ELSEVIER

Biophysical Chemistry 75 (1998) 115–128

Biophysical Chemistry

Mean-field analysis of protein–protein interactions

Mark A. Olson*

Molecular Modeling Laboratory, and Department of Cell Biology and Biochemistry, USAMRIID, 1425 Porter Street, Frederick, MD 21702-5011, USA

Received 1 May 1998; received in revised form 18 August 1998; accepted 18 August 1998

Abstract

Calculations were performed on the D1.3–E5.2 antibody–antibody complex estimating the binding affinities of the wild-type and 16 alanine substitutions. Analyzed were structural models of the interfacial region containing a zinc ion and crystallographic waters. A continuum approach was used to evaluate the electrostatic free energies and the hydrophobic effect was calculated by employing a buried molecular surface area relationship. Estimates of the absolute binding affinity reproduced the experimental value within the uncertainty of assessing entropic and strain energy contributions. The best correlation for mutants with experimental data was achieved when the hydrophilicity of created cavities were considered, and yielded a correlation coefficient of 0.7 and an average error of ± 1.4 kcal/mol. Empirically fitting the free energy function produced a smaller error of ± 1.0 kcal/mol. Depending on the electrical potential and electrostatic reorganization, scaling the 'protein dielectric constant' to ~ 10 may improve the accuracy of continuum models for evaluating amino acid substitutions.

Keywords: Protein–protein interaction; Continuum model; Binding free energy; Electrostatics; Hydrophobic effect; Molecular association

1. Introduction

The prediction of biological specificity underlying the molecular association of proteins or nucleic acids with other macromolecules is a problem of long-standing interest to theoretical structural biology. A common thermodynamic measure of specificity is the Gibbs free energy of complex formation. Current theoretical methods of estimating free energies can be separated into two general approaches, explicit or implicit. Explicit models are microscopic treatments of the macromolecules and solvent molecules by using an atomic force-field description, and are centered on sampling conformational space by employing either molecular dynamics or Monte Carlo simulation techniques. Implicit or continuum models are semi-microscopic treatments of the macro-

* Tel.: +1 301 6194236; fax: +1 301 6192348; e-mail: molson@ncifcrf.gov

0301-4622/98/$ - see front matter
PII S0301-4622(98)00201-4

molecular local environment, while the solvent is modeled by using bulk physical properties. In the framework of continuum models, electrostatic contributions to binding free energies are commonly obtained from finite-difference solutions to mean-field equations calculated from the Poisson–Boltzmann formula [1,2]. Hydrophobic effects are modeled by employing a free-energy relationship derived from the interfacial surface area.

While explicit methods developed from free-energy perturbation theory [3] have yielded impressive results on estimating relative binding free energies for small protein–ligand complexes [4], they are typically for large protein assemblies computationally prohibitive. On the other hand, continuum methods are computationally tractable, requiring only modest computer resources, and are thought to provide an alternative approach for predicting binding affinities of macromolecular assemblies. A recent application demonstrating the potential of continuum methods is the work of Novotny et al. [5], where they estimated changes in binding free energies of 10 lysozyme mutants of the HyHEL-10 antibody–lysozyme complex. These workers found a positive, yet only moderate, degree of correlation between calculated free energy values and the experimentally observed values. To achieve a significantly better agreement with experiments, they linearly fitted their free energy function and obtained a set of mean-field parameters scaling the energetic contributions arising from electrostatics, hydrophobic and entropic terms. Possible improvements in predictive capabilities of continuum methods by using these empirically fitted parameters suggested by Novotny et al. [5] depend on whether they can be applied consistently to other protein complexes. Equally challenging for continuum methods is the determination of absolute binding free energies of macromolecular assemblies, as recently illustrated in the works of Jackson and Sternberg [6] and Honig and co-workers [7].

This paper further examines the general applicability of continuum approaches applied to estimating interactions of macromolecules in solution and the effects of amino acid substitutions. Additional modeling benchmarks are clearly needed to gauge the overall accuracy of these theoretical techniques prior to their routine use in a priori predictions of biological specificities. Several issues remain to be clarified in the use of continuum models, in particular, the 'uniqueness' of mean-field parameters for scaling the free energy contributions and the effect of incorporating interfacial ions and crystallographic waters into implicit macromolecular descriptions. Calculations are presented here analyzing the macromolecular complex of the D1.3 monoclonal antibody bound with the anti-lysozyme antibody E5.2 [8]. The wild-type and relative binding affinities of 16 alanine substitutions were evaluated by using the X-ray crystal structure of the D1.3–E5.2 complex. This highly resolved structure contains an interfacial zinc ion and many crystal water molecules, and is well suited for theoretical studies. Comparisons with experimentally determined affinities are presented along with discussions illustrating several shortcomings of continuum models applied to substitutions resulting in the creation of large hydrophobic or non-hydrated hydrophilic cavities.

2. Theory and methods

2.1. Formalism

The thermodynamic analysis for evaluating binding affinities is based on the calculational framework that the free energy of molecular association can be decomposed into independent contributions arising from electrostatic (ΔG_{ele}) and non-polar (ΔG_{np}) interactions, a strain-energy (ΔG_{strain}) plus configurational entropic ($T\Delta S$) contributions [1,5–7]:

$$\Delta G_{bind} = \Delta G_{ele} + \Delta G_{np} + \Delta G_{strain} - T\Delta S \qquad (1)$$

where $T\Delta S \approx T\Delta S_{sc} + T\Delta S_{t,r}$, combining the terms of side-chain torsional and translational–rotational, respectively, and T is the absolute temperature. The electrostatic interaction compo-

nent can be expressed as a sum of three terms [1,5–7] given by

$$\Delta G_{ele} = \Delta\Delta G^{D}_{sol} + \Delta\Delta G^{E}_{sol} + \Delta G^{D-E}_{int} \quad (2)$$

where the superscripts D and E denote antibodies D1.3 and E5.2, respectively; the term $\Delta\Delta G_{sol}$ corresponds to the loss of solute–solvent interaction energy through the partial desolvation of the electrostatically charged D and E on binding; and ΔG_{int} is the electrostatic interaction energy between the two proteins in solvent water. The electrostatic process of association is schematically depicted in Fig. 1 [1,6], and involves the removal of the high dielectric solvent (ε_s) from the space occupied by the binding partner and replacing it with the low dielectric medium of the protein (ε_p). The net solvation loss for each protein is evaluated from the solvation interaction free energies [1]

$$\Delta\Delta G^{D}_{sol} = \Delta G^{D-E}_{sol} - \Delta G^{D}_{sol} \quad (3a)$$

$$\Delta\Delta G^{E}_{sol} = \Delta G^{E-D}_{sol} - \Delta G^{E}_{sol} \quad (3b)$$

where each reaction field is determined by bringing the solvent boundary from infinity to the solvent accessible surface of a given protein, i.e.

$$\Delta G_{sol} = G_{sol}(\varepsilon_p, \varepsilon_s) - G_{sol}(\varepsilon_p, \varepsilon_p) \quad (4)$$

The reaction fields ΔG^{D-E}_{sol} and ΔG^{E-D}_{sol} for the associated states can be determined by setting the atomic charges of the binding partner to zero (Fig. 1). The ΔG_{int} term of Eq. (2) is evaluated from the pair-wise Coulomb interaction

$$\Delta G^{D-E}_{int} = \int \rho^{D}(\mathbf{r})\varphi^{E}(\mathbf{r})\mathrm{d}\nu \quad (5)$$

where φ is the potential generated by the protein E interacting with the bound protein D of charge density ρ in volume ν, and $\mathbf{r}$ is the position vector. From a continuum representation, the electrostatic free energies in Eqs. (3a),(3b)–(5) can be obtained from the electrostatic field potential calculated by the linearized Poisson–Boltzmann formula

$$\nabla[\varepsilon(\mathbf{r})\nabla\varphi(\mathbf{r})] - \varepsilon(\mathbf{r})\kappa(\mathbf{r})^2\varphi(\mathbf{r}) + 4\pi\rho(\mathbf{r}) = 0 \quad (6)$$

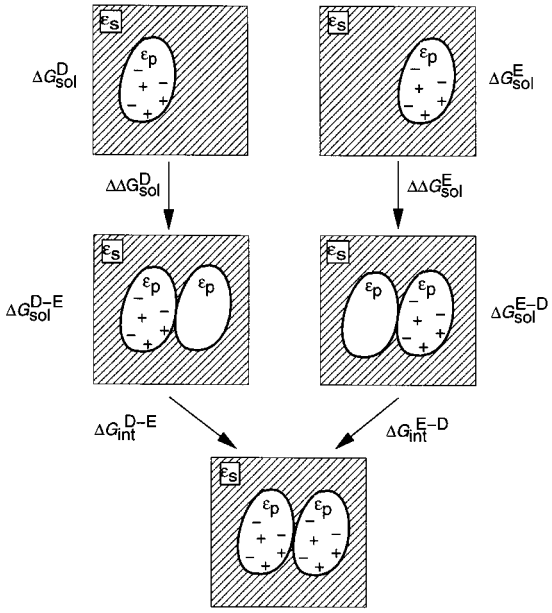

Fig. 1. Electrostatic process of molecular association [1,6] of the D1.3 and E5.2 antibodies. The total electrostatic solvation free energy is given by $\Delta\Delta G_{sol}$ and is the difference between reaction field energies, ΔG_{sol}, calculated for bound and unbound proteins. The association process represents the removal of the high dielectric medium of the solvent (ϵ_s) and is replaced with the low dielectric medium of the protein (ϵ_p). Electrostatic interactions between the two proteins embedded in the solvent medium is ΔG_{int}.

where κ is a function of the Debye length and ionic strength of the bulk solution.

The non-polar contribution of Eq. (1) is partitioned as

$$\Delta G_{np} = \Delta G_{vdW} + \Delta G_{cav} \quad (7)$$

with free energies arising from van der Waals (ΔG_{vdW}) interactions and the hydrophobic effect (solvent cavitation free energy, ΔG_{cav}). The contribution of the change in van der Waals interactions at the interface of the complex can be neglected by invoking an enthalpy–entropy compensation phenomenon argument [6,9]. Simply stated, the change in dispersion energy between atoms making interactions at the interface of the complex and atoms contacting water in the dissociated state is equal to the loss in side-chain

conformational entropy upon binding. This cancellation is given by

$$\Delta H_{\mathrm{disp}}^{\mathrm{D-E}} - \left(\Delta H_{\mathrm{disp}}^{\mathrm{D-wtr}} + \Delta H_{\mathrm{disp}}^{\mathrm{E-wtr}}\right) - T\Delta S_{\mathrm{sc}} = 0 \quad (8)$$

where the quantities ΔH_{disp} are the dispersion energies and the terms in parenthesis are the interactions at the interfacial surface of the complex with water in the dissociated state. Although the assertion of Eq. (8) is physically sound, its accuracy is difficult to gauge by simulation techniques. An alternative is to assume that the differences in van der Waals forces between the bound and free states are likely to be small and, moreover, their differentiation of the wild-type and mutants is weak and embedded in the solvent cavitation free energy (see further below). In combination with this alternative approach is to estimate explicitly the changes in the side-chain conformational freedom upon association by using either an empirical scale [10] or the formula [5]

$$T\Delta S_{\mathrm{sc}} = RT\ \ln\left(\frac{N_{\mathrm{bound}}}{N_{\mathrm{free}}}\right) \quad (9)$$

where R is the universal gas constant, N_{bound} and N_{free} are the number of side-chain torsions available in the bound and free state, respectively. For the work reported here estimating the absolute free energy of molecular association for the wild-type complex, Eq. (8) will be applied. For comparison purposes with the work of Novotny et al. [5], the entropic term using Eq. (9) will be implemented for estimating changes in binding free energies due to the alanine substitutions. Clearly Eq. (8) is equally valid for the mutants, however, the approach of Eq. (9) allows an evaluation of the importance of explicitly incorporating the side-chain entropic term and, as described later in the paper, linear-scaling of the continuum model permits a readjustment in the magnitude of this contribution to the net free energy change.

For the cavitation free energy of Eq. (7), the following linear relationship is commonly used

$$\Delta G_{\mathrm{cav}} = \gamma \Delta A \quad (10)$$

where on complex formation ΔA is the change in either the molecular surface area or the solvent accessible surface area, and γ is the surface tension. It has been proposed by Jackson and Sternberg [6,11] that the molecular surface is a better descriptor of the hydrophobic effect in continuum modeling of protein–protein association. Thermodynamic arguments suggest that, unlike the solvent accessible surface, the molecular surface area yields a free energy independent of shape. Thus for calculations of the D1.3–E5.2 antibody complex, the molecular surface area was chosen as the dependent variable and the corresponding surface tension was set at the 'macroscopic' value of $\gamma = 69$ cal/mol/Å^2 [11]. Discussion of surface tension values in the microscopic and macroscopic regimes can be found in the papers of Honig and co-workers [9,12] and Jackson and Sternberg [11].

In evaluating Eq. (1), two additional approximations are generally implemented in implicit models. The first assumes from theoretical estimates that the rotational and translational entropy loss on complex formation is in the range of 7–15 kcal/mol [13–16]. This ansatz is typified in the recent works of Jackson and Sternberg [6] and Froloff et al. [7]. A further approximation is to neglect $\Delta G_{\mathrm{strain}}$. This unfavorable energetic term can, in principle, be estimated from a conformational relaxation search by using molecular dynamics or Monte Carlo methods of the unbound and bound protein structures. These calculations are, however, problematic in conformational sampling as well as eliminating crystallographic artifacts in obtaining reliable strain energies and solvation free energies resulting from the conformational changes during association.

2.2. Computational approach

The tertiary structure of the complex between the Fv fragments of D1.3 and E5.2 (Brookhaven Data Base file 1dvf) employed in the energy calculations is illustrated in Fig. 2. Placement of the

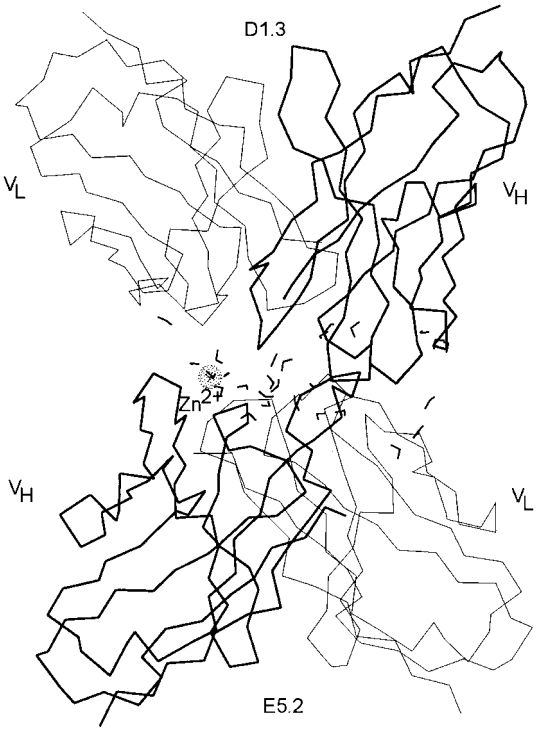

Fig. 2. Tertiary structure of the D1.3–E5.2 complex. Shown are the α-carbon backbones of the variable light (V_L) chains depicted as thin lines and the heavy (V_H) chains as thick lines. The zinc ion and crystal water molecules near the protein–protein interface are also illustrated.

hydrogen atom coordinates was optimized by a combination of molecular dynamics simulation and energy minimization using the CFF91 molecular mechanics force field [17]. Heavy atoms of the complex were fixed at their initial crystallographic positions during optimization. Molecular dynamics was first performed for 1000 iterations at a temperature of 300 K, followed by 500 cycles of energy minimization using a steepest descent algorithm. Force-field cutoffs were set at 5 Å and a distance-dependent dielectric constant of $\varepsilon = 4r$ was applied.

In calculating the electrostatic free energies implementing the above formalism, a finite-difference Poisson–Boltzmann method [18,19] was employed by using the program DelPhi [20] (Molecular Simulations, Inc.; release 95.0). The electrostatic potentials for each protein were calculated by using the solvent accessible surfaces to define regions of low dielectric medium (modeled as $\varepsilon_p = 4$) embedded in a high dielectric solvent water ($\varepsilon_s = 80$) of ionic strength set at 0.145 M. Three model systems were investigated, each representing different local environment descriptions of the binding interface of the complex embedded in the solvent continuum. The models included (1) interfacial region modeled without explicit zinc ion or crystallographic waters, (2) zinc ion explicitly included without crystallographic waters, and (3) no zinc ion but crystallographic waters.

The PARSE parameter set was used to represent the charges and atomic radii [21]. This included modeling the explicit crystallographic waters with charges of -0.980 and 0.490 for oxygen and hydrogens, respectively, and atomic radii of 1.4 and 1.0 Å. Parameters for the interfacial zinc ion were estimated and set at a charge of $+2$ and radius of 1.4 Å. DelPhi calculations were carried out on a cubic grid of resolution 1.82 grid points/Å with a border spacing of 10.0 Å. Boundary conditions were set at full Coulombic for all calculations. Dummy atoms were implemented to retain an identical scale and position on the grid for the complex and each individual protein.

The cavitation free energy was determined by the Connolly algorithm [22] with the solvent probe radius set at 1.4 Å. Atomic radii (in Å) were set at (depending on connecting atoms): C (1 H) = 1.85; C (2 H) = 1.90; C (3 H) = 1.95; C (aromatic or carbonyl) = 1.80; C (1 H, aromatic) = 1.90; O (carbonyl) = 1.60; O (hydroxyl) = 1.70; N (0 or 1 H) = 1.65; N (2 H) = 1.70; N (3 H) = 1.75; S = 1.90; and H = 1.20.

For modeling alanine substitutions in the D1.3–E5.2 complex, each mutant was constructed from the wild-type crystal structure. No energy minimization or molecular dynamics was performed on the mutant structures. Calculations of the electrostatic and cavitation free energies were performed similar to that of the wild-type complex. The side-chain entropic contribution was estimated by solving Eq. (9) with a procedure comparable to that suggested by Novotny et al. [5].

3. Results and discussion

In this section, the absolute binding free energy for the wild-type complex is presented for the model calculations and a comparison is made with the experimentally determined value. This is followed by a similar analysis of calculations on 16 single-site mutants.

3.1. Wild-type binding affinity

Table 1 shows the results for the calculated total free energy of molecular association (denoted as ΔG_{calc}), partitioned according to solvation effects, protein–protein interactions, and the hydrophobic effect. The term $\Delta\Delta G_{sol}$ represents the net solvation contribution to the binding process from both proteins. Free energy estimates are reported for $T\Delta S$.

The calculations reveal that the simplest model of 'absorbing' the explicit zinc ion or crystallographic waters into the solvent continuum provides a thermodynamic description that yields a ΔG_{calc} showing a physically reasonable estimate in the range of -16 to -24 kcal/mol. The corresponding experimental value (ΔG_{expt}) is -10.9 kcal/mol. It is further reassuring that when a ΔG_{strain} contribution is added to ΔG_{calc}, the predicted absolute free energy would become more positive and closer to ΔG_{expt}. Depending on the plasticity of both proteins, ΔG_{strain} will be less than the free energy of unfolding, which is typically 5–20 kcal/mol for globular proteins [23,24]. For calculations incorporating a zinc ion modeled with an atomic charge of $+2$, ΔG_{calc} is within the range of -8 to -16 kcal/mol and predicts a better agreement with ΔG_{expt} prior to the addition of the unfavorable ΔG_{strain}. The difference between the two calculations is, as expected, due to the electrical potential of the binding interface, with the zinc ion contributing greater favorable interaction energy, while at the same time, offset in part by a larger desolvation cost. Note that while the PARSE parameter set was developed strictly to reproduce the transfer free energies of amino acid side-chain and peptide backbone analogs [21], calculations without a detailed parametrization study of incorporating additional moieties are still capable of providing a first-order accurate approach to estimating the overall effect of electrostatics on the binding affinity. Although not a goal of the present study, a better determination of the charge of the zinc ion complexed with side-chain ligands can be achieved through ab initio quantum mechanical calculations [25]. Nonetheless, the error in ΔG_{calc} for either model system is comparable to errors reported by other studies of various protein–protein and protein–ligand complexes employing continuum models [6,7,26].

In contrast with the first two model systems, the calculation of incorporating crystallographic waters shows a positive value for ΔG_{calc}. This unfavorable free energy arises directly from an increase in the solvation penalty of both proteins. To bring the ΔG_{ele} cost down in an attempt to obtain a more realistic ΔG_{calc} in better agreement with ΔG_{expt} requires an increase in the so-called 'protein dielectric constant' from a value of $\varepsilon_p = 4$ to approx. 7; a result that, in effect, reduces the overestimation of charge–charge interactions plus the desolvation penalty of individual polar atoms. As discussed by Warshel and co-workers [27,28], and described further below,

Table 1
Comparison between the absolute free energy of association calculated using three structural models and the experimentally determined value

Model[a]	$\Delta\Delta G_{sol}$	ΔG_{int}	ΔG_{ele}	ΔG_{cav}	$T\Delta S$	ΔG_{calc}	ΔG_{expt}
1	54.2	−24.6	29.6	−60.5	~11 ± 4	−20 ± 4	−10.9
2	82.0	−44.1	37.9	−61.3	~11 ± 4	−12 ± 4	−10.9
3	149.3	−86.7	62.6	−60.5	~11 ± 4	+13 ± 4	−10.9

[a] Models are designated as (1) protein–protein interface without zinc ion or crystallographic waters; (2) interface with zinc ion but without waters; and (3) interface containing only waters. Estimates are reported for $T\Delta S$. Energy units are kcal/mol.

ε_p is merely an adjustable parameter that represents the contributions that are not treated explicitly (i.e. gas phase with $\varepsilon_p = 1$) in continuum models, rather than the true physical dielectric constant of the protein and local environment.

It is interesting to note that while the calculations yielding realistic values for ΔG_{calc} in Table 1 show the electrostatic contribution opposes binding ($\Delta G_{ele} > 0$), the hydrophobic effect compensates ($\Delta G_{cav} < 0$), producing a favorable complex. This thermodynamic offset predicted of macromolecular complexes by continuum models [6,7,26] is also fundamentally similar to the results of continuum analyses of hydrophilic interactions involved in protein folding [29,30]. Despite the importance of the hydrophobic effect in the molecular association process, hydrophobicity contributes little to molecular specificity. Rather in general, specificity among different assembly configurations occurs on an electrostatic free energy surface that is unfavorable, and the configurational sampling that takes place in nature searches this surface for the net, minimum electrostatic penalty.

3.2. Free-energy changes for mutants without zinc ion or crystallographic waters

Table 2 presents the relative free energy changes ($\Delta\Delta G$) determined for 16 single-residue mutants of the D1.3–E5.2 complex with alanine substitutions in the both the variable light- and heavy-chains of the D1.3 antibody. The location of each side-chain mutation within the native crystal structure is illustrated in Fig. 3. Results for $\Delta\Delta G_{calc}$ are presented first without either the zinc ion or crystallographic waters, and discussion of these two additional models will be deferred until later in the following section. The free energies are defined similarly as above, with the exception of the entropic term, which is calculated only for changes upon substitution in the side-chain torsional freedom ($T\Delta S_{sc}$). Also reported in Table 2 is the absolute difference between $\Delta\Delta G_{calc}$ and the experimental $\Delta\Delta G_{expt}$.

The comparison of predicted $\Delta\Delta G$s with experiment shows several values that are within 1.0 kcal/mol (Y50A, W92A, T30A, D58A and D100A), while the remaining mutants display large errors. The overall average absolute error of the calculations is ± 2.2 kcal/mol, which is less than the average error of ± 3.4 kcal/mol achieved by Novotny et al. [5] in their study of 10 lysozyme mutants of the HyHEL-10 complex. The difference of ~ 1 kcal/mol between the two set of calculations is likely due to substitutions of side chains predominately other than alanine in the lysozyme complex. Presented in Fig. 4a is a scatter plot of $\Delta\Delta G_{calc}$ vs. $\Delta\Delta G_{expt}$, showing a correlation coefficient of 0.5.

Many of the large errors observed in $\Delta\Delta G_{calc}$ can be reconciled with $\Delta\Delta G_{expt}$ if one considers the hydrophilicity of the cavities created from the alanine substitutions. The question becomes the following: are these cavities from mutations hydrated, and if they are not, can a simple consistent model be developed that reduces the overestimation of desolvation energetics? A number of factors are thought to stabilize buried water molecules, including polarization, entropy, and hydrogen bonding. Clearly, in a mean-field approach these factors are absorbed into the implicit description of the protein/solvent dielectric boundary. A simple continuum model to test the hydrophilicity of the cavities is to construct a constant dielectric boundary for mutants equivalent to the wild-type protein. This idea is illustrated in Fig. 5 for the wild-type and the light-chain mutant Y32A. The depicted mesh shown (Fig. 5a) for the wild-type protein (colored white) represents the molecular surface (or alternatively, the solvent accessible surface) and hence the protein/solvent dielectric boundary. Whereas for Y32A (Fig. 5b), the mesh represents only the molecular surface for calculating the hydrophobic effect, but the dielectric boundary separating ε_p and ε_s for electrostatics remains the same as the wild-type structure. In effect this fills the empty cavity with ε_p, rather than ε_s.

Results for $\Delta\Delta G_{calc}$ from this 'constant dielectric (const-ε) boundary' model are listed in Table 2 and clearly show an improvement in several mutants as compared with the solvent accessible surface dielectric (SAS-ε) boundary calculations. Overestimates in $\Delta\Delta G_{sol}$ are significantly reduced for light-chain mutants Y32A, Y49A, Y50A, and

Table 2
Relative free energy changes for D1.3 mutants[a]

Mutant	$\Delta\Delta G_{sol}$	$\Delta\Delta G_{int}$	$\Delta\Delta G_{cav}$	$T\Delta S_{sc}$	$\Delta\Delta G_{calc}$	$\Delta\Delta G_{expt}$	Difference
Solvent accessible surface dielectric boundary							
Light chain							
H30A	−0.2	−0.1	0.1	0.0	−0.2	1.7	1.9
Y32A	−5.1	1.6	2.3	0.0	−1.2	2.0	3.2
Y49A	−1.3	−0.6	1.9	0.0	0.0	1.7	1.7
Y50A	−2.0	0.9	0.9	0.0	−0.2	0.7	0.9
W92A	−2.7	1.0	2.9	0.0	1.2	0.3	0.9
S93A	0.0	−0.1	0.0	0.0	−0.1	1.2	1.3
Heavy chain							
T30A	−0.3	0.0	0.2	0.0	−0.1	0.9	1.0
Y32A	−0.7	−0.1	0.0	0.0	−0.8	1.8	2.6
W52A	−3.8	1.5	3.7	0.0	1.4	4.2	2.8
D54A	−1.3	3.0	0.8	−0.4	2.1	4.3	2.2
N56A	−0.7	0.1	0.8	−0.4	−0.2	1.2	1.4
D58A	−1.5	3.6	0.2	0.0	2.3	1.6	0.7
E98A	2.2	8.1	1.3	0.0	11.6	4.2	7.4
R99A	−2.7	1.0	2.5	0.0	0.8	1.9	1.1
D100A	−1.2	4.8	0.6	−1.3	2.9	2.8	0.1
Y101A	−6.7	2.7	3.0	0.0	−1.0	4.0	5.0
Constant dielectric boundary							
Light chain							
H30A	−0.2	0.1	0.1	0.0	0.0	1.7	1.7
Y32A	−1.2	1.0	2.3	0.0	2.1	2.0	0.1
Y49A	−0.6	−1.0	1.9	0.0	0.3	1.7	1.4
Y50A	−0.6	0.1	0.9	0.0	0.4	0.7	0.3
W92A	−0.4	0.6	2.9	0.0	3.1	0.3	2.8
S93A	0.0	0.0	0.0	0.0	0.0	1.2	1.2
Heavy chain							
T30A	−0.3	0.0	0.2	0.0	−0.1	0.9	1.0
Y32A	−0.2	0.0	0.0	0.0	−0.2	1.8	2.0
W52A	−0.2	0.4	3.7	0.0	3.9	4.2	0.3
D54A	−1.7	3.0	0.8	−0.4	1.7	4.3	2.6
N56A	1.0	0.1	0.8	−0.4	1.5	1.2	0.3
D58A	−1.5	3.6	0.2	0.0	2.3	1.6	0.7
E98A	0.4	8.1	1.3	0.0	9.8	4.2	5.6
R99A	−0.6	1.1	2.5	0.0	3.0	1.9	1.1
D100A	−0.5	4.6	0.6	−1.3	3.4	2.8	0.6
Y101A	−1.7	1.8	3.0	0.0	3.1	4.0	0.9

[a]Calculations performed employing the continuum model without zinc ion or crystallographic waters. Energy units are kcal/mol.

for heavy-chain W52A, Y32A, N56A and Y101A; thus suggesting that the created cavities for these substitutions are more hydrophobic than hydrophilic. An example of the improvement is the light-chain Y32A, exhibiting a change from $\Delta\Delta G_{calc}$ of −1.2 to +2.1 kcal/mol, while $\Delta\Delta G_{expt}$ is 2.0 kcal/mol. This model failed for several mutants (most notably the light-chain W92A) and for others was relatively insensitive to changes in the treatment of the dielectric boundary. The overall performance of the model is a reduction in the average absolute error, now showing a value of ±1.4 kcal/mol. Collecting the most accurate calculated free energies from the two dielectric models (a 'hybrid' approach) yields an average error of ±1.2 kcal/mol. The correspond-

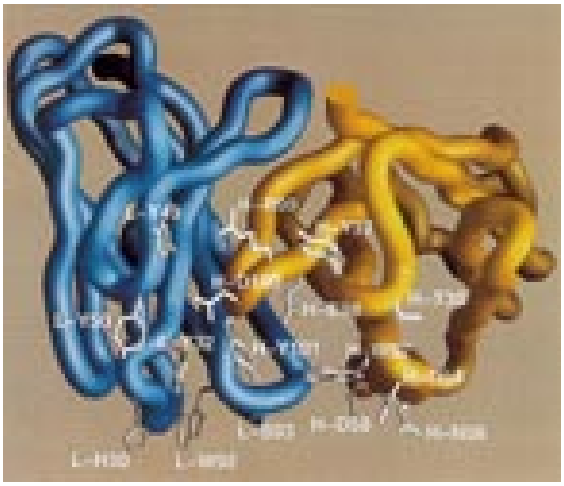

Fig. 3. Molecular representation of the D1.3 antibody showing the location of amino acid side-chains replaced with alanines. The light chain (L) is colored blue and the heavy chain (H) is gold.

ing scatter plot for this hybrid analysis is shown in Fig. 4b, where the filled circles represent the predicted hydrophilic cavities of W92A, D54A and D100A by using the SAS-ε boundary model. The correlation coefficient for this system is now 0.7. Given the many caveats of the calculations, in particular, the rigid-body assumption, the ability to reproduce $\Delta\Delta G_{expt}$ within an accuracy tolerance of roughly less than 1.5 kcal/mol is very encouraging.

Applying the premise that the protein dielectric constant is an adjustable parameter that smears or averages out the explicit electrostatic details of the molecule into a homogenous dielectric continuum medium, it is possible to fit $\Delta\Delta G_{expt}$ using ε_p, and in fact, all three free energy terms can be scaled according to

$$\Delta\Delta G_{expt} = \xi_1\Delta\Delta G_{ele} + \xi_2\Delta\Delta G_{cav} + \xi_3 T\Delta S_{sc} \quad (11)$$

where the variables ξ are fitting parameters using a Marquardt–Levenberge algorithm [31]. Values for $\Delta\Delta G_{ele}$, $\Delta\Delta G_{cav}$ and $T\Delta S_{sc}$ are taken from Table 2. New continuum parameters for ε_p and γ are calculated from ξ_1 and ξ_2. Fitting either the const-ε boundary model or the hybrid data extracted from both models yields an ε_p in the range of 9–10, a value for γ of 65–68 cal/mol/Å^2, and a conformational entropic term of zero. This fitted entropic value is tantamount to assuming

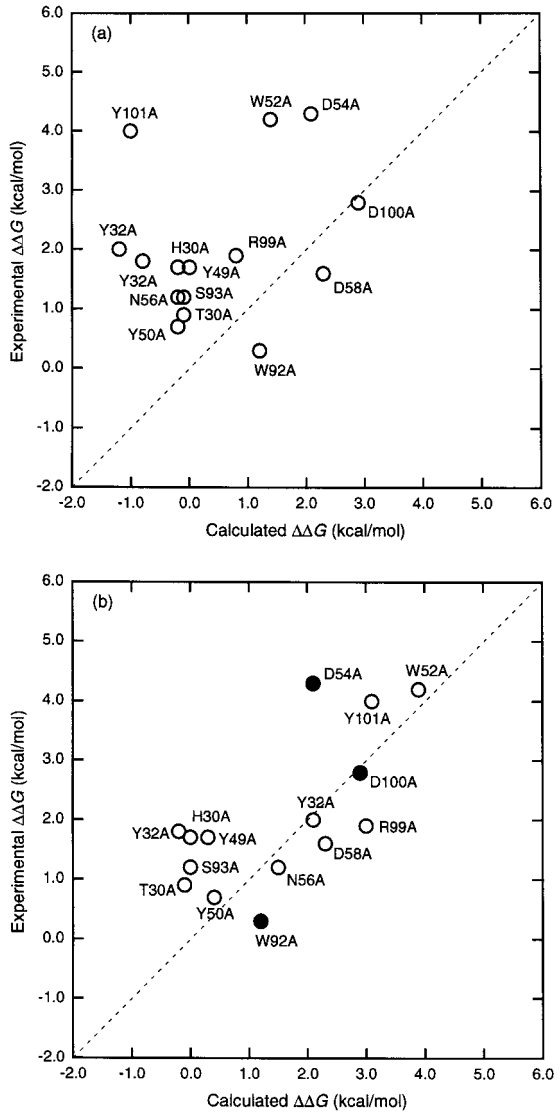

Fig. 4. Scatter plots showing comparisons between the calculated free changes for the mutants vs. the experimental values. (a) Calculations with the solvent accessible surface dielectric boundary. (b) Combined calculations with the constant dielectric boundary model (○) plus three hydrophilic mutants using data from the above plot (●). Data for the mutant E98A is located off scale.

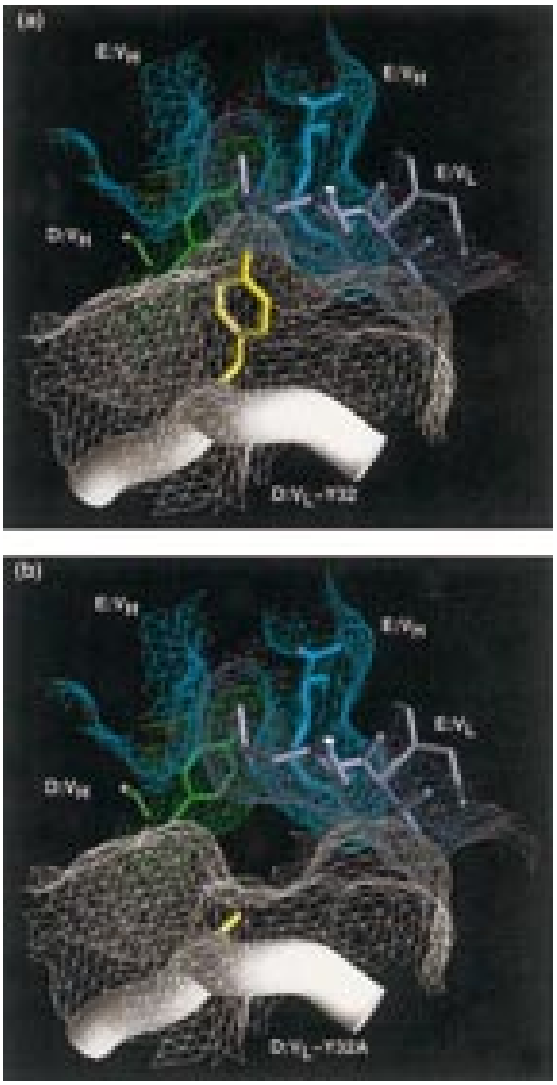

Fig. 5. Molecular surface and the dielectric boundary. (a) Wild-type complex showing the molecular surface (mesh) of the D1.3 light chain centered around residue Y32 (colored yellow) packed against neighboring residues of both antibodies. (b) Light-chain mutant Y32A illustrating the cavity created from the alanine substitution. The molecular surface for Y32A is given by the white mesh, while the dielectric boundary is the same as the molecular surface shown above for the wild-type protein.

the enthalpy–entropy compensation phenomenon. It is noteworthy that given the range in the literature on choosing a value for γ [9,11,12], the value obtained from fitting is nearly equivalent to the initial value of 69 cal/mol/Å^2, and not the 'microscopic' range of 20–25 cal/mol/Å^2, which is typically applied in continuum modeling using solvent accessible surface areas [32]. The quality of the fit for the hybrid approach is an average absolute error of ± 1.0 kcal/mol, with a correlation coefficient remaining at roughly 0.7. The scatter plot for this analysis is displayed in Fig. 6 and shows a much smaller error for E98A than calculated from either of the two models. Fitting the SAS-ε boundary model yields an ε_p of 13, γ of 84 cal/mol/Å^2, and zero for entropy. The average error for this model is now at ± 1.1 kcal/mol and a correlation coefficient of 0.6 is obtained. A fitted ε_p in the range of 9–10 is very similar with the fitting calculations reported by Novotny et al. [5], while their value for γ (~ 42 cal/mol/Å^2) is significantly lower than found here. The initial value for γ employed in their calculations was set at 70 cal/mol/Å^2, and the observed reduction may be needed to account for possible structural reorganization of the lysozyme complex due to substitutions. These workers also found that increasing the conformational entropy penalty was needed to obtain a proper fit. Despite these differences, the consistent values obtained for ε_p by both studies of completely different

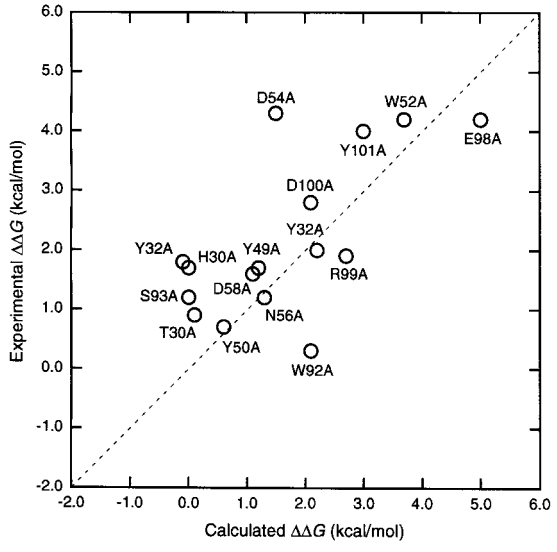

Fig. 6. Scatter plot for calculations fitting the free energy function using the constant dielectric boundary model.

protein complexes (although, perhaps similar atomic complementarity due to lysozyme) certainly suggests a common value for scaling the electrostatic interactions when applied to continuum models evaluating protein mutations. However, it will be shown below that including the zinc ion requires even greater effective damping of the charge–charge interactions.

3.3. Effect of incorporating zinc ion or crystallographic waters

Table 3 presents values for $\Delta\Delta G_{calc}$ determined for the model system of incorporating the zinc ion into the protein–protein interface using both dielectric boundary treatments described above. The zinc ion is bound by the heavy-chain D1.3 residue

Table 3
Relative free energy changes for D1.3 mutants calculated by the continuum model incorporating the zinc ion but without crystallographic waters

Mutant	$\Delta\Delta G_{sol}$	$\Delta\Delta G_{int}$	$\Delta\Delta G_{cav}$	$T\Delta S_{sc}$	$\Delta\Delta G_{calc}$	$\Delta\Delta G_{expt}$	Difference
Solvent accessible surface dielectric boundary							
Light chain							
H30A	−0.2	0.1	0.1	0.0	0.0	1.7	1.7
Y32A	−7.6	2.7	2.3	0.0	−2.6	2.0	4.6
Y49A	−4.5	−0.5	1.9	0.0	−3.1	1.7	4.8
Y50A	−7.5	2.5	0.9	0.0	−4.1	0.7	4.8
W92A	−2.9	1.0	2.9	0.0	1.0	0.3	0.7
S93A	0.0	−0.1	0.0	0.0	−0.1	1.2	1.3
Heavy chain							
T30A	−0.3	−0.1	0.2	0.0	−0.2	0.9	1.1
Y32A	−1.0	−0.1	0.0	0.0	−1.1	1.8	2.9
W52A	−3.8	0.5	3.7	0.0	0.4	4.2	3.8
D54A	−1.3	3.6	0.8	−0.4	2.7	4.3	1.6
N56A	−0.7	0.1	0.8	−0.4	−0.2	1.2	1.4
D58A	−1.5	4.0	0.2	0.0	2.7	1.6	1.1
E98A	2.2	10.7	1.3	0.0	14.2	4.2	10.0
R99A	−3.4	−1.8	2.5	0.0	0.8	1.9	1.1
D100A	−6.0	30.1	1.3	−1.3	24.1	2.8	21.3
Y101A	−7.8	3.6	3.0	0.0	−1.2	4.0	5.2
Constant dielectric boundary							
Light chain							
H30A	−0.2	0.1	0.1	0.0	0.0	1.7	1.7
Y32A	−1.2	0.9	2.3	0.0	2.0	2.0	0.0
Y49A	−0.6	−2.1	1.9	0.0	−0.8	1.7	2.5
Y50A	−0.6	0.3	0.9	0.0	0.6	0.7	0.1
W92A	−0.4	0.6	2.9	0.0	3.1	0.3	2.8
S93A	0.0	0.0	0.0	0.0	0.0	1.2	1.2
Heavy chain							
T30A	−0.3	−0.1	0.2	0.0	−0.2	0.9	1.1
Y32A	−0.2	−0.1	0.0	0.0	−0.3	1.8	2.1
W52A	−0.2	0.5	3.7	0.0	4.0	4.2	0.2
D54A	−1.7	3.6	0.8	−0.4	2.3	4.3	2.0
N56A	1.0	0.0	0.8	−0.4	1.4	1.2	0.2
D58A	−1.5	4.1	0.2	0.0	2.8	1.6	1.2
E98A	0.4	10.7	1.3	0.0	12.4	4.2	8.2
R99A	−0.6	−2.5	2.5	0.0	−0.6	1.9	2.5
D100A	−0.8	30.9	1.3	−1.3	30.1	2.8	27.3
Y101A	−1.7	2.2	3.0	0.0	3.5	4.0	0.5

D100 and two protein side chains of E5.2 plus a water molecule. Calculations employing the SAS-ε boundary model show significant errors in $\Delta\Delta G_{calc}$, with the largest being, as expected, that of D100A. The average absolute error for this model is significant at ± 4.2 kcal/mol. For the const-ε boundary model, the difference between $\Delta\Delta G_{calc}$ and $\Delta\Delta G_{expt}$ improves somewhat, showing a ± 3.4 kcal/mol average error, but falls far short in terms of accuracy as compared with the non-zinc ion calculations.

Using Eq. (11), both models were fitted to $\Delta\Delta G_{expt}$ yielding an ε_p in the range of 40–47, and γ of 69–77 cal/mol/Å^2. As found above, the conformational entropy was set at zero. Both models returned an average absolute error of ± 1.1 kcal/mol, although the correlation coefficients obtained were only 0.5. It appears that fitting either model shows no advantage of one over the other. Fig. 7 presents the comparison between $\Delta\Delta G_{calc}$ and $\Delta\Delta G_{expt}$ when using the SAS to calculate the dielectric boundary. The much larger increase observed in the fitted ε_p obtained from including the zinc ion is not surprising, and is simply an effect of overcompensating for solvation and interaction energetics of the stronger electrical potential. As for the larger values in the fitted ε_p parameters compared with those employed in determining the absolute binding free energy (Table 1), calculations on the various mutants lack reorganization of the permanent dipoles found initially in the wild-type protein, thus requiring a greater ε_p than the initial value of 4.

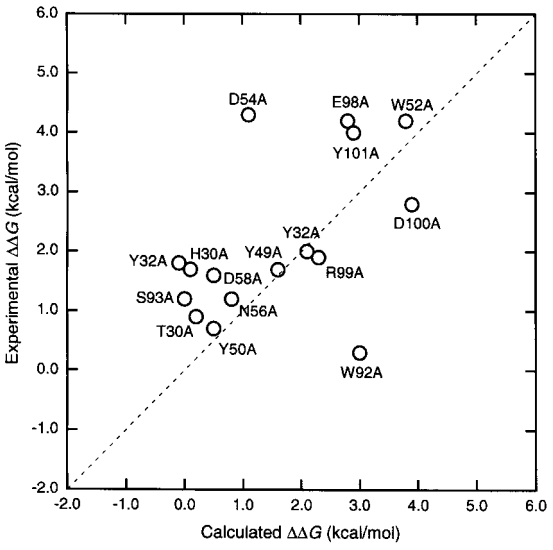

Fig. 7. Scatter plot for fitting calculations incorporating the interfacial zinc ion using the solvent accessible surface dielectric boundary model.

Finally, listed in Table 4 are $\Delta\Delta G_{calc}$ determined for the complex with interfacial crystallographic waters. In the initial structure, several buried waters were hydrogen bonded with side chains of residues replaced with alanines (light chain Y32A, and heavy chain T30A, N56A and Y101A); consequently, upon substitution, these waters are thought to be removed. Overall, including crystal waters increases the magnitude of error in the calculations, and only marginally improves several mutants (H30A, W92A, S93A of the light chain, and D54A and Y101A of the heavy chain). Fitting the free energy function yields $\varepsilon_p = 37$ and $\gamma = 87$ cal/mol/Å^2, and produces an absolute error of ± 1.2 kcal/mol with a correlation coefficient of 0.5. In contrast with the other models, the discontinuity in the molecular surface introduced by the crystal water molecules now shows the largest fitted γ value.

4. Conclusion

The mean-field analysis of the D1.3–E5.2 antibody–antibody complex presented in this paper indicates that the application of continuum methods for estimating binding free energies is sensitive to the treatment of interfacial ions and crystal water molecules. Moreover, the modeling of interfacial mutations suggests that the hydrophilicity of cavities created from alanine substitutions play an important role in determining realistic free energy changes. The overall success of continuum models clearly depend on a priori values of ε_p and γ. Although a reasonable absolute binding free energy was obtained with $\varepsilon_p = 4$, the best results for substitutions were achieved when ε_p was scaled to ~ 10. This scaled ε_p is larger than the range of 2–5 commonly used in

Table 4
Relative free energy changes for D1.3 mutants calculated by the continuum model incorporating only crystallographic waters and employing the solvent accessible surface dielectric boundary

Mutant	$\Delta\Delta G_{sol}$	$\Delta\Delta G_{int}$	$\Delta\Delta G_{cav}$	$T\Delta S_{sc}$	$\Delta\Delta G_{calc}$	$\Delta\Delta G_{expt}$	Difference
Light chain							
H30A	−0.2	0.1	0.1	0.0	0.0	1.7	1.7
Y32A	−14.5	3.5	2.3	0.0	−8.7	2.0	10.7
Y49A	−2.2	−0.6	1.9	0.0	−0.9	1.7	2.6
Y50A	−5.9	1.2	0.9	0.0	−3.8	0.7	4.5
W92A	−4.0	1.2	2.9	0.0	0.1	0.3	0.2
S93A	0.0	−0.1	0.0	0.0	−0.1	1.2	1.3
Heavy chain							
T30A	−1.1	−0.2	0.2	0.0	−1.1	0.9	2.0
Y32A	−2.2	0.0	0.0	0.0	−2.2	1.8	4.0
W52A	−4.7	1.4	3.7	0.0	0.4	4.2	3.8
D54A	−3.1	4.8	0.8	−0.4	2.1	4.3	2.2
N56A	−4.0	0.3	0.8	−0.4	−3.3	1.2	4.5
D58A	−2.4	5.7	0.2	0.0	3.5	1.6	1.9
E98A	1.6	11.4	1.3	0.0	14.3	4.2	10.1
R99A	−4.9	2.7	2.5	0.0	0.3	1.9	1.6
D100A	−4.3	8.5	0.6	−1.3	3.5	2.8	0.7
Y101A	−18.8	3.1	3.0	0.0	−12.7	4.0	16.7

continuum models and underscores the notion that the 'protein dielectric constant' represents the contributions that are not treated explicitly. Possible universality in ε_p for evaluating substitutions in protein–protein complexes may be near the value of 10, albeit it undoubtedly depends on the electrical potential of the interface, as observed from calculations incorporating the zinc ion. Additionally, the magnitude of ε_p will vary according to the amount of electrostatic reorganization incurred from mutations (rigid body vs. conformational relaxation). Scaling the parameter γ in the range of 65–68 cal/mol/Å^2 gave the best reproduction of experimental data. Without empirically fitting the free energy function, including the zinc ion or crystal waters reduces the accuracy of reproducing the experimental data. Once the mean-field parameters are determined, continuum calculations provide meaningful physical insights into the interactions occurring at the interface and the origins of variations in $\Delta\Delta G$ values.

Acknowledgements

Gratitude is expressed to Dr Lilee Cuff for computational assistance in the earlier stages of this work and for many valuable discussions. The author acknowledges the Biomedical Supercomputing Center of the Frederick Cancer Research and Development Center for a generous grant of computer time.

References

[1] M.K. Gilson, B. Honig, Proteins 4 (1988) 7.
[2] K.A. Sharp, B. Honig, Annu. Rev. Biophys. Biophys. Chem. 19 (1990) 151.
[3] W.F. van Gunsteren, in: W.F. van Gunsteren, P.K. Weiner (Eds.), Computer Simulation of Biomolecular Systems, ESCOM, Leiden, 1989, p. 27.
[4] P.A. Kollman, Chem. Rev. 93 (1993) 2395.
[5] J. Novonty, R.E. Bruccoleri, M. Davis, K.A. Sharp, J. Mol. Biol. 268 (1997) 401.
[6] R.M. Jackson, M.J.E. Sternberg, J. Mol. Biol. 250 (1995) 258.
[7] N. Froloff, A. Windemuth, B. Honig, Protein Sci. 6 (1997) 1293.
[8] W. Dall'Acqua, E. Goldman, E. Eisenstein, R.A. Mariuzza, Biochemistry 35 (1996) 9667.
[9] A. Nicholls, K.A. Sharp, B. Honig, Proteins 11 (1991) 281.
[10] A.J. Doig, M.J.E. Sternberg, Protein Sci. 4 (1995) 2247.
[11] R.M. Jackson, M.J.E. Sternberg, Protein Eng. 7 (1994) 371.

[12] K.A. Sharp, A. Nicholls, R. Friedman, B. Honig, Biochemistry 30 (1991) 9686.
[13] M.I. Page, W.P. Page, Proc. Natl. Acad. Sci. 68 (1971) 1678.
[14] A.V. Finkelstein, J. Janin, Protein Eng. 3 (1989) 1.
[15] H.P. Erickson, J. Mol. Biol. 206 (1989) 465.
[16] B. Tidor, M. Karplus, J. Mol. Biol. 238 (1994) 405.
[17] J.R. Maple, M.-J. Hwang, T.P. Stockfisch, U. Dinur, M. Waldman, C.S. Ewig, A.T. Hagler, J. Comput. Chem. 15 (1994) 195.
[18] J. Warwicker, H.C. Watson, J. Mol. Biol. 157 (1982) 671.
[19] I. Klapper, R. Hagstrom, R. Fine, K.A. Sharp, B. Honig, Proteins 1 (1986) 47.
[20] M.K. Gilson, K.A. Sharp, B. Honig, J. Comput. Chem. 9 (1988) 327.
[21] D. Stifoff, K.A. Sharp, B. Honig, J. Phys. Chem. 98 (1994) 1978.
[22] M.L. Connolly, Molecular Surface Program 429 (1981). The Quantum Chemistry Program Exchange, Indiana University, Bloomington, Indiana 47405, USA.
[23] C.N. Pace, CRC Crit. Rev. Biochem. 3 (1975) 1.
[24] P.L. Privalov, Adv. Protein Chem. 33 (1979) 167.
[25] S.C. Hoops, K.W. Anderson, K.M. Merz, Jr., J. Am. Chem. Soc. 113 (1991) 8262.
[26] M.A. Olson, L. Cuff, J. Mol. Recogn. 10 (1998) 277.
[27] Y.Y. Sham, Z.T. Chu, A. Warshel, J. Phys. Chem. 101 (1997) 4458.
[28] I. Muegge, H. Tao, A. Warshel, Protein Eng. 10 (1997) 1363.
[29] B. Honig, A. Nicholls, Science 258 (1995) 1144.
[30] Z.S. Hendsch, B. Tidor, Protein Sci. 3 (1994) 211.
[31] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes, Cambridge University Press, Cambridge, UK, 1986.
[32] T. Zhang, D.E. Koshland, Jr., Protein Sci. 5 (1997) 348.