflecting the fact that the dominant contributions to ΔG_{inter} come from the pH-independent $\Delta G_{nonpolar}$ term (see below). On the other hand, there are slight but noticeable changes in the distributions (see in particular the loss of the secondary cluster obtained at pH 3), indicating that the most favorable structures will be dependent on pH to some extent. This result is to be expected, given that lysozyme can crystallize in different space groups under different solution conditions (Velev et al., 1998). The distribution of the various energetic components among the 200 structures is shown in Table 2. Interestingly, the ordering of mean ΔG_{inter} values is the same as the ordering of the calculated B_{22} values: for example, the

FIGURE 3   Dependence of the calculated B_{22} value for lysozyme at [100 mM, pH 6] as a function of γ. The *dashed line* indicates the experimental value (Velev et al., 1998).

TABLE 1

Value	100 mM pH 3	300 mM pH 3	100 mM pH 6	300 mM pH 6	100 mM pH 9	300 mM pH 9
B_{22} (exp.)	3.95	1.0*	−2.46	−3.60	−4.41	−5.28
B_{22} (best)	10.21	3.78	1.63	−3.77	−2.93	−6.71
B_{22} (Velev)	10.10	N.R.	1.77	−4.02	−1.00	−5.14

incorrect ordering of the [300 mM, pH 6] and [100 mM, pH 9] results remains. In one sense this is not too surprising: the exponential nature of the theoretical expression for B_{22} (Eq. 1), means that the more favorable structures make a disproportionately large contribution to the result (Neal et al., 1998). In another sense, however, this is actually encouraging, because it suggests that conducting further analysis on a small, representative set of structures might be a useful approach to understanding the discrepancies between calculation and experiment, thus avoiding the lengthy recalculation of the integral.

The mean values of the different energetic components of ΔG_{inter} also show some interesting trends. The ordering of the mean $\Delta G_{nonpolar}$ values, for example, is identical with the ordering of the total ΔG_{inter} values, suggesting (perhaps surprisingly) that this is the primary determinant of the final B_{22} values calculated. The ordering of the mean $\Delta G_{elec,inter}$ values, on the other hand, is in line with what one would expect if the electrostatic energy of interaction were determined solely by the net charge on the two protein molecules. Increasing the pH so that it approaches the pI of lysozyme ($\sim$11) decreases the net charge on the proteins and thus decreases the electrostatic repulsion. Increasing the ionic strength, on the other hand, simply causes a screening of the electrostatic interactions. In these respects, the behavior exhibited by lysozyme is exactly that outlined by Velev et al. (1998). The ordering of the mean $\Delta G_{elec,desol}$ values can also be understood straightforwardly, since it is common to find in binding situations that the desolvation energy is greatest when the screened electrostatic interaction energy is the most favorable (see for example, Hendsch and Tidor, 1994; Majeux et al., 1999; Elcock et al., 1999). It is also worth noting that rigorous calculation of the electrostatic solvation energies of the isolated protein molecules shows the solvation energy at pH 3 to be greater (i.e., more favorable) than at pH 9 (data not shown). Therefore, the desolvation energy involved in complex formation is not simply related to the negative of the overall solvation energy of each protein, since this would suggest that the desolvation energy should be greatest at pH 3. Instead, the desolvation energy involved in complex formation reflects only the solvation properties of the residues at or near the protein-protein interface.

Chymotrypsinogen A

Fig. 6 shows the difference between the experimental and calculated values of B_{22} as a function of γ for chymotryp-

