Free Energy of Ligand Binding to Protein: Evaluation of the Contribution of Water Molecules by Computational Methods

Pietro Cozzini[^], Micaela Fornabaio[°], Anna Marabotti[#], Donald J. Abraham[§], Glen E. Kellogg[§] and Andrea Mozzarelli[*°]

[^]*Laboratory of Molecular Modeling, Department of General and Inorganic Chemistry,* [°]*Department of Biochemistry and Molecular Biology and National Institute for the Physics of Matter, University of Parma, 43100 Parma, Italy*

[#]*Laboratory of Bioinformatics, Institute of Food Science, National Research Council 83100 Avellino, Italy*

[§]*Department of Medicinal Chemistry and Institute for Structural Biology and Drug Discovery, Virginia Commonwealth University, Richmond, Virginia, 23298-0540 USA*

Abstract: One of the more challenging issues in medicinal chemistry is the computation of the free energy of ligand binding to macromolecular targets. This allows for the screening of libraries of chemicals for fast and inexpensive identification of lead compounds. Many attempts have been made and several algorithms have been developed for this purpose. Whereas enthalpic contributions are evaluated using methods and equations for which there is a reasonable consensus among researchers, the entropic contribution is evaluated using very different, and, in some cases, very approximate methods, or it is entirely ignored. Entropic contributions are of primary importance in the formation of many ligand-protein complexes, as well as in protein folding. The hydrophobic interaction, associated with the release of water molecules from the protein active site and the ligand, plays a significant role in complex formation, predominantly contributing to the total entropy change and, in some cases, to the total free energy of binding. There are distinct approaches for the evaluation of the contribution of water molecules to the free energy of binding based on Newtonian mechanics force fields, multi-parameter empirical scoring functions and experimental force fields. This review describes these methods – discussing both their advantages and limitations. Particular emphasis will be placed on HINT (Hydropatic INTeractions), a "natural" force field that takes into account in a unified way enthalpic and entropic contributions of all interacting atoms in protein-ligand complexes, including released and structured water molecules. As a case-study, the contribution of water molecules to the binding free energy of HIV-1 protease inhibitors is evaluated.

Keywords: binding free energy, computational chemistry, water molecules, protein crystallography, drug discovery, in silico screening, scoring function, LogP

INTRODUCTION

The determination of the genome from Homo sapiens [1] and several other organisms [2-5] has opened a new era in the investigation of the molecular basis of diseases [6]. Pharmacogenomics [7], chemogenomics [8,9], proteomics [10,11], and metabolomics [12] are just some of the so-called "omics" sciences that were triggered by the genome project. These sciences have allowed a more detailed characterization and understanding of the biomolecules that play a role in the transformation from normal to pathological conditions [13]. As a consequence of this extraordinary effort, the potential targets for drugs are expected to increase from about 400 (today) to over 4000. This explosion requires the determination of the three-dimensional structures of a large number of proteins by X-ray crystallography and nuclear magnetic resonance. Indeed, Structural Genomics Initiatives have been already launched in several countries [14-16]. More protein targets of known three-dimensional structure will encourage the design of more compounds that bind to their active sites, acting either as inhibitors/antagonists or activators/agonists [17-19]. In addition, some of the "old" protein targets undergo amino acid mutations under the selective pressure of efficient drugs and bacteria express new enzyme activities in response to the overuse of antibiotics. The result is a decrease in the ability of a number of drugs to bind to their intended target protein, thus becoming less effective disease cures. A classical example of the first situation is HIV-1 protease where the rate of amino acid mutations in the enzyme active site challenges the capability of researchers to keep pace. For the second situation, penicillin degradation by lactamase is paradigmatic of the growing problem of antibiotic resistance [20].

Considering these issues and the available resources, it is becoming increasingly critical to accelerate the discovery of new leads and to develop new strategies for "moving" targets [21]. The traditional methods of screening proprietary libraries of compounds obtained either by synthesis or isolated from natural sources, as well as the recently developed approaches of testing millions of chemical species produced by combinatorial chemistry [22,23], the so-called

*Address correspondence to this author at the Department of Biochemistry and Molecular Biology, University of Parma, *via* Parco delle Scienze 23/A, 43100 Parma, Italy. Tel: 39-0521-905138; Fax: 39-0521-905151; E-mail: biochim@unipr.it

0929-8673/04 $45.00+.00

"the needle in the haystack search", now appear to be too slow to meet this goal. Alternative "rational" approaches are structure-based drug design [24-29], integrated with combinatorial chemistry [30-32], NMR-based screening [33-35] and "in silico" screening [36-42]. The key features of the latter approach are: i) the availability of high resolution three-dimensional structures of the targets; ii) the ability to "dock" ligands designed or optimized by computer modeling within protein active sites; and iii) the evaluation of free energy of the interaction between ligands and proteins by computational methods. Several excellent reviews have been published to describe the molecular bases of protein-ligand recognition [43], ligand docking, interaction forces and physico-chemical or empirical models, as well as the pitfalls [44-48]. Comprehensive and thoughtful descriptions of methods aimed at the prediction of the binding affinity of ligands to proteins have also appeared [49-53]. In the goal of obtaining either relative or absolute affinities of ligands, the evaluation of the contribution to the free energy of binding played by water molecules either bound to active site residues or to the ligand is "hard to determine and hard to model" [54]. This is initially quite surprising given the ubiquitous presence of water molecules in biological systems. Indeed, Ladbury's statement emphasizes the difficulties that researchers have encountered in the quantitative description of such a "simple" process as the association of a ligand with a protein in an aqueous medium [55-57].

A WATER "WORLD"

Life processes depend on water at the macroscopic and molecular levels [58]. For the former it is sufficient to state that the year 2003 was declared by the United Nations as the Year of Water [59]. Also, water plays a variety of cellular functions, being the solvent of most biological molecules, the substrate and product of enzymatic catalysis, a building block of macromolecules, and a "lubricant" *via* the formation of networks linking distant residues [60-62]. Moreover, the thermodynamic balance associated with attaining the native state of nucleic acids and proteins is strongly dependent on water, with hydrophobic interactions together with hydrogen bonds being crucial in shaping and stabilizing biological macromolecules [63]. The thermodynamics of ligand-protein and protein-protein recognition is water-dependent with both enthalpic and entropic contributions. These multiple roles played by water are associated with its unusual and amazingly unique properties. Its small size, the dipolar nature caused by its charge distribution, the highly directional hydrogen bonds as both donor and acceptor and the entropic gain associated with the release to bulk solution of water molecules bound to biomacromolecules and ligands

Fig. (1). Representative example of a protein structure with water molecules detected by X-ray crystallography. The protein is the αβ dimer of tryptophan synthase with bound a transition-state analogue of the α subunit, 4-(2-hydroxyphenylthio)-1-butenylphosphonic acid (green), solved at 2.3 Å [263]. Water molecules are shown surrounding the protein, near the ligand, and at the subunit interface.

are all key features that help explain water-water and water-solute interactions [64, and references therein]. Here, we will first briefly outline the information and theories for the structure, dynamics and function of water bound to proteins. On this basis, the encounter between a protein and ligands will be described at the molecular level, as part of a long-standing effort to derive predictive models for the strength of the interaction between proteins and ligands. The multiplicity of the proposed models reveals the complexity of properly taking into account the energetics associated with the movement of water molecules bound to proteins and ligands in the formation of a complex, i.e., the hydration and dehydration processes.

Because Nature does not like empty spaces, water molecules generally occupy all the available space not occupied by other (protein and/or ligand) atoms. Water molecules may even occupy sites that appear to be energetically unfavorable for them. X-ray and neutron diffraction and nuclear magnetic resonance studies have provided most of the information on the distribution of water molecules localized either within or on the surface of a protein (Fig. (**1**)). In a typical protein there are about 200 water molecules [65]. However, this number strongly depends on the quality of the structural determination [63], being one water molecule per residue at 2.0 Å and 1.6-1.7 at 1.0 Å resolution [65,66]. The analysis of 873 protein crystal structures determined at room temperature and 33 structures determined at low temperature indicated that there is a weak dependence between the number of water molecules and the fraction of polar/apolar surface, but there is no apparent influence by temperature [66]. This first result is somewhat surprising since water is expected to preferentially interact with polar residues. These findings indicate that the localization and count of water molecules within a protein matrix may vary from one protein to another depending on the resolution. Therefore, protein hydration remains ill-defined and ill-understood as there is clearly "no rigid shell of water around a protein molecule but rather there is a fluctuating cloud of water molecules that are thermodynamically affected more or less strongly by the protein molecule" [67]. Water molecules interacting with a protein have been classified and modeled depending on their localization and residence time. The concepts of first and second hydration shells [68,69], water bound to flexible or fixed side chains in wide, deep or narrow crevices, small or big cavities in the interior of the protein [65], distribution hierarchies [70] and proximal or perpendicular radial distribution functions [71-73] are among many put forward to link simulation and experimental evidence. Buried and tightly bound waters exhibit residence times of the order of hundreds of picoseconds [74-77], whereas water molecules that are more on the surface and in contact with bulk waters exhibit residence times of the order of 5-50 ps [71,77-81]. Computational simulations can provide details on water networks within a protein, pathways that allow water to move from one site to another, and pathways that provide access to sites located deep in the protein core [82].

The interaction of a protein with a ligand is a coordinated process that involves the breakage and formation of several hydrogen bonds, including the reorganization of water molecules around the ligand and within the protein active site (Fig. (**2**)). A detailed molecular description and a thermodynamic evaluation of these events are critical for drug design [54,83]. A topological analysis of high resolution structures of protein-ligand complexes indicated that water molecules mediate recognition *via* formation of hydrogen bonds [84]. Strongly bound water molecules in the active site of a protein are not easily displaced in ligand binding, thus structurally modifying the shape of the protein surface recognized by a ligand [85]. It was also found that some water molecules are conserved within the active site of homologous proteins [86]. Careful analyses of the three-dimensional structure of proteins in the absence and presence of ligands have allowed a partial, yet valuable, understanding of the role played by individual water molecules in complex stabilization. Comparison of the structure of the Fv fragment of the anti-hen egg white lysozyme antibody D1.3 between the free and bound forms indicated the presence of four buried water molecules and others creating a network that bridge the antibody to the antigen [87]. Of these bridging water molecules, ordered water molecules present in the free form of the protein are retained in the complex, while others are only present in the complex. This suggests that complex formation is associated with solvation or a decrease in water mobility, making these water molecules detectable in the complex and undetectable in the protein free form. On the basis of this structural evidence, the thermodynamic forces that lead to complex

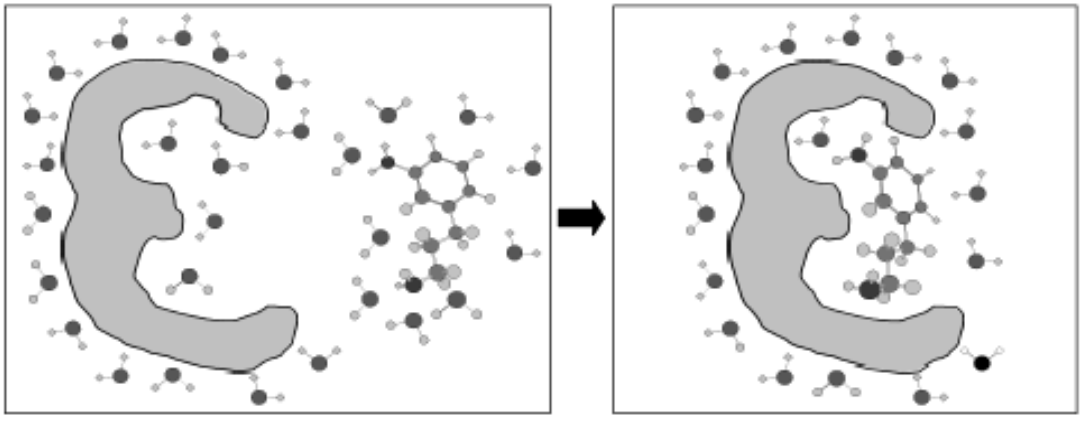


Fig. (2). The cartoon represents water molecules bound to a protein active site and a ligand, before and after formation of a ligand-protein complex.

formation were suggested to be hydrogen bonding, van der Waals interactions and enthalpy of hydration rather than hydrophobic interactions, consistent with calorimetric measurements. Similar studies were carried out on the same D1.3 antibody-hen egg white lysozyme, on D1.3 antibody bound to anti-lysozyme antibody E5.2, and on barnase bound to barstar [88]. Water molecules bridging the two proteins in each set accounted for about 25% of the total binding affinity. The effects of amino acid replacement on the binding energy of complex formation are explained both by direct changes associated with residue mutation and indirect changes due to water molecule reorganization near the mutation site. Furthermore, buried water molecules were found to be key elements in the energetics of the interaction between a bacterial oligopeptide transporter protein and its ligands [89], in mediating DNA-tryptophan repressor recognition [90], and in the specificity and binding affinity of saccharides to lectin [91].

The incorporation of a water molecule into a protein-ligand interface can lead to a free energy change of –1.67 kcal/mol at 298 K [54,92]. However, there are examples where water incorporation into a binding interface is believed to be energetically unfavourable and thus inhibitors have been designed to displace these waters [93-96]. The computational method GRID was developed to locate water molecules as well as other groups within a protein matrix [97]. The interaction energy between the probe and the protein is evaluated using an empirical energy function based on Lennard-Jones, electrostatic and hydrogen bond contributions. The agreement between predicted and structurally determined sites of water molecules and other groups has often been found to be within 0.3-0.5 Å. This finding suggests that within the protein active site there are energetically favored locations for water molecules. These molecules are termed "conserved waters" and can be assumed to be an integral part of a protein site, thus modifying its actual shape and functional and topological profiles. Several improvements were applied to the original GRID functions in order to design selective ligands incorporating the presence of water molecules [98-102]. The placement of single water molecules was also carried out in the FLEXX docking algorithm using a "particle concept" [103]. Other methods were also developed to locate water molecules using knowledge-based water positions derived from X-ray crystal structures [104,105] and a hybrid K-nearest-neighbors genetic algorithm, *Consolv* [106].

BINDING FREE ENERGY OF PROTEIN-LIGAND COMPLEXES

"Free energy is arguably the most important general concept in physical chemistry" wrote Peter Kollman in 1993 [107]. The free energy for the formation of a protein-ligand complex is equal to:

$$\Delta G_{bind} = \Delta H - T\,\Delta S, \quad (1)$$

that, under equilibrium conditions, is equal to:

$$\Delta G^{\circ} = \Delta H^{\circ} - T\,\Delta S^{\circ} = - R\,T\,\mathrm{Log}\,(K_{eq}) \quad (2)$$

The binding free energy contains both enthalpic and entropic contributions that in many reactions of biological systems compensate each other [108]. Although enthalpy-entropy compensation has in the past been regarded as an 'ubiquitous property of water', this appears to be a property of all weak intermolecular interactions, of which hydrogen bonding in aqueous solution is merely one of the most frequently encountered in biochemical reactions. The thermodynamic rationale for enthalpy-entropy compensation is based on the fact that, as the binding becomes stronger, enthalpy becomes more negative and entropy concomitantly tends to decrease due the formation of a tight complex. On the contrary, as the binding becomes weaker, enthalpy becomes less negative and entropy tends to increase due the formation of a loose complex [108]. To explicitly include the contribution of the role of water molecules associated with binding, equation (1) can be written as:

$$\Delta G_{bind} = \Delta G_{interact} - \Delta G_{solv,L} - \Delta G_{solv,R} \quad (3)$$

where $\Delta G_{interact}$ is the interaction free energy of the complex, $\Delta G_{solv,L}$ is the free energy of desolvation of the Ligand and $\Delta G_{solv,R}$ is the free energy of shielding the Receptor (protein active site) from water molecules [109]. The estimation of the individual contributions is a difficult task due to limited knowledge of the relative relevance of the interacting forces that are themselves strongly dependent on the character of individual complexes. Moreover, each contributing term can be large whereas the net result is usually small, thereby imposing a large uncertainty on that result. In some cases, the enthalpy gain can be derived from making extra water-mediated hydrogen bonds greater than the entropic penalty that must be paid for immobilizing water molecules [54].

