ELSEVIER

Biophysical Chemistry 78 (1999) 219–231

Biophysical Chemistry

Calculation of short-range interactions between proteins

D. Asthagiri, B.L. Neal[1], A.M. Lenhoff*

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

Received 13 January 1999; received in revised form 11 February 1999; accepted 11 February 1999

Abstract

Macromolecular association is an integral component of numerous cellular and technologically relevant processes. Most molecular theories of such association neglect the explicit solvent structure and rely on continuum concepts such as surface energies for calculating short-range interactions. We present a new such method for calculating the non-electrostatic component of the interaction-free energy, based on formalisms for calculating dispersion interactions between macromolecules. The interactions are separated into a short-ranged component that is treated atomistically, and a longer range component that is treated within the continuum Lifshitz–Hamaker approach. This description avoids the singularities inherent in the continuum dispersion formulation, and its effectiveness in characterizing the shape complementarity between interacting surfaces is shown to be comparable to that of surface area-based methods of similar parametric complexity. An advantage of the new method is that it allows facile calculation of the interaction free energy as a function of intermolecular separation, including steric effects; this makes it suitable for use in simulations of protein solutions.

Keywords: Macromolecular association; Dispersion interactions; Solvation; Binding free energy; Complementarity

1. Introduction

Macromolecular association mediates phenomena such as protein–nucleic acid and protein–protein binding that are critical to the integrity of the cellular structure. In some instances, though, macromolecular association is detrimental to the organism, as in the case of β-amyloid plaque formation. From a technological perspective, macromolecular association facilitates numerous separation processes; crystallization of proteins is one example, and is also important in characterizing the macromolecule itself. Thus, understanding macromolecular association

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

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

0301-4622/99/$ - see front matter
PII: S0301-4622(99)00028-9

is an essential step in understanding numerous cellular and technologically relevant processes.

Macromolecular interactions are governed principally by non-covalent interactions arising from a limited set of fundamental interactions that are understood fairly well at the level of individual atoms [1]. Attempts to understand such associations at a molecular level are, to a large extent, limited by our understanding of how the collection of atoms comprising the macromolecules interacts with the usual aqueous medium, where the presence of the hydrogen bonded structure of water adds an additional level of complexity. In principle, the system comprising the solutes (macromolecules) and the medium can be described by considering the interaction of each atom with all other atoms comprising the system. In practice, though, such an approach is infeasible for all but very small systems. Thus, there is a need for simplified representations of the macromolecular interactions in the solvent, retaining physical realism to the greatest extent possible.

The interaction between two protein molecules can be considered by accounting for different contributions to the interaction free energy. The non-electrostatic contributions are typically van der Waals (dispersion) interactions and the free energy changes associated with the modification of the water structure and its influence on the solute–solute interaction. The electrostatic contributions, including to some degree hydrogen bonding effects, are typically considered within the framework of the macroscopic Poisson and Poisson–Boltzmann equations [2,3]. The electrostatic contributions are fairly well-understood and have been used to describe numerous molecular phenomena with considerable success; for example, protein interactions in biological processes [3,4], electrostatic contributions to solvation [4,5], pK shifts of amino acid residues [6,7], and electrostatically driven protein adsorption [8,9] are some of the diverse phenomena that have been adequately described using continuum electrostatics. The non-electrostatic interactions, on the other hand, have proven to be much more difficult to treat theoretically, but it is precisely these interactions that are expected to dominate the interaction between macromolecular solutes at very small separations ($\ll 10$ Å).

Among the short-range interactions, such as van der Waals and solvation interactions, van der Waals interactions are the best understood. The van der Waals interactions comprise Keesom (dipole–dipole), Debye (dipole-induced dipole) and London dispersion interactions [10]. The most important amongst these, for the usual case of atomic groups with permanent dipole moments less than approximately 1 debye, is the London dispersion interaction, which arises due to the correlated fluctuation of electron density in the interacting atoms.

At the atomic level van der Waals interactions between atoms i and j are usually described by the Lennard–Jones (LJ) form

$$U_{ij} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{1}$$

which captures the attractive nature of the van der Waals interactions via the sixth power dependence, and the very short-range Born repulsion due to the overlap of electron clouds via the 12th power dependence [1,10]. Here $-\epsilon_{ij}$ is the minimum interaction energy and σ_{ij} is the collision diameter. In describing a collection of non-bonded atoms using the LJ form of dispersion interactions, the approximation of pairwise additivity is usually made. Within this approximation the interaction of atom pairs is described by the above equation, but the influence of other atoms on their interaction is neglected.

The other short-range interaction, the solvent-mediated interaction, is less well understood. Such interactions arise due to changes in the structure of the solvent as the separation between the macromolecules is varied. The macromolecules themselves may be driven to associate due to the resulting decrease in the free energy of the solvent (hydrophobic association), or the macromolecular association due to purely dispersion interactions may change the solvent structure, which in turn influences the macromolecular interaction. In any case, the important physical picture that emerges is one of a cooperative inter-

action among numerous solvent molecules and the solute molecules.

In general, if one were to consider the description of solute atoms and solvent atoms in a simulation box, the LJ description would still be assumed to hold for the pairwise interaction between the constituent atoms. As mentioned earlier, however, such a description is infeasible for macromolecular systems as the number of solvent molecules required for describing the system would be orders of magnitude greater than the number of solute atoms. To circumvent this problem, approaches treating the solvent implicitly as opposed to atomistically have evolved. If one considers the insertion of a small apolar solute into water, it is intuitively reasonable to expect the solvation to be proportional to the energy required to create a cavity of an appropriate size in water. This leads to the expectation that the solvation energy should scale with the solvent-accessible surface area. This reasoning has been found to be true in the case of small molecular solutes [11,12]; the molecular basis for this phenomenological model of relating the solvation-free energy to the surface area of the solute can be found in the scaled particle theory [13,14] (see also Hummer and Garde [15] for a development based on perturbation theory).