sinogen A in four experimental combinations of salt concentration and pH. As with lysozyme, we find that higher values of γ are required to fit the experimental results for the pH 3 cases than for pH 7. Overall, the optimum value of γ in this case is 0.0219 kcal/mol/Å^2, which is very similar to the value obtained with lysozyme. This is probably the most important result in the work reported here. The calculated values of B_{22} obtained with this value of γ are compared with the experimental values in Table 3. Again, there is a strong resemblance between our results and those obtained by Velev et al. (1998). The unusual experimental result obtained at pH 7 is qualitatively reproduced by both sets of calculations, but the quantitative agreement is poor: the B_{22} calculated at [100 mM, pH 7] should be much lower than at [300 mM, pH 7]. As with lysozyme, we find that the ordering of the mean ΔG_{inter} values calculated for the best 200 structures is identical with that of the overall calculated B_{22} values (Table 4). Interestingly, the small difference between the mean ΔG_{inter} values at [100 mM, pH 7] and [300 mM, pH 7] appears to be due to partial cancellation of opposing influences: although the electrostatic components are considerably more favorable at 100 mM, the nonpolar contributions are more favorable at 300 mM. Surprisingly, better reproduction of these two results is obtained from virial coefficient calculations that omit electrostatic desolvation entirely (data not shown); such a model produces worse results for other cases, however, and so is not considered in further detail.

Ionization state changes accompany complex formation

As discussed in Methods, our calculations take account of the effects of different pH values on the ionization states of the separated proteins: in the case of lysozyme, for example, the net charge on the protein decreases from +13.50e to +5.84e as the pH changes from 3 to 9 at 100 mM ionic strength. Our model does not, however, account for the possibility that further ionization state changes occur when two proteins associate to form a complex. Recently, Sharp (1998) has shown the importance of considering changes in ionization states for understanding the effects of mutations on lysozyme-antibody binding affinities (Sharp, 1998; see also McDonald et al., 1995). It seems reasonable to assume that neglecting such changes may also affect our calculated B_{22} values.

This is particularly likely to be true, given that for both lysozyme and chymotrypsinogen, a noticeably higher value
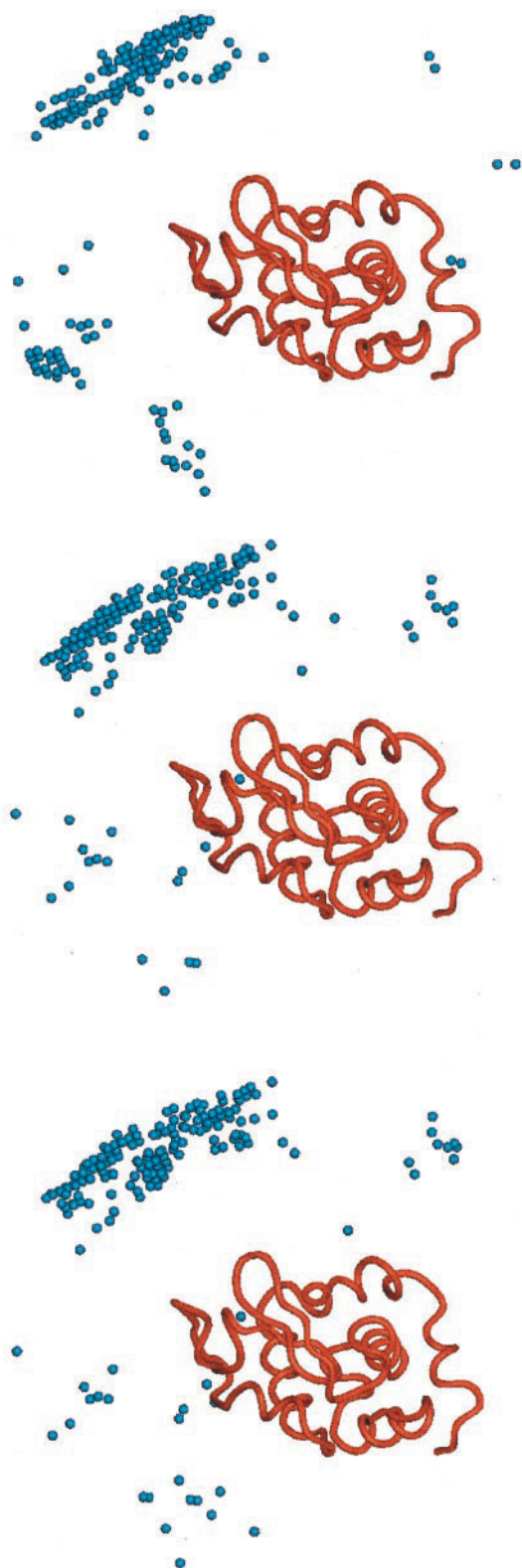
FIGURE 5 Distribution of the 200 most favorable structures obtained in calculations on lysozyme at 100 mM ionic strength. The position of one lysozyme molecule is indicated by the red ribbon; the positions of the center of mass of the second lysozyme molecule are indicated by blue spheres. Results are shown for pH 3 (*top*), pH 6 (*middle*), and pH 9 (*bottom*).