The hope for accurate measurements of the terms in equation (3) is to combine high resolution structural details with accurate thermodynamic data. In many biomolecular interactions removing water from a binding site is energetically favorable. This is due to the entropic gain when surface-associated solvent molecules are released into bulk solvent. With thermodynamic arguments it is possible to explain the fact that water can provide an overall favorable contribution to the free energy of binding of the interface by providing its maximal number of hydrogen bonds at the right proximity and orientation. Analysis of conserved water binding sites indeed indicates that these are sites where the interactions are highly favorable in terms of hydrogen bond lengths, positioning, and electrostatic and Lennard-Jones potentials.

Due to the relevance of water in structure-based drug design it is essential to address the following issues: i) if a water molecule is already present on the binding surface in the free (unbound) state of interacting molecules, will the formation of the complex provide sufficient additional hydrogen bonds (or sufficiently strengthen existing hydrogen bonds) to obtain a net favorable free energy change?, ii) can the ligand surface be designed in such a way that the existing water molecules can make an optimal arrangement of hydrogen bonds?, iii) how large is the entropic cost associated with the binding of water molecules to proteins or other macromolecules? Dunitz determined this contribution starting with data on the transfer of a single water molecule to a site where it cannot move [92]. The standard entropy of liquid water is 16.7 cal mol^{-1} at 298 K. What happens in proteins? Water molecules in a crystalline protein are unlikely to be bound more tightly than a water molecule in a

crystalline hydrate salt, where its entropy contribution is about 10 cal mol^{-1} K^{-1}. Thus, the entropy cost of transferring a water molecule from the liquid to the protein has an upper bound of this difference, i.e., about 7 cal mol^{-1} K^{-1}, corresponding to a free energy cost of 0-2 kcal mol^{-1} at 300 K, the upper values referring to water molecules that are firmly bound (conserved waters) [92].

Protein-ligand as well as protein-protein recognition is analyzed on the basis of the different forces involved in the process. Whereas hydrogen bonding and van der Waals contributions are mainly relevant in achieving specificity between interacting molecules, hydrophobicity plays a major role in the stabilization of protein-protein complexes [110]. The role of hydrophobicity in protein-ligand complexes may also be significant, depending on the hydrophilic or lipophilic nature of the interacting species. One key concept developed for the evaluation of the hydrophobic contribution to biomolecular interaction is the molecular surface area accessible to solvent (SASA) for the interacting groups. SASA, introduced to quantitatively treat the role played by the solvent when a complex is formed, is "the area on the surface of a sphere with radius R, on each point of which the center of a solvent molecule can be placed in contact with the atom without penetrating any other atoms of the molecule" [111]. SASA is correlated with the hydrophobic free energy because it should depend on the number of water molecules that are released during complex formation, and thus it measures the entropic gain derived from these water molecules that are no longer coordinated to polar groups. This contribution may account for about 25 cal mol^{-1} per $Å^2$ of accessible surface area in the case of protein residues [110,112,113] or as much as 45 cal mol^{-1} per $Å^2$ [114]. In the case of protein-protein complexes, the reduction of surface area accessible to solvent can be about 1000-1500 $Å^2$, resulting in a hydrophobic free energy of 25-40 kcal mol^{-1} [110]. This approach was considered to be oversimplified because it did not take into account the polarity of the groups [68]. The contribution of protein-solvent interactions to the free energy of binding or during folding was assumed to be the sum over all atoms involved in binding or folding (with the exception of hydrogen atoms) of the product of SASA (A) and ASP ($\Delta\sigma$), the latter being the atomic solvation parameter that measures the transfer of an atom from the interior of a protein to aqueous solution [68], i.e.,

$$\Delta G_i = \Sigma \Delta \sigma_i A_i \qquad (4)$$

The polarity of the interior of the protein was assumed to be that of 1-octanol. A key assumption is that the solvation free energy results from the additive contributions of all atoms. This concept originated from studies carried out by Hine and Mookerjee [115] and Leo *et al*. [116].

Other models were developed based on weighted atomic pairwise surface burial [88] and on a weighted solvent accessible surface area (WSAS) [117]. In the latter model, the solvation free energy is evaluated using the equation:

$$\Delta G_{WSAS} = \sum_{i=0}^{m} \sum_{j=0}^{n_i} W_i S_j \qquad (5)$$

where m is the number of atom types for a given model and n_i are the number of atoms with atom type i in a molecule; w_i is the solvation free energy weight of atom type i and S_j is the solvent accessible surface area of atom j. Fitting procedures were applied to define training sets of compounds in order to derive the weights w_i. The model was used to predict the free energy of five binding modes of efavirenz complexed to HIV-1 reverse transcriptase, determining that the most favorable conformation exhibits a binding free energy of 10 kcal mol^{-1} lower than other binding modes [117]. Another method estimated the binding free energy of peptidic inhibitors of HIV protease mutants using a dielectric continuum solvation approach to calculate electrostatic hydration contribution, exposed surface area, semiempirical quantum chemistry to determine the protonation state of active site residues, and molecular mechanics for the determination of relative binding energies [118].

Several methods were developed to take into account solvation in molecular docking. DOCK [119] corrected the electrostatic interaction energy with a ligand electrostatic solvation energy, and the van der Waals component of the interaction energy with a non-polar parameter for ligand solvation [109]. This method was applied to screen the Available Chemicals Directory for ligands to bind to thymidilate synthase, dihydrofolate reductase and T4 lysozyme leading to an improvement of the ranking of known ligands and selected molecules with reduced charge and size. The solvation correction also significantly improved the agreement between calculated and experimental free energies. However, these DOCK energy solvation-corrected calculations still over-predicted free energy by three to five kcal mol^{-1} compared with the experimental values [109]. Another docking program, FLOG (Flexible Ligands Oriented Grid) [120], makes use of several potentials for electrostatic, hydrogen bonding, van der Waals and hydrophobic contributions. Only conserved water molecules are considered as their presence modifies the protein active site structure. Ligand flexibility was included by allowing up to 25 explicit conformations for each structure. The program was applied in the search from a database for inhibitors of metallo-β-lactamase [121], HIV-1 protease [122] and dihydrofolate reductase from *L. casei*, in the presence of NADP and two active site conserved waters, water 201 and water 253, whereas other water molecules were removed [120]. In the third version of the program AutoDock, developed by Olson and co-workers [123], structural water heterogeneity is incorporated in an automated docking procedure [124]. Protein mobility, a challenging issue for docking ligands, is also taken into account. The program was applied to evaluate 21 peptidomimetic inhibitors of HIV-1 protease where a structural water, important for complex stability, is present in 20 of the considered structures and is displaced in the remaining one by a cyclic urea inhibitor.

FREE ENERGY AND THERMODYNAMIC ADDITIVITY

One fundamental principle that has to be taken into account in calculating free energy of binding is that the free energy is a state function and thermodynamic additivity can be applied only if components (groups or atoms A, B, C…) or contributions (ΔG_{ionic}, $\Delta G_{hydrogen\ bonding}$,

$\Delta G_{\text{van der Waals}}$, $\Delta G_{\text{non-polar solvation}}$,….) are independent. As pointed out by Dill [125], "if tetraalanine binds a protein, does the free energy equal four times that of the binding of alanine? Is protein folding the sum of oil-to-water transfers of each amino acid? Can we add surface area-based solvation terms to molecular dynamics force fields? Are the conformational entropies of biopolymers simple sums of the monomer entropies or of backbone plus side chain entropies?" In many models, additivity is erroneously assumed, likely leading to incorrect determination of the free energy of binding. However, deviations from non ideal behavior might be small and predicted values may be within a few kcal mol^{-1} with respect to the experimental data, a penalty that may be acceptable in exchange for fast screening of many thousands of compounds.

MOLECULAR MECHANICS, MOLECULAR DYNAMICS AND MONTE CARLO SIMULATIONS

Prediction of the strength of noncovalent binding between molecules and prediction of the 3D structure of the corresponding complexes have been the overriding goals in computational chemistry for many years. During the past decade the advances made in computer-aided drug design and computational studies have had a large impact on the drug design process and in the search for new pharmaceuticals. Methodologies based on force field calculations, such as molecular mechanics (MM), molecular dynamics (MD), and Monte Carlo (MC) simulations, have been key to these developments. These topics are treated in specific sections below.

Molecular mechanics force fields represent the cornerstone of all simulations of complex chemical systems. They are extensively used in calculating structural and thermodynamics properties of biomolecules. Some of the force fields more commonly applied to biomolecules are AMBER [126,127], CHARMM [128] GROMOS [129] OPLS-AA [130], DISCOVER [131], TRIPOS [132], MMFF [133], MM2 [134], MM3 [135] and MM4 [136]. In particular, the AMBER, CHARMM and GROMOS force fields are routinely used in calculating relative solvation free energies (SFEs) of small organic molecules and relative binding free energies of ligands, whereas the MMFF force field is aimed at simulating protein-ligand complexes [52,137]. The typical potential energy function of molecular mechanics force fields is:

$$E_{\text{total}} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{R^{12}_{ij}} - \frac{B_{ij}}{R^{6}_{ij}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] \quad (6)$$

A characteristic of all current force fields applied to biopolymers is that they are two-body additive, meaning that the potential energy function (eq. 6) is a function of atom pairs. With these models it is possible to implicitly include many-body effects in the parameterization.

Solvent Models

The influence of the solvent on the electrostatic interactions that play a relevant role in many biological processes has been deeply investigated. This requires modeling the solute and the solvent and the interactions between them [138-141]. Molecular mechanics describes solute or solvent molecules or both on the basis of a force field and intermolecular interactions are taken into account by non-bonded van der Waals and Coulombic electrostatic terms. Solvent effects are described by calculating the mutual interactions of a large number of molecules and averaging these over many solvent configurations [138]. Biological molecules are surrounded by water and solvation is a parameter that influences several properties of biomolecules such as solubility, reaction rate equilibria, partition coefficients, and enzyme-substrate and ligand-receptor binding [137]. In free energy calculations solvent can be represented using either an explicit or an implicit solvent model, the choice depending on the available computer resources and/or the quality required in the result. Many implicit solvation models for proteins have been developed that combine an empirical molecular mechanics force field for the intramolecular interactions in vacuum with a solvation correction. These models can be used for analysis of the contributions of solvation and desolvation [142]. Alternative approaches of implicit solvation involving continuum or macroscopic models in which solvent properties are described in terms of average values [138] may also be applied. Relative solvation free energies can often be obtained with fair accuracy with implicit solvent models such as the Generalized Born/surface area (GB/SA) method [143], or with greater accuracy at larger expense with continuum methods that involve solution of the Poisson-Boltzmann equation [144,145].

While implicit solvent models are generally most useful in macromolecular systems where long simulations are required to reach convergence, in many cases explicit solvent models can have significant advantages. Potential functions for water have been under development for more than 30 years [52,137, and references therein]. The explicit solvent models currently most often used were developed in the 1980s and are called TIP3P, TIP4P [146], SPC [147], and SPC/E [148]. The empirical adjustment of the parameters in these models allows the user to reproduce the enthalpy of vaporization and the density of water. In particular, because TIP3P, SPC, and SPC/E solvent models are simple and relatively fast, and provide good results, they are often used in computer simulations of both small and large molecules. The SPC/E model very accurately reproduces experimental properties of water, such as the dielectric constant and diffusion coefficient. Jorgensen recently developed a TIP5P model that better describes the temperature dependence of the density of water [149]. While for many applications the explicit treatment of solvent molecules and mobile ions is not easily accessible, the improved accuracy of calculated relative binding affinities of ligands to proteins may be worth the computational expense [137].

Calculations of the relative solvation and binding free energies between two similar molecules are usually carried out by MC or MD simulations in conjunction with the

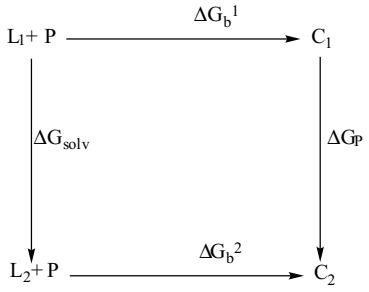
thermodynamic cycle perturbation approach with explicit solvent [137]. The methodology gave satisfactory agreement with some experimental results within ± 1.0 kcal mol^{-1}, but a potential problem for solvation and binding free energy calculations is the difficulty in reaching simulation convergence. Proper convergence of the free energy calculations requires simulation times much longer than previously estimated [150-152].

Simulations on Protein-Ligand Interactions

Molecular dynamics (MD) and Monte Carlo (MC) methods have provided dynamic and atomic insights into biological systems. Among the methods for protein-ligand binding free energy prediction, the rigorous free energy perturbation (FEP) and thermodynamic integration (TI) methods have been successfully applied, but they are computationally expensive [107,153,154]. Other less rigorous methods have been developed to evaluate binding free energies, such as the linear interaction energy (LIE) [155] and the molecular mechanics/Poisson Boltzmann surface area (MM/PBSA) methods [52].

Free Energy Perturbation and Thermodynamic Integration Methods

From a thermodynamic point of view, the most rigorous methods among those currently available for prediction of the relative free energy of binding of ligands to proteins are free-energy-perturbation (FEP) or thermodynamic integration (TI) methods, with explicit consideration of solvent molecules and flexibility of both the receptor and the ligand [156,157]. The explicit consideration of water is usually implemented by solvating with a sphere around the active site with water molecules. The free energy perturbation approach for calculating relative free energies of binding between two ligands (L_1 and L_2) and a given receptor (P) employs the thermodynamic cycle shown in Scheme (**1**) [52,107,153-155]:



Scheme 1. Thermodynamic cycle used in FEP procedure for the calculation of the relative binding free energy between two ligands bound to the same protein.

Thus,

$$\Delta\Delta G = \Delta G_b^1 - \Delta G_b^2 = \Delta G_{solv} - \Delta G_P \qquad (7)$$

where ΔG_b^1 and ΔG_b^2 are the binding free energies for ligands L_1 and L_2, respectively, and ΔG_{solv} and ΔG_P are the non-physical transmutation free energies from ligand L_1 to ligand L_2 in free and bound states representing the differences in solvation free energy between L_1 and L_2 in water and when bound to the solvated receptor site. If ligand 1 and 2 are similar to each other, then ΔG_{solv} and ΔG_P are usually easier to calculate than ΔG_b^1 and ΔG_b^2 because the mutation from ligand 1 to ligand 2 is assumed to cause only localized changes, and FEP or TI is used to calculate ΔG_{solv} and ΔG_P. The following equation (8) is used in FEP calculations,

$$\Delta G = -RT \sum_{i=1}^{N-1} \ln \left\langle \exp\left[- \frac{H(\lambda_{i+1}) - H(\lambda_i)}{RT} \right] \right\rangle_{\lambda_i} \qquad (8)$$

where ΔG is the free energy difference between two states, A and B. λ_i varies from 0 (state A) to 1 (state B), $H(\lambda_i)$ is the Hamiltonian of the system at λ_i and $\langle\ \rangle_{\lambda i}$ indicates an ensemble average. In the TI method, the average of derivatives of the Hamiltonian at each λ, $H(\lambda)$, is calculated and the numerical integration over λ is used to determine the free energy difference between two states,

$$\Delta G = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle d\lambda \qquad (9)$$

where λ has the same meaning as in FEP [52,158]. With the FEP method it is possible to calculate the free energies associated with the two "unphysical" vertical paths $L_1 \rightarrow L_2(aq)$ and $C_1 \rightarrow C_2(aq)$ corresponding to a mutation of L_1 into L_2 in the free and bound states. The ensemble averages along these paths are calculated by MD or MC simulations. The paths are typically divided into a number of steps, each represented by a separate potential energy function built up as a linear combination of the initial and final state potentials [155].