Based on this accessible surface area (ASA) dependence of solvation, numerous procedures have been developed that have attempted to include the molecular characteristics in describing macromolecular solvation, notable among them being the solvation models of Eisenberg and McLachlan [16] and of Ooi et al. [17]. In either model, the change in free energy F for the process of solvation is approximated as a sum of atomic terms

$$\Delta F = \sum_i \gamma_i A_i \qquad (2)$$

where the areas A_i are the solvent-accessible areas and the γ_i's are the appropriate atomic solvation parameters (ASPs). (We use the symbol F throughout to denote the free energy; the Gibbs and Helmholtz free energies are essentially equivalent in the systems of interest here.) The difference in ASP between the models is due to the different processes considered: Eisenberg and McLachlan consider the transfer of a protein side chain from the protein interior to water, whereas Ooi et al. consider the transfer from air to water. In these formulations, the effects of charges are usually included in the ASP. In cases where the electrostatic interactions are treated separately, the ASPs reflect just the non-polar contributions due to the constituent atoms [18].

If this analysis is extended to binding of solutes in solvent, the binding can be described by the changes in ASA upon binding, scaled by the appropriate ASPs. This procedure can be further refined by partitioning the interaction free energy into a suitably weighted sum of polar and nonpolar contributions, each of which is calculated as a change in ASA scaled by the appropriate ASP [19]. Similar ideas are used in analyzing the energetics of protein folding. There are other measures that can be used to describe macromolecular binding, for example the 'gap index' [20], which is a ratio of the volume enclosed between the two molecules and the interfacial ASA; thus, the lower the gap index, the greater the complementarity of surfaces, and hence the greater the binding strength.

There are several concerns regarding the implementation of such approaches; for example, the size dependence of the solvation free energy of individual macromolecules [21–23], and inconsistencies among the atomic solvation parameter sets themselves [24]. Despite these unresolved issues, procedures describing macromolecular binding based on changes in ASA should be seen mainly in terms of the dependence of the strength of binding on the geometric characteristics of the interface, particularly scaling with quantities that measure complementarity of the interfaces. As early as 1940, Pauling and Delbrück recognized such complementarity as an important determinant of the 'specific attraction of molecules and the enzymatic synthesis of molecules' [25].

Despite its successful use in various contexts, the validity and utility of the ASA approach can be questioned in several respects. Within this formalism the macromolecular interactions are determined purely by the solvent-exposed atoms,

and the contribution of atoms that are just below the surface is largely neglected (we are assuming that macromolecules associate as rigid bodies). As can be seen from the LJ expression, this contribution need not be negligible. The ASA-based approach is also not fully satisfactory in computing the interaction free energy of macromolecules at various separations, as would be required, for example, in a dynamic simulation of macromolecular association. Firstly, there is no steric component in the description to prevent overlap. Secondly, at contact, the ASP should reflect the energy of transfer from water to the protein interior, but before molecular contact in the region where solvent is excluded, the ASP should reflect transfer from water to vacuum. Thus, using the same ASP at all separations may be inappropriate, although some authors have used this approach [26,27]. In such cases, the ASP should be regarded as an empirical model parameter, and no physical process of transfer between phases should be ascribed to it.

In this communication we present an alternative method for calculating interaction energies that also treats solvent effects implicitly, but that addresses some of the issues addressed above. This method is based on the descriptions used for van der Waals interactions, but in a form reparameterized empirically to account for other short-range interactions, such as solvation, as well.

2. Theory and model formulation

As discussed earlier, the van der Waals approach utilizing the LJ formulation (Eq. (1)) requires the atomistic description of all atoms comprising the system. An alternative formulation of dispersion interactions is via the Lifshitz–Hamaker approach [28]. This approach is valid for calculating the dispersion interaction between macromolecular solutes when the medium and the macromolecules can be treated as continua, i.e. at macromolecular separations large enough that the medium is essentially a bulk solvent and the local atomic details of the macromolecule are inconsequential. This (usually attractive) interaction is described in terms of the interaction free energy ΔF_{12} by

$$\Delta F_{12} = -\frac{A_{132}}{\pi^2}\int_{v_1}\int_{v_2}\frac{1}{r_{12}^6}\mathrm{d}v_1\ \mathrm{d}v_2 \tag{3}$$

where r_{12} is the distance between two volume elements dv_1 and dv_2 in the two macromolecules 1 and 2, and A_{132} is the Hamaker constant, which depends on the material properties of the two macromolecules and the intervening medium. The Hamaker constant is most rigorously calculated based on the continuum Lifshitz theory, using dielectric/absorption data for the various media [28]. The calculated protein–vacuum–protein interaction Hamaker constant is 23.4 kT, and the protein–water–protein Hamaker constant is 3.1 kT [29] (k is the Boltzmann constant, and T is the absolute temperature). Thus the role of the solvent is accounted for in Eq. (3) via the value used for the Hamaker constant.

As is evident from Eqs. (1) and (3), the dispersion interactions are extremely short-ranged. However, in Eq. (3) there is no allowance for macromolecular contact (via the Born repulsion term, for example). This leads to the mathematical breakdown of the continuum formulation at very short-range, and at contact ($r_{12} = 0$) the interaction free energy diverges to $-\infty$. Furthermore, the continuum assumption underlying Eq. (3) breaks down for the parts of the two macromolecules that are closest together.

The LJ description does not suffer from the singularities inherent in the Lifshitz–Hamaker approach, but it suffers from the need to include solvent molecules explicitly. Thus, although Eqs. (1) and (3) are intended to describe the same physical behavior, the two treatments are not easily reconciled, but the form of the equations suggests an ad hoc marriage. In this new hybrid method, the LJ formulation is used at short-range, where the Lifshitz–Hamaker approach is expected to fail, and the Lifshitz–Hamaker approach is used at long-range, where the solvent may be adequately described as a continuum.

Completing the description requires specification of both arbitrary and empirical parameters that account for the relative merits and shortcomings of the two contributions. Firstly, the demarcation of short- vs. long-range is based on the

diameter of a water molecule (approx. 3 Å). When the interacting surfaces are closer than approximately 3 Å, the atomic groups comprising the surfaces are essentially interacting across a vacuum. Groups that are separated by (surface-to-surface) distances greater than 3 Å can do so across at least one water molecule. Thus the LJ formulation is used for parts of the molecules closer than 3 Å surface-to-surface separation, and the Lifshitz–Hamaker approach as applied to realistic protein shapes [29], with a Hamaker constant of 3.1 kT, for all other groups (Fig. 1). As the implementation of a surface-to-surface delineation is less efficient computationally than specification of center-to-center separations, the separation can be equivalently specified by a 6-Å center-to-center distance, based on the van der Waals diameters of C, N, and O atoms (in the range of 3.0–3.4 Å). Secondly, the ϵ and σ parameters for the LJ formulation can be taken from any molecular mechanics parameter set; in our treatment, we use the OPLS parameter set [30], as this has been specially developed for peptides and small organic molecules.

Finally, a means to reconcile the continuum Hamaker and the atomistic LJ methods is required. Within the Hamaker approach at long-range, the role of solvent is predominantly that of modulating the protein–protein dispersion interactions, and this is adequately captured by the protein–water–protein Hamaker constant. The LJ interaction at short-range is deficient in not accounting for solvent effects at all. However, we postulate that because of the strong geometric dependence of the short-range interactions, the LJ approach correctly captures the *trends* in the interaction energy, but we include an empirical parameter α to correct the *magnitude* empirically.

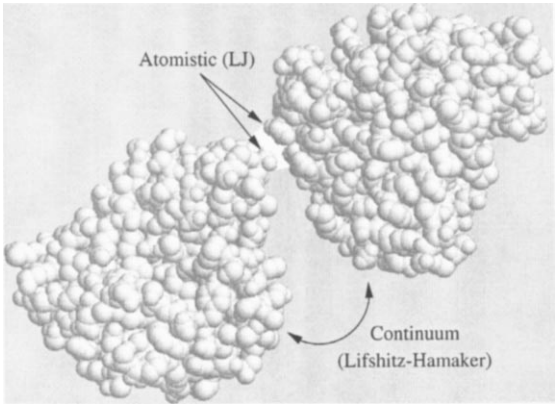

Fig. 1. Schematic showing the atomistic LJ and continuum Lifshitz–Hamaker approaches based on demarcation of short- and long-range between atom pairs. The shaded part represents the region where water is present, and the unshaded part represents the region sterically inaccessible to water.

The final form of our non-electrostatic contribution (ΔF_{ne}) to the potential of mean force, used at all separations, is thus

$$\Delta F_{\mathrm{ne}} = \Delta F_{\mathrm{Ham}} + \alpha \Delta F_{\mathrm{LJ}} \tag{4}$$

where ΔF_{Ham} and ΔF_{LJ} represent the contributions of atom pairs described by the Lifshitz–Hamaker equation and by the LJ equation, respectively.

The electrostatic contribution to the binding free energy, ΔF_{elec}, is modeled by solving the continuum Poisson equation for the potential in the protein interior and the linearized Poisson–Boltzmann equation for the potential in the exterior aqueous solution [3,4,31]. As in the case of the non-electrostatic component, the solvent is described as a continuum with a high dielectric constant ($=80$). In a continuum solvent, all solvation effects attributable to electrostatics alone are implicitly described by this formulation. Specifically, for example, if some charged groups are buried in the interface in the binding process, the desolvation penalty would be reflected as a positive contribution to the electrostatic component of the interaction free energy. For a spherical solute with a net charge at its center, the above formulation leads to the Born solvation model, if the dielectric constant of the solute is taken to be 1.

The final form of the overall calculated binding free energy is then of the form

$$\begin{aligned}\Delta F_{\mathrm{calc}} &= \Delta F_{\mathrm{ne}} + \Delta F_{\mathrm{elec}} + \Delta F_{\mathrm{ent}} \\ &= \alpha \Delta F_{\mathrm{LJ}} + \Delta F_{\mathrm{Ham}} + \Delta F_{\mathrm{elec}} + \Delta F_{\mathrm{ent}}\end{aligned} \tag{5}$$

ΔF_{ent} is the free energy loss due to entropic effects, arising from the reduction in the transla-

tional and rotational degrees of freedom upon binding and the concomitant changes in the vibrational modes and degrees of freedom upon binding. This contribution is difficult to calculate directly, but has been variously estimated as 6.2 kcal/mol [19], 7–15 kcal/mol [32], 20 kcal/mol [33] and 15 kcal/mol [34,35], all based on a 1 M standard state. In this study, we estimate its value empirically.

To summarize, our formulation has two adjustable parameters, α and ΔF_{ent}. Their values are determined by fitting the calculated binding free energy ΔF_{calc} to the corresponding experimental value ΔF_{expt} for representative macromolecular binding pairs.

3. Methods

3.1. Electrostatics

Electrostatic interactions in proteins arise from the ionizable sites on amino acid residues, with the charge at each of these sites being dependent on the solution pH. As mentioned above, these interactions are modeled within the macroscopic Poisson equation for the protein interior and the linearized Poisson–Boltzmann equation — the Debye–Hückel approximation — for the potential in the exterior aqueous electrolyte medium. Charges are assigned to ionizable side chains using intrinsic pK values [36,37], with the charge locations and the molecular geometry obtained using the atomic coordinates from the Protein Data Bank (PDB). The governing electrostatic equations are then solved using the boundary element technique [31,38–40]. The electrical potential calculated in this way is used to compute the electrostatic contribution to the binding free energy [40].

Although it is possible to describe the molecular geometry accurately [8,40], in this work we have adopted a simplified geometrical representation of the protein for reasons of computational efficiency. The protein molecules are represented as spheres of low dielectric constant ($=4$), with the radii scaled according to their molecular volumes. The angular distribution of charges is similar to that in the protein, but the radial positions are scaled such that the distance of each charge below the dielectric interface is similar to that in the actual protein, which for most charges is approximately 1.7 Å. Although this description represents a significant approximation, it is adequate for the present study, where the dominant interactions are short-ranged and non-electrostatic: as is shown later, the electrostatic contributions are typically only approximately 10% of the short-range component. Furthermore, the approximation captures the key trends found from calculations using the full structural details [41].

3.2. Short-range interactions

The short-range interactions are calculated using the full structural details of the proteins from their PDB files. The calculations are based on use of molecular groups [42]; in the LJ calculations, in the context of the OPLS parameterization, the molecular groups are referred to as united atoms [30]. For calculating the Hamaker contribution for groups separated by >6 Å center-to-center separation, the procedure of Roth et al. [29] is followed, with a Hamaker constant of 3.1 kT. Eq. (1) is used directly with the OPLS parameter set for groups separated by <6 Å. For calculating the interaction free energy at various separations, the macromolecules are translated along the line connecting their centers.

Calculations were also performed in which the LJ and Lifshitz–Hamaker approaches were compared directly for interactions in vacuum. For this situation the same method was used for all atom pairs regardless of separation distance. In the case of the Lifshitz–Hamaker calculations, a Hamaker constant of 23.4 kT was used, corresponding to protein–protein interactions in vacuum [29].

4. Results and discussion

The parameter α in Eq. (4) characterizes the weighting of the LJ contribution to the short-range, non-electrostatic interactions to reconcile the LJ description with the Lifshitz–Hamaker one and to account for such effects as solvation. One indication of the value of α comes from a comparison of the predictions of the LJ and Ha-

maker formulations in vacuum. Here both the LJ formulation and the Hamaker formulation with the protein–vacuum–protein Hamaker constant should be applicable at separations where the macromolecule can be characterized as a continuum. A comparison of the vacuum-mediated interactions between crystallographically-related protein pairs (Figs. 2 and 3) shows that for gap distances greater than 2–3 Å, the distance dependence of the LJ-derived potential is quite similar to that of the Hamaker-based potential. This same behavior was seen for the interaction of a F_{ab} fragment with a protein antigen [29]. As seen from the figures, the Hamaker energy tends to $-\infty$ at molecular contact. However, at separations greater than approximately 2 Å, its magnitude is approximately half that of the LJ energy. Thus if the LJ energies are scaled by $1/2$, the two methods agree, suggesting that α should have a value of $1/2$.

This approach does not, of course, account for any solvation effects. A more complete way to determine both α and ΔF_{ent} in Eqs. (4) and (5) is to regress these parameters against experimen-

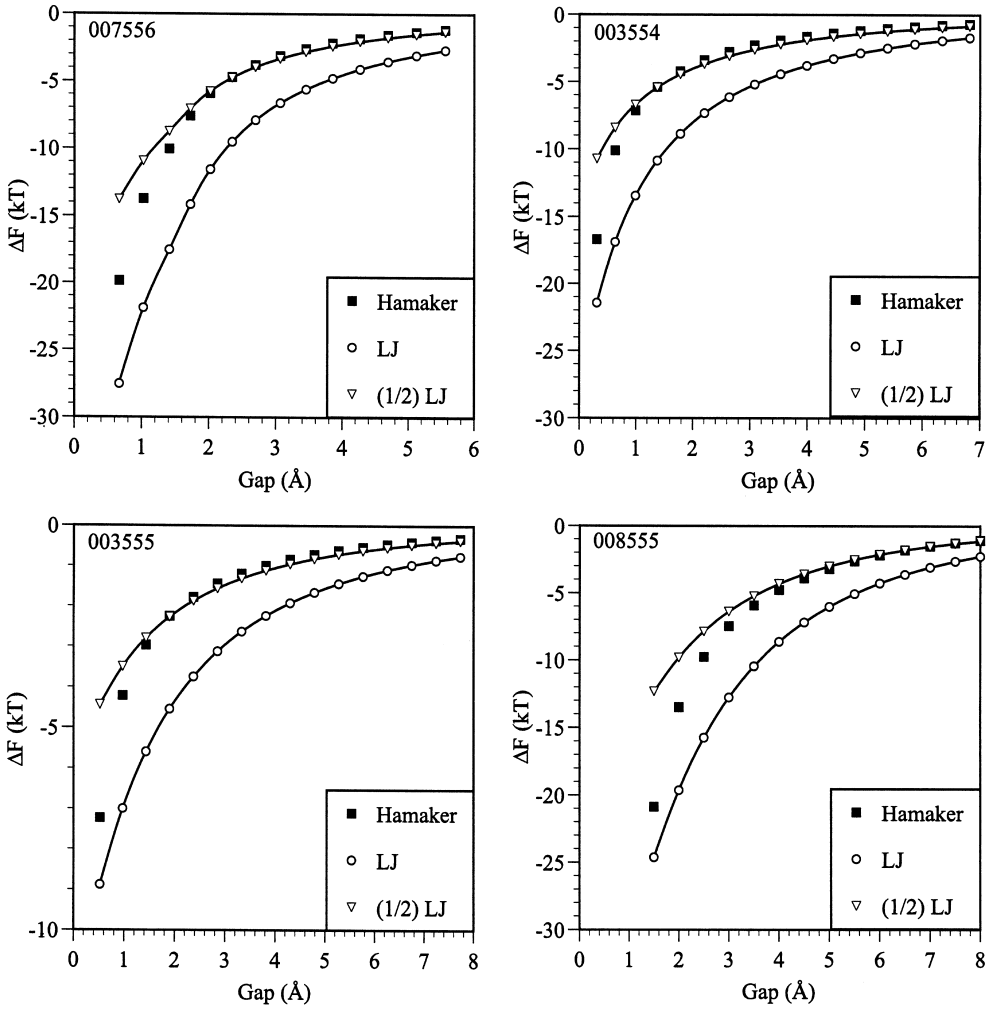

Fig. 2. Interaction of lysozyme molecules in crystallographically-related protein pairs. The gap distance is measured relative to the crystallographic position. The symmetry transformations are given by nnnMMM, where n is the serial number of the symmetry operator, and MMM represents the concatenated cell translations along x, y, z with respect to the base number 555 [52].

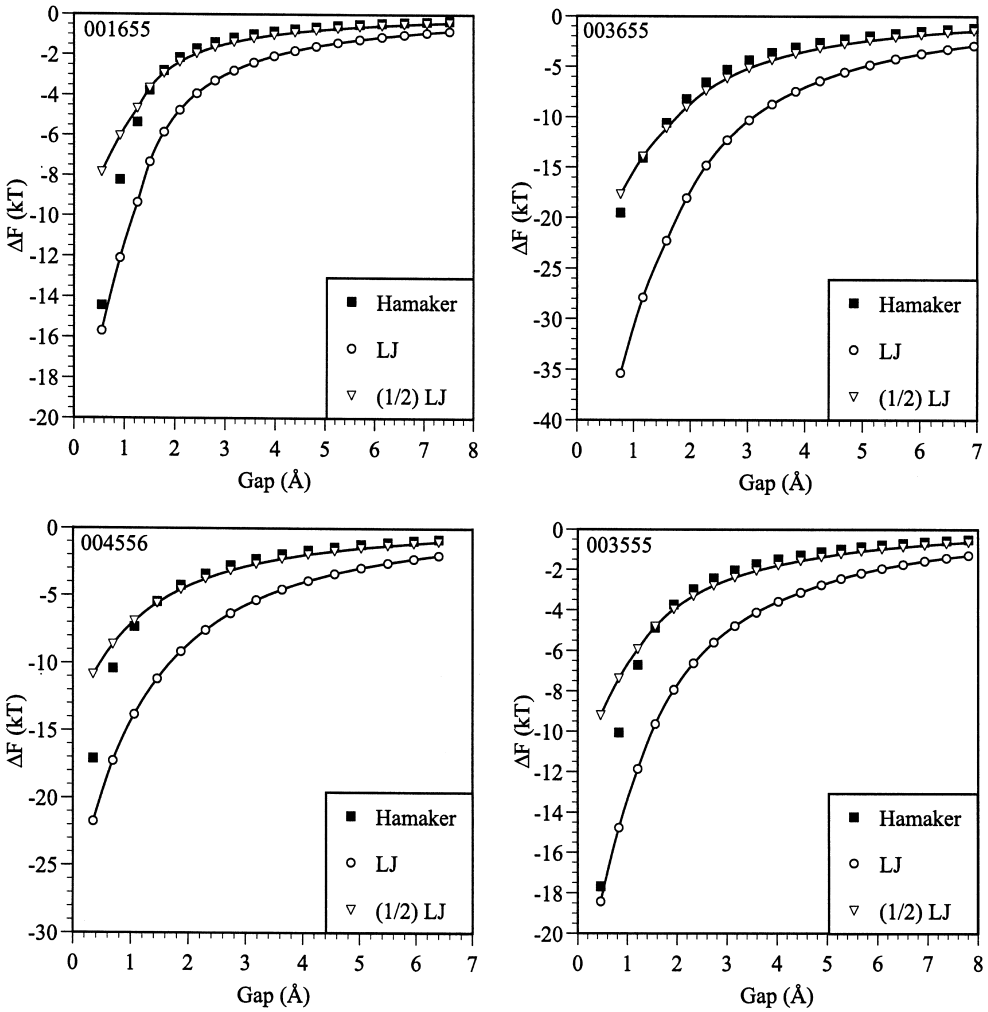

Fig. 3. Interaction of subtilisin molecules in crystallographically-related protein pairs. Additional information as in Fig. 2.

tally observed binding free energies, since all contributions to binding are then accounted for empirically. We have used the same set of systems as used by Horton and Lewis [19] in order to allow comparison between their method and ours for calculating free energies. The calculated contributions to the free energy of binding are shown in Table 1. They show the LJ contribution to be the dominant one, and confirm electrostatic interactions as being relatively small. As discussed earlier, the latter observation justifies the use of economical approximations in the electrostatic calculations.

Because of the scatter in the experimental data (Fig. 4), a value of $1/2$ lies within any reasonable confidence interval for the regressed value of α. We thus set $\alpha = 1/2$, which then leads to a least-squares estimate $\Delta F_{\text{ent}} = 12.8\ kT$ (7.6 kcal/mol) for the second fitting parameter. The regression coefficient, $R^2 = 0.78$, is similar to that for the two-parameter model of Horton and Lewis [19]. Their fit improved when they introduced an additional parameter via the separation of the non-entropic contribution into a polar and a non-polar contribution. The agreement in α based on the regression in Fig. 4 and that inferred

Table 1
Coordinates and binding free energies with $\alpha = 1/2$

PDB file	ΔF_{expt}	ΔF_{Ham}	$\alpha \cdot \Delta F_{LJ}$	ΔF_{elec}	ΔF_{calc}
3HFL	−24.1	−5.4	−33.2	−2.6	−28.4
1FDL	−19.4	−4.4	−30.6	2.6	−19.6
3SSI	−27.2	−5.9	−27.5	0.0	−20.6
3SGB	−25.0	−4.6	−31.8	0.5	−23.1
2PTC	−31.0	−5.4	−40.7	5.0	−28.3
1TPA	−30.0	−5.4	−47.4	4.2	−35.8
2KAI	−21.0	−5.9	−22.2	−3.0	−21.2
2TPI	−10.0	−1.9	−18.9	−1.0	−9.0
2TPI	−31.0	−5.5	−45.0	0.5	−37.2
4CPA	−17.0	−5.0	−28.0	−2.0	−22.3
1CHO	−27.0	−5.6	−37.5	0.0	−30.3

The interaction free energies are in units of kT. All the experimental values are taken from Horton and Lewis [19] except the data for 1FDL [50].
Abbreviations: 3HFL, lysozyme-Hy/Hel-5 F_{ab}; 1FDL, lysozyme-D1.3 F_{ab} [50]; 3SSI, *Streptomyces* subtilisin inhibitor dimer; 3SGB, proteinase B-third domain of turkey ovomucoid inhibitor (OMTK3); 2PTC, trypsin-bovine pancreatic trypsin inhibitor (BPTI); 1TPA, anhydrotrypsin-BPTI; 2KAI, kallikrein A-BPTI; 1CHO, α-chymotrypsin-OMTKY3; 2TPI, trypsinogen(+BPTI) isoleucylvaline (IV); 2TPI, trypsinogen(+IV)-BPTI; 4CPA, carboxypeptidase A-potato carboxypeptidase A inhibitor.

from the vacuum-mediated potential (Figs. 2 and 3) is surprisingly good. Considering that the two estimates were obtained by completely unrelated methods, this agreement appears fortuitous. The (unfavorable) entropic contribution to the association is at the lower end of the range commonly found in the literature (0–40 kcal/mol; see end of Section 2) [43].

The magnitudes of the energies calculated using the new hybrid method and those using the ASA method can be compared by calculating the ratio of the energy in the hybrid method interaction to the corresponding change in the ASA (Table 2; the energy minimum in the hybrid method calculation was used). Except for the lysozyme interaction in pairs 3 and 4, the ratio is fairly constant, varying between 0.013 and 0.029 $kT/Å^2$ (7.7–17.2 cal/mol, $Å^2$). The relative constancy of this value emphasizes that geometric complementarity, an essential component in macromolecular association, can be captured using either a surface area description or a space-filling description such as our hybrid method.

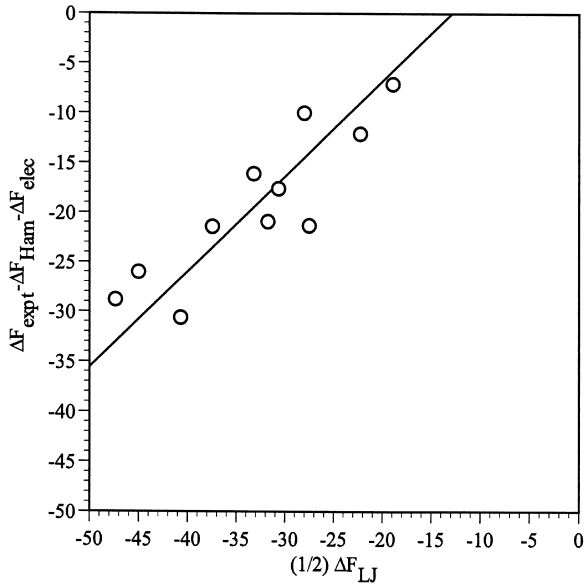

Fig. 4. Correlation of calculated binding free energy with experimental data to allow estimation of ΔF_{ent}. Line has a slope of 1 and an intercept of 12.8 kT.

The more general predictions of the hybrid method (Eq. (4)) and the surface area method can be compared for pairs of molecules; we show results for crystallographically-related protein pairs of subtilisin and lysozyme. Figs. 5–8 show the calculated interaction energy profiles as the molecular pairs are translated relative to their crystallographic positions, defined as a 0 Å gap. For the hybrid method, although the transition between the Hamaker formulation and the LJ formulation is taken to be discontinuous, the interaction energy curves are fairly smooth as a function of distance because of the large number of atom pairs involved. Except for the lysozyme pair 3 interaction (Table 2, Fig. 8), the minimum in the interaction profile always lies close to zero gap distance, i.e. the experimentally determined crystallographic configuration. For the case involving the lysozyme pair 3 interaction, the minimum is at approximately 1.8 Å, and at zero gap the interaction is calculated to be within the Born repulsion region. This is a consequence of overlaps of atoms of the two molecules for this symmetry pair, which may be due to the limited resolution of the atoms comprising this interface.

Table 2
Protein association in crystal complexes

PDB file	Protein pair	Symmetry relation	vdW Minimum (kT)	ΔASA Å^2	Ratio (kT/Å^2)
1LYZ	1	007556	−21.0	1193.4	0.018
	2	003554	−10.2	560.1	0.018
	3	003555	−1.8	211.4	0.009
	4	008555	−6.4	638.3	0.010
SBT	1	001655	−5.0	203.6	0.024
	2	003655	−26.4	904.1	0.029
	3	004556	−13.8	789.6	0.018
	4	003555	−4.6	259.3	0.019
1LIB	1	002454	−15.5	910.5	0.017
8DFR	1	002656	−1.3	91.4	0.014
1DBS	1	002656	−2.2	169.4	0.013
5PEP	1	011555	−15.2	808.4	0.019

Notes. The specified ratio is that of the minimum in the energy profile calculated using the hybrid method, to the change in accessible surface area at the minimum. The accessible surface areas were evaluated using Connolly's program [51]. The proteins are referenced by their four letter PDB code (coordinates for subtilisin (SBT) were provided by Dr Travis Gallagher of NIST, Gaithersburg, MD, USA). The symmetry transformations are given by nnnMMM, where n is the serial number of the symmetry operator, and MMM represents the concatenated cell translations along x, y, z with respect to the base number 555 [52].
Abbreviations: 1LYZ, hen egg-white lysozyme; SBT, subtilisin BPN′ (s88) mutant; 1LIB, lipid binding protein; 8DFR, dihydrofolate reductase; 1DBS, dethiobiotin synthetase; 5PEP, pepsin.

Figs. 5–7 show the hybrid potential well to be quite narrow, being concentrated within approximately 1 Å of the surface. It is interesting to note that the phase behavior of proteins and small colloidal particles has been successfully modeled using a sticky hard sphere potential [44,45], which is characterized by a very short-range attractive potential. The potential based on the hybrid method does suggest that such an approximation is qualitatively appropriate. In Figs. 5–8, the interaction energy profiles calculated using the ASA method are scaled by the ratios given in the last column of Table 2. The ASA curves also vary smoothly, but they differ from those using the hybrid method. First, the apparent interaction reflected is of somewhat longer range. More importantly, the ASA method provides no indication of the location of the crystallographic contact, and the calculated free energy would continue to decrease if the molecules were allowed to overlap. This would be a drawback in Monte Carlo and Brownian dynamics simulations, and indeed, in typical applications of Brownian dynamics, only

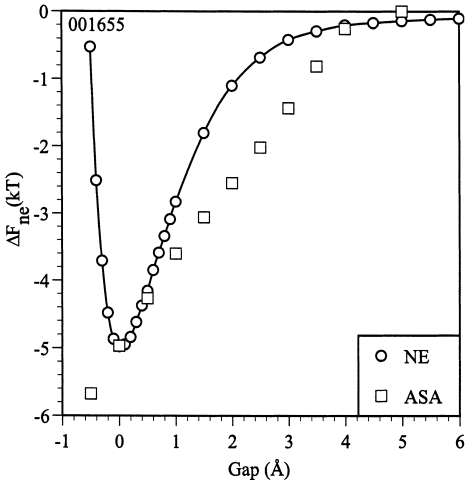

Fig. 5. Non-electrostatic interaction for subtilisin interaction for symmetry-related pair 1. $\Delta F_{ASA} = \gamma \cdot \Delta ASA$; $\gamma = 0.024$ kT/Å^2 (Table 2).

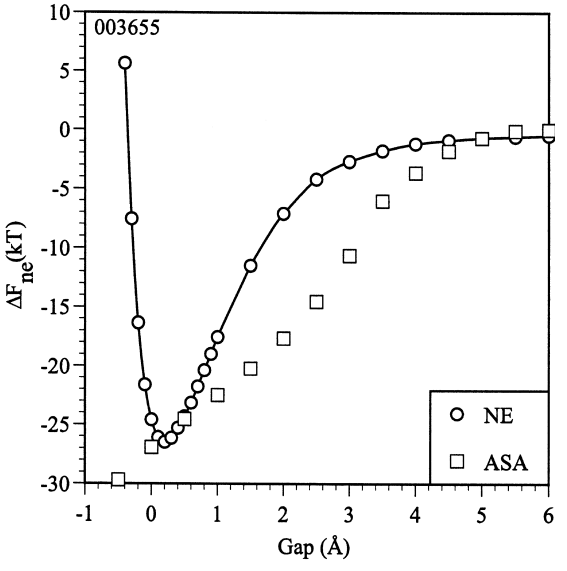

Fig. 6. Non-electrostatic interaction for subtilisin pair 2. $\gamma =$ 0.029 kT/Å^2 (Table 2).

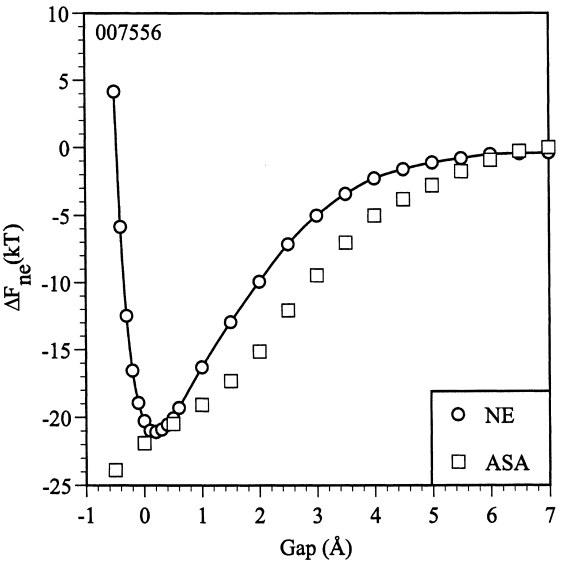

Fig. 7. Non-electrostatic interaction for lysozyme pair 1. $\gamma = 0.018\ kT/Å^2$ (Table 2).

electrostatic interactions and short-range excluded volume repulsion are considered [46,47]. The hybrid method is suitable for accounting for short-range attraction in such applications; this effect is likely to be important because of the Boltzmann weighting effect on the encounter rates [47,48].

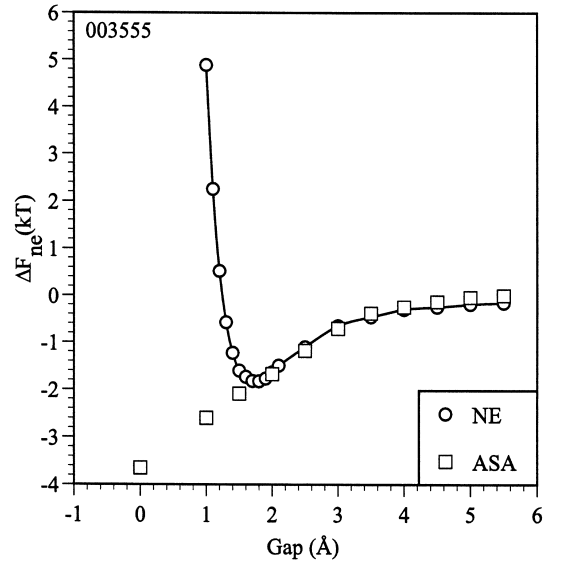

Fig. 8. Non-electrostatic interaction for lysozyme pair 3. $\gamma = 0.009\ kT/Å^2$ (Table 2).

Two approximations are worth mentioning in addition to those inherent in the interaction energy functions used. First, we have assumed that the molecules associate as rigid objects, and have made no allowance for any changes in conformation that might occur in the binding process. In the set of macromolecular complexes that we have chosen and in many others, this is not expected to be a crucial issue [19,49], but conformational changes may be more appreciable in other cases. Including the conformational effects in the calculations is, of course, much more demanding. The second approximation enters in our description of solvation. Whereas the non-electrostatic contributions are included in the short-range interactions, the electrostatic contributions to solvation are accounted for in our solution of the Poisson and Poisson–Boltzmann equations, where the use of a spherical representation for a protein is approximate. However, since this contribution typically represents a minor part of the interaction energy, the error introduced by the approximation is small.

5. Conclusions

A new hybrid method for calculating short-range interactions between protein pairs has been developed. Although the formulation is based on descriptions of dispersion interactions, it does not suffer from the singularity at molecular contact observed for the Hamaker formulation and it is parameterized empirically to account for solvation contributions. It shares with methods based on accessible surface area the capability of describing geometric complementarity, an important factor in macromolecular association. However, it allows molecular approach, including steric repulsion, to be captured more naturally, which gives it greater utility than surface area-based methods in rigid-body simulations of protein solutions.

Acknowledgements

We gratefully acknowledge the support of the National Science Foundation (grants BCS-

9210401 and BES-9510420). We thank Dr Travis Gallagher for making available the coordinates of subtilisin BPN′ (s88) mutant.

References

[1] J.O. Hirschfelder, C.F. Curtiss, R.B. Bird, Molecular Theory of Gases and Liquids, Wiley, New York, 1964.
[2] B. Honig, K. Sharp, A.-S. Yang, Macroscopic models of aqueous solutions: biological and chemical applications. J. Phys. Chem. 97 (1993) 1101–1109.
[3] H. Nakamura, Roles of electrostatic interaction in proteins. Q. Rev. Biophys. 29 (1996) 1–90.
[4] B. Honig, A. Nicholls, Classical electrostatics in biology and chemistry. Science 268 (1995) 1144–1149.
[5] A.A. Rashin, Hydration phenomena, classical electrostatics, and the boundary element method. J. Phys. Chem. 94 (1990) 1725–1733.
[6] D. Bashford, M. Karplus, Electrostatic effects of charge perturbations introduced by metal oxidation in proteins. J. Mol. Biol. 203 (1988) 507–510.
[7] D. Bashford, M. Karplus, pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry 29 (1990) 10219–10225.
[8] B.J. Yoon, A.M. Lenhoff, Computation of the electrostatic interaction energy between a protein and a charged surface. J. Phys. Chem. 96 (1992) 3130–3134.
[9] C.M. Roth, A.M. Lenhoff, Electrostatic and van der Waals contributions to protein adsorption: Computation of equilibrium constants. Langmuir 9 (1993) 962–972.
[10] J. Israelachvili, Intermolecular and Surface Forces, Academic Press, London, 1991.
[11] R.B. Hermann, Theory of hydrophobic bonding: II. The correlation of hydrocarbon solubility in water with solvent cavity surface area. J. Phys. Chem. 76 (1972) 2754–2759.
[12] C. Chothia, Hydrophobic bonding and accessible surface area in proteins. Nature 248 (1974) 338–339.
[13] F. Stillinger, Structure of aqueous solutions of nonpolar solutes from the standpoint of the scaled-particle theory. J. Soln. Chem. 2 (1973) 141–158.
[14] A. Ben-Naim, Statistical thermodynamics for chemists and biochemists, Plenum, New York, 1992.
[15] G. Hummer, S. Garde, Cavity expulsion and weak dewetting of hydrophobic solutes in water. Phys. Rev. Lett. 80 (1998) 4193–4196.
[16] D. Eisenberg, A.D. McLachlan, Solvation energy in protein folding and binding. Nature 319 (1986) 199–203.
[17] T. Ooi, M. Oobatake, G. Nemethy, H.A. Scheraga, Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 3086–3090.
[18] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, Semi-analytical treatment of solvation for molecular mechanics and dynamics. J. Am. Chem. Soc. 112 (1990) 6127–6129.
[19] N. Horton, M. Lewis, Calculation of the free energy of association for protein complexes. Protein Sci. 1 (1992) 169–181.
[20] S. Jones, J.M. Thornton, Principles of protein–protein interactions. Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 13–20.
[21] A. Ben-Naim, R.M. Mazo, Size dependence of solvation free energies of large solutes. J. Phys. Chem. 97 (1993) 10829–10834.
[22] A. Ben-Naim, Solvation: from small to macro molecules. Curr. Opin. Struct. Biol. 4 (1994) 264–268.
[23] A. Ben-Naim, R.M. Mazo, Size dependence of solvation Gibbs energies: a critique and a rebuttal of some recent publications. J. Phys. Chem. B 101 (1997) 11221–11225.
[24] A.H. Juffer, F. Eisenhaber, S.J. Hubbard, D. Walther, P. Argos, Comparison of atomic solvation parameter sets: Applicability and limitations in protein folding and binding. Protein Sci. 4 (1995) 2499–2509.
[25] L. Pauling, M. Delbrück, The nature of intermolecular forces operative in biological processes. Science 92 (1940) 77–79.
[26] Y. Fukunishi, M. Suzuki, Reproduction of the potential of mean force by a modified solvent-accessible surface method. J. Phys. Chem. 100 (1996) 5634–5636.
[27] Y. Fukunishi, M. Suzuki, Potential of mean force calculation of solute macromolecules in water by a modified solvent-accessible surface method. J. Comp. Chem. 18 (1997) 1656–1663.
[28] D.B. Hough, L.R. White, The calculation of Hamaker constants from Lifshitz theory with applications to wetting phenomena. Adv. Colloid Interface Sci. 14 (1980) 3–41.
[29] C.M. Roth, B.L. Neal, A.M. Lenhoff, Van der Waals interactions involving proteins. Biophys. J. 70 (1996) 977–987.
[30] W.L. Jorgensen, J. Tirado-Rives, The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. 110 (1988) 1657–1666.
[31] B.J. Yoon, A.M. Lenhoff, A boundary element method for molecular electrostatics with electrolyte effects. J. Comp. Chem. 11 (1990) 1080–1086.
[32] H.P. Erickson, D. Pantaloni, The role of subunit entropy in cooperative assembly: Nucleation of microtubules and other two-dimensional polymers. Biophys. J. 34 (1981) 293–309.
[33] B. Tidor, M. Karplus, The contribution of vibrational entropy to molecular association: the dimerization of insulin. J. Mol. Biol. 238 (1994) 405–414.
[34] A.V. Finkelstein, J. Janin, The price of lost freedom: entropy of bimolecular complex formation. Protein Eng. 3 (1989) 1–3.
[35] J. Janin, Elusive affinities. Protein Struct. Funct. Genet. 21 (1995) 30–39.
[36] G. Zubay, Biochemistry, Addison-Wesley, Reading, 1983.
[37] L. Stryer, Biochemistry, W.H. Freeman, New York, 1988.

[38] R.J. Zauhar, R.S. Morgan, The rigorous computation of the molecular electric potential. J. Comp. Chem. 9 (1988) 171–187.
[39] A.H. Juffer, E.F.F. Botta, B.A.M. van Keulen, A. van der Ploeg, H.J.C. Berendsen, The electric potential of a macromolecule in a solvent: a fundamental approach. J. Comp. Phys. 97 (1991) 144–171.
[40] H.-X. Zhou, Boundary element solution of macromolecular electrostatics: interaction energy of two proteins. Biophys. J. 65 (1993) 955–963.
[41] B.L. Neal, D. Asthagiri, A.M. Lenhoff, Molecular origins of osmotic second virial coefficients of proteins. Biophys. J. 75 (1998) 2469–2477.
[42] A. Bondi, Van der Waals volumes and radii. J. Phys. Chem. 68 (1964) 441–451.
[43] G.P. Brady, K.A. Sharp, Entropy in protein folding and in protein–protein interactions. Curr. Opin. Struct. Biol. 7 (1997) 215–221.
[44] D. Rosenbaum, P.C. Zamora, C.F. Zukoski, Phase behavior of small attractive colloidal particles. Phys. Rev. Lett. 76 (1996) 150–153.
[45] D. Rosenbaum, C.F. Zukoski, Protein interactions and crystallization. J. Cryst. Grow. 169 (1996) 752–758.
[46] R.R. Gabdoulline, R.C. Wade, Simulation of the diffusional association of barnase and barstar. Biophys. J. 72 (1997) 1917–1929.
[47] R.C. Wade, R.R. Gabdoulline, B.A. Luty, Species dependence of enzyme-substrate encounter rates for triose phosphate isomerases. Protein Struct. Funct. Genet. 31 (1998) 406–416.
[48] H.-X. Zhou, Brownian dynamics study of the influence of electrostatic interaction and diffusion on protein–protein kinetics. Biophys. J. 64 (1993) 1711–1726.
[49] D.R. Davies, G.H. Cohen, Interactions of protein antigens with antibodies. Proc. Natl. Acad. Sci. USA 93 (1996) 7–12.
[50] J. Novotny, R.E. Bruccoleri, F.A. Saul, On the attribution of binding energy in antigen–antibody complexes McPC 603, D1.3, and HyHel-5. Biochemistry 28 (1989) 4735–4749.
[51] M. Connolly, Analytical molecular surface calculation. J. Appl. Cryst. 16 (1983) 548–558.
[52] Brookhaven National Laboratory, Protein Data Bank contents guide, 2nd ed., 1996.