of γ is required to fit the experimental results at pH 3 than at other pH values (see Figs. 4 and 6). This result suggests that the electrostatic component of ΔG_{inter} in our calculations is probably too unfavorable at pH 3. Note that this cannot be due to errors in the net charges assigned to the proteins at this pH, since these are, if anything, underestimated by our calculations (see Methods). Instead, we believe for the following reasons that the effect results from the neglect of ionization state changes. At pH 3, many of the acidic residues on the surface of the proteins are likely to be protonated. For example at [100 mM, pH 3], 4 of the 10 acidic residues in an isolated lysozyme molecule are calculated to be more than 50% neutralized, and the mean charge on the acidic residues is −0.54e. However, if on complexation a neutralized acidic residue were to approach a basic residue on the surface of a second protein, it is likely that it will become deprotonated: the pKa values of acidic residues involved in salt bridges, for example, are often shifted downward by as much as 3 units (i.e., to <3; Anderson et al., 1990). This change in ionization state would result in a more favorable electrostatic interaction between the two proteins, but this effect is missed in our calculations of B_{22}. Of course, similar effects could in principle occur at other pHs, but in practice these are less likely because at pH 6, 7, and 9, there are fewer groups that can undergo ionization state changes on complexation, because few groups titrate at these pH values: only the N-terminal amino group (pKa ∼7.5) and histidine residues (pKa ∼6.3) titrate in these pH regions. There are only one and two histidines in lysozyme and chymotrypsinogen, respectively.

To assess whether ionization state changes are likely to affect the calculated B_{22} values, we performed the following study. For each set of solution conditions, the 200 most favorable structures were subjected to explicit calculations of their protonation states (see Methods; Antosiewicz et al., 1994). The protonation states calculated in the complexed forms were then compared with the protonation states in the uncomplexed forms. The mean changes in charge that accompany complexation for each set of 200 structures calculated for lysozyme are shown in Table 5. In all cases, the net change in charge is negative, indicating that throughout the set of structures, there is a net loss of protons when the complex forms. More importantly, however, the net change in charge is much larger for pH 3 than for either of the other two pH values: at [100 mM, pH3] for example, the mean change is −0.79e, with one of the structures even showing a net change of −1.84e. A large mean change is also observed at [300 mM, pH3], but the magnitude is decreased, a result that makes sense given the increased electrostatic screening that occurs at the higher salt concentration. At pH 6 on the other hand, the mean change corresponds to only −0.05e at both salt concentrations investigated. This result is consistent with the argument proposed above, that the number of groups capable of undergoing ionization state changes is decreased at this pH.

TABLE 2

Value	100 mM pH 3	300 mM pH 3	100 mM pH 6	300 mM pH 6	100 mM pH 9	300 mM pH 9
$\langle \Delta G_{inter} \rangle$	−5.21	−6.02	−7.09	−7.28	−7.25	−7.34
$\langle \Delta G_{elec,inter} \rangle$	1.60	0.98	0.18	0.08	0.06	0.03
$\langle \Delta G_{elec,desol} \rangle$	3.05	3.22	3.31	3.45	3.49	3.50
$\langle \Delta G_{nonpolar} \rangle$	−9.88	−10.22	−10.59	−10.81	−10.80	−10.87

These changes in charge have consequences for the calculated free energy of interaction. As with the overall charge, we can compare ΔG_{inter} values calculated when ionization state changes are allowed to occur following complexation with those calculated under the assumption that no re-ionization occurs (see Methods), in order to estimate the mean change in ΔG_{inter} that occurs after re-ionization. These calculated $\Delta\Delta G$ values for lysozyme are also reported in Table 5. At pH 3 and pH 9, allowing re-ionization causes a mean decrease in ΔG_{inter}, suggesting that for these cases, the electrostatic interaction energy computed in our virial coefficient calculations is too unfavorable. This effect is especially pronounced at pH 3, in line with the argument outlined above, but is also considerable at pH 9, where, although the mean change for the 200 structures is small, individual structures show changes of approximately −0.5 kcal/mol. At pH 6, on the other hand, only very minor changes occur, and in fact there is, on average, a very slight increase in the ΔG_{inter}. Given these effects, it is reasonable to suggest that the neglect of re-ionization may provide an explanation not only for the inflated B_{22} values calculated at pH 3, but also for the incorrect ordering of the [300 mM, pH 6] and [100 mM, pH 9] results.

Similar results are obtained when we repeat the process for chymotrypsinogen (see Table 6). Again, much larger changes in charge occur at pH 3 than at pH 6.8, and again at both pH values studied, larger changes are on average found at the lower salt concentrations. In energetic terms, this again means that the ΔG_{inter} values computed in the virial coefficient calculations at pH 3 are not as favorable as they should be. The magnitudes of these changes are again consistent with the trends in the γ values seen in Fig. 6, where the highest value is required at [100 mM, pH 3] and the next highest at [300 mM, pH 3]. Note, however, that the absence of re-ionization does not appear to be the cause of the unrealistic closeness of the calculated B_{22} values for [100 mM, pH 6.8] and [300 mM, pH 6.8]; this poor result remains unexplained.

DISCUSSION

The second virial coefficient, B_{22}, provides a convenient measure of the strength of macromolecular interactions in solution. As such, it offers a potential role in characterizing weak protein-protein interactions that cannot currently be fulfilled by structural techniques. The work reported here has aimed at exploiting experimental B_{22} data to develop a theoretical method capable of describing generic protein-protein interactions. In some respects the study has been successful, in others it has not. The most important result is that for intermediate pH values, very similar values of γ are obtained for two quite different protein systems; this is encouraging, because it hints that a transferable, albeit empirically derived, energetic model may have been attained. Because our energetic model is in many respects only approximate (see Methods), it should be clear that the only proper way to verify this is by application to other protein systems. Certainly, a degree of caution is necessary anyway given the fact that B_{22}, as an integral over many configurations, can in principle be correctly reproduced by an
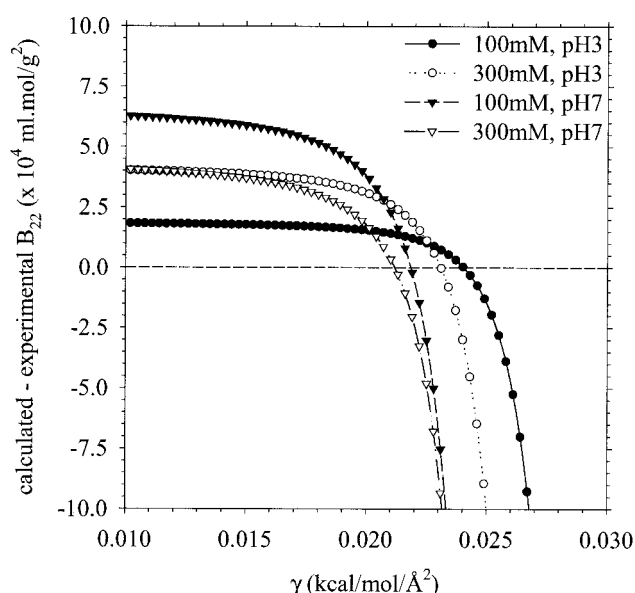

FIGURE 6 Dependence of the difference between calculated and experimental B_{22} values for chymotrypsinogen A in different solution conditions as a function of γ.

TABLE 3

Value	100 mM pH 3	300 mM pH 3	100 mM pH 6.8	300 mM pH 6.8
B_{22} (exp.)	2.50	−1.40	−4.10	−2.05
B_{22} (best)	3.71	0.48	−4.30	−4.09
B_{22} (Velev)	3.21	0.08	−2.52	−2.43
B_{22} (elec.)	4.36	2.73	2.40	2.22

TABLE 4

Value	100 mM pH 3	300 mM pH 3	100 mM pH 6.8	300 mM pH 6.8
$\langle \Delta G_{inter} \rangle$	−6.12	−6.90	−7.94	−7.80
$\langle \Delta G_{elec,inter} \rangle$	1.05	0.59	−1.08	−0.58
$\langle \Delta G_{elec,desol} \rangle$	3.65	3.98	4.95	5.02
$\langle \Delta G_{nonpolar} \rangle$	−10.82	−11.46	−11.81	−12.23

energetic model that provides incorrect (but compensating) descriptions of short- and long-range interactions. Although this is possible, the fact that we describe long-range electrostatic interactions by the well proven effective charge method (Gabdoulline and Wade, 1996, 1997) should minimize this possibility.

Interestingly, the optimal value we obtain for γ ($\sim$0.023 kcal/mol/Å^2) is in close correspondence with the value of 0.025 kcal/mol/Å^2 obtained by several groups in correlating SASA values with solvation energies of amino acids and proteins (Chothia, 1974; Simonson and Brunger, 1994). Since our γ is expected to account not only for solvation energy changes, but also for van der Waals interactions and all other factors not treated explicitly in the model, including conformational flexibility, there is no a priori reason why our derived value should be similar to values derived only to reproduce hydration energy data. It is, however, good to see that this value is also similar to proportionality constants derived from fitting SASA changes to the affinity of protein-protein complexes (Horton and Lewis, 1992).

Although the derivation of an apparently consistent energetic model should be considered a considerable success in the present work, other aspects of the results are less pleasing. The most important of these is that it has not proven possible to arrive at a single value of γ that can quantitatively describe the variation of B_{22} over the full range of solution conditions. This failing should not be overstated. It reflects in part the tremendous sensitivity of B_{22} to the details of the energetic model, which results directly from the disproportionately large contributions made by favorable protein-protein geometries. At least some of the discrepancies with experimental results do, however, appear to be attributable to a failing of the model, namely, its neglect of the changes in protonation states that can accompany complexation. We have shown this by explicitly calculating protonation states for the 200 best complex structures obtained at each solution condition. Unfor-

tunately, it is impossible, for two reasons, to demonstrate with certainty that reionization effects are capable of completely correcting our computed B_{22} values. First, the 200 best structures that we analyzed were obtained under the assumption of no reionization; it is possible that these might not be completely representative of the best structures that would be obtained if reionization were allowed. Second, we cannot estimate the effect of reionization on the millions of other structures that contribute to the integral: we can claim, at least, that given the exaggerated importance of the most favorable structures in determining the integral, it seems highly probable that the effects we see in 200 structures will be the dominant effects.

If inclusion of ionization state changes proves to be essential for describing B_{22} in different solution conditions, a much more rapid means of calculating protonation states will be required: the protonation state calculations conducted here, for example, took many hours to complete on fast computer workstations. A rather promising development in this direction has recently been reported by Havranek and Harbury (1999), who have resurrected Tanford and Kirkwood's (1957) electrostatic model and demonstrated the possibility of conducting extremely rapid and accurate pKa calculations in proteins. Such a method may ultimately prove useful for the present application.

There are, of course, other deficiencies of our model. A pressing requirement for the future is to take account of flexibility in the proteins. Structural changes accompany all protein-protein association events, even if they are only minor side-chain reorientations; models that treat proteins as rigid bodies neglect this aspect. Unfortunately, the inclusion of flexibility raises very severe computational difficulties, which is why, as stated in the Introduction, the general protein docking problem has not been solved (Sternberg et al., 1998). Aside from noting again that the incorporation of flexibility is much more of a problem for strongly bound complexes, we can also envisage that it is likely to be important for only a fraction of the millions of structures generated in the computation of B_{22}. It should be possible in future, therefore, to develop a hybrid scheme in which rigid models are used to describe geometries in which the proteins are separated by several Ångstroms, with flexible models being used only for the smaller subset of structures in which the proteins are in direct (and already favorable) contact.

In addition to the details of the energetic model, there are other aspects of the calculations that can probably be im-

TABLE 5

Value	100 mM pH 3	300 mM pH 3	100 mM pH 6	300 mM pH 6	100 mM pH 9	300 mM pH 9
Δq (mean)	−0.79	−0.61	−0.05	−0.05	−0.13	−0.07
Δq (min.)	−1.84	−1.41	−0.15	−0.16	−0.45	−0.38
Δq (max.)	−0.46	−0.29	−0.01	0.00	−0.05	−0.01
$\Delta\Delta G$ (mean)	−1.81	−1.33	0.01	0.03	−0.10	−0.04
$\Delta\Delta G$ (min.)	−4.48	−2.85	−0.03	−0.03	−0.66	−0.56

TABLE 6

Value	100 mM pH 3	300 mM pH 3	100 mM pH 6.8	300 mM pH 6.8
Δq (mean)	−0.81	−0.52	−0.08	−0.05
Δq (min.)	−1.49	−1.16	−0.24	−0.52
Δq (max.)	−0.38	−0.10	0.09	0.09
$\Delta\Delta G$ (mean)	−1.16	−0.69	0.10	0.05

proved. The integration itself, for example, might be dramatically accelerated using Fast Fourier Transform (FFT) methods of the kind used already in a number of protein docking algorithms (Katchalski-Katzir et al., 1992; Gabb et al., 1997; Ten Eyck et al., 1995). Our energetic formulation is deliberately well suited to this purpose, in that all three of the energetic terms can be readily computed as convolutions. The use of FFT methods would allow much finer resolution searches to be conducted in both translational and rotational space. They are not, however, easily reconciled with including internal flexibility in the macromolecules, so thought would have to be given to how the two aspects could be profitably combined. One indication that the use of FFT methods might be less pressing than the other issues discussed above is that despite the fact that B_{22} is so sensitive to the energetic model, the calculated γ values appear to be surprisingly insensitive to the resolution of the integration procedure. For both protein systems, we find that almost identical values of γ are obtained from B_{22} calculations that use a much coarser grid (3 Å spacing), with many fewer rotations (192) (data not shown).

This last result may also help to explain why our more structurally detailed model produces results that are in some ways not significantly better than the simpler DLVO-type model used by Velev et al. (1998) (see Tables 1 and 3). Their model treats the proteins as spheres interacting only by charge-charge, charge-dipole, dipole-dipole, and dispersion interactions. Two parameters are adjusted to fit the experimental results: (i) the effective radius of the protein, which determines the excluded volume contribution, and (ii) the Hamaker constant, which determines the magnitude of the dispersion interactions. That such a simple model performs almost as well as our model suggests that, at least for reproducing the trends of B_{22} values for the two proteins studied here, the results are not strongly dependent on the fine structural details of the interactions. On the other hand, two observations suggest that a general description of B_{22} values will require an atomic level description. First, the present work has shown that B_{22} can be sensitive to protonation state changes, which result directly from interactions between individual amino acids. Second, it has very recently been demonstrated that single amino acid substitutions can cause appreciable changes in B_{22}, even when such substitutions do not result in changes in the protein's net charge (Chang et al., 2000). Because of these points, we can suggest that the atomically detailed basis of the current

model (which uses only one adjustable parameter) makes it suitable for a whole range of applications where the spherical approximation fails. In particular, the contributions made by individual residues to weak protein-protein interactions can be straightforwardly investigated using the present model.

Finally, it is worth noting that the model developed here lends itself naturally to inclusion in BD simulations of the type that have been used extensively (Gabdoulline and Wade, 1997; Elcock et al., 1999; Northrup et al., 1988). BD simulations have been used to provide considerable insight into the factors governing the kinetics of macromolecular association events, but have up to now been used only sparingly in applications that depend on thermodynamics (Thomasson et al., 1997). This has been the case simply because it has not been certain that the models used (which typically model only electrostatic interactions) reproduce the thermodynamics correctly. The novelty of the present methodology is that it provides a rather complete framework in which electrostatic interactions, electrostatic desolvation and nonpolar contributions are all included, and it has been proven to reproduce a key thermodynamic property of protein-protein interactions. Although there are obviously other ways of developing approximate energetic models with the same goal in mind (Asthagiri et al., 1999; Camacho et al., 2000), the present model ought to be useful in a variety of different situations.

REFERENCES

Albright, R. A., M. C. Mossing, and B. W. Matthew. 1998. Crystal structure of an engineered Cro monomer bound nonspecifically to DNA: possible implications for nonspecific binding by the wild-type protein. *Protein Sci.* 7:1485–1494.

Anderson, D. E., W. J. Becktel, and F. W. Dahlquist. 1990. pH-induced denaturation of proteins: a single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4-lysozyme. *Biochemistry.* 29:2403–2408.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1994. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* 238:415–436.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1996. The determinants of pKas in proteins. *Biochemistry.* 35:7819–7833.

Asthagiri, D., B. L. Neal, and A. M. Lenhoff. 1999. Calculation of short-range interactions between proteins. *Biophys. Chem.* 78:219–231.

Betts, M. J., and M. J. E. Sternberg. 1999. An analysis of conformational changes on protein-protein association: implications for predictive docking. *Protein Eng.* 12:271–283.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comp. Chem.* 4:187–217.

Camacho, C. J., S. R. Kimura, C. DeLisi, and S. Vajda. 2000. Kinetics of desolvation-mediated protein-protein binding. *Biophys. J.* 78: 1094–1105.

Chang, R. C., D. Asthagiri, and A. M. Lenhoff. 2000. Measured and calculated effects of mutations in bacteriophage T4 lysozyme on interactions in solution. *Proteins Struct. Funct. Genet.* 41:123–132.

Chothia, C. 1974. Hydrophobic bonding and accessible surface area in protein. *Nature.* 248:338–339.

Elcock, A. H., and J. A. McCammon. 1996. The Low Dielectric Interior of Proteins is Sufficient to Cause Major Structural Changes in DNA on Association *J. Am. Chem. Soc.* 118:3787–3788.

Elcock, A. H., and J. A. McCammon. 1998. Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.* 280:731–748.

Elcock, A. H., R. R. Gabdoulline, R. C. Wade, and J. A. McCammon. 1999. Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin. *J. Mol. Biol.* 291:149–162.

Gabb, H. A., R. M. Jackson, and M. J. E. Sternberg. 1997. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* 272:106–120.

Gabdoulline, R. R., and R. C. Wade. 1996. Effective charges for macromolecules in solvent. *J. Phys. Chem.* 100:3868–3878.

Gabdoulline, R. R., and R. C. Wade. 1997. Simulation of the diffusional association of barnase and barstar. *Biophys. J.* 72:1917–1929.

Gilson, M. K. 1993. Multiple-site titration and molecular modelling: two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins Struct. Func. Genet.* 15:266–282.

Havranek, J. J., and P. B. Harbury. 1999. Tanford-Kirkwood electrostatics for protein modeling. *Proc. Natl. Acad. Sci. USA.* 96:11145–11150.

Hendsch, Z. S., and B. Tidor. 1994. Do salt bridges stabilize proteins? A continuum electrostatics analysis. *Protein Sci.* 3:211–216.

Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science.* 268:1144–1149.

Horton, N., and M. Lewis. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci.* 1:169–181.

Jackson, R. M., H. A. Gabb, and M. J. E. Sternberg. 1998. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol.* 276:265–285.

Janin, J. 1995. Protein-protein recognition. *Prog. Biophys. Mol. Biol.* 64:145–166.

Janin, J., and C. Chothia. 1990. The structure of protein-protein recognition sites. *J. Biol. Chem.* 265:16027–16030.

Jones, S., and J. M. Thornton. 1996. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* 93:13–20.

Katchalski-Katzir, E., I. Shariv, M. Eisenstein, A. A. Friesen, C. Aflalo, and S. J. Wodak. 1992. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U.S.A.* 89:2195–2199.

MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnik, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D.Yin, and M. Karplus. 1998. *J. Phys. Chem. B.* 102:3586–3616.

Madura, J. D., M. E. Davis, M. K. Gilson, R. C. Wade, B. A. Luty, and J. A. McCammon. 1994. Biological applications of electrostatics calculations and Brownian dynamics simulations. *Rev. Comp. Chem.* 5:229–267.

Madura, J. D., J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon. 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Comp. Phys. Comm.* 91:57–95.

Majeux, N., M. Scarsi, J. Apostolakis, C. Ehrhardt, and A. Caflisch. 1999. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins Struct. Funct. Genet.* 37:88–105.

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.

McDonald, S. M., R. C. Willson, and J. A. McCammon. 1995. Determination of the pKa values of titratable groups of an antigen-antibody complex, HyHEL-5-hen egg lysozyme. *Protein Eng.* 8:915–924.

McMillan, W. G., and J. E. Mayer. 1945. The statistical thermodynamics of multicomponent systems. *J. Chem. Phys.* 13:276–305.

Minton, A. P. 1983. The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Mol. Cell Biochem.* 55:119–140.

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

Neal, B. L., D. Asthagiri, and A. M. Lenhoff. 1998. Molecular origins of osmotic second virial coefficients of proteins. *Biophys. J.* 75:2469–2477.

Northrup, S. H., J. Boles, and J. Reynolds. 1988. Brownian dynamics of cytochrome c and cytochrome c peroxidase association. *Science.* 241:67–70.

Ovádi, J. 1991. Physiological significance of metabolic channeling. *J. Theor. Biol.* 152:1–22.

Pelletier, H., and J. Kraut. 1992. Crystal structure of a complex between electron-transfer partners, cytochrome-c peroxidase and cytochrome-c. *Science.* 258:1748–1755.

Ree, F. H., and W. G. Hoover. 1967. Seventh virial coefficients for hard spheres and hard disks. *J. Chem. Phys.* 46:4181–4197.

Roberts, V. A., and M. E. Pique. 1999. Definition of the interaction domain for cytochrome c on cytochrome c oxidase. *J. Biol. Chem.* 274:38051–38060.

Sept, D., A. H. Elcock, and J. A. McCammon. 1999. Computer simulations of actin polymerization can explain the barbed-pointed end asymmetry. *J. Mol. Biol.* 294:1181–1189.

Sharp, K. A. 1998. Calculation of HyHEL10-lysozyme binding free energy changes: effect of ten point mutations. *Proteins.* 33:39–48.

Simonson, T., and A. T. Brunger. 1994. Solvation free energies estimated from microscopic continuum theory: an accuracy assessment. *J. Phys. Chem.* 98:4683–4694.

Srere, P. A., A. D. Sherry, C. R. Malloy, and B. Sumegi. 1997. Channelling in the Krebs tricarboxylic acid cycle. *In* Channelling in Intermediary Metabolism. Agius, L., and Sherratt, H. S. A., eds. Portland Press, London.

Sternberg, M. J. E., H. A. Gabb, and R. M. Jackson. 1998. Predictive docking of protein-protein and protein-DNA complexes. *Curr. Opin. Struct. Biol.* 8:250–256.

Tanford, C., and J. G. Kirkwood. 1957. Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* 79:5333–5339.

Tanford, C., and M. L. Wagner. 1954. Hydrogen ion equilibria of lysozyme. *J. Am. Chem. Soc.* 76:3331–3336.

Ten Eyck, L. F., J. Mandell, V. A. Roberts, and M. E. Pique. 1995. *In* Proceedings of the 1995 ACM/IEEE Supercomputing Conference, San Diego, December 3–8, 1995. Hayes, A., and Simmons, M., eds. IEEE Computer Society Press, Los Alamitos, CA.

Thomasson, K. A., I. V. Ouporov, T. Baumgartner, J. Czaplinski, T. Kaldor, and S. H. Northrup. 1997. Free energy of nonspecific binding of cro repressor protein to DNA. *J. Phys. Chem. B.* 101:9127–9136.

Velev, O. D., E. W. Kaler, and A. M. Lenhoff. 1998. Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen. *Biophys. J.* 75:2682–2697.

Winkler, F. K., D. W. Banner, C. Oefner, D. Tsernoglou, R. S. Brown, S. P. Heathman, R. K. Bryan, P. D. Martin, K. Petratos, and K. S. Wilson. 1993. The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* 12:1781–1795.