The FEP approach is generally regarded as the most important technique for free energy calculations by MD or MC simulations, but frequently a number of problems are encountered. First, the most severe limitation is the extensive conformational sampling usually required in order to obtain convergent results from the simulations, so that the results can be very sensitive to the length of the simulation. Second, FEP methods also depend on the accuracy of the applied force field [107,153] and require large computational resources for force field parametrization as well as for the requisite MC/MD calculations. Third, FEP methods only allow for minor chemical differences in the ligands to predict reliable relative free energies, i.e., the "perturbations" involved in FEP cannot be too drastic. Fourth, most of the computer time is spent on "uninteresting" configurations corresponding to an unphysical "mixture" of L_1 and L_2 [155,159,160]. Fifth, the results depend on the protocols used for the simulation. Finally, before applying this methodology for predicting binding affinities of new analogs of a lead ligand, the length of simulation, the number of windows and force field parameters need to be carefully validated for ligands with known experimental data [137]. However, free energy perturbation methods have the advantage of very accurately predicting differences in binding affinity to a receptor for structurally similar ligands, which is an extremely important application for drug discovery. Most of the results suggest that existing methodologies provide good agreement with experimental results [137]. The energetic cost of desolvation is evaluated by solvation free energy calculations, which can

also help in optimizing the ligand by highlighting structural changes that minimize the desolvation penalty. Furthermore, separation of the free energy into enthalpy and entropy terms and for each individual chemical group provides information about the dominant contributions and relevant residues in the binding process.

To predict relative binding free energies accurately, the FEP method must calculate the relative difference of solvent free energies between two ligands, as well as the free energy differences for the complexes [137]. Desolvation free energy is an important factor and it is neglected in most approximate methods to determine relative binding affinities. Its importance is evident using the FEP method. There have been several applications of the thermodynamic cycle (Scheme (**1**)) to protein-ligand interactions [52,107,137, and references therein]. One of the earliest was McCammon's perturbation study of substituted benzamidines interacting with trypsin [161] that was later also investigated by Bash *et al*. [162]. Another success of the FEP calculations was to show the influence of desolvation involved in the binding of transition state inhibitors to thermolysin [163], their prediction being later experimentally confirmed [164]. More recently, calculations of relative binding free energies of closely (structurally) related ligands to dihydrofolate reductase were accurate enough for quantitative conclusions [165].

Absolute binding energies can in principle also be calculated with FEP methods by setting the interaction potential of the ligand to zero in one of the states. In order to enhance conformational sampling in macromolecular simulations, Ota *et al*. [166] combined non-Boltzmann sampling of configuration space with TI (NBTI). This so-called "umbrella" sampling is advantageous because it increases the sampling of the conformational space by enhancing the ligand flexibility by reduced energy barriers to rotation. The relative binding free energies of trypsin to the two inhibitors benzamidine and benzylamine were determined using NBTI, obtaining smaller deviations between calculated and experimental relative free energies than the classical TI.

Linear Interaction Energy Method

The linear interaction method (LIE) is a semi-empirical approach, originally proposed by Åqvist *et al*. to estimate absolute binding free energies [155,160]. This method was developed to obtain useful information on binding energetics from simulations of the only physically relevant states (free and bound) of the ligand [155,160]. LIE does not require simulations of any transformation processes and allows comparisons of compounds with very different structures. The method uses time averages from two simulations, one for the ligand in water (free state), the other for the ligand bound inside the macromolecule surrounded by water (bound state). The binding free energy is evaluated from the difference between them [160,167]. This method is based on the assumption of a linear response of the solvent polarization to changes in the electrostatic field exerted by the solute, and a single solvent dielectric constant [52,160]. LIE divides the interaction between the ligand and its environment into polar (electrostatic) and non-polar (van der Waals) parts. The binding free energy is estimated as:

$$\Delta G_{bind} = \Delta G^{el}_{bind} + \Delta G^{vdw}_{bind} = \alpha\left(\left\langle V^{el}_{l\text{-}s:bound}\right\rangle - \left\langle Vl\text{-}s_{:free}{}^{el}\right\rangle\right) + \beta\left(\left\langle V_{l\text{-}s:bound}{}^{vdw}\right\rangle - \left\langle V_{l\text{-}s:free}{}^{vdw}\right\rangle\right) \quad (10)$$

where $V^{el}_{l\text{-}s:bound}$ and $V^{vdw}_{l\text{-}s:bound}$ are the electrostatic and van der Waals interaction energies between the ligand and the solvated protein from an MD trajectory with ligand bound to protein; $V^{el}_{l\text{-}s:free}$ and $V^{vdw}_{l\text{-}s:free}$ are the electrostatic and van der Waals interaction energies between the ligand and the water from an MD trajectory with the ligand in water; $\langle\ \rangle$denotes an ensemble average, and α and β are two empirically-determined parameters. The main strength of the method is the explicit treatment of the water reference state and its treatment, at a microscopic level, of details that can be important for binding, such as interactions with active site water molecules. Åqvist and co-workers have applied this method to calculate absolute binding free energies of several protein-ligand complexes. They found that $\alpha = 0.5$ and $\beta = 0.16$ gave good agreement between calculated binding free energies with experimental data. The coefficient $\alpha = 0.5$ is a consequence of assuming a linear response of the surrounding to the electrostatic field, while β is an empirical parameter that relates the average van der Waals interaction energies to a corresponding nonpolar ("hydrophobic") binding contribution that can be fitted to experimental binding free energies. The calibration set used to determine the optimum value of β was comprised of a series of endothiapepsin inhibitors with known binding constants [160]. This method was also successfully used to calculate absolute binding free energies of HIV-1 protease inhibitors and two charged benzamidine inhibitors bound to trypsin [167,168]. An additional correction term for long-range electrostatic contribution to the binding free energy was included in these studies. In both the simulations of the complexes and of the ligands in aqueous solution, spherical systems with restrained boundaries were used and water was represented by the SPC model [147] with the surface molecules of the sphere subjected to radial and polarization restraints according to the Surface Constrained All-Atom Solvent model [169]. More recently, the LIE method has been revised [170] and, subsequently, used in investigating protein ligands, including inhibitors of dihydrofolate reductase [171] and human thrombin [172].

From the numerous studies that have successfully applied the LIE method a question arose whether one set of α and β can be used in different protein-ligand complexes to give reasonable estimates of binding free energies. The use of distinct force fields cannot explain the difference in α and β found in simulations. This issue was further investigated in seven different complex systems [173] and it was found a linear correlation between the value of β and the hydrophobicity of both the ligand and the receptor binding site. In effect, larger values of β were found in the presence of a higher number of hydrophobic groups buried after binding. Within the constraints and considerations mentioned above, LIE is a useful method for estimating absolute binding free energies for protein-ligand systems and it does have the advantage of being somewhat computationally more efficient than the FEP and TI method.

Methods Based on Molecular Mechanics/Poisson-Boltzmann Surface Area

A hybrid method termed MM/PBSA (Molecular Mechanics/Poisson-Boltzmann Surface Area) combining molecular mechanics and continuum solvent calculations has recently been developed to analyze the free energies of binding and relative free energies of different conformations [52,174-179]. The MM/PBSA method extracts solute conformations or snapshots from a MD trajectory carried out with explicit solvent, typically a periodic box with water and counter-ions. For each snapshot, solvent molecules are removed to obtain the molecular mechanics energy of the solute, using the same molecular mechanics potential as in the simulation, but in the absence of cut-offs in order to evaluate the non-bonded interactions. The binding free energy is calculated as:

$$\Delta G_b=\Delta G_{MM}+\Delta G^{LP}{}_{solv}-\Delta G^{L}{}_{solv}-\Delta G^{P}{}_{solv}-T\Delta S \qquad (11)$$

where ΔG_b is the binding free energy in water, $\Delta G^{LP}{}_{solv}$, $\Delta G^{L}{}_{solv}$ and $\Delta G^{P}{}_{solv}$ represent the solvation free energies for the complex, the ligand, and the protein, respectively, and $-T\Delta S$ is the conformational entropy contribution to the binding, usually estimated by quasi-harmonic or normal-mode analysis of the MD trajectory. ΔG_{MM} represents the interaction energy between the ligand and the protein and it is calculated from molecular mechanics interaction energies, resulting from the sum of electrostatic and van der Waals contributions:

$$\Delta G_{MM} = \Delta G^{el}{}_{int} + \Delta G^{vdw}{}_{int} \qquad (12)$$

where $\Delta G^{el}{}_{int}$ and $\Delta G^{vdw}{}_{int}$ are electrostatic and van der Waals interaction energies, respectively. The free energy of solvation, ΔG_{solv}, is the sum of electrostatic and non-polar contributions:

$$\Delta G_{solv} = \Delta G^{el}{}_{solv} + \Delta G^{nonpolar}{}_{solv} \qquad (13)$$

where the electrostatic solvation term, $\Delta G^{el}{}_{solv}$, is calculated with the Poisson-Boltzmann (PB) approach, computed in a continuum solvent using a finite difference Poisson-Boltzmann model [138,180,181]. The nonpolar solvation term, $\Delta G^{nonpolar}{}_{solv}$, can be derived from the solvent-accessible surface area.

The ensemble of structures for the uncomplexed reactants can be generated either by running separate MD simulations of the free species or by using the trajectory of the complex and then removing the atoms of the protein and the ligand. The MM energies and solvation free energies are computed for each snapshot of the solute and then averaged to compute the difference in free energies. The free energy difference can be computed to obtain the absolute binding or the relative binding for different mutants. Kuhn and Kollmann applied the MM/PBSA method to study biotin and its derivatives binding with avidin/streptavidin, obtaining relative binding free energies in very good agreement with experimental values. Then, comparing MM/PBSA and LIE methods in calculating binding free energies for diverse avidin and streptavidin ligands [177], their calculations could reproduce experimental ΔG_{bind} with better correlation coefficient (r^2 = 0.92) than the results from the LIE method (r^2 = 0.55) with fixed parameters (α = 0.5 and β = 1).

MM/PBSA combines an explicit molecular mechanical model for the solute with a continuum method for the solvation free energy. This method is more efficient than traditional free energy methods, even if less accurate. Its peculiarity is the applicability to systems which differ substantially in structure. Even though MM/PBSA is fast relative to conventional free energy perturbation methods, further reductions of the computational effort per ligand can be made. Two approaches that can enable MM/PBSA to efficiently rank a large set of ligands are: i) replacing the PB continuum calculation – one of the more time consuming steps in the MM/PBSA approach - with a Generalized Born (GB) solvent model [143] for the calculation of polar interaction energies, and ii) instead of calculating individual trajectories for each ligand, design a simulation in which all the ligands simultaneously interact with the protein [177].

Summary

FEP and TI are the most rigorous methods in calculating free energies, but suffer several limitations. The LIE method allows calculation of absolute binding free energies and can be used to study specific complex systems by applying suitable empirical parameters. MD simulations can be accelerated by replacing explicit water molecules with a solvent continuum, as in the case of MM/PBSA, thus enabling direct calculation of binding free energies [52].

KNOWLEDGE-BASED POTENTIALS

A second evolving category of methods that use empirical and knowledge-based scoring functions for binding affinity estimation is becoming increasingly useful for rapid docking and virtual screening of potential ligands. Knowledge-based potentials, also known as potentials of mean force (PMF) or structure-derived potentials, have only recently been developed in order to achieve the goal of obtaining rapid prediction of binding affinities between ligands and receptors of known 3D structure [50]. They have been applied to the detection of errors in protein structures, the threading of proteins, fold recognition and the *ab initio* prediction of protein structure [182]. The basic assumption of this approach is that experimentally derived molecular structures contain a large amount of information on physical properties of protein-ligand interactions. Consequently, statistical analysis applied on a sufficiently representative sample of all interactions should be able to reveal rules and principles governing the stability of complexes. Practically, PMF are derived by calculating experimentally observed frequencies of non-bonded pairs of atoms in sets of training complexes. Because the native structure that corresponds to the lowest energy state at equilibrium can be described by energy functions of intra- and intermolecular atomic pair interactions [182], the starting assumption is that atoms interact in a ligand-protein complex as in the gas phase, at temperature T. Under these conditions, the interaction free energy is calculated by the equilibrium pairwise density between two atom types at a distance **r** by using the Boltzmann equation. Probably the most interesting characteristic of knowledge-based potentials is the fact that they incorporate all forces acting between two atom pairs, and, moreover, they also contain to some extent the

influence of the surrounding medium. This latter point is very important, because having a function that is capable of rapidly estimating reliable solvation free energy has become an increasingly important goal [142].

Solvent-residue potentials can be explicitly determined by assuming desolvation of residues prior to their association as a reference state for a protein-ligand complex [183]. The most serious problem with these approaches is that protein structures derived from X-ray and NMR studies are not representative of amino acids in the gas phase, and, thus, several approximations are made. The most significant approximation is that each amino acid pair in the system is assumed to be independent of all other pairs, which is an obvious fault because residues in proteins interact with each other across secondary, tertiary and quaternary structural features and are also covalently linked with each other in specific sequences. The principal weakness of these studies is that knowledge-based potentials depend on chain length and composition, and they may be not able to quantitatively reflect the true energy that causes amino acid pairing in proteins, even if they often correctly rank the relative strengths of inter-residue interactions [184]. However, despite these limitations and the controversial relationship between knowledge-based potential scoring functions and free energy [185], several studies have been carried out to correlate free energies of binding estimated using PMF methods and experimental binding affinities for different sets of protein-ligand complexes. The results obtained by these approaches seem to be of the same or even better quality than those obtained by other more rigorous approaches [50]. We review these studies here with particular attention to the methods used in treating the solvent.

The first knowledge-based approach to estimate binding affinities from a data set of proteins was made by Verkhivker *et al*. [186]. A distance-dependent knowledge-based pairs potential was derived from a data set of 30 HIV-1, -2, and SIV-protease-inhibitors complexes by defining 12 atom types. A comprehensive thermodynamic analysis was performed including: i) an investigation of the balance between hydrophobic and H-bond interactions, ii) the energetic factors that regulate desolvation of polar groups in the active site, and iii) the validity of enthalpy/entropy compensation effects. The starting point was a simplified model of the complex, in which ligand-solvent and ligand-protein interfaces were analyzed separately. The ligand-solvent interface of the complex was considered as the portion of the total ligand surface covered by solvent, and the ligand-protein interface was considered as the ligand surface covered by a protein environment. To evaluate ligand and protein desolvation contributions to the binding free energies, an empirical solvation scale was used in which the contributions of non-polar and polar components were separated. The atomic solvation parameters for vacuum-water transfer were determined as mean values of atomic solvation parameters previously obtained [180,187].

To take into account desolvation contributions of non-polar and polar groups, the ligand-solvent interface of the complex was considered as that portion of the total ligand surface covered by solvent, including a solvation mean field potential that measures the propensity for specific ligand atom types to be covered by a protein environment. Moreover, to evaluate ligand and protein desolvation contributions to the binding free energy, an empirical solvation scale was added and the contributions of non-polar and polar components were separated. The atomic solvation parameters were determined as mean values of atomic solvation parameters [180,187]. Finally, entropy loss arising from favorable interaction of discrete water molecules located in the active site in fixed positions was taken into account by a mean fixed value [186]. The results of these free energy calculations as compared with experimental binding affinities were quite encouraging, since not only the agreement of the estimated affinities of the complexes with observed affinities was good, but also significant differences in interaction energies of different inhibitors across subsites were explained. In particular, it was found that the most important contribution to the overall binding energy comes from hydrophobic interactions as a result of the entropic contribution of desolvation of hydrophobic groups at the ligand-protein interface. Ligand-specific interactions contribute to the total energy, thus justifying binding energy differences between two inhibitor molecules. Desolvation of polar groups contributes a normally large and unfavorable component. Finally, it is important to note that the crystallographic complexes used in the analysis included crystallographically-located water molecules at the ligand-protein interface. Simulations [186] show that an unfavorable entropic contribution to the total binding energy is due to "freezing" water in the active site, thus contributing to the loss of translational and rotational entropy of the system. The weakness of the PMF approach in this study is that no reasonable estimation of absolute binding affinity is possible, and that the starting data set is limited to a single class of protein-ligand complexes, limiting the generalization of the method to other complexes.

Similar, but more general, approaches were developed almost simultaneously by Muegge and Martin [188] and Mitchell and colleagues [189,190]. Both of these groups describe knowledge-based potentials extracted from PDB crystallographic structures by analyzing large numbers of complexes. In the former investigation, the PMF was calculated on 697 complexes, whereas the latter two "generations" of the PMF, called BLEEP-1 and 2, were calculated on 351 and 188 PDB entries, respectively, in order to create general, reliable and fast scoring functions. Moreover, both methods used more atom type definitions for proteins and ligands (16 protein and 34 ligand atom types [188], and 2 hydrogen, 38 non-metal and 18 metal atom types [190]). With this enhancement it is possible to more precisely model a wider variety of interactions. Both approaches are also based on a calculation of distance distribution of specific atom type interactions and implicitly consider entropic contributions to binding free energy inasmuch they are encoded within the crystallographic structures on which the PMF models are based. Additional terms added to include rotational and/or translational entropy contributions were not seen to improve the resulting correlation and, thus, were not adopted in the final expression of PMF [188]. With respect to water interactions, these two implementations are designed differently. Muegge and Martin's approach [188] incorporates solvation entropy into the PMF expression by evaluating the "degree of ligand penetration" into a cavity of the protein and using it as an

implicit recognition of solvation effects. A large cutoff (>12 Å) for atom pair interactions was implemented in PMF calculations in order to take into account the presence or absence of protein atoms around the ligand. Hydrogen atoms were omitted since they are not present in PDB structures, and water was not treated explicitly. The dependence of the resulting score on the interaction with discrete water molecules considered as a part of the protein was evaluated in a few cases and was found to be negligible. A different approach was adopted by Mitchell *et al*. [190]. Two "generations" of their knowledge-based potential were implemented within BLEEP-2. It included explicit interactions of protein and ligand with crystallographically detected water molecules and also incorporated the effect of water at the interfaces between bulk solvent and protein, and between solvent and the ligand surface. The first hydration shell of water around the complex was added with the program AQUARIUS 2 [104,191] in such a way that atoms at the surface of the protein were considered to interact with water molecules. Protein-water, ligand-water and water-water interactions were then used, in addition to ligand-protein interactions, to derive distance pair potentials. In addition, with BLEEP-2, an optional calculation was attempted to determine solvent-inclusive interaction energies. To add the explicit contribution of the bulk medium the BLEEP-2 algorithm includes terms to estimate changes in the solvent-solvent energy. However, because the direct calculation of energies on all water molecules was too expensive in terms of CPU time, an approximation was used that considered a hydrogen bond between two solvent molecules to be formed for every two hydrogen bonds lost between protein or ligand and solvent. By calculating the number of hydrogen bonds formed and by multiplying it with the hydrogen bond water-water potential, it was possible to estimate explicitly energies of solvation for a chosen interaction sphere.

Results from both studies are quite encouraging, despite some surprising issues. Muegge and Martin [188] tested their PMF against several sets of ligand-protein complexes and found that their score is especially effective in ranking affinities of similar molecules for the same protein target, although not capable of giving a reliable estimate of the binding free energy. Their scoring function is general, since a test on 77 different complexes showed a linear correlation with an $R^2 = 0.64$ and another test with a protein and 33 inhibitors gave a result with $R^2 = 0.74$. The most encouraging result was the linear relationship between estimated and experimental binding affinities for a test of 16 serine protease-inhibitor complexes, with an $R^2 = 0.92$. Moreover, by comparing the PMF score with other scoring functions, the scores were found to be reliable in more diverse test sets. The evaluation of ligand volume in perturbing the PMF score was tested and found significant, especially for those complexes where the relationships between PMF score and binding affinity were not as well defined [192]. Results from Mitchell and collegues [189,190] appeared to be promising. They evaluated their PMF on a test set of 90 complexes extracted from the PDB and not part of the learning set used to derive the knowledge-based potential. The correlation coefficient between experimental binding affinities and the BLEEP score show that the second generation of PMF (BLEEP-2), which implicitly includes solvent effects, performs better than BLEEP-1 that does not contain solvent information. However, the calculations with BLEEP-2 and the addition of explicit solvent effects in a sphere of water molecules of 8 Å radius gave a worse result ($R^2 = 0.74$ without explicit solvent, $R^2 = 0.63$ with explicit solvent). The same effect was also found in another test set of nine serine protease-inhibitor complexes ($R^2 = 0.71$ without explicit solvent, $R^2 = 0.54$ with explicit solvent). The suggested probable cause is that approximations included in the solvation model calculation lead to an accumulation of errors that overwhelms the good correlation of the PMF itself. Therefore, the problem of an explicit water-water energy term cannot be considered solved using this approach. Finally, BLEEP-2 does not allow one to directly calculate experimental binding affinities because the absolute magnitude of the "energies" predicted is significantly different from the experimental values.

Another PMF, named DrugScore, was derived by calculating distance-dependent pair potentials retrieved from 1376 ligand-protein complexes [193]. In this work, the distance-dependent PMF was associated with a solvent-accessible surface-dependent singlet potential, in order to take into account the entropy-dependent solvent contribution. The first term is parameterized to describe specific short-range distances, whereas the second term calculates solvent-accessible surface regions of both protein and solvent that become buried after association. This approach acknowledges the fact that binding between ligand and protein occurs in an aqueous environment, and it is necessary to consider the entropic effect caused by the release of solvent molecules from the binding cavity and the reorganization of surrounding solvent molecules after ligand binding. The resulting PMF reflects the contributions arising from differences in the solvent–accessible surface for each atom pair in a fully solvated state, thus avoiding the average of this function over all atom types. Conformation of both proteins and ligands free in solution were directly derived from the X-ray crystallographic complexes and assumed to be identical. The solvent-accessible surface was calculated by adapting an algorithm previously developed [194]. The implicit description of solute/solvent interactions and solvent entropic effects along with involved enthalpic contributions resulting from interatomic forces is incorporated in the formalism to derive the PMF. DrugScore software was tested to evaluate its reliability in identifying near-native protein-ligand binding motifs out of a set of different complex geometries and its reliability to rank affinities for ligands that are closely to crystallographic geometry. Two sets of 91 and 68 protein-ligand complexes were used. The results with DrugScore are satisfactory in discriminating between "well-docked" binding modes. It was found that scoring values obtained by DrugScore for protein-ligand complexes can be scaled to experimentally determined pKi values [195]. Recently, DrugScore was implemented in the AutoDock Lamarckian genetic algorithm [123] in order to compare its reliability with the AutoDock scoring function in searching for favorable ligand binding modes, and found to be of comparable quality [196]. Finally, DrugScore was used in a new approach to adopt knowledge-based potentials specifically for one protein by considering ligand-based information in a CoMFA-type approach [197].

Another knowledge-based potential of mean force was developed by Wallqvist and colleagues [198] in a slightly different way from those previously presented. A systematic analysis of atomic interactions between adjacent buried surfaces within a binding interface was carried out and the parameters for this PMF were extracted from contact frequencies between atomic surfaces for each member of the complex. In this case, the interactions between aminoacids are assumed to be short-ranged [184] and are approximated using a contact potential. The atom-based statistical preferences are then produced by normalizing the results with the product of buried surfaces of individual atoms. The underlying approach assumes that several thermodynamic quantities exhibit a strong correlation with solvent-exposed surfaces, as found in previous studies [199-201], and solvent exposure has proven to be a highly effective way of developing an understanding of empirical observations. Therefore, the free energy change of desolvating the appropriate region of a bound enzyme-inhibitor complex and the free energy change for association of these surfaces are assumed to be proportional to the jointly buried surface areas. In this way, it is possible to calculate atom-atom binding coefficients and to interpret in terms of favorable/unfavorable processes the release of water and subsequent association between two atom types. As an example, the binding parameter for a C_{aliph}-C_{aliph} association is calculated to be 23 cal mol^{-1} per $Å^2$. This means that removal of these hydrophobic surfaces from solvent exposure is energetically favored. This PMF was applied to the analysis of ten HIV-1 protease inhibitors. The results suggest that these complexes are characterized by a small principal and common set of protein-ligand contacts that contribute most to the total binding energy, and by a set of more variable and specific interactions that can further stabilize the complexes. The correspondence between measured and predicted binding energy for these complexes is also good. In addition, the identification of a common binding motif may allow for the improved design of inhibitors by taking into account more relevant features of binding.

Another interesting PMF method, called SMoG, was developed using contact potentials [202]. A simplified description of interactions between ligand and protein in a complex was obtained, producing a "coarse-grained" potential function, characterized by atomic contacts considered either "on" (i.e., 1) or "off" (i.e., 0) depending on whether atom pairs distances are within a cutoff radius distance of 5 Å. The core assumption is that changes in solvation entropy resulting from complex formation are influenced by loss or gain of solvent order that is related to the surface potential of the complex atoms exposed to solvent. The probability of observing a specific contact includes the effect of an average over the contribution of solvation entropy to the global free energy. The formation of each contact also involves energetic costs for desolvation. These effects are taken into account by the choice of a reference state in which formation of contacts more frequently observed in the database is favored, and formation of contacts rarely observed is penalized. Tests and applications of SMoG were divided in two subsets: when binding of a ligand occurs at the protein surface, and when a ligand is deeply buried in a pocket. In fact, significantly different interactions are present in both situations, due to the different extent of solvent exposure, and, consequently, different contributions are observed in resulting solvation/desolvation terms. The surface set included 17 complexes, and the non-surface set 109 complexes, all taken from the PDB. Results were evaluated in three different ways. First, the authors tested the ability of SMoG to discriminate the native ligand geometry from alternative conformations generated by docking programs. Second, it was evaluated whether there was a correlation with the CHARMM interaction energy score. No correlation was found, even though one encouraging result was that the complexes scoring as the best candidates for CHARMM were also described as good candidates by SMoG. The loss of correlation was attributed to the fact that CHARMM estimates only vacuum enthalpy and does not take into account entropic factors like solvent effects. This indicates again the importance of a global evaluation of all factors affecting binding free energy. Finally, for the third evaluation, the correlation with binding free energy was tested. To do this, SMoG was applied to three protein-ligand systems: purine-nucleoside phosphorylase, SH3 Domain and HIV-1 protease. Initially, several negative results were observed, but these were explained on the grounds that the experimental IC_{50} or K_i of the complexes in the datasets were tested at varying concentrations of phosphate ions, and several of these molecules had been shown to have high sensitivity to phosphate concentration. By excluding these molecules from the data set, a better correlation was reported.

Summary

Knowledge-based potentials are promising methods for correlation of estimated binding free energy with experimental data, despite their occasionally unclear physical meaning. They appear as reliable as other scoring functions in correlating estimated and experimental binding affinities Since the starting point for these methods takes into account the influence of several factors ignored by traditional regression methods, knowledge-based methods can include, to some extent, entropic factors, that are in general not considered or underestimated with other methods. The solvation effect is analysed and incorporated in several ways in the calculations; however, the explicit solvent effect has not yet been adequately simulated. The method developed by Verkhivker *et al*. shows a comprehensive thermodynamic analysis but its parameterization is strictly linked to a limited starting data set. Wallqvist *et al* created a PMF in which contact frequencies between atomic surfaces are the starting point for parameterization, but again the training set is numerically limited, although more general than that of the previous PMF. BLEEP-1 and 2 are based on a robust training data set and they are the first to implicitly include solvation entropy, but a quite surprising result is the loss of predictivity resulting by the explicit inclusion of solvent effects. DrugScore is probably the most general PMF in parameterization and its score can be related to experimentally determined pK_i values. Finally, SMoG was developed using contact potentials that describe the interactions in a ligand-protein complex by meaning of a "coarse-grained" potential function. Analysis of correlation with binding free energy tested on three protein-ligand

systems was found quite satisfactory. With the increasing availability of structural information and binding data between protein-ligand complexes, further advances in parameterization could be achieved and water treatment improved.

EMPIRICAL FORCE FIELDS

In order to obtain fast scoring for thousands of compounds in the search for a lead candidate, several empirical force fields have been developed. Within these approaches, the free energy of binding is expressed as the sum of partial free energy terms corrected by a weighing coefficient ΔG_i:

$$\Delta G_{binding} = \Sigma \Delta G_i f_i (r_l, r_p) \quad (15)$$

where f_i is a function of the ligand and protein coordinates r_l and r_p, respectively [203]. Each term corresponds to defined contributions to the free energy accounting for electrostatic, ionic, hydrophobic interactions, solvation, flexibility, rotatable bonds, etc. The coefficients are obtained by fitting the experimental data derived from a training set of ligand-protein complexes for which binding affinities and three-dimensional structures are available. SCORE1 is a four-term function developed by Böhm using a training set of 45 complexes, and implemented in the popular docking software LUDI [194]. The standard deviation for SCORE1 over the 45 complexes was 1.9 kcal mol^{-1}. In a more recent version, SCORE2, water molecules in the binding site were explicitly taken into account to obtain some estimation of desolvation using a combination of energy minimization and molecular dynamics simulation [204]. The training set included 82 complexes with seven adjustable parameters and the standard deviation was 1.7 kcal mol^{-1}. Several other empirical functions have also been developed. One of them, VALIDATE, combines empirical functions and force fields and makes use of a training set of 51 complexes with 12 terms [205] obtaining a standard deviation of 1.6 kcal mol^{-1}. A four-term function was developed with the coefficients derived by a training set of 82 complexes, leading to a cross-validated error of 2.1 kcal mol^{-1} [206]. Further development of this function has made use of Bayesian statistics and a hydrogen bond term that includes water-mediated contacts [207]. The Jain scoring function contains 17 fitting parameters including terms for entropic and solvation effects [208]. The standard error in the prediction of affinities with the Jain function is about 1.4 kcal mol^{-1}. Another method called SCORE uses a function that includes a term for desolvation effects. The SCORE training set was based on 170 protein-ligand complexes and the cross-validated deviation was 1.5 kcal mol^{-1} [209]. Other empirical functions have been developed using a specific protein as a training set [210-214], thus limiting their general applicability.

Summary

The main advantage of empirical force fields for the determination of free energy of binding is the reduced time required for the classification of compounds. However, in most cases the solvent is only implicitly taken into account and the applicability to search lead compounds for a defined target is strongly dependent of the training set. Another source of approximation derives from the separation of the binding free energy in different contributions, an assumption not generally thermodynamically acceptable (*vide supra*) [125].

THE "NATURAL" FORCE FIELD

As shown above, understanding and computationally analyzing macromolecule-ligand interactions with molecular mechanics is somewhat hampered by the lack of well-conditioned interaction terms in the force fields beyond the (Coulombic) electrostatic and the (Lennard-Jones) London forces terms. The bulk of the terms in force fields are designed to reliably describe intramolecular (bond path) structure, i.e., bond lengths, angles, torsions, etc. All force fields perform fantastically well at this task in the vast proportion of cases. Indeed, it is only an increasingly small number of cases where quantum mechanics must be invoked to accurately model a small molecule structure. However, the intermolecular (through-space) interaction terms in most force fields are not as well developed due to a number of complexities. The two key intermolecular equations used in virtually all implementations of molecular mechanics, Coulomb's Law and the Lennard-Jones equation, are not particularly well suited to the complex biological environment. First, the fact that there is not a single, simple dielectric constant is a major problem. While there are accepted dielectric values and applied modification schemes, e.g. distance dependence, these are not easy to calibrate with experiment. Second, it must be remembered that the Lennard-Jones equation is simply the mathematical representation of the London force behavior; it is not a physical law itself. Ascribing other, more emergent, effects such as the hydrophobic effect to this simple mathematical function is problematic at best. Third, and most importantly, biological intermolecular interactions are **concerted**, that is, they arise from a unified attraction between two species, and are a complex and unknowable mix of multiple effects, including electrostatic forces, London forces, hydrophobic forces, as well as solvation/desolvation energies and the effect of entropy. Dill [125] states that there is no reason to expect that summing these disparate terms of free energy to derive an overall free energy of interaction for a system is valid.

A different approach is to exploit available free energy measurements as the basis of an interaction model. While there is a large set of ΔG values derived from experimental measurements of binding ligands to protein, the independent variables that might be used to rationalize this set are extremely complex arising from specific molecular structure, measurements conditions, and several features unique to each complex. In addition, these measurements themselves carry a relatively high degree of uncertainty. Another, more accessible and equally biochemically relevant, set of free energy data is the very large validated collection of LogP data. P is the partition coefficient for water to solvent ligand transfer, which is a thermodynamic free energy parameter. Generally the solvent is 1-octanol, thus, $LogP_{o/w}$. Because of this parameter's wide-spread applicability in QSAR and other property predictions, $LogP_{o/w}$ data have been collected

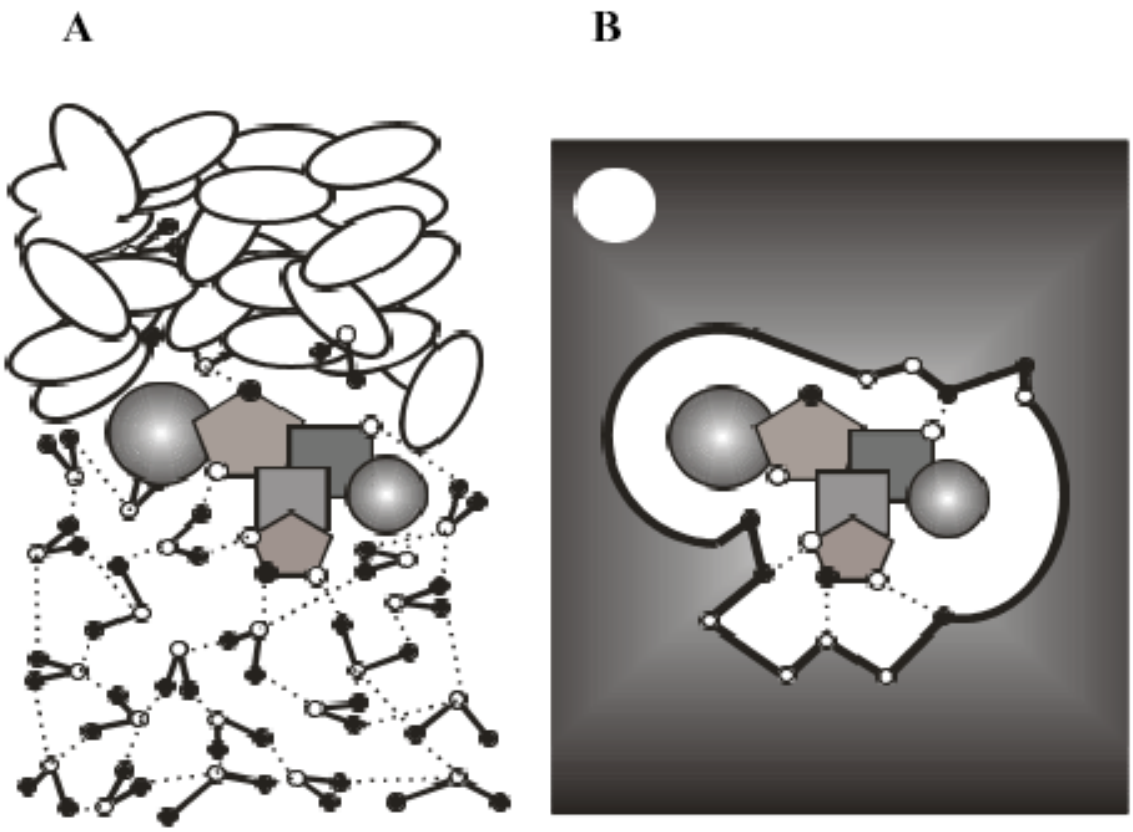


Fig. (3). (A) The molecule is in the LogP "shake flask" surrounded by water and 1-octanol (ovoids). Curved shapes represent hydrophobic regions, filled circles are H-bond donors, unfilled circles are H-bond acceptors. Note that some water is present in the octanol "layer". B) In the protein "site" the molecule ligand makes similar interactions – hydrophobic regions interact with hydrophobic pockets, polar regions of ligand are making hydrogen bonds with complementary functionalities of the site residues in polar pockets. Hydrophobic-polar mismatches between the ligand and the site can represent the energetic cost of desolvation.

and archived by medicinal chemists for many years [215]. This archive includes virtually all types of biologically important molecules and functional groups, and particularly emphasizes molecules of interest in drug discovery and design. The relevance of $LogP_{o/w}$ for the purpose we envision, a computational free energy model, is summarized in (Fig. (**3**)). In effect, the water solvent is a model environment for a "polar" pocket in a biomolecular receptor/ligand binding site. Water possesses both hydrogen bond donors and hydrogen bond acceptors. The 1-octanol solvent is similarly a model environment for a "hydrophobic" pocket. In effect, a small molecule targets a specific macromolecular binding site for the same energetic principles that it uses to partition between these two solvents. The associated computational interaction model, which is called HINT (for Hydropathic INTeractions), seeks to exploit the interaction information implicit in $LogP_{o/w}$ as the basis for describing and quantitating all interactions in the biological environment [216-220]. The HINT model describes a "natural" force field in that the totally empirical energetic terms are defined by a real experiment that encodes in $LogP_{o/w}$ all of the types of interactions including Coulombic, hydrogen bond and hydrophobic expected between two molecules in the biological environment. In addition, and most importantly, HINT encodes a free energy force field and thus includes entropy and solvation/desolvation in addition to the other enthalpic terms.

The interaction equation of HINT is as follows:

$$b_{ij} = a_i S_i a_j S_j T_{ij} r_{ij} + R_{ij} \quad (16)$$

where b_{ij} is the interaction score for the interaction of atoms i and j, that are on different molecules in the case of a ligand-protein binding calculation. a_i and a_j are the hydrophobic atom constants for atoms i and j. The hydrophobic atom constant is the contribution of that atom to the total $LogP_{o/w}$. S_i and S_j are the solvent accessible surface areas for the atoms. T_{ij} is a discriminant function for polar-polar interactions [217]. In most studies, r_{ij}, the hydropathic dependent distance function, is the simple exponential, e^{-r}, where r is the distance between atom i and j, and R_{ij}, is an implementation of the Lennard-Jones equation [217,219,221].

It has been shown previously that $LogP_{o/w}$ can be directly related to the free energy (ΔG) of molecule/solvent partitioning and that since $\Sigma a_i = LogP_{o/w}$, each a_i is a partial free energy, δg [217]. In this model these a_i parameters encode the propensity for (hydropathic) interaction and when they are correlated, as in the product $a_i a_j$, the result is that b_{ij} is proportional to δg for a specific atom-atom interaction and that the double sum $\Sigma \Sigma b_{ij}$ over all i and j is similarly related to ΔG for the entire intermolecular interaction [217]. This report is not meant to be a review of the many methods of predicting $LogP_{o/w}$ from structure and connectivity. There are many such papers

in the literature [222-225]. We will make two comments: 1) Because of its intimate relation to the Pomona MedChem data base, the method of Leo, as encoded in CLOG-P [226], deals most comprehensively with the types of molecules encountered in typical drug discovery scenarios and has several features of major importance with respect to the HINT model. Notably, the CLOG-P method deals rationally with complex cases of multiple polar groups on molecules (the "Polar Proximity Effect") and has a crucial factor for intramolecular hydrogen bond formation. 2) Validation studies of $LogP_{o/w}$ predictions typically ignore a very cogent fact – arguably the most important feature of prediction for drug design is how well the system works in relating structures within a single family of compounds (i.e., the project) and not how well the system works in predicting a diverse set of test compounds. In this regard, nearly all of the $LogP_{o/w}$ prediction methods perform very well. $LogP_{o/w}$ predictions for the HINT model are made by an adaptation of the CLOG-P method.

The HINT model has largely been validated through comparisons between HINT score calculations and measured free energies. A fairly large number of biomolecular systems have been evaluated this way, including protein-protein interactions in native and mutant hemoglobins [227,228], intercalation of antibiotic anticancer drugs into DNA [229], protein-sugar interactions in wheat germ agglutinin [230], RNA-aminoglycoside interactions [231], docking, scoring and drug design in HIV-1 reverse transcriptase [232,233] and cyclin-dependent kinase [234], and a number of other systems. A comprehensive free energy validation study was performed on the HINT score in a collection of 53 structurally well-characterized protein-ligand complexes for which accurate binding data was available [235] and, recently, in a collection of 76 complexes [236].

Two of these studies are reviewed here. First, 23 mutant hemoglobin tetramers were modeled from the starting native hemoglobins in both the deoxygenated (T) state and the oxygenated (R) state [228]. The free energies for dimer-dimer association for these mutant proteins ($\Delta G_{d\rightarrow t}$) were calculated and correlated with experimental data with fair success [237]. However, there are a significant number of water molecules bridging the dimer-dimer interface, and these contribute significantly to $\Delta G_{d\rightarrow t}$. Each mutation caused a unique perturbation to this field of water molecules. To account for this effect the GRID program [97] with the water probe was applied to region surrounding each mutation site and the energetically viable water molecules were then added to the models. Recalculation of $\Delta\Delta G_{d\rightarrow t}$ for each case yielded very robust correlations between the experimental and calculated changes in free energy: r = 0.79, standard error = 1.4 kcal mol^{-1} for deoxy hemoglobin mutants; r = 0.87, standard error = 0.8 kcal mol^{-1} (after the chemically reasonable deletion of three outliers) for the oxy hemoglobin mutants. As part of this work a conversion factor for arbitrary HINT score units to free energy was invoked. This factor was one kcal mol^{-1} is approximately equivalent to 515 HINT score units [228] .

The second study was an analysis of protein-ligand binding through a large array of systems, ranging in properties from very polar binding sites to very hydrophobic binding sites. The available complex structures in the Protein Data Bank were filtered for quality (X-ray resolution), metals at active site (structures excluded), and presence of water at active site (structures also excluded for this study). The corresponding binding data were collected from the literature and converted to free energies. Molecular models for the resulting 53 complexes were built and optimized in a number of ways that did not, in general, affect heavy atom placement [235]. In particular hydrogen bonds were optimized by rotation around R-XH bonds to place the donor hydrogens more appropriately for interaction with their acceptors (this procedure does not move any heavy atoms). Also, a number of cases where protonation or deprotonation of ionizable functional groups on the protein and/or ligand were noted. These structures were then modified appropriately. The HINT scores for the molecular models optimized as described were then correlated with measured free energies of binding to obtain a simple linear equation: $\Delta G^{\circ} = -0.00195\ H_{TOTAL} - 5.543$ (kcal mol^{-1}), where H_{TOTAL} is the total HINT score for the protein-ligand interaction. This interaction has r = 0.54 and standard error = 2.6 kcal mol^{-1}. The conversion factor between arbitrary HINT score units and free energy is one kcal mol^{-1} approximately equivalent to 513 HINT score units [235]. Within groups of ligands bound at the same protein the standard errors were much tighter, approaching ± 1 kcal mol^{-1}.

Five themes have emerged from this and related work: 1) There is a correlation between X-ray diffraction quality, i.e., resolution, and the ability of the models to predict accurate free energy. The standard errors of our correlations improve when more structures of intermediate resolution are excluded from the analysis. 2) Molecular mechanics force fields are not always completely appropriate for optimizing biomolecular structure. Most notably, the hydrophobic effect is penalized by Newtonian force fields because the partial group charge of hydrophobic entities is generally positive thus causing charge repulsion between them that is only compensated by the London forces. 3) The genesis of most entropy in biomolecular systems and the cause of the hydrophobic effect is the exclusion of water from the "unoccupied" active site. Water molecules still present at the active site and bridging between the interacting species make a significant energetic contribution and in a sense give some energetic clues about the makeup of the pre-binding active site. 4) The ionization state of protic functional groups in the active site can not be generally assumed "pH neutral". Each local environment, i.e., other functional groups, water molecules, buffer ions, effectively influences pK_a values. However, protons are not static, and the "real" case is likely to be a hybrid of several models. 5) Local minima in active site model building with force field energy minimization can be manifested in small but significant ways that can alarmingly affect free energy estimates. It is not unusual for an –OH on an amino acid residue to be arbitrarily oriented away from its likely hydrogen bond partner and for it to be "stuck" in a local minimum potential well. Similarly, water molecules, even when present at an active site, can be poorly oriented after energy minimization.

The first two of these themes are actually two parts of the same issue. The most obvious consequence of lower resolution crystallographic data is that more errors in atom placement would be expected, and this would certainly be manifested in computational predictions of binding free

energy based on these data. However, there is another, more subtle, consequence that also should be discussed. When higher quality diffraction data is available for crystallographic model refinement, the model building and fitting of the electron density is driven more by the actual data and less by the underlying molecular mechanics force field. In other words, there is less ambiguity of atom positioning and second order effects, such as hydrophobic-hydrophobic attractions, are not penalized in refinement. In contrast, with lower quality diffraction data, the electron density envelopes are not as well defined, and force field methods must be invoked to propel the refinement. While we have not, as of yet, performed a detailed analysis of this phenomenon, we have noted that we achieve much better correlations between our predictions and measured free energy when the data set excludes more lower resolution structures. (Fig. (**4**)) is a modified version of a figure from Cozzini *et al*. [235] that now includes more complexes and has been segregated to indicate those with crystallographic structures collected at < 2.5 Å for consideration in the regression analysis [236]. The standard error for this correlation is 1.8 kcal/mol and r = 0.85. A plausible explanation may involve a number of factors, but we believe that one of them will be the lack of any hydrophobic interaction terms in the molecular mechanics force fields. Unfavorable hydrophobic interactions arise from polar atoms in close proximity to apolar atoms. These HINT interactions can be interpreted as a desolvation penalty [218]. Also of potential concern is that molecular mechanics force fields can calculate certain of these interactions, e.g. between a carboxylate oxygen and a methyl, as being favorable because of the opposite sign of the partial charges on these atoms. This may also be a contributing factor to our ability to predict free energy with lower quality X-ray structures.

In most cases there is little information about the actual binding event. We are normally relegated to examination of the resulting, ligand-bound, structure with no information of how it got there. In particular we know nothing about the movement of water during ligation. Even when both unligated and ligand-bound structures are available, we do not know with certainty which water molecules were pushed out by the ligand, and which ones were carried in with the ligand. The third theme indicates that modeling water at the active site, particularly in cases where the water bridges between the ligand and receptor, is crucial for accurate free energy predictions with our technology. Even though our ligand-receptor data set [235] was filtered to remove cases where water was presumably a large factor, there were several cases where nearby water molecules appeared to be impacting the models and distorting the free energy predictions. In the section to follow, we present portions of an analysis of the binding site in HIV-1 protease where the water molecule(s) present and bridging between the protein and ligand are critical to the system. Indeed, proteins, ligands and water must be considered as systems. Each piece has a role to play in the binding event.

We have been interested in the impact of functional group ionization in the context of ligand binding for some time. In fact, in one of the first applications of HINT (unpublished results), an analysis of the HIV-1 enzyme bound with the Abbott ligand A74704, we saw a strong indication that one of the two active site aspartates must be protonated for the structure to make sense. Recently we have been quantitating this effect in a procedure we call *Computational Titration* [238]. Basically the idea is that we build a family of models for each protonation level, i.e., a

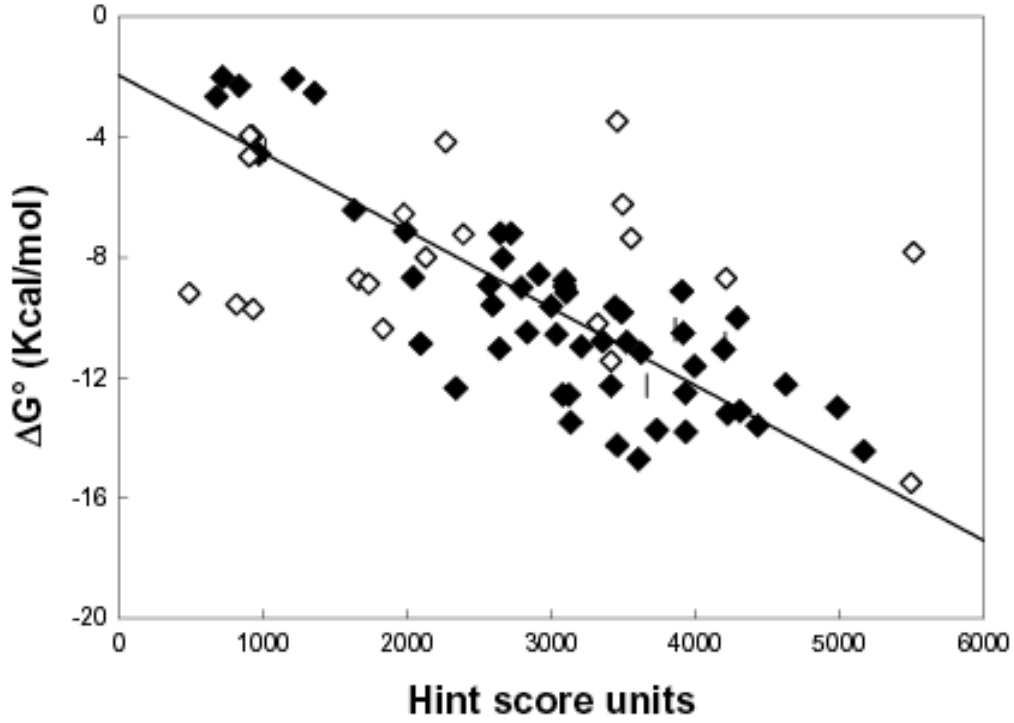


Fig. (4). Correlation between the experimental free energy of binding and HINT score units. The straight line through data points is the best least-squares fit, carried out for 56 complexes with resolution < 2.5 Å, and at least 3 ligands for each protein (closed diamonds), showing R = 0.85, R^2 = 0.72 and SE = 1.8 kcal mol^{-1}. For 76 protein-ligand complexes with resolution <3.2 Å (open and closed diamonds), the fitting (not shown) gives R = 0.69, R^2 = 0.48, SE = 2.4 kcal mol^{-1}.

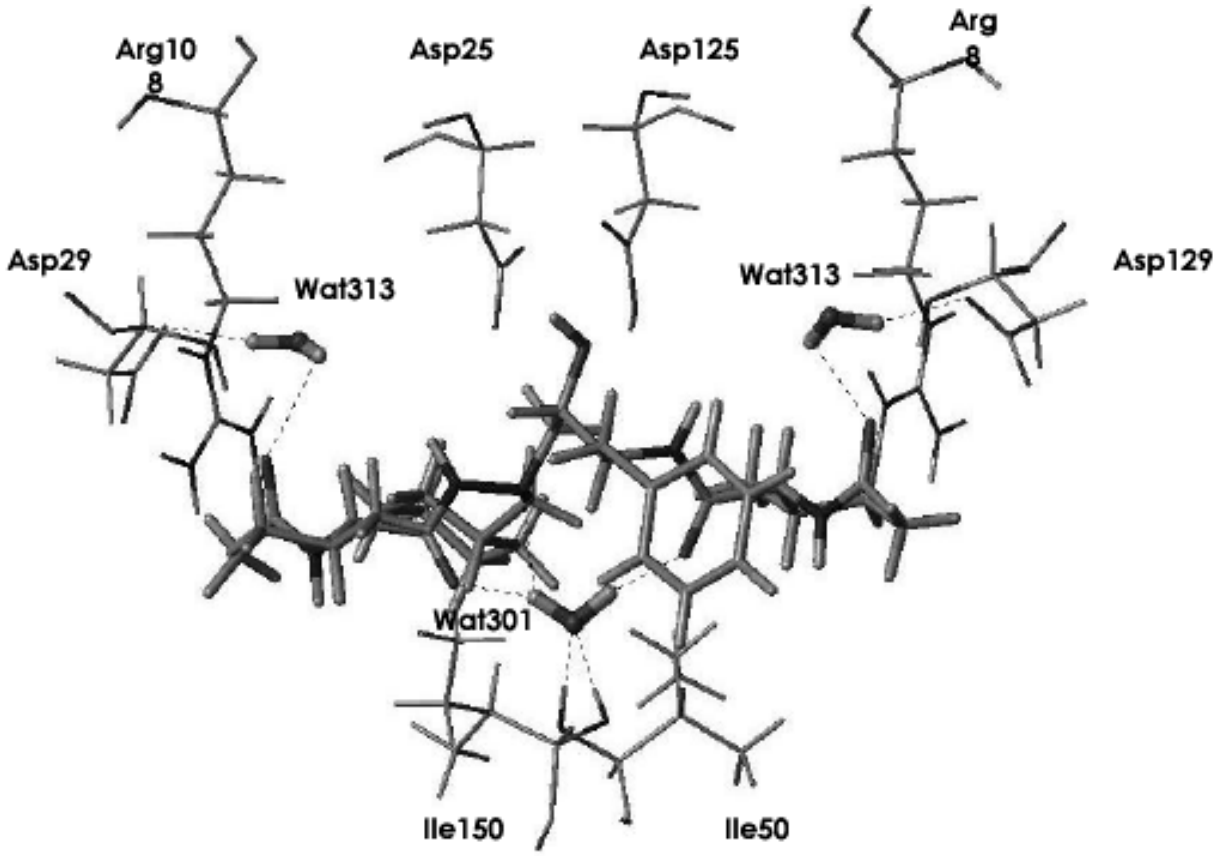


Fig. (5). Structure of the active site of HIV-1 protease, complexed with CGP53820, a psudosymmetric inhibitor [244]. The conserved water 301 and the semi-conserved water 313 and 313' are hydrogen bonded with the ligand and protein residues.

proton may often be added to the system in multiple locations – at ionizable amino acid residues and ligand groups. The scores for all accessible models at each protonation level are then averaged. A plot of these averages as a function of protonation level indicates the optimum protonation as a peak in average score. The key point here is that modeling a protein-ligand interaction and attempting to computationally predict free energy without careful examination of the protonation state at the active site will be likely inaccurate. It should also be noted that in many cases the experimental binding was not necessarily measured at the optimum pH, and that the X-ray crystals were not necessarily grown at the same pH as the binding was measured. These effects must also be factored into predictions.

The problem of local minima in molecular mechanics structure optimization is well known. For the purposes of the HINT score algorithm, we have been particularly interested in the very large effects on score arising from a number of small structural perturbations involving only hydrogen positions. When hydrogen atoms are added to structurally determined atoms with the algorithms in Sybyl, InsightII or any other modeling package they are added to their parent heavy atoms with appropriate and consistent geometries for the heavy atom. However, no accounting is made of the through space interactions that hydrogen may be involved with. Structure optimization of the hydrogen set corrects the vast majority of the resultant van der Waals clashes, etc., but in a number of cases, usually involving –OH groups on serine, threonine, and tyrosine or on the ligand, an R-XH_n bond is rotated away from the expected acceptor and what should have been a hydrogen bond is scored as a much weaker interaction. This is only, as we mentioned above, a manifestation of the local minima problem, but it has a large impact on free energy predictions. Similarly, water molecules are routinely trapped in local minima, and their contribution to free energy predictions are not exploited. Thus, in our model building we routinely scan for, and manually correct by rotation about the R-XH_n bond, these situations.

The above five themes are driving our current research and development of the HINT model. We are envisioning a principle of "fixing the important bits" for free energy prediction, and are in the process of computationally automating "fixes" for these issues. For example, our code now will scan the active site for ionizable residues, acids or bases, and optimize for highest HINT score the protonation state of these residues or functional groups. The code also identifies and corrects the local minima described in the fifth theme. Finally, we are automatically optimizing water molecules at the active site and are identifying those that would appear to contribute to free energy. We should emphasize that the vast majority of these "fixes" are crystallographically indistinct. That is, even after these optimizations, the resulting models would still fit within the original experimental electron density envelope. We are also exploring how the HINT force field can interface with crystallographic energy refinement to improve the quality of poorer resolution structures.

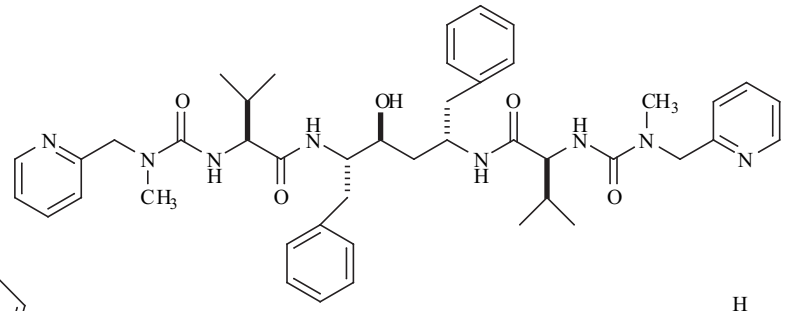
HIV-1 PROTEASE, INHIBITORS, WATER AND FREE ENERGY OF BINDING

HIV-1 protease belongs to aspartate protease family, because the active site contains two aspartate residues that play a key role in the catalytic hydrolysis of peptidic bonds [239]. The protease is one of the three fundamental enzymes of the retroviral HIV, the others being reverse transcriptase and integrase, and is vital for the HIV replication cycle. For this reason, HIV protease has been a target for the design of enzyme inhibitors to be used as drugs in AIDS therapy since its structure was first reported [240]. Also, because of its therapeutic importance and the large volume of experimental research into its structure, ligand binding and inhibitors [241,242], HIV-1 has been the target of many computational studies, including a number reviewed here. The enzyme is a dimer formed by two identical polypeptide chains. The dimer possesses a two-fold symmetry axis leading to a single symmetric active site (Fig. (**5**)) containing residues from both subunits [243]. Key features of HIV protease are: a conserved water molecule, water 301, placed on the symmetry axis and bridging the two subunits, and the "catalytic" water 300, observed in the free form and coordinated to Asp 25 and Asp125, one of these residues being protonated and the other unprotonated. Water 301 is hydrogen bonded to the N-amide of Ile50 and Ile150, and, in the presence of specific peptidic inhibitors to two carbonyl oxygens, as shown in the case represented in (Fig. (**5**)) [244]. These inhibitors were designed based on the shape of the active site as modified by the presence of the conserved water 301. Alternatively, a number of inhibitors were designed to displace water 301 [93]. Examples of these are cyclic urea compounds that place a carbonyl oxygen in the position occupied by water 301, thus forming hydrogen bonds with Ile50 and Ile150. In some, but not all, complexes the crystallographic analysis of HIV-1 protease detected two other water molecules, water 313 and water 313' (Fig. (**5**)).

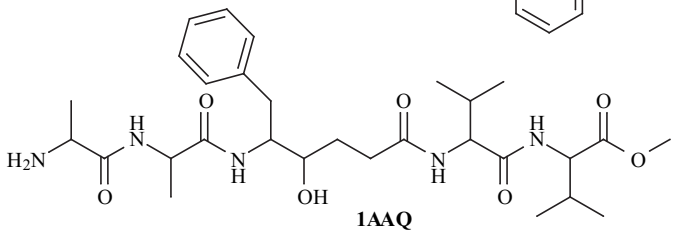
Table 1. Crystallographic Resolution, Binding Affinity and HINT Score in the Absence and Presence of the Contribution of Water 301 for 23 HIV-1-Inhibitor Complexes

PDB	Resolution (Å)	ΔG° (Kcal/mol)	HS_pl[a]	HS_pl+lw[b]
1HBV	2.30	-8.68	2042	2819
1AJV	2.00	-10.52	3916	3916
1SBG	2.30	-10.56	3037	4163
1AJX	2.00	-10.79	3357	3357
1G2K	1.95	-10.82	3525	3525
1HIH	2.20	-10.97	3210	4290
1HTF	2.20	-11.04	2641	3367
1G35	1.80	-11.06	4198	4198
1AAQ	2.50	-11.45	3416	4049
1HVL	1.80	-12.27	3416	4669
1HIV	2.00	-12.27	3660	4986
4PHV	2.10	-12.51	3932	4721
1HPV	1.90	-12.57	3080	4138
1HPS	2.30	-12.57	3124	3953
1DMP	2.00	-12.99	4988	4988
7HVP	2.40	-13.11	4311	5540
1HTG	2.00	-13.20	4226	5498
1HXB	2.30	-13.49	3135	4184
1HVI	1.80	-13.74	3734	4945
1HVK	1.80	-13.80	3935	4999
1HVJ	1.80	-14.25	3460	4663
1QBT	1.80	-14.44	5170	5170
1HXW	1.80	-14.71	3607	5061

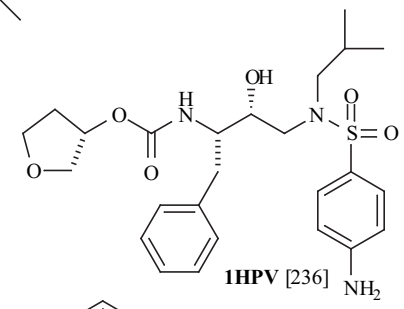
[a] HS_pl is the HINT score for the protein-ligand interaction. [b]HS_pl+lw is the sum of the HINT score for the protein-ligand and ligand-water301 interactions.



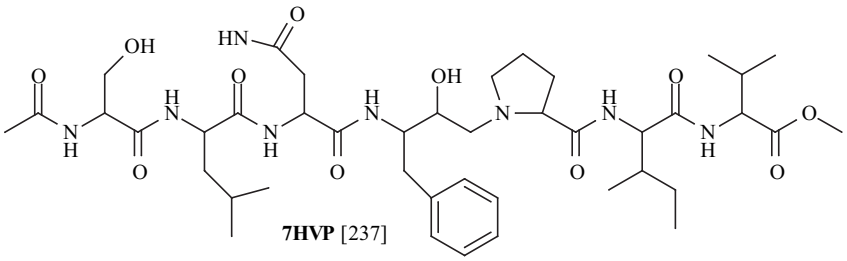
1HVJ [234]



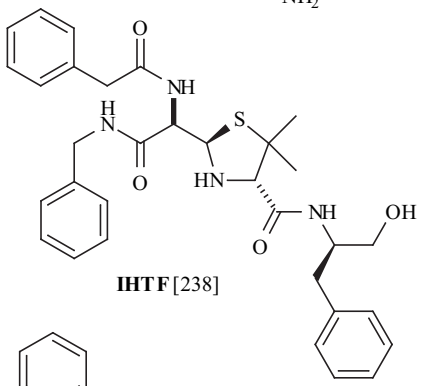
1AAQ [235]



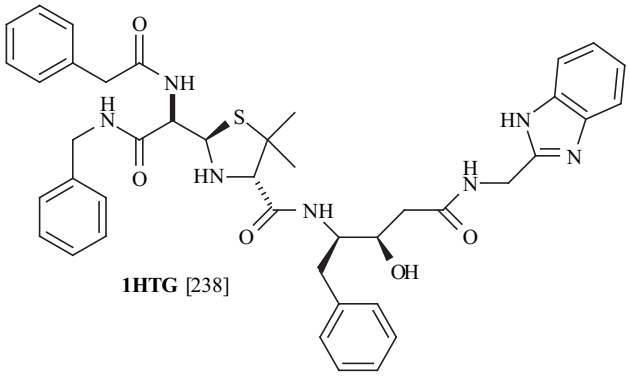
1HPV [236]



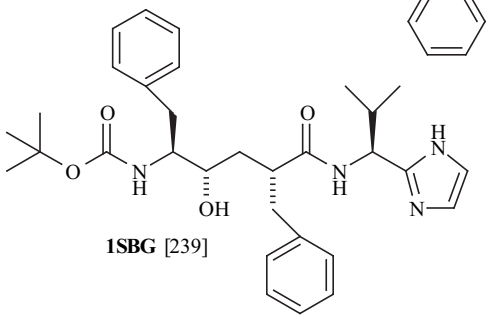
7HVP [237]



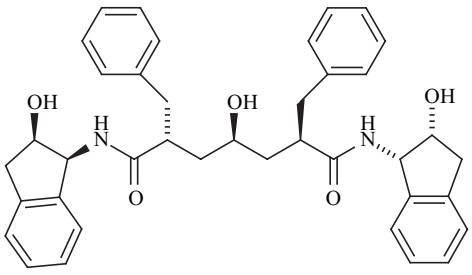
IHTF [238]



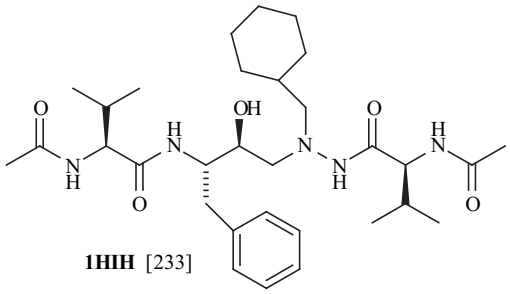
1HTG [238]



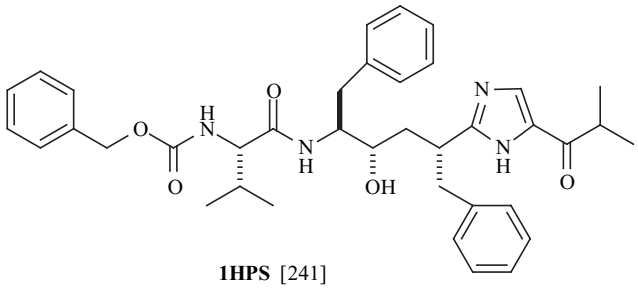
1SBG [239]



4PHV [240]

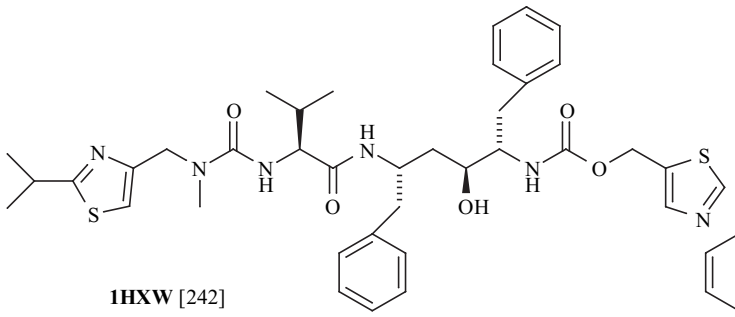


1HIH [233]

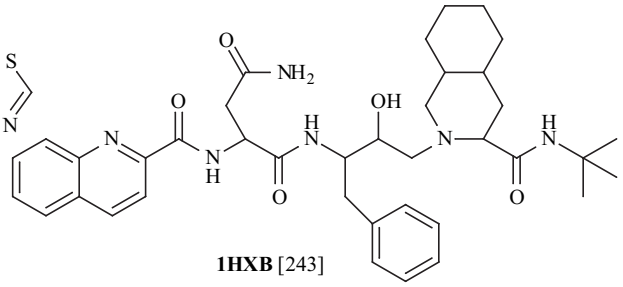


1HPS [241]

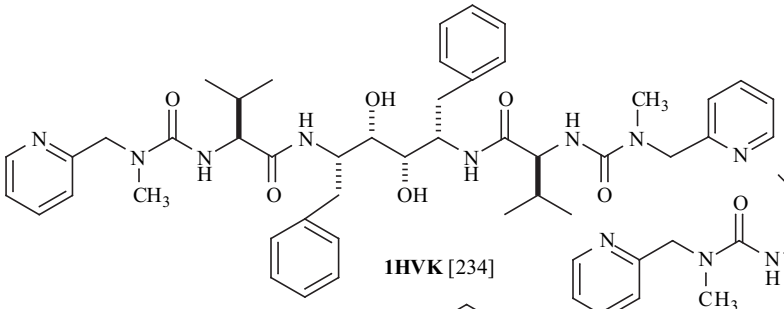
Scheme 2). contd.....



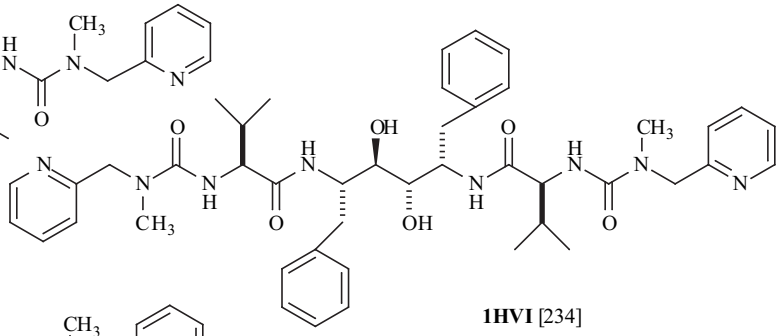
1HXW [242]



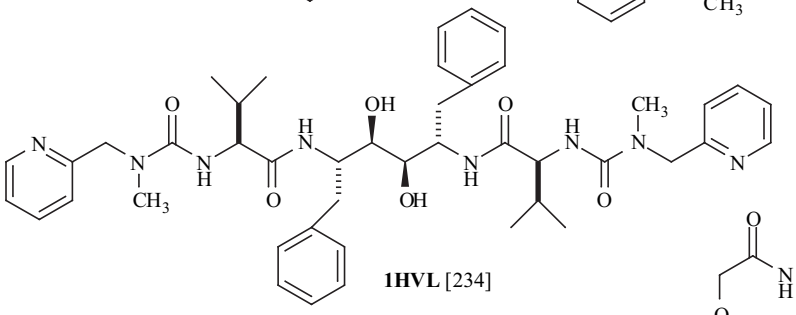
1HXB [243]



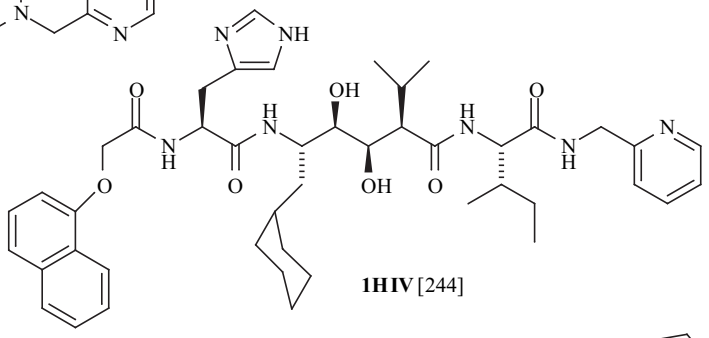
1HVK [234]



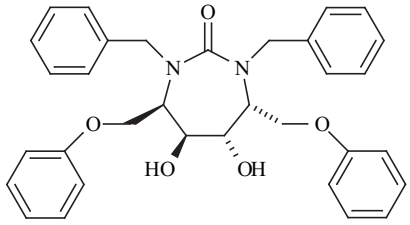
1HVI [234]



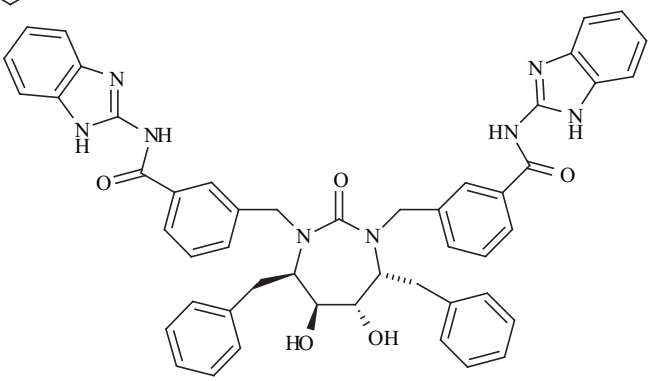
1HVL [234]



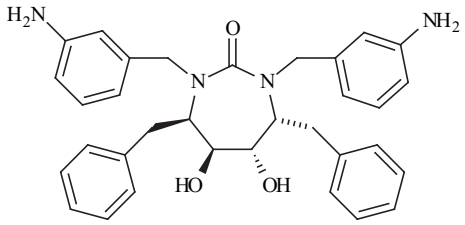
1H1V [244]



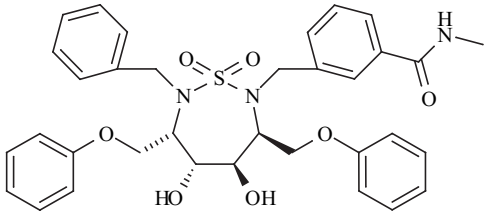
1AJX [245]



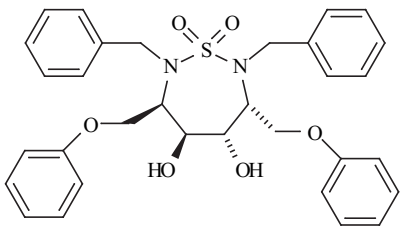
1QBT [246]



IDMP [247]

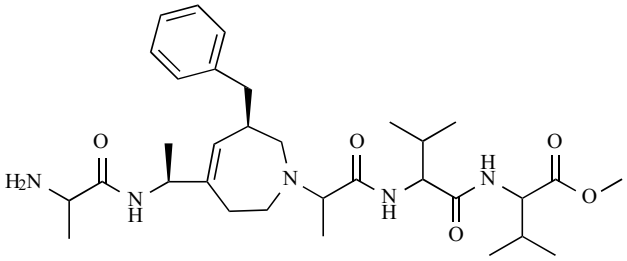


1G2K [248]

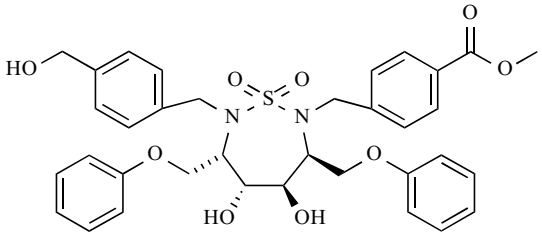


IAJV [245]

(Scheme 2). contd.....



1HBV [249]



1G35 [248]

Scheme 2. Chemical structure of the 23 HIV-1 protease inhibitors.

As part of our efforts to understand the role of water in ligand binding, we are currently addressing the question of the contribution of the crystallographically detected water molecules to the free energy of binding of inhibitors to HIV-1 protease. We selected 23 inhibitor-HIV-protease complexes of known three-dimensional structure and binding affinity (Table (**1**)). These inhibitors are characterized by different chemical structures, peptidomimetic diols, and hydroxyethylenic, cyclic ureidic and cyclic sulfonamide derivatives (Scheme (**2**)). Their interaction with the active site was evaluated by HINT either with or without the inclusion of water 301 (Table (1)), following procedures previously described [235]. When water 301 is not taken into account, the total HINT score is equal to the HINT score of the protein-ligand interaction (H_{P-L}). When water 301 is taken into account, the total HINT score is equal to $H_{P-L} + H_{L-W}$, where H_{L-W} is the HINT score of the interaction of the ligand with the conserved water. The correlation between HINT score and $\Delta G°$ significantly improves when the energetic balance includes the water-ligand interaction, as shown in (Fig. (**6**)), the calculated standard error being 1.3 and 1.0 kcal mol^{-1}, in the absence and presence of water, respectively. We further evaluated the contribution to the binding free energy of the more mobile waters 313 and 313' on a subset of ligand-protease complexes. When these water molecules were not observed in the crystal structures of the complexes, the GRID software [97] was used to place them. We should note that, as a validation, GRID correctly located water 313 and 313', as well as 301, in the active sites of structures where these water molecules had been crystallographically observed and manually removed. The overall contribution of water 313 and 313' to the total HINT score is much less than that of water 301. The correlation between $\Delta G°$ and HINT scores with and without water 313 and 313' only slightly improved. This analysis, extended to inhibitors of HIV-2 protease and endothiapepsin, confirmed the relevant contribution of conserved water molecules to the free energy of binding (unpublished results). The application of this procedure to a more complete set of protein-ligand complexes of three-dimensional structures with a comparable quality will allow us to correlate the energetic contribution of active site water molecules with different mobility to the free energy of ligand binding.

CONCLUSIONS

The quantitative modeling of water in biological systems is a surprisingly difficult problem. Understanding and parsing the many roles that water plays to stabilize biological molecules and complexes is a large part of the problem. Second, while it is generally the foundation of computational biomolecular modeling, crystallographic data has a number of limitations that, at best, add complexity to calculations in the biological milieu. Of particular interest here, the presence and positions of water molecules are highly dependent on the resolution of the crystallographic data collection. Also, since hydrogen atoms are virtually never located in X-ray analyses, the presence and description of crucial structural features such as hydrogen bonds are often more of a modeling result than an experimental result. However, despite these constraints, a quantitative energetic description of the recognition between ligand/drug and biomacromolecule/protein remains a crucial and fundamental goal for both medicinal chemists and biochemists. A wide array of approaches and models have been developed, of which a few also try to capture the elusive but essential contribution of water molecules to the free energy of binding. The key question is: how do these methods perform in free energy prediction? The best results indicate errors in the order of a few kcal mol^{-1}, thus serving relatively well for *de novo* drug design and *in silico* screening of chemical libraries in the search of new leads.

The follow-up of the genomics-proteomics revolution leads to many new challenges for the discovery of more potent and efficient drugs. The design of drugs that interfere with the protein-protein recognition supporting signal transduction-mediated diseases is just one challenge [245]. Whereas structural characterization and physico-chemical description of ligand-protein are in a reasonably advanced state, the prediction of protein-protein and protein-DNA complexes are still mostly inadequate. In these processes, water molecules play a key and yet poorly understood role as ordered water at the interface may also mediate inter-protein signals, as shown in the case of dimeric hemoglobin [246]. Most, if not all, of the drug discovery computational methods have been developed for the evaluation of the interactions of small ligands in narrow and well-defined active sites. How will they perform in quantitative evaluation of the interaction between protein-protein

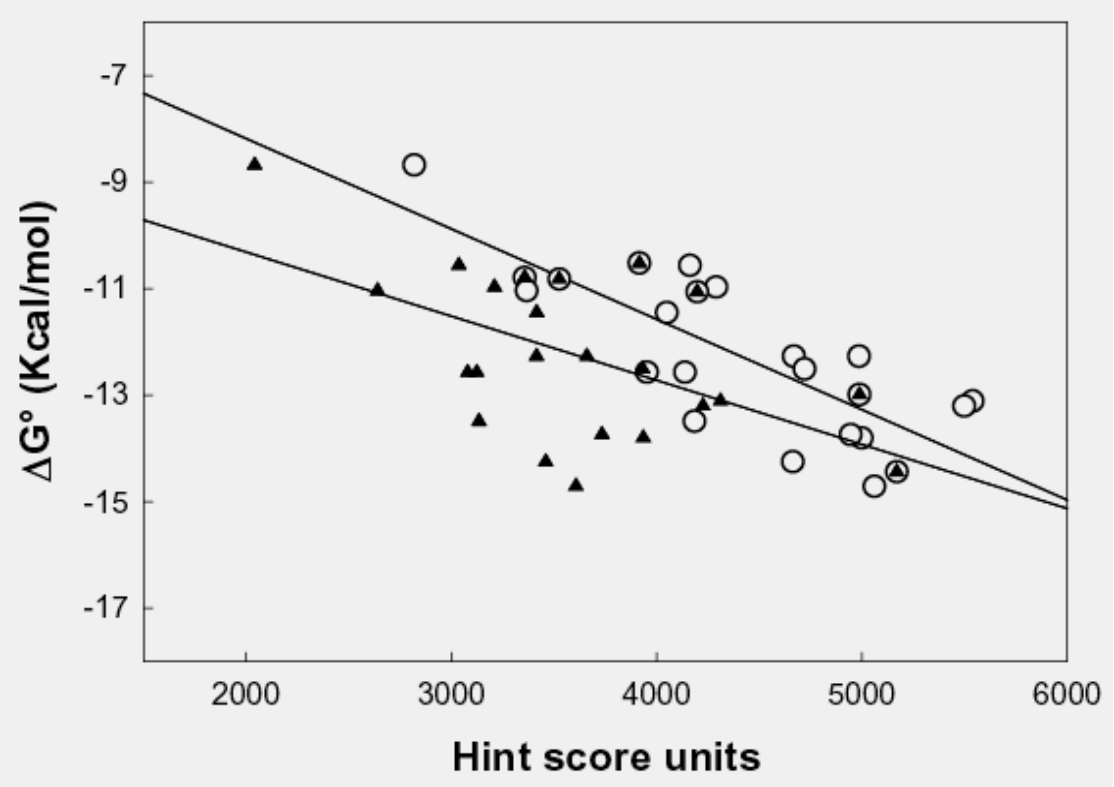


Fig. (6). Correlation between experimental free energy and HINT score units for 23 HIV-1 protease-lignand complexes with (open circles) and without (closed triangles) the contribution of water 301. The straight line through data points is the bes least squares fit with SE of 1.3 and 1.0 kcal mol^{-1}, in the absence and presence of water 301, respectively.

complexes formed by recognition of extensive, widely dispersed, and shallow active sites? In these cases, in the absence of high resolution crystallographic data, the application of reliable computational methods to validate and/or create a set of water molecules may be a reasonable first step. Several complexes have already been analyzed using approaches that include the contribution of interface water molecules. However, comprehensive and in depth analyses have not yet been attempted. Certainly, the functional and thermodynamic role of water molecules trapped between partner proteins and proteins and nucleic acids will be a major research emphasis in the near future.

ACKNOWLEDGEMENTS

We are grateful to Dr. Gene Lamm for careful reading the manuscript and helpful discussion. A. M. gratefully acknowledges the support of Italian Ministry of Instruction, University and Research (Grants COFIN 2003 and FIRB2003) and the National Institute for the Physics of Matter.

ABBREVIATIONS

SASA = Solvent accessible surface area

MM = Molecular mechanics

MD = Molecular dynamics

MC = Monte Carlo

GB/SA = Generalized Born/surface area

FEP = Free energy perturbation

TI = Thermodynamic integration

LIE = Linear interaction energy

MM/PBSA = Molecular mechanics/Poisson-Boltzmann surface area

PMF = Potential of mean force

REFERENCES

[1] Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J.P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J.C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R.H.; Wilson, R.K.; Hillier, L.W.; McPherson, J.D.; Marra, M.A.; Mardis, E.R.; Fulton, L.A.; Chinwalla, A.T.; Pepin, K.H.; Gish, W.R.; Chissoe, S.L.; Wendl, M.C.; Delehaunty, K.D.; Miner, T.L.; Delehaunty, A.; Kramer, J.B.; Cook, L.L.; Fulton, R.S.; Johnson, D.L.; Minx, P.J.; Clifton, S.W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J.F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M.; Gibbs, R.A.; Muzny, D.M.; Scherer, S.E.; Bouck, J.B.; Sodergren, E.J.; Worley, K.C.; Rives, C.M.; Gorrell, J.H.; Metzker, M.L.; Naylor, S.L.; Kucherlapati, R.S.; Nelson, D.L.; Weinstock, G.M.; Sakaki, Y.; Fujiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.;

Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissenbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Pelletier, E.; Robert, C.; Wincker, P.; Rosenthal, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Yang, H.M.; Yu, J.; Wang, J.; Huang, G.Y.; Gu, J.; Hood, L.; Rowen, L.; Madan, A.; Qin, S.Z.; Davis, R.W.; Federspiel, N.A.; Abola, A.P.; Proctor, M.J.; Myers, R.M.; Schmutz, J.; Dickson, M.; Grimwood, J.; Cox, D.R.; Olson, M.V.; Kaul, R.; Raymond, C.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G.A.; Athanasiou, M.; Schultz, R.; Roe, B.A.; Chen, F.; Pan, H.Q.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W.R.; de la Bastide, M.; Dedhia, N.; Blocker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J.A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; Brown, D.G.; Burge, C.B.; Cerutti, L.; Chen, H.C.; Church, D.; Clamp, M.; Copley, R.R.; Doerks, T.; Eddy, S.R.; Eichler, E.E.; Furey, T.S.; Galagan, J.; Gilbert, J.G.R.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.; Hokamp, K.; Jang, W.H.; Johnson, L.S.; Jones, T.A.; Kasif, S.; Kaspryzk, A.; Kennedy, S.; Kent, W.J.; Kitts, P.; Koonin, E.V.; Korf, I.; Kulp, D.; Lancet, D.; Lowe, T.M.; McLysaght, A.; Mikkelsen, T.; Moran, J.V.; Mulder, N.; Pollara, V.J.; Ponting, C.P.; Schuler, G.; Schultz, J.R.; Slater, G.; Smit, A.F.A.; Stupka, E.; Szustakowki, J.; Thierry-Mieg, D.; Thierry-Mieg, J.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y.I.; Wolfe, K.H.; Yang, S.P.; Yeh, R.F.; Collins, F.; Guyer, M.S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K.A.; Patrinos, A.; Morgan, M.J. *Nature* **2001**, *409*, 860-869.

[2] O'Brien, S.J.; Menotti-Raymond, M.; Murphy, W.J.; Yuhki, N. *Annu. Rev. Genet.* **2002**, *36*, 657-686.

[3] Waterston, R.H.; Lindblad-Toh, K.; Birney, E.; Rogers, J.; Abril, J.F.; Agarwal, P.; Agarwala, R.; Ainscough, R.; Alexandersson, M.; An, P.; Antonarakis, S.E.; Attwood, J.; Baertsch, R.; Bailey, J.; Barlow, K.; Beck, S.; Berry, E.; Birren, B.; Bloom, T.; Bork, P.; Botcherby, M.; Bray, N.; Brent, M.R.; Brown, D.G.; Brown, S.D.; Bult, C.; Burton, J.; Butler, J.; Campbell, R.D.; Carninci, P.; Cawley, S.; Chiaromonte, F.; Chinwalla, A.T.; Church, D.M.; Clamp, M.; Clee, C.; Collins, F.S.; Cook, L.L.; Copley, R.R.; Coulson, A.; Couronne, O.; Cuff, J.; Curwen, V.; Cutts, T.; Daly, M.; David, R.; Davies, J.; Delehaunty, K.D.; Deri, J.; Dermitzakis, E.T.; Dewey, C.; Dickens, N.J.; Diekhans, M.; Dodge, S.; Dubchak, I.; Dunn, D.M.; Eddy, S.R.; Elnitski, L.; Emes, R.D.; Eswara, P.; Eyras, E.; Felsenfeld, A.; Fewell, G.A.; Flicek, P.; Foley, K.; Frankel, W.N.; Fulton, L.A.; Fulton, R.S.; Furey, T.S.; Gage, D.; Gibbs, R.A.; Glusman, G.; Gnerre, S.; Goldman, N.; Goodstadt, L.; Grafham, D.; Graves, T.A.; Green, E.D.; Gregory, S.; Guigo, R.; Guyer, M.; Hardison, R.C.; Haussler, D.; Hayashizaki, Y.; Hillier, L.W.; Hinrichs, A.; Hlavina, W.; Holzer, T.; Hsu, F.; Hua, A.; Hubbard, T.; Hunt, A.; Jackson, I.; Jaffe, D.B.; Johnson, L.S.; Jones, M.; Jones, T.A.; Joy, A.; Kamal, M.; Karlsson, E.K.; Karolchik, D.; Kasprzyk, A.; Kawai, J.; Keibler, E.; Kells, C.; Kent, W.J.; Kirby, A.; Kolbe, D.L.; Korf, I.; Kucherlapati, R.S.; Kulbokas, E.J.; Kulp, D.; Landers, T.; Leger, J.P.; Leonard, S.; Letunic, I.; Levine, R.; Li, J.; Li, M.; Lloyd, C.; Lucas, S.; Ma, B.; Maglott, D.R.; Mardis, E.R.; Matthews, L.; Mauceli, E.; Mayer, J.H.; McCarthy, M.; McCombie, W.R.; McLaren, S.; McLay, K.; McPherson, J.D.; Meldrim, J.; Meredith, B.; Mesirov, J.P.; Miller, W.; Miner, T.L.; Mongin, E.; Montgomery, K.T.; Morgan, M.; Mott, R.; Mullikin, J.C.; Muzny, D.M.; Nash, W.E.; Nelson, J.O.; Nhan, M.N.; Nicol, R.; Ning, Z.; Nusbaum, C.; O'Connor, M.J.; Okazaki, Y.; Oliver, K.; Larty, E.O.; Pachter, L.; Parra, G.; Pepin, K.H.; Peterson, J.; Pevzner, P.; Plumb, R.; Pohl, C.S.; Poliakov, A.; Ponce, T.C.; Ponting, C.P.; Potter, S.; Quail, M.; Reymond, A.; Roe, B.A.; Roskin, K.M.; Rubin, E.M.; Rust, A.G.; Santos, R.; Sapojnikov, V.; Schultz, B.; Schultz, J.; Schwartz, M.S.; Schwartz, S.; Scott, C.; Seaman, S.; Searle, S.; Sharpe, T.; Sheridan, A.; Shownkeen, R.; Sims, S.; Singer, J.B.; Slater, G.; Smit, A.; Smith, D.R.; Spencer, B.; Stabenau, A.; Strange-Thomann, N.S.; Sugnet, C.; Suyama, M.; Tesler, G.; Thompson, J.; Torrents, D.; Trevaskis, E.; Tromp, J.; Ucla, C.; Vidal, A.U.; Vinson, J.P.; von Niederhausern, A.C.; Wade, C.M.; Wall, M.; Weber, R.J.; Weiss, R.B.; Wendl, M.C.; West, A.P.; Wetterstrand, K.; Wheeler, R.; Whelan, S.; Wierzbowski, J.; Willey, D.; Williams, S.; Wilson, R.K.; Winter, E.; Worley, K.C.; Wyman, D.; Yang, S.; Yang, S.P.; Zdobnov, E.M.; Zody, M.C.; Lander, E.S. *Nature* **2002**, *420*, 520-562.

[4] Nature 2000, 408, 796-815.

[5] Couzin, J. *Science* **2003**, *301*, 1176-1177.

[6] DeFife, K.M.; Wong-Staal, F. *Curr. Opin. Drug Discov. Devel*. **2002**, *5*, 683-689.

[7] Licinio, J.; Wong, M.-L. *Pharmogenomics: The Search for Individualized Therapies*; Wiley-VCH, **2002**.

[8] Bleicher, K.H. *Curr. Med. Chem.* **2002**, *9*, 2077-2084.

[9] Sehgal, A. *Curr. Opin. Drug Discov. Devel.* **2002**, *5*, 526-531.

[10] Tyers, M.; Mann, M. *Nature* **2003**, *422*, 193-197.

[11] Ryan, T.E.; Patterson, S.D. *Trends Biotechnol*. **2002**, *20*, S45-51.

[12] Watkins, S.M.; German, J.B. *Curr. Opin. Biotech*. **2002**, *13*, 512-516.

[13] Hanash, S. *Nature* **2003**, *422*, 226-232.

[14] Hou, J.T.; Sims, G.E.; Zhang, C.; Kim, S.H. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2386-2390.

[15] Yokoyama, S.; Hirota, H.; Kigawa, T.; Yabuki, T.; Shlrouzu, M.; Terada, T.; Ito, Y.; Matsuo, Y.; Kuroda, Y.; Nishimura, Y.; Kyogoku, Y.; Miki, K.; Masui, R.; Kuramitsu, S. *Nature Struct. Biol*. **2000**, *7*, 943-945, Suppl.

[16] Sali, A.; Glaeser, R.; Earnest, T.; Baumeister, W. *Nature* **2003**, *422*, 216-225.

[17] Buchanan, S.G. *Curr. Opin. Drug Discov. Devel.* **2002**, *5*, 367-381.

[18] Kuhn, P.; Wilson, K.; Patch, M.G.; Stevens, R.C. *Curr. Opin. Chem. Biol.* **2002**, *6*, 704-710.

[19] Pellecchia, M.; Sem, D.S.; Wuthrich, K. *Nature Rev. Drug Discov*. **2002**, *1*, 211-219.

[20] Sandanayaka, V.P.; Prashad, A.S. *Curr. Med. Chem.* **2002**, *9*, 1145-1165.

[21] Rosin, C.D.; Belew, R.K.; Walker, W.L.; Morris, G.M.; Olson, A.J.; Goodsell, D.S. *J. Mol. Biol.* **1999**, *287*, 77-92.

[22] Kingston, D.G.I.; Newman, D.J. *Curr. Opin. Drug Discov. Develop.* **2002**, *5*, 304-316.

[23] Geysen, H.M.; Schoenen, F.; Wagner, D.; Wagner, R. *Nature Rev. Drug Discov*. **2003**, *2,* 222-230.

[24] Goodford, P.J. *J. Med. Chem.* **1984**, *27*, 558-64.

[25] Abraham, D.J. I*ntra-Science Chem. Rept.* **1974**, *8*, 4-9.

[26] Wade, R.C. *Structure* **1997**, *5*, 1139-1145.

[27] Whittle, P.J.; Blundell, T.L. *Annu. Rev. Biophys. Biomol. Struct*. **1994**, *23*, 349-375.

[28] Rowland, R.S. *Curr. Opin. Drug Discov. Devel.* **2002**, *5*, 613-619.

[29] Ooms, F. *Curr. Med. Chem.* **2000**, *7,* 141-158.

[30] Salemme, F.R.; Spurlino, J.; Bone, R. *Structure* **1997**, *5*, 319-324.

[31] Schneider, G. *Curr. Med. Chem*. **2002**, *9*, 2095-2101.

[32] Adang, A.E.P.; Hermkens, P.H.H. *Curr. Med. Chem.* **2001**, *8*, 985-998.

[33] Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. *Quart. Rev. Biophys.* **1999**, *32*, 211-40.

[34] Vogtherr, M.; Fiebig, K. *Exs*. **2003**, 183-202.

[35] Hicks, R.P. *Curr. Med. Chem.* **2001**, *8,* 627-650.

[36] Darvas, F.; Keseru, G.; Papp, A.; Dorman, G.; Urge, L.; Krajcsi, P. *Curr. Top. Med. Chem*. **2002**, *2*, 1287-1304.

[37] Bajorath, J. *Nature Rev. Drug Discov*. **2002**, *1*, 882-94.

[38] Toledo-Sherman, L.M.; Chen, D. *Curr. Opin. Drug Discov. Develop.* **2002**, *5*, 414-421.

[39] Waszkowycz, B. C*urr. Opin. Drug. Discov. Devel.* **2002**, *5*, 407-413.

[40] Charifson, P.S.; Walters, W.P. *Molecular Diversity* **2000**, *5*, 185-197.

[41] Pearlman, D.A.; Charifson, P.S. *J. Med. Chem.* **2001**, *44*, 3417-3423.

[42] Stahura, F.L.; Bajorat, J. *Curr. Med. Chem.* **2003**, *10*, 707-715.

[43] Bohm, H.J.; Klebe, G. *Angew. Chem. Int. Ed*. **1996**, *35*, 2588-2614.

[44] Muegge, I.; Rarey, M. In *Rev. Comp. Chem.;* D. B. Boyd, K. B. *Lipkowