

Binding Affinity and Specificity from Computational Studies

Themis Lazaridis*

Department of Chemistry, City College of the City University of New York, 138th St. & Convent Ave, New York, NY 10031, USA

Abstract: Computational methods available for the calculation of relative and absolute binding affinities (free energy simulations, continuum electrostatics, linear interaction energy approximations, and empirical solvation models) are reviewed together with recent applications to biological systems. The decomposability of the binding free energy into physically meaningful components is examined and results obtained for these components are presented. Some of these components, such as the direct interactions, the translational/rotational entropy loss, and the desolvation free energy are well recognized. Recent calculations have shown that the translational/rotational entropy loss is not as large as some theoretical calculations have previously suggested because of substantial residual movements in the bound complex. Recent work also points to the importance of contributions that are often neglected in binding affinity calculations, such as the protein reorganization energy and, for flexible ligands, the ligand reorganization energy. Future work should concentrate on the improvement of the energy functions and simulation protocols for the achievement of more precise and accurate predictions.

1. INTRODUCTION

A biological cell contains thousands of macromolecular components moving in a sea of water, lipids, and ions. The function of a cell depends critically on specific interactions between these macromolecules. For example, an enzyme must act on specific substrates. Information along a signal transduction pathway flows via specific interactions between proteins. Transcription initiation occurs by binding of proteins to specific sites on DNA among thousands of similar sites. Foreign substances in an organism must be selectively isolated and neutralized by antibodies. The highly selective binding between molecules in solution is termed molecular recognition. Although a cell in a non-equilibrium system, the equilibrium binding free energy is a useful measure for the tendency of two macromolecules to associate in vivo [1]. The equilibrium free energy of binding between two molecules is termed binding affinity. The relative binding affinity of a molecule towards one of many potential partners is termed binding specificity. Molecular recognition requires a sufficiently high binding affinity and very high binding specificity.

Aside from its fundamental interest, molecular recognition has many practical applications, one of the most important of which is rational drug design. Most known drugs bind specifically to a disease-causing biomolecule and inhibit its function. Drug discovery is increasingly becoming more systematic and rational [2]. The emerging paradigm of rational, structure-based drug design consists of three steps: identification of a disease target, determination of its three-dimensional structure by x-ray crystallography or NMR spectroscopy, and design/discovery of an appropriate

small molecule which binds to the active site of the target with high affinity. At the same time, the designed drug must not bind and inhibit the action of other, beneficial biomolecules; i.e. it has to be specific. Other applications of molecular recognition include the design of sensors [3] or separation techniques [e.g. 4, 5].

Theory and simulation holds considerable promise in analyzing the fundamentals and helping in the practical applications of molecular recognition. The key in this endeavor is the accurate calculation of binding free energies from the structure of the binding partners. In the field of rational drug design, the ability to predict the binding affinity of a candidate molecule without having to synthesize it would save a lot of time and resources. In this review we examine the latest methods available for prediction of binding affinities and the current understanding of the physical contributions to binding affinity. Earlier reviews on this topic are available [6, 7].

2. CALCULATION OF RELATIVE BINDING FREE ENERGIES

In many cases it is sufficient to calculate *relative* binding free energies, i.e., the difference in binding free energy between different inhibitors binding to the same protein or the same inhibitor binding to variants of one protein. A procedure to calculate such differences in free energy using molecular simulations was proposed in the 80s [8]. This procedure is based on thermodynamic cycles and the fact that the free energy is a state function. Instead of calculating directly the free energies of binding of two ligands and taking the difference, one could calculate the free energy of transforming one ligand into the other in solution and in the receptor and taking the difference. The transformations are done by use of a "coupling parameter" such that $\lambda=0$ corresponds to one ligand and $\lambda=1$ to the other. MD

*Address correspondence to this author at the Department of Chemistry, City College of the City University of New York, 138th St. & Convent Ave, New York, NY 10031, USA. Tel: (212) 650-8364; Fax: (212) 650-6107; E-mail: themis@sci.cuny.edu; http://blakey.sci.cuny.edu

simulations are used to sample the conformational space and statistical mechanical formulas (Free Energy Perturbation or Thermodynamic Integration) are used to obtain the free energy change [9, 10].

Many applications of this approach to binding have appeared over the last decade [11, 12]. For example, the binding of trimethoprim derivatives to dihydrofolate reductase [13], binding of substrate to tyrosyl-tRNA synthetase mutants [14], sulfonamide inhibitors of carbonic anhydrase [15], benzamidine inhibitors to trypsin [16], distamycin analogs to DNA [17], inhibitors to different mutants of ribonuclease T1 [18], and drugs to DNA [19]. Archontis *et al.* [20] calculated a large difference in binding free energy of Asp and Asn to aspartyl-tRNA synthetase favoring the natural substrate. This system was particularly challenging because it involved the creation of a charge, so special techniques had to be employed for proper accounting of long range interaction contributions to the free energy. Free energy simulations on a series of inhibitors to p38 MAP kinase showed the best correlation with experiment compared to more empirical methods [21]. Similar calculations have been presented for metal binding to macrocyclic hosts [22-26]. A variant of free energy simulations called " γ -dynamics" simulates a hybrid of many ligands and yields directly estimates of their relative binding affinity [27-30].

The above methods employ fully atomistic simulations, i.e. the solvent is represented in full atomic detail by explicit solvent molecules. This makes them computationally expensive. Alternative methods are based on a continuum description of the solvent. The best studied continuum model is based on the Poisson-Boltzmann (PB) equation [31]. Since the PB approach computes only the electrostatic part of the free energy, it is usually combined with a surface-area model for the non-electrostatic part. One limitation of this approach is that the results are very sensitive to the value for the dielectric constant used for the protein and ligand interiors, the chosen atomic radii, and the definition of the dielectric boundary. From a different angle, this limitation can be seen as a strength, since by adjusting these parameters any experimental value can be reproduced.

Although these calculations can be used to obtain absolute binding free energies (see below), many applications focused on just the electrostatic contribution and compared that for different ligands or proteins, hence they provided relative binding free energies. Jackson & Sternberg combined PB calculations with a surface-area model for the cavitation free energy (solvation free energy with dispersion and electrostatic interactions subtracted) and an empirical model for the loss of sidechain conformational entropy to evaluate alternative binding modes that arise in docking simulations [32]. They found that use of the molecular surface instead of the more common solvent accessible surface improved the results significantly. Zhang & Koshland [33] combined PB calculations with surface-area based estimates of nonpolar contributions to different substrates of mutant isocitrate dehydrogenases to calculate relative binding free energies. The results showed good correlation with experiment. A similar approach was applied to inhibitors of HIV-1 protease [34]. Some of these studies include a term for the loss of

sidechain conformational entropy upon binding, often estimated by the empirical scale of Pickett and Sternberg, obtained by analysis of the database of protein structures [35]. Novotny *et al.* noted an improvement in the correlation between calculated and experimental binding affinities for antibody-lysozyme mutant complexes when the PB equation was used for electrostatics, the molecular surface for the hydrophobic effect, and explicit sampling of torsional degrees of freedom for the sidechain conformational entropy [36].

Sometimes the PB equation is applied to multiple structures obtained from a MD simulation, as opposed to a single structure. This combination of explicit MD and PB calculations was applied to obtain relative free energies of hapten binding to the germ line and mature forms of an antibody [37], different dimers of the HIV protease [38], ligands to matrix metalloprotein MMP-1 [39], an inhibitor to HIV and FIV protease [40], and peptide binding to DnaK [41]. Archontis *et al.* [42] revisited the aspartyl-tRNA synthetase system and compared the continuum approach to their earlier FE simulations. Good agreement was found when an interior dielectric of 4 was used. PB combined with surface-based estimates for nonpolar solvation and a molecular mechanics force field for the bonded terms was used to evaluate binding configurations resulting from docking simulations [43].

PB calculations usually show that the electrostatic contribution to binding is unfavorable due to the large desolvation penalty of polar and charged groups which is not sufficiently compensated by the direct ligand-receptor interactions [44, 45]. However, this is not always the case. Tidor and coworkers developed an elegant optimization scheme where the charge distribution of the ligand or the receptor is adjusted so as to minimize the electrostatic binding free energy [46, 47]. It was shown that with reasonable adjustments the electrostatic binding free energy can become favorable. The method was initially applied to idealized ligand geometries [48] and more recently to realistic systems [49, 50]. The approach has also been extended to optimize specificity rather than binding affinity [51] and has obvious applications to rational ligand design.

3. CALCULATION OF ABSOLUTE BINDING FREE ENERGIES

Absolute binding free energies are more difficult to obtain. Some difficulties have been conceptual: earlier calculations did not properly account for the standard state dependence of the binding free energy [52, 53]. Another important difficulty is that the binding free energy contains significant entropic contributions which are not trivial to calculate. A number of approaches have been employed, ranging from rigorous atomistic chemical physics to entirely empirical, knowledge-based methods.

Free Energy Simulations

The most detailed and rigorous approach to absolute binding free energies is through full atomistic simulations.

Hermans & Wang proposed a method combining free energy simulations with restraints with analytical results that incorporate the choice of standard state and calculated the binding free energy of benzene in a cavity of a mutant T4 lysozyme [54]. A similar methodology was applied to the binding of camphor to cyt P450 [55] and of water into the bacteriorhodopsin proton channel [56]. The binding free energy of biotin to streptavidin, a system initially studied by Miyamoto & Kollman [57], was calculated by this methodology [58].

Continuum Electrostatics and MM-PBSA

For larger systems explicit solvent simulations are prohibitively expensive and researchers have turned to implicit solvation methods. As mentioned above, continuum electrostatics can provide the electrostatic component of the binding free energy. Addition of an estimate of the nonelectrostatic contribution and the entropic components, such as translational, rotational, and conformational entropy changes, gives an estimate of the total binding free energy.

Froloff *et al.* [59] used continuum electrostatics together with a surface area model for the nonpolar contribution to obtain the interaction and desolvation contributions to the binding free energy of peptides to a protein. The translational/rotational entropy and the protein reorganization upon binding were neglected, and, as expected, the calculated binding free energies were more negative than the experimental values. The same methodology was applied to trypsin-inhibitor complexes [60].

The above calculations were done on single configurations (either the crystal or energy-minimized structures). To account for thermal fluctuations, a combination of explicit solvent and continuum calculations has been proposed. An explicit solvent simulation is used to generate many snapshots of the complex and subsequently the Poisson-Boltzmann equation and a surface-area equation are used to estimate the electrostatic and nonpolar contributions to binding, respectively [61]. This method (MM/PBSA) has been applied to calculate absolute binding free energies of a p53 peptide and its alanine mutants to oncoprotein Mdm2 [62], human growth hormone [63], RNA-protein binding [64], and various ligands to avidin and streptavidin [65]. A similar approach has been applied to the evaluation of candidate protein folds [66-69].

Other Implicit Solvation Models

In the above methods the entropic terms are estimated based on empirical models or simplifying assumptions. If they are to be calculated from first principles, some sampling of the configurational space available to the binding partners is required. Although PB calculations are more efficient than explicit solvent simulations, they still incur a significant computational cost. In this case, analytical implicit solvation models are the method of choice. A popular implicit solvation model is the Generalized Born model for the electrostatic free energy Generalized Born model for the

electrostatic free energy supplemented by a surface-area term for the nonpolar contributions (GB/SA) [70-72].

A rigorous method to compute binding affinities using implicit representation of the solvent was proposed by Gilson and coworkers [73]. Their approach involves the approximate calculation of configurational integrals for the free and bound states. The method with the GB/SA solvation model was applied to adenine binding to small, synthetic receptors [74] and to cyclic ureas binding to HIV-1 protease [75]. In the latter study it was found that the GB/SA model gives smaller desolvation penalties compared to PB calculations and cannot reproduce well the experimental binding affinity trends. Purisima and coworkers had earlier used a similar approach to compute relative binding free energies of different thrombin inhibitors [76]. Beveridge and coworkers [77-79] studied protein-DNA binding using the GB/SA model incorporating salt effects and a counterion release model on single configurations. The reorganization energies were either calculated from the known structures of complexed and uncomplexed DNA [77] or were obtained as the difference between the calculated contributions and the experimental affinity [78].

Another implicit solvation model (EEF1) [80] has recently been applied by our group to estimate binding affinities of biotin and other ligands to avidin and streptavidin [81]. MD simulations of the complexes and the unbound molecules were performed and the average effective energy (intramolecular energy + solvation free energy) was calculated directly from the simulations. The various energetic and entropic components were evaluated individually (see below). The resulting binding affinities were found to be in the correct range, but the large statistical uncertainty precluded precise predictions.

Linear Interaction Energy Methods

In search of faster methods for affinity prediction, Aqvist *et al.* [82, 83] proposed calculating the van der Waals and electrostatic interactions of the ligand with the protein in the complex and with water in its unbound state. The difference in these interactions between bound and solvated states was then scaled by empirical factors and summed to give the binding free energy. This was termed the Linear Interaction Energy method. The electrostatic interaction was scaled by the factor $\epsilon=0.5$, which can be theoretically justified based on the linear response approximation of continuum electrostatics. For the scaling factor for the van der Waals interactions the value $\epsilon=0.16$ was found optimal in the original study [82] but different values were found optimal for other systems. For example, $\epsilon=1$ works much better for avidin ligands [84] and a correlation exists between the value of ϵ and the hydrophobicity of the binding site [85]. Jorgensen and coworkers included the change in surface area and a different set of coefficients in their study of nonnucleoside inhibitors of HIV reverse transcriptase [86] and FKBP12 inhibitors [87]. Warshel *et al.* employed similar linear response-based methods in the study of cyclic urea inhibitors of HIV protease [88].

Empirical Methods and Scoring Functions

Other, more empirical computational methods have been proposed [89]. Freire and coworkers [90] developed an empirical method for predicting the binding free energy from the polar and nonpolar surface area buried upon binding. Another empirical method combines electrostatic interactions calculated by a MM force field with a surface area-based solvation model [91, 92].

Even simpler functions are intended for use by various docking programs [93-97]. These functions are based on intuitive ideas about contributions to binding and various practical considerations, rather than rigorous theoretical arguments. The venerable DOCK program uses a MM-type sum of van der Waals and Coulomb interactions calculated on a grid for computational efficiency [98]. The program GOLD [99] uses a function that accounts for hydrogen bonding, weighted based on the geometry, soft van der Waals interactions between protein and ligand atoms, and the intramolecular torsional and steric energy of the ligand. The program FlexX [100] uses Bohm's function [101], which includes terms for hydrogen bonding, ionic interactions, hydrophobic interactions and conformational entropy loss, adding an extra term for aromatic interactions. Each of these interactions is represented by a constant (adjustable parameter) multiplied by a weighting factor that depends on the geometry of the contact. The program AUTODOCK [102] uses a similar approach: terms for van der Waals, hydrogen bond, screened electrostatics, torsional angle entropy, and desolvation, all weighted by adjustable free energy parameters. The program SEED for docking molecular fragments [103] uses a more sophisticated approach: the generalized Born model for the receptor-fragment interaction and desolvation of the fragment, and a PB-based method for the desolvation of the receptor. The GB model [104] and a modified Born equation [105] have also been used to improve scoring functions for docking.

4. DECOMPOSITION OF THE BINDING FREE ENERGY

The purpose of theoretical studies is not only to provide quantitative estimates, but also, and mainly, to enhance our understanding. Understanding often involves breaking up a complex phenomenon into simpler elements. In the case of the binding affinity between two large molecules, it involves breaking up the binding free energy into contributions from the atoms that constitute the binding interface and various energetic and entropic factors. Better understanding of these components has practical consequences as well, since many of the binding free energy calculation methods mentioned above employ separate evaluation of such free energy components.

The very concept of "contributions" to the binding free energy implies that the free energy can be separated into additive components. The validity of this idea is not obvious. A number of authors argued that entropy is a collective property of a many-body system and cannot be broken up into contributions from specific atoms [106, 107].

Therefore, the concept of decomposition and the assumptions it involves will be considered briefly below.

An exact (and trivial) decomposition of the binding free energy is the thermodynamic decomposition into enthalpy and entropy, $G = H - TS$. Usually binding enthalpies are obtained by isothermal titration calorimetry and entropies from the difference between free energy and enthalpy. Consideration of enthalpies and entropies of binding separately often gives added insights into the physical origin of the binding affinity. One difficulty is that these quantities include contributions from solvation and these contributions are often quite unpredictable. The breakup of solvation free energy into enthalpy and entropy is sensitive to the geometry of the binding interface, possibly the presence of bound waters, etc. It has been customary to consider hydrophobic interactions entirely entropic. However, this is not necessarily so because the entropy-enthalpy balance in hydrophobic solvation is sensitive to the geometry of the binding interface [108, 109]. Therefore, the observation of enthalpy-driven binding does not necessarily mean that the driving force is not hydrophobic in origin.

More detailed, extra-thermodynamic, decompositions are possible. They fall into two categories: decompositions into types of contributions, such as desolvation, hydrophobic interactions, translational/rotational entropy etc., and decompositions into group contributions.

Different types of decomposition are suggested by different theoretical formulations. In free energy thermodynamic integration calculations, an integral is calculated over the average derivative of the Hamiltonian with respect to a coupling parameter:

$$A = \int_0^1 \left\langle \frac{U(\lambda)}{d} \right\rangle d\lambda \quad (1)$$

where $U(\lambda) = (1 - \lambda)U_{\text{initial}} + \lambda U_{\text{final}}$. Since the solute-solvent energy can be written as the sum of residue or group-solvent interactions, the integral, and thus the free energy, can be decomposed into a sum of contributions [110]. It has been recognized that these components depend on the exact path taken in transforming one molecule into another [111, 112]. However, this does not eliminate their usefulness [113, 114].

Another type of decomposition is suggested by the Poisson-Boltzmann approach. In the linearized PB equation the electrostatic potential at any point is the sum of contributions from individual charges [42, 115]. Hence, the contribution of a charge j to the electrostatic free energy is the sum of one half of all other charges times the potential generated by charge j . The contribution of j to the electrostatic binding free energy is equal to its contribution in the complex minus its contribution in the unbound molecules. This contribution can be further subdivided into solvation, direct, and indirect terms. The solvation contribution is the loss of solvation free energy of group j upon binding. The direct contribution is the solvent-screened Coulomb interactions of group j with the charges of the other molecule. The indirect (or intramolecular) term is the

change in the interactions of j with the other charges of the molecule it belongs to due to the presence of the other molecule. The latter term is favorable to binding because the presence of the second molecule reduces the screening of electrostatic interactions within the first molecule. This term was found to make a significant contribution to binding [115].

Other theoretical formulations lead to different decompositions. In our recent work [81] we proposed a formalism where the free energy of a system with two macromolecules and solvent is written as

$$A = A^0 + kT \ln \left(\frac{3M}{V8^2} \right) + \int p(\mathbf{R}) d\mathbf{R} \int p(\mathbf{q}^A, \mathbf{q}^B | \mathbf{R}) W(\mathbf{q}^A, \mathbf{q}^B, \mathbf{R}) d\mathbf{q}^A d\mathbf{q}^B \\ + kT \int p(\mathbf{R}) \ln p(\mathbf{R}) d\mathbf{R} + kT \int p(\mathbf{R}) d\mathbf{R} \int p(\mathbf{q}^A, \mathbf{q}^B | \mathbf{R}) \ln p(\mathbf{q}^A, \mathbf{q}^B | \mathbf{R}) d\mathbf{q}^A d\mathbf{q}^B \quad (2)$$

The first term is the free energy of pure solvent and the second term the ideal contribution from translational motion. These terms are constant and make no contribution to the binding free energy. The third term is the average effective energy, the fourth term the relative translational/rotational entropy, and the last term the conformational entropy. The above equation is exact. However, the separation of the translational/rotational entropy from the conformational entropy is not unique because it depends on the choice for \mathbf{R} [52].

In general, entropic terms can be further broken down if we assume that certain degrees of freedom are independent (uncoupled) from each other. For example, if we can assume that the translational and rotational degrees of freedom are decoupled, that is

$$p(\mathbf{R}) = p(\mathbf{r}) p(\omega) \quad (3)$$

then the translational/rotational entropy term can be further split into translational and rotational terms:

$$kT \int p(\mathbf{r}) \ln p(\mathbf{r}) d\mathbf{r} + kT \int p(\omega) \ln p(\omega) d\omega \quad (4)$$

Also, the conformational entropy (last) term in Eq. 2 can be broken into two terms, one for each molecule, if we assume that their degrees of freedom are uncoupled ($p(\mathbf{q}^A, \mathbf{q}^B | \mathbf{R}) = p(\mathbf{q}^A | \mathbf{R}) p(\mathbf{q}^B | \mathbf{R})$).

For the unbound state the expression can be simplified, since \mathbf{q}^A and \mathbf{q}^B are now decoupled ($p(\mathbf{q}^A, \mathbf{q}^B | \mathbf{R}) = p(\mathbf{q}^A) p(\mathbf{q}^B)$) and $p(\mathbf{R})$ is constant (from normalization

$p(\mathbf{R}) = \frac{1}{V8^2}$). Also, the two macromolecules do not interact, so that $W = W_A + W_B$. Therefore,

$$A(A, B, \text{in } V) = A^0 + kT \ln \left(\frac{3M}{V8^2} \right) + kT \int p(\mathbf{q}^A) W_A d\mathbf{q}^A + kT \int p(\mathbf{q}^B) W_B d\mathbf{q}^B \\ + kT \frac{1}{V8^2} \ln \frac{1}{V8^2} d\mathbf{R} + kT \int p(\mathbf{q}^A) \ln p(\mathbf{q}^A) d\mathbf{q}^A + kT \int p(\mathbf{q}^B) \ln p(\mathbf{q}^B) d\mathbf{q}^B \quad (5)$$

The standard free energy change is then the difference between Eq. 2 and 5:

$$G = W - T \Delta S_{\text{tr/rot}} - T \Delta S_{\text{conf}}^A - T \Delta S_{\text{conf}}^B \quad (6)$$

where the terms on the right hand side are differences in the corresponding terms of Eqs. 2 and 5 and decoupling of the conformational degrees of freedom has been assumed.

The change in effective energy can be further decomposed. W contains two terms: E and G^{slv} . Most molecular mechanics force fields that are used for E

are pairwise additive. Therefore, E can be split into $E_{\text{inte}} + E_{\text{intra1}} + E_{\text{intra2}}$, where E_{inte} is the interaction energy between the two partners and the other two terms are the reorganization energies upon binding. Due to the pairwise additivity of the energy function, each of them can be further decomposed into group contributions. The decomposition of the solvation free energy is in principle a more complicated issue (a formal scheme for group decomposition of the solvation free energy has been presented [1]). In practice, most popular solvation models are decomposable. For example, the SASA model [116] assumes that each group contributes to the solvation free energy an amount proportional to its solvent accessible surface area. EEF1 [80] also assumes group additivity. Assuming group-decomposability of the solvation free energy, G^{slv} can be written as $G^{\text{slvA}} + G^{\text{slvB}}$ and each of the G^{slv} terms can be further decomposed into group contributions. This separation is also possible for the PB and GB approaches [115].

5. COMPONENTS OF THE BINDING FREE ENERGY

In this section the most common and intuitively appealing components of the binding affinity are listed and results for them obtained by different theoretical approaches are reviewed.

Direct Interactions

The direct interactions between the binding partners is the most obvious contribution to the binding affinity. They include van der Waals, electrostatic, and hydrogen bonding interactions. In most MM energy functions these interactions are pairwise additive. Therefore, one can easily define

contributions from particular atoms or groups. For all actually occurring binding complexes the direct interactions are expected to be favorable to binding. It is possible to have unfavorable direct interactions for putative binding complexes if, for example, like-charged groups are juxtaposed. Such binding modes would be highly unlikely and would be screened out by most, if not all, scoring functions. One ambiguity that may arise with respect to the definition of this contribution is whether the direct electrostatic interactions are screened or unscreened (as in the gas phase). For example, in the group decomposition based on continuum electrostatics, the direct term refers to solvent-screened interactions [115]. Alternatively, one could define the direct interactions as those occurring in the gas phase and include any solvent screening effects in the solvation term.

Desolvation

Binding usually involves the complete removal of the groups at the binding interface from water. The loss of interactions with the solvent is an important contribution to the binding free energy. Nonpolar groups have a positive solvation free energy [117]. Therefore, for such groups, both desolvation and the direct van der Waals interactions are favorable, so that nonpolar contacts (hydrophobic interactions) are a driving force to binding. The solvation free energy of polar groups is highly favorable. Therefore, desolvation of polar groups is unfavorable but the direct hydrogen bonds or electrostatic interactions between such groups are favorable. Therefore, the contribution of polar contacts can be favorable, unfavorable, or neutral, depending on the exact balance of desolvation and direct contributions.

Desolvation is often modeled by a surface area model [116]. That is, a model that assumes that the solvation free energy lost is proportional to the buried surface multiplied by a surface-tension-like coefficient that depends on the type of atom. Nonpolar groups have positive and polar groups negative "surface tensions". There are two types of parameters in such models: a) parameters derived from transfer of model compounds from water to a nonpolar phase [116]. These parameters should be used when the direct intermolecular van der Waals interactions are neglected because these interactions are implicitly accounted for by the solvation model, and b) parameters derived from transfer of model compounds from water to the gas phase [118]. This is a "pure" solvation model, appropriate for adding to molecular mechanics force fields.

Another method similar in spirit but different in form is the EEF1 solvation model [80]. This model is based on the idea that the solvation free energy of each group is modified by the presence of surrounding groups by an amount dependent on the volume and the distance of the surrounding groups. The distance dependence of the solvation free energy density was described by a Gaussian function, such that about 85% of the solvation free energy arises from the first solvation shell. In addition to this term, the ionic groups in proteins are neutralized and the linear distance-dependent dielectric constant is used for the electrostatic interactions. EEF1 gives modest deviations from crystal structures upon molecular dynamics simulations at room temperature,

unfolding pathways in agreement with explicit solvent simulations, and it discriminates native conformations from misfolded decoys [119]. It has been used to determine the folding free energy landscape of a α -hairpin [120], exploration of partially unfolded states of α -lactalbumin [121], studies of protein unfolding [122-125], identification of stable building blocks in proteins [126], and analysis of the energy landscape of polyalanine [127]. In EEF1 and surface area models the desolvation of polar groups dominates in magnitude the desolvation of nonpolar groups and makes the overall desolvation free energy unfavorable.

A more sophisticated evaluation of electrostatic desolvation effects is provided by the Poisson-Boltzmann method. As mentioned above, to obtain to full desolvation free energy one should combine PB with a model for the non-electrostatic part of solvation. Electrostatic desolvation is invariably unfavorable.

One effect that should be included in the desolvation term is the enhancement of intramolecular electrostatic interactions on binding [115]. When two molecules bind, the removal of solvent from the binding interface reduces the screening of electrostatic interactions within the two molecules. Although this sounds like an intramolecular energy effect and should at first sight be listed with the reorganization energies (see below), it occurs independent of any conformational adaptation. It is essentially a solvation effect.

Binding of highly charged molecules, such as nucleic acids, is accompanied by displacement and/or release of counterions, which could make a significant contribution to the binding free energy. One way to account for this is by use of the Poisson-Boltzmann equation. This method showed salt effects to favor association in one case [128] and disfavor it in another [129]. An alternative method treats counterions explicitly, assigns them a fractional charge based on counterion condensation theory, and assumes a certain entropic gain from the release of each counterion into the bulk [77]. In this method the average effect of counterions of 40 protein-DNA complexes is unfavorable [78].

Ligand Reorganization Energy

If the bound conformation of the ligand is different from the conformation of the free ligand in solution, the intramolecular energy of the ligand will make a contribution to the binding free energy. Ligand reorganization energies have been calculated using a Molecular Mechanics force field and the Generalized Born/Surface Area model [130]. For small ligands the reorganization energy (excluding solvation) was less than 3 kcal/mol. For larger ligands the authors felt that the energy function and/or the experimental structures are not reliable enough. The ligand reorganization energy was also considered by Vajda *et al.* [91], but the reported results include the transfer free energy of the ligand from water to octanol. Vieth *et al.* [131] compared receptor-bound structures of small ligands to energy minima resulting from systematic search of conformational space in solution, but their emphasis was on structure rather than energetics.

In our study of ligands of avidin and streptavidin [81], simulations of the ligands alone and bound to the protein were used to compare their intramolecular energies in the two states. For biotin, which is small and binds without significant rearrangements, the ligand reorganization energy was essentially zero. For the peptides the ligand reorganization energy was substantial (up to 10 kcal/mol) because when unbound the peptides can relax their structures to maximize the interactions between their own atoms. When they are bound to the protein, maximizing the interactions with the protein atoms leads to a compromise in the intramolecular energies.

Another issue that possibly needs to be considered is any ligand-ligand (or protein-protein) dissociation energy. For example, if the ligands do not exist entirely as monomers but partially associate in the unbound state, binding to a protein would require breaking ligand-ligand interactions. In this case, the true binding affinities would be lower than one would predict assuming monomeric ligand. In principle, such self-association can be detected by experiments at varying ligand concentrations.

Protein Reorganization Energy

Protein reorganization energy is the energy cost of adapting the conformation of the protein for optimal interactions with a ligand. This adaptation most often involves rearrangement of sidechains and closure of flexible loops. Crystallographic observations of protein reorganization upon ligand binding have been reported [e.g., 132].

Protein reorganization energies are most difficult to calculate because they correspond to a small difference between large numbers. Most work on binding affinities so far has either assumed a rigid protein or neglected the contribution of protein reorganization. Honig and coworkers recognized the potential importance of this quantity [59, 60] but did not calculate it. Horton & Lewis [133] attributed the difference in binding energy of BPTI to trypsin vs trypsinogen (about 10 kcal/mol) to the energy required to order the active site residues in trypsinogen. Hermans' method [54] includes implicitly the contribution of protein reorganization but will not be easy to apply to more complex, flexible solutes. Gilson and coworkers [74] have not so far considered protein or receptor flexibility. In Kollman's work [65] identical configurations are used for the bound and the unbound protein, therefore the protein reorganization energy is zero. However, in a study of protein-RNA binding, where the structures of both the RNA-bound and free protein are known, the MM/PBSA method was used to calculate a 10 kcal/mol reorganization energy for the protein [64]. Jayaram *et al.* [77] applied a similar method to calculate the energy of reorganization of DNA upon binding a protein and obtained the value 63 kcal/mol. Baginski *et al.* obtained a DNA reorganization energy of +32.2 kcal/mol upon binding of an intercalating antibiotic [134]. Noskov and Lim found that protein reorganization is important, but did not explicitly calculate protein reorganization energies [135].

In our study of biotin/streptavidin [81] we found that the protein reorganization energy makes a large and positive contribution to the binding free energy, from about 10 to 30 kcal/mol depending on the ligand. This results from the fact that binding to the ligand causes some compromise in intra-protein interactions. These large values are a sum of many van der Waals interactions and hydrogen bonds, such as the reorientation of a serine sidechain to hydrogen bond to the protein backbone. The largest contributions to W^{prot} come from van der Waals, electrostatic, and desolvation interactions. The protein reorganization energy tends to increase with ligand size. This scaling compensates for the scaling of the favorable W^{inte} with ligand size. Another unfavorable contribution that scales with ligand size is the ligand conformational entropy and ligand reorganization energy. These compensating favorable (W^{inte}) and unfavorable (W^{prot} , W^{lig} , $S_{\text{conf,lig}}$) contributions combine into a ΔG that shows little correlation with ligand size.

Translational, Rotational, And Vibrational Entropy (S_{trans/Rot})

When two molecules bind they lose the freedom to translate and rotate independently from each other. It has long been recognized that this makes an unfavorable contribution to binding [136]. Numerous estimates of the translational and rotational entropy loss upon binding have been published. Early work assumed that one of the partners loses entirely its translational and rotational entropy, which were estimated from ideal gas expressions [137]. It was soon recognized that this leads to overestimation of the magnitude of the entropy loss because part of the translational and rotational entropy is transformed into vibrational entropy in the complex [138]. Tidor and Karplus (TK) [139] used normal mode analysis to estimate that the contribution of residual vibrations to the dimerization of insulin at 300 K is about 7 kcal/mol. Taking that into account, they estimated the association entropy cost at 300 K to be about 20 kcal/mol. This approach was also applied to other systems [65, 135]. Finkelstein and Janin (FJ) [140] followed a different approach. They assumed a certain residual amplitude of translational and orientational motion in the bound complex and obtained the entropy from the ratio of conformational space available in the bound and free state, for the given standard state. They estimated a translational and rotational entropy loss similar to that of Tidor & Karplus, about 15 kcal/mol.

Others obtained much smaller estimates. Murphy *et al.* [141] deduced the magnitude of $S^{\text{rot,tr}}$ by subtracting estimates of all other contributions from experimental values. They obtained much smaller values than FJ and TK and suggested that the cratic correction ($R \ln x$, where $x=1/55.5$ is the mole fraction of the associating molecules in solution) is a better estimate for the loss of translational entropy upon association. Novotny *et al.* [142] also employed the cratic correction. However, the cratic correction has been shown to be theoretically untenable [52, 143]. The translational entropy should depend on the concentration of the solute but not on the density of the solvent [52]. Some researchers [141, 144, 145] argued that use of gas phase

equations is not appropriate because in the condensed phase the motions of the associating molecules are restricted by the solvent molecules. However, this effect is part of the solvation entropy [146]. Horton & Lewis [133] fitted data for a number of protein-protein complexes in terms of a constant contribution from $S^{\text{rot, tr}}$ and a desolvation/interaction contribution estimated based on polar and nonpolar buried surface area [116]. They obtained a value for $T S^{\text{rot, tr}}$ about 6 kcal/mol, significantly smaller than FJ and TK.

Rashin [147] also suggested that the estimates of FJ and TK are too large based on the following argument. Experiments have shown that when the protein barnase is cleaved at an appropriate site (surface or turn site) it can easily refold [148]. The stability of the cleaved protein (9 kcal/mol at 1M standard state) is only slightly lower than the stability of the intact protein (about 10 kcal/mol). In order to fold, the cleaved protein must overcome, in addition to its conformational entropy, the translational and rotational entropy of the two fragments. Since the stability of a protein is typically about 10 kcal/mol, if the translational and rotational entropy cost is 15-20 kcal/mol at 1M, the cleaved protein should not be able to refold at all. Similar experiments have been reported on other proteins [149, 150]. The experiment of Tamura & Privalov [151] is the opposite of cleaving a protein. They introduced a crosslink into a noncovalently dimerizing protein and compared the thermodynamics of unfolding of the wild type and the crosslinked dimer. They found the crosslinked dimer to be only slightly more stable than the natural dimer and estimated the entropic cost of crosslinking to be about 5 e.u. at 1 M standard state (1.5 kcal/mol at 300 K).

This "paradox" can be resolved in the following way (see also [152]). Fig. 1 shows schematically unfolding of the intact and cleaved protein. We are considering only "internal" protein entropy changes; i.e., excluding solvation entropy. For the intact protein the change in entropy is entirely conformational. For the cleaved protein, it is a) conformational for the two fragments, and b) translational and rotational. The argument of Rashin and of Tamura and Privalov implies that the only difference between processes 1 and 2 is $S^{\text{rot, tr}}$ which, according to FJ and TK is of the order of 15-20 kcal/mol. Therefore, the cleaved protein should never fold if these estimates were correct. What is being overlooked in the above argument is that the conformational entropy contribution is not the same in the two processes, i.e., $S_1^{\text{conf}} > S_2^{\text{conf}}$; some of the translational and rotational entropy in process 2 is equivalent to conformational entropy in process 1. In other words, $S_1^{\text{conf}} > S_2^{\text{conf}}$. The two sections of the intact protein corresponding to the fragments of the cleaved protein certainly have a lot of freedom to translate and rotate with respect to each other in the unfolded state, despite the covalent attachment between them. This, in process 1, is included in S_1^{conf} , whereas in process 2 it is included in $S^{\text{rot, tr}}$. Therefore, the difference in entropy change between processes 1 & 2 is smaller than the $S^{\text{rot, tr}}$ one would expect.

Recently, estimates of $T S^{\text{trans/rot}}$ have been obtained by more rigorous calculations based on evaluation of configurational integrals of free energy simulations. Luo and

Gilson calculated 7 kcal/mol for the binding of adenine to synthetic adenine receptors [74]. Hermans and Wang also obtained this value for the binding of benzene to a lysozyme cavity [54]. Brady & Sharp [153] estimated the entropy of association to be 5.8 kcal/mol (1 M standard state) based on an analysis of cyclic dipeptide crystals. This value was calculated by subtracting the intramolecular vibrational entropy from the experimental sublimation entropy. In our study of biotin/streptavidin [81] we followed the FJ approach but obtained the amplitudes of translational and rotational motion from the MD trajectories and corrected for a non-flat probability distribution within the allowed bounds. For the orientational entropy we calculated the distribution of orientations of the ligand with respect to the binding site. The resultant translational (1M standard state) and rotational entropy losses were found to be between 2-3 kcal/mol each, opposing binding. For the translational and rotational entropy Finkelstein and Janin estimated 6.9 kcal/mol and 7.2 to 9 kcal/mol, respectively because the amplitudes of motion assumed by these authors are smaller than those found in our simulations. They assumed a translational amplitude of 0.25 Å, whereas we find typical translational amplitudes of 4-6 Å. They also assume a rotational amplitude of 2-5 degrees, whereas we find much broader distribution of the Euler angles.

Thus, a number of recent theoretical results suggests that the translational/rotational entropy loss is smaller than previously calculated, not because the use of gas phase equations is inappropriate, but because the residual movements in the bound complex were underestimated. The values obtained by normal mode analysis may be too large due to the harmonic approximation or to the fact that the analysis is done in vacuum, which may underestimate the range of movements available for the ligand in the complex.

Ligand Conformational Entropy

A flexible ligand that is free to sample many conformations in solution and is forced to adopt one of these conformations when bound to a receptor loses conformational entropy. That this makes an unfavorable contribution to binding is well known in the drug design community, who always try to fix a ligand in its productive conformation. A quantitative estimate of this contribution can be made by enumerating the possible conformations of a ligand in solution and applying the Boltzmann expression: $S = -k \ln W$. This expression assumes that all conformations have the same probability. A more accurate expression that does not make this assumption is the Gibbs expression: $S = -k \ln \sum_i P_i \ln P_i$. Systematic explorations of conformational space can provide accurate values for small ligands [74].

In practice, because effective sampling of all the conformational space of a ligand is very difficult, the estimation of conformational entropy loss is usually based on the number of rotatable torsional angles of the ligand. Pickett & Sternberg examined the variation of sidechain angles for each amino acid in the database of protein structures and developed a popular empirical scale for sidechain entropy loss [35]. Their empirical scale suggests 0.54 kcal/mol (at 300 K) for fixing a single bond. Melting of liquid alkanes suggests 0.45 kcal/mol for the same quantity [154].

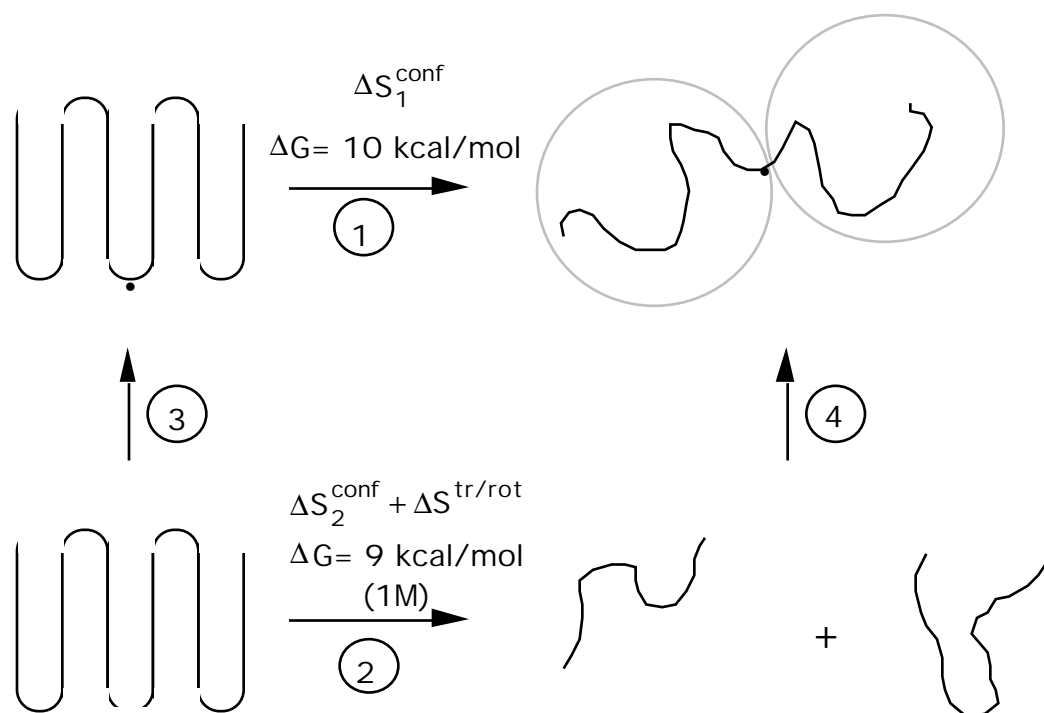


Fig. (1). Effect of chain cleavage on binding/folding thermodynamics

Theoretical calculations gave 0.3 kcal/mol [74] and 0.48 kcal/mol [76] for the conformational entropy contribution per rotatable bond. Similar values have been obtained by Monte Carlo simulations of peptides [155] and energy maps of small organic molecules [156]. An average value of entropy loss per torsional degrees of freedom is 0.4 kcal/mol.

Protein Sidechain Conformational Entropy

Upon binding, several protein sidechains that were exposed in the free protein become buried in the binding interface. This may cost entropy if the sidechains could sample many rotamers in the free state and have to select one rotamer in the bound state. In most binding affinity calculations this contribution is estimated empirically from the number of torsional angles in the sidechain (see above).

In our biotin/streptavidin study the conformational entropy change for protein sidechains upon ligand binding was estimated by comparing the distribution of χ angles in the complex and in the free protein for those sidechains that had atoms within 5 Å from any ligand atom. The integrals of $p \ln(p)$ of the two distributions were used to estimate the entropy change, assuming that the χ_1 and χ_2 distributions were independent from each other. The total entropy was obtained by summing the contributions of each dihedral angle and was found to be rather small, in agreement with other studies [59]. The entropy change was found to be negative, except for thiobiotin, where it was small and positive. In the latter case the protein sidechains have more conformational freedom in the complex than in the apo-protein. Such increased disorder upon binding has been observed in some NMR studies [157, 158].

Global Entropic Contributions

One important recent development was the use of NMR relaxation to probe the changes in the dynamics of the protein backbone and sidechains upon ligand binding at the ps-ns timescale [159, 160]. These changes in dynamics can be related to entropy changes. For example, the binding of a pheromone to a protein was found to increase the backbone flexibility of the protein [157]. Specifically, 68 residues, many of which were not in direct contact with the peptide, showed reductions in the motional order parameter, which indicates increased disorder. A quantitative estimate of the entropic consequence of this increased flexibility gave a value comparable to the total free energy of binding. Another study observed an increase in sidechain flexibility of a signal transduction protein upon binding of a 46-residue peptide [158]. The increase was more pronounced for residues away from the peptide binding interface, while methyl groups right at the interface experienced a restriction of mobility. In contrast, in the binding of a peptide to calmodulin, the backbone dynamics was found to be unaffected but the sidechains experienced restricted motion [161]. The estimated contribution to the change in entropy from these restricted motions was very large (-35 kcal/mol at 35 °C). These observations raise questions of validity for most theoretical calculations which assume conformational entropy changes localized near the binding site. If binding indeed incurs changes in the global dynamics of a protein, new theoretical methods will have to be developed that include such entropic effects in binding free energy calculations. That will require accurate reproduction of the dynamics of unbound and bound proteins. Even more difficult would be the inclusion of long time scale motions (μ s-ms) in the entropy calculations. Such calculations do not seem to be feasible at this point.

6. CONCLUSIONS AND FUTURE WORK

A multitude of methods at various levels of rigor and speed are available today for the estimation of relative and absolute binding affinities. However, all methods are fraught with difficulties. Simple and fast methods make highly simplifying approximations and may neglect important components of the binding free energy. More sophisticated methods rely on sampling of the conformational space of the complex and the free molecules. Prediction accuracy requires this sampling to be realistic, which in turn requires accurate force fields and simulation protocols. It is not certain that currently available force fields are adequate for this task. In addition, certain methods require taking the small difference between large numbers and therefore suffer from large statistical uncertainties. The good agreement with experiment often reported should not give one a false sense of confidence about the capabilities of current affinity prediction methods.

Our physical understanding of binding thermodynamics has been significantly enhanced over the past decade. Much of this understanding comes from analyzing the components of binding affinity. The decomposition of the binding free energy differs depending on the theoretical framework used. Certain components, such as translational/rotational entropy, are well recognized and have been debated for over 50 years. Recent work seems to lead to a consensus that this contribution at 1M standard state is about 6-8 kcal/mol. Other components, such as the protein reorganization energy, are often neglected. This quantity may be similar when different ligands bind to the same protein and thus cancel in calculations of relative binding free energies. However, it is certainly not negligible in absolute affinity calculations. Future studies should focus on further characterization of protein and ligand reorganization in different systems and with different force fields and implicit solvation models.

One issue that still needs to be clarified is the role of ordered interfacial water molecules in bound complexes. Typically, in biomolecular complexes water is completely removed from the interface between the binding partners. However, this is not always the case [162, 163]. In the *trp* repressor-operator complex no direct hydrogen bonds to the nucleic acid bases were observed that could explain the binding specificity [164]. Instead, a number of water-mediated polar contacts to the bases were observed, as well as six water-mediated hydrogen bonds to the phosphate groups of DNA. The structure of lysozyme complexed with a monoclonal antibody revealed four water molecules buried and others forming a hydrogen bonding network around the interface [165]. The L-arabinose binding protein binds L-arabinose and D-galactose similarly, except for a water molecule that is incorporated at the binding interface in the case of arabinose [166]. Interestingly, the binding affinity is slightly higher for arabinose. The oligopeptide binding protein OppA binds two to five-residue peptides regardless of sequence. The crystal structure revealed that the peptide backbone binds directly to the protein but the sidechains are enclosed in hydrated cavities [167]. The amount of water in the cavity varies with the size of the sidechain but the affinity remains about the same. In fact, the affinity is slightly larger for the Ala sidechain, which incorporates 7 water molecules, than for the *Trp* sidechain, which incor-

porates 4 water molecules [168]. The specificity of binding of EcoRI protein to DNA has been found to depend strongly on osmotic pressure [169]. This suggests that water molecules are involved at the DNA - protein interface. On the other hand, displacement of bound water by ligands is sometimes observed to lead to stronger binding. The best-known example is the design of cyclic urea inhibitors of HIV protease [170]. In crystal structures of this protein with other inhibitors an ordered water molecule was observed near the binding site. The cyclic urea inhibitors were designed to displace and mimic the interactions of this water molecule based on the reasoning that entropy would be gained by the release of this bound water into the bulk. Indeed, the new inhibitors bound more strongly to the protein. In another system, the presence of an ordered water molecule at the interface of a resistant mutant of DNA gyrase with the antibiotic novobiocin was thought to be responsible for the weaker binding of the drug to the mutant protein [171]. It is unclear whether current theoretical and computational methods can reproduce and rationalize the above results.

Another issue that needs to be experimentally clarified is the extent of dynamics of ligands in protein binding sites. Do the ligands remain close to the crystal structures or they experience large fluctuations in their positions? There is some evidence that ligand binding is not as static as it appears in crystal structures. For example, application of temperature-jump relaxation spectroscopy with time-resolved fluorescence on the binding of NADH to lactate dehydrogenase showed that there are multiple, structurally different bound states at equilibrium [172]. Similar observations were reported for NADP⁺ binding to dihydrofolate reductase using NMR [173]. Ladbury *et al.* [174] found that the pY residue of a phosphotyrosine peptide bound to a SH2 domain is fixed in its binding pocket but the rest of the peptide is very mobile. Sometimes multiple binding modes are visible in crystal structures as well [175-179]. Brem & Dill used an exactly solvable lattice model to show that multiple ligand binding modes can lead to large errors in the calculation of binding affinities [180].

The clarification of such fundamental questions can have beneficial influence on rational drug design. For example, better understanding of the components of binding free energies and their magnitude could affect the design of scoring functions for docking. Characterization of the conformational adaptation that proteins undergo upon binding will likely be useful in docking studies that incorporate protein flexibility. Understanding of the role of ordered water molecules will give guidance as to when these should be displaced and when they should be retained. Finally, further improvement of these energy functions will lead to more faithful simulations and more accurate binding thermodynamics, which, in conjunction with longer simulations and more systematic exploration of conformational space should lead to improvements in binding affinity predictions.

ACKNOWLEDGEMENTS

This work was supported by grant 34680-G4 from the Petroleum Research Fund, administered by the American Chemical Society and by the National Science Foundation (DBI-9974621).

ABBREVIATIONS

- MD : molecular dynamics
 MM : Molecular mechanics
 PB: : Poisson-Boltzmann
 SA : surface area
 FE : free energy

REFERENCES

- [1] Lazaridis, T.; Karplus, M., In E. DiCera (Ed.): Thermodynamics in biology, Oxford University Press, Oxford **2001**, p. 3.
- [2] Gane, P.J.; Dean, P.M. *Curr. Opin. Struct. Biol.*, **2000**, *10*, 401.
- [3] de Silva, A.P.; Gunaratne, H.Q.N.; Gunnlaugsson, T.; Huxley, A.J.M.; McCoy, C.P.; Rademacher, J.T.; Rice, T.E. *Chem. Rev.*, **1997**, *97*, 1515.
- [4] Deng, Q.; German, I.; Buchanan, D.; Kennedy, R.T. *Anal. Chem.*, **2001**, *73*, 5415.
- [5] Sellergren, B. *J. Chromatog. A*, **2001**, *906*, 227.
- [6] McCammon, J.A. *Curr. Opin. Struct. Biol.*, **1998**, *8*, 245.
- [7] Ajay.; Murcko, M.A. *J. Med. Chem.*, **1995**, *38*, 4953.
- [8] Tembe, B.; McCammon, J.A. *Comput. & Chem.*, **1984**, *8*, 281.
- [9] Beveridge, D.L.; DiCapua, F.M. *Ann. Rev. Biophys. Chem.*, **1989**, *18*, 431.
- [10] McCammon, J.A. *Curr. Opin. Struct. Biol.*, **1991**, *1*, 196.
- [11] Kollman, P.A. *Acc. Chem. Res.*, **1996**, *29*, 461.
- [12] Kollman, P.A. *Chem. Rev.*, **1993**, *93*, 2395.
- [13] Fleischman, S.H.; Brooks, C.L., III *Proteins*, **1990**, *7*, 52.
- [14] Lau, F.T.K.; Karplus, M. *J. Mol. Biol.*, **1994**, *236*, 1049.
- [15] Merz, K.M. Jr.; Murcko, M.A.; Kollman, P.A. *J. Am. Chem. Soc.*, **1991**, *113*, 4484.
- [16] Essex, J.W.; Severance, D.L.; Tirado-Rives, J.; Jorgensen, W.L. *J. Phys. Chem. B*, **1997**, *101*, 9663.
- [17] Singh, S.B.; Ajay.; Wemmer, D.E.; Kollman, P.A. *Proc. Natl. Acad. Sci. USA*, **1994**, *91*, 7673.
- [18] Hirono, S.; Kollman, P.A. *Prot. Engin.*, **1991**, *4*, 233.
- [19] Singh, S.B.; Ajay.; Wemmer, D.E.; Kollman, P.A. *Proc. Natl. Acad. Sci. USA*, **1994**, *91*, 7673.
- [20] Archontis, G.; Simonson, T.; Moras, D.; Karplus, M. *J. Mol. Biol.*, **1998**, *275*, 823.
- [21] Pearlman, D.A.; Charifson, P.S. *J. Med. Chem.*, **2001**, *44*, 3417.
- [22] Sun, Y.; Caldwell, J.W.; Kollman, P.A. *J. Phys. Chem.*, **1995**, *99*, 10081.
- [23] Sun, Y.; Kollman, P.A. *J. Am. Chem. Soc.*, **1995**, *117*, 3599.
- [24] Salvatierra, D.; Sanchez-Ruiz, X.; Garduno, R.; Cervello, E.; Jaime, C.; Virgili, A.; Sanchez-Ferrando, F. *Tetrahedron*, **2000**, *56*, 3035.
- [25] McDonald, N.A.; Duffy, E.M.; Jorgensen, W.L. *J. Am. Chem. Soc.*, **1998**, *120*, 5104.
- [26] Jorgensen, W.L.; Nguyen, T.B. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 1194.
- [27] Kong, X.; Brooks, C.L., III *J. Chem. Phys.*, **1996**, *105*, 2414.
- [28] Banba, S.; Guo, Z.; Brooks, C.L., III *J. Phys. Chem. B*, **2000**, *104*, 6903.
- [29] Jarque, C.; Tidor, B. *J. Phys. Chem. B*, **1997**, *101*, 9362.
- [30] Guo, Z.; Brooks, C.L., III *J. Am. Chem. Soc.*, **1998**, *120*, 1920.
- [31] Honig, B.; Nicholls, A. *Science*, **1995**, *268*, 1144.
- [32] Jackson, R.M.; Sternberg, M.J.E. *J. Mol. Biol.*, **1995**, *250*, 258.
- [33] Zhang, T.; Koshland, D.E. Jr. *Protein Sci.*, **1996**, *5*, 348.
- [34] Tawa, G.J.; Topol, I.A.; Burt, S.K.; Erickson, J.W. *J. Am. Chem. Soc.*, **1998**, *120*, 8856.
- [35] Pickett, S.D.; Sternberg, M.J.E. *J. Mol. Biol.*, **1993**, *231*, 825.
- [36] Novotny, J.; Brucoleri, R.E.; Davis, M.; Sharp, K.A. *J. Mol. Biol.*, **1997**, *268*, 401.
- [37] Chong, L.T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P.A. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 14330.
- [38] Wang, W.; Kollman, P.A. *J. Mol. Biol.*, **2000**, *303*, 567.
- [39] Domini, O.A.T.; Kollman, P.A. *J. Med. Chem.*, **2000**, *43*, 4180.
- [40] Dominy, B.N.; Brooks, C.L., III *Proteins*, **1999**, *36*, 318.
- [41] Kasper, P.; Christen, P.; Gehring, H. *Proteins*, **2000**, *40*, 185.
- [42] Archontis, G.; Simonson, T.; Karplus, M. *J. Mol. Biol.*, **2001**, *306*, 307.
- [43] Caflisch, A.; Fischer, S.; Karplus, M. *J. Comput. Chem.*, **1997**, *18*, 723.
- [44] Misra, V.K.; Hecht, J.L.; Yang, A.-S.; Honig, B. *Biophys. J.*, **1998**, *75*, 2262.
- [45] Misra, V.K.; Honig, B. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 4691.
- [46] Lee, L.-P.; Tidor, B. *J. Chem. Phys.*, **1997**, *106*, 8681.
- [47] Kangas, E.; Tidor, B. *J. Chem. Phys.*, **1998**, *109*, 7522.

- [48] Chong, L.T.; Dempster, S.E.; Hendsch, Z.S.; Lee, L.-P.; Tidor, B. *Protein Sci.*, **1998**, *7*, 206.
- [49] Lee, L.-P.; Tidor, B. *Protein Sci.*, **2001**, *10*, 362.
- [50] Lee, L.-P.; Tidor, B. *Nature Struct. Biol.*, **2001**, *8*, 73.
- [51] Kangas, E.; Tidor, B. *J. Chem. Phys.*, **2000**, *112*, 9120.
- [52] Gilson, M.K.; Given, J.A.; Bush, B.L.; McCammon, J.A. *Biophys. J.*, **1997**, *72*, 1047.
- [53] Janin, J. *Proteins*, **1996**, *24*, i-ii.
- [54] Hermans, J.; Wang, L. *J. Am. Chem. Soc.*, **1997**, *119*, 2707.
- [55] Helms, V.; Wade, R.C. *J. Am. Chem. Soc.*, **1998**, *120*, 2710.
- [56] Roux, B.; Nina, M.; Pomes, R.; Smith, J.C. *Biophys. J.*, **1996**, *71*, 670.
- [57] Miyamoto, S.; Kollman, P.A. *Proteins*, **1993**, *16*, 226.
- [58] Dixit, S.B.; Chipot, C. *J. Phys. Chem. A*, **2001**, *105*, 9795.
- [59] Froloff, N.; Windemuth, A.; Honig, B. *Prot. Sci.*, **1997**, *6*, 1293.
- [60] Polticelli, F.; Ascenzi, P.; Bolognesi, M.; Honig, B. *Prot. Sci.*, **1999**, *8*, 2621.
- [61] Srinivasan, J.; Cheatham, T.E., III; Cieplak, P.; Kollman, P.A.; Case, D.A. *J. Am. Chem. Soc.*, **1998**, *120*, 9401.
- [62] Massova, I.; Kollman, P.A. *J. Am. Chem. Soc.*, **1999**, *121*, 8133.
- [63] Huo, S.; Massova, I.; Kollman, P.A. *J. Comp. Chem.*, **2002**, *23*, 15.
- [64] Reyes, C.M.; Kollman, P.A. *J. Mol. Biol.*, **2000**, *297*, 1145.
- [65] Kuhn, B.; Kollman, P.A. *J. Med. Chem.*, **2000**, *43*, 3786.
- [66] Lee, M.R.; Duan, Y.; Kollman, P.A. *Proteins*, **2000**, *39*, 309.
- [67] Lee, M.R.; Baker, D.; Kollman, P.A. *J. Am. Chem. Soc.*, **2001**, *123*, 1040.
- [68] Vorobjev, Y.N.; Hermans, J. *Biophys. Chem.*, **1999**, *78*, 195.
- [69] Vorobjev, Y.N.; Almagro, J.C.; Hermans, J. *Proteins*, **1998**, *32*, 399.
- [70] Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T. *J. Am. Chem. Soc.*, **1990**, *112*, 6127.
- [71] Qiu, D.; Shenkin, P.S.; Hollinger, F.P.; Still, W.C. *J. Phys. Chem. A*, **1997**, *101*, 3005.
- [72] Dominy, B.N.; Brooks, C.L., III. *J. Phys. Chem. B*, **1999**, *103*, 3765.
- [73] Head, M.S.; Given, J.A.; Gilson, M.K. *J. Phys. Chem. A*, **1997**, *101*, 1609.
- [74] Luo, R.; Gilson, M.K. *J. Am. Chem. Soc.*, **2000**, *122*, 2934.
- [75] Mardis, K.L.; Luo, R.; Gilson, M.K. *J. Mol. Biol.*, **2001**, *309*, 507.
- [76] Wang, J.; Szewczuk, Z.; Yue, S.-Y.; Tsuda, Y.; Konishi, Y.; Purisima, E.O. *J. Mol. Biol.*, **1995**, *253*, 473.
- [77] Jayaram, B.; McConnell, K.J.; Dixit, S.B.; Beveridge, D.L. *J. Comput. Phys.*, **1999**, *151*, 333.
- [78] Jayaram, B.; McConnell, K.J.; Dixit, S.B.; Das, A.; Beveridge, D.L. *J. Comput. Chem.*, **2002**, *23*, 1.
- [79] Blakaj, D.M.; McConnell, K.J.; Beveridge, D.L.; Baranger, A.M. *J. Am. Chem. Soc.*, **2001**, *123*, 2548.
- [80] Lazaridis, T.; Karplus, M. *Proteins*, **1999**, *35*, 133.
- [81] Lazaridis, T.; Masunov, A.; Gandolfo, F. *Proteins*, **2002**, *47*, 194.
- [82] Åqvist, J.; Medina, C.; Samuelson, J.-E. *Prot. Engin.*, **1994**, *7*, 385.
- [83] Hansson, T.; Aqvist, J. *Prot. Engin.*, **1995**, *8*, 1137.
- [84] Wang, J.; Dixon, R.; Kollman, P.A. *Proteins*, **1999**, *34*, 69.
- [85] Wang, W.; Wang, J.; Kollman, P.A. *Proteins*, **1999**, *34*, 395.
- [86] Smith, R.H. Jr.; Jorgensen, W.L.; Tirado-Rives, J.; Lamb, M.L.; Janssen, P.A.J.; Michejda, C.J.; Smith, M.B.K. *J. Med. Chem.*, **1998**, *41*, 5272.
- [87] Lamb, M.L.; Tirado-Rives, J.; Jorgensen, W.L. *Bioorg. Medic. Chem.*, **1999**, *7*, 851.
- [88] Sham, Y.Y.; Chu, Z.T.; Tao, H.; Warshel, A. *Proteins*, **2000**, *39*, 393.
- [89] Vajda, S.; Sippl, M.; Novotny, J. *Curr. Opin. Struct. Biol.*, **1997**, *7*, 222.
- [90] Bardi, J.S.; Luque, I.; Freire, E. *Biochemistry*, **1997**, *36*, 6588.
- [91] Vajda, S.; Weng, Z.; Rosenfeld, R.; Delisi, C. *Biochemistry*, **1994**, *33*, 13977.
- [92] Weng, Z.; Vajda, S.; Delisi, C. *Prot. Sci.*, **1996**, *5*, 614.
- [93] Jain, A.N. *J. Comp.-Aided Mol. Des.*, **1996**, *10*, 427.
- [94] Böhm, H.-J. *J. Comp.-Aided Mol. Des.*, **1998**, *12*, 309.
- [95] Eldridge, M.D.; Murray, C.W.; Auton, T.R.; Paolini, G.V.; Mee, R.P. *J. Comp.-Aided Mol. Des.*, **1997**, *11*, 425.
- [96] Tame, J.R.H. *J. Comp.-Aided Mol. Des.*, **1999**, *13*, 99.
- [97] Gohlke, H.; Klebe, G. *Curr. Opin. Struct. Biol.*, **2001**, *11*, 231.
- [98] Meng, E.C.; Shoichet, B.K.; Kuntz, I.D. *J. Comp. Chem.*, **1992**, *13*, 505.
- [99] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. *J. Mol. Biol.*, **1997**, *267*, 727.

- [100] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.*, **1996**, *261*, 470.
- [101] Bohm, H.J. *J. Comp-Aided Mol. Des.*, **1994**, *8*, 243.
- [102] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. *J. Comput. Chem.*, **1998**, *19*, 1639.
- [103] Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. *Proteins*, **1999**, *37*, 88.
- [104] Zou, X.; Sun, Y.; Kuntz, I.D. *J. Am. Chem. Soc.*, **1999**, *121*, 8033.
- [105] Shoichet, B.K.; Leach, A.R.; Kuntz, I.D. *Proteins*, **1999**, *34*, 4.
- [106] Dill, K.A. *J. Biol. Chem.*, **1997**, *272*, 701.
- [107] Mark, A.E.; van Gunsteren, W.F. *J. Mol. Biol.*, **1994**, *240*, 167.
- [108] Southall, N.T.; Dill, K.A. *J. Phys. Chem. B*, **2000**, *104*, 1326.
- [109] Lazaridis, T. *Acc. Chem. Res.*, **2001**, *34*, 931.
- [110] Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. *Science*, **1989**, *244*, 1069.
- [111] Simonson, T.; Brünger, A.T. *Biochemistry*, **1992**, *31*, 8661.
- [112] Brooks, C.L.I., In C. R. A. Catlow, S. Parker, M. Allen (Eds.): Computer modeling of Fluids Polymers and Solids, Kluwer Academic Publishers, Dordrecht **1990**, p. 289.
- [113] Boresch, S.; Archontis, G.; Karplus, M. *Proteins*, **1994**, *20*, 25.
- [114] Boresch, S.; Karplus, M. *J. Mol. Biol.*, **1995**, *254*, 801.
- [115] Hendsch, Z.S.; Tidor, B. *Prot. Sci.*, **1999**, *8*, 1381.
- [116] Eisenberg, D.; McLachlan, A.D. *Nature*, **1986**, *319*, 199.
- [117] Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.*, **1984**, *81*, 2016.
- [118] Wesson, L.; Eisenberg, D. *Protein Science*, **1992**, *1*, 227.
- [119] Lazaridis, T.; Karplus, M. *J. Mol. Biol.*, **1999**, *288*, 477.
- [120] Dinner, A.R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 9068.
- [121] Paci, E.; Smith, L.J.; Dobson, C.M.; Karplus, M. *J. Mol. Biol.*, **2001**, *306*, 329.
- [122] Lazaridis, T.; Karplus, M. *Science*, **1997**, *278*, 1928.
- [123] Inuzuka, Y.; Lazaridis, T. *Proteins*, **2000**, *41*, 21.
- [124] Paci, E.; Karplus, M. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 6521.
- [125] Paci, E.; Karplus, M. *J. Mol. Biol.*, **1999**, *288*, 441.
- [126] Kumar, S.; Sham, Y.Y.; Tsai, C.-J.; Nussinov, R. *Biophys. J.*, **2001**, *80*, 2439.
- [127] Levy, Y.; Jortner, J.; Becker, O.M. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 2188.
- [128] Zacharias, M.; Luty, B.A.; Davis, M.E.; McCammon, J.A. *Biophys. J.*, **1992**, *63*, 1280.
- [129] Misra, V.K.; Hecht, J.L.; Sharp, A.; Honig, B. *J. Mol. Biol.*, **1994**, *238*, 264.
- [130] Bostrom, J.; Norrby, P.-O.; Liljefors, T. *J. Comp.-Aided Mol. Design*, **1998**, *12*, 383.
- [131] Vieth, M.; Hirst, J.D.; Brooks, C.L., III. *J. Comp.-Aided Mol. Des.*, **1998**, *12*, 563.
- [132] Fritz, T.A.; Tondi, D.; Finer-Moore, J.S.; Costi, M.P.; Stroud, R.M. *Chem. & Biol.*, **2001**, *8*, 981.
- [133] Horton, N.; Lewis, M. *Prot. Sci.*, **1992**, *1*, 169.
- [134] Baginski, M.; Fogolari, F.; Briggs, J.M. *J. Mol. Biol.*, **1997**, *274*, 253.
- [135] Noskov, S.Y.; Lim, C. *Biophys. J.*, **2001**, *81*, 737.
- [136] Brady, G.P.; Sharp, K.A. *Curr. Opin. Struct. Biol.*, **1997**, *7*, 215.
- [137] Doty, P.; Myers, G.E. *Disc. Farad. Soc.*, **1953**, *13*, 51.
- [138] Steinberg, I.Z.; Scheraga, H.A. *J. Biol. Chem.*, **1963**, *238*, 172.
- [139] Tidor, B.; Karplus, M. *J. Mol. Biol.*, **1994**, *238*, 405.
- [140] Finkelstein, A.V.; Janin, J. *Prot. Engin.*, **1989**, *3*, 1.
- [141] Murphy, K.P.; Xie, D.; Thompson, K.S.; Amzel, L.M.; Freire, E. *Proteins*, **1994**, *18*, 63.
- [142] Novotny, J.; Bruccoleri, R.E.; Saul, F.A. *Biochemistry*, **1989**, *28*, 4735.
- [143] Holtzer, A. *Biopolymers*, **1995**, *35*, 595.
- [144] Amzel, L.M. *Proteins*, **1997**, *28*, 144.
- [145] Privalov, P.L.; Tamura, A. *Prot. Engin.*, **1999**, *12*, 187.
- [146] Lazaridis, T. *J. Phys. Chem.*, **1998**, *102*, 3531.
- [147] Rashin, A.A. *Prog. Biophys. Mol. Biol.*, **1993**, *60*, 73.
- [148] Sancho, J.; Fersht, A.R. *J. Mol. Biol.*, **1992**, *224*, 741.
- [149] Neira, J.L.; Davis, B.; Ladurner, A.G.; Buckle, A.M.; de Prat Gay, G.; Fersht, A.R. *Fold. & Des.*, **1996**, *1*, 189.
- [150] Georgescu, R.E.; Braswell, E.H.; Zhu, D.; Tasayco, M.L. *Biochemistry*, **1999**, *38*, 13355.
- [151] Tamura, A.; Privalov, P.L. *J. Mol. Biol.*, **1997**, *273*, 1048.
- [152] Karplus, M.; Janin, J. *Protein Engineering*, **1999**, *12*, 185.
- [153] Brady, G.P.; Sharp, K.A. *Biophys. J.*, **1997**, *72*, 913.
- [154] Nicholls, A.; Sharp, K.A.; Honig, B. *Proteins*, **1991**, *11*, 281.
- [155] Creamer, T.P. *Proteins*, **2000**, *40*, 443.

- [156] D'Aquino, J.A.; Freire, E.; Amzel, L.M. *Proteins*, **2000**, Suppl. 4, 93.
- [157] Zidek, L.; Novotny, M.V.; Stone, M.J. *Nature Struct. Biol.*, **1999**, 6, 1118.
- [158] Loh, A.P.; Pawley, N.; Nicholson, L.K.; Oswald, R.E. *Biochemistry*, **2001**, 40, 4590.
- [159] Forman-Kay, J.D. *Nature Struct. Biol.*, **1999**, 6, 1086.
- [160] Wand, A.J. *Nature Struct. Biol.*, **2001**, 8, 926.
- [161] Lee, A.L.; Kinnear, S.A.; Wand, A.J. *Nature Struct. Biol.*, **2000**, 7, 72.
- [162] Ladbury, J.E. *Chem. & Biol.*, **1996**, 3, 973.
- [163] Janin, J. *Structure*, **1999**, 7, R277.
- [164] Otwinowski, Z.; Schevitz, R.W.; Zhang, R.-G.; Lawson C.L.; Joachimiak, A.; Marmorstein, R.Q.; Luisi, B.F.; Sigler, P.B. *Nature*, **1988**, 335, 321.
- [165] Bhat, T.N.; Bentley, G.A.; Boulot, G.; Greene, M.I.; Tello, D.; Dall'Acqua, W.; Souchon, H.; Schwarz, F.P.; Mariuzza, R.A.; & Poljak, R.J. *Proc. Natl. Acad. Sci. USA*, **1994**, 91, 1089-93.
- [166] Quioco, F.A.; Wilson, D.K.; Vyas, N.K. *Nature*, **1989**, 340, 404.
- [167] Tame, J.R.H.; Murshudov, G.N.; Dodson, E.J.; Neil, T.K.; Dodson, G.G.; Higgins, C.F.; Wilkinson, A.J. *Science*, **1994**, 264, 1578.
- [168] Tame, J.R.H.; Sleigh, S.H.; Wilkinson, A.J.; Ladbury, J.E. *Nature Struct. Biol.*, **1996**, 3, 998.
- [169] Robinson, C.R.; Sligar, S.G. *J. Mol. Biol.*, **1993**, 234, 302.
- [170] Lam, P.Y.S.; Jadhav, P.K.; Eyermann, C.J.; Hodge, C.N.; Ru, Y.; Bachelier, L.T.; Meek, J.L.; Otto, M. J.; Rayner, M.M.; Wong, Y.N.; Chang, C.H.; Weber, P.C.; Jackson, D.A.; Sharpe, T.R.; & Erickson- Viitanen, S. *Science*, **1994**, 263, 380-4.
- [171] Holdgate, G.A.; Tunnicliffe, A.; Ward, W.H.J.; Weston, S.A.; Rosenbrock, G.; Barth, P.T.; Taylor, I.W.F.; Pauptit, R.A.; Timms, D. *Biochemistry*, **1997**, 36, 9663.
- [172] Deng, H.; Zhadin, N.; Callender, R. *Biochemistry*, **2001**, 40, 3767.
- [173] Birdsall, B.; Feeney, J.; Tendler, S.J.B.; Hammond, S.J.; Roberts, G.C.K. *Biochemistry*, **1989**, 28, 2297.
- [174] Ladbury, J.E.; Hensmann, M.; Panayotou, G.; Campbell, I.D. *Biochemistry*, **1996**, 35, 11062.
- [175] Zhou, G.; Ferrer, M.; Chopra, R.; Kapoor, R.M.; Strassmaier, T.; Weissenhorn, W.; Skehel, J.J.; Oprian, D.; Schreiber, S.L.; Harrison, S.C.; Wiley, D.C. *Bioorg. Medic. Chem.*, **2000**, 8, 2219.
- [176] Montfort, W.R.; Perry, K.M.; Fauman, E.B.; Finer-Moore, J.S.; Maley, G.F.; Hardy, L.; Maley, F.; Stroud, R.M. *Biochemistry*, **1990**, 29, 6964.
- [177] Murthy, K.H.M.; Winborne, E.L.; Minnich, M.D.; Culp, J.S.; Debouck, C. *J. Biol. Chem.*, **1992**, 267, 22770.
- [178] Wojtczak, A.; Luft, J.; Cody, V. *J. Biol. Chem.*, **1992**, 267, 353.
- [179] Decanniere, K.; Transue, T.R.; Desmyter, A.; Maes, D.; Muyldermans, S.; Wyns, L. *J. Mol. Biol.*, **2001**, 313, 473.
- [180] Brem, R.; Dill, K.A. *Prot. Sci.*, **1999**, 8, 1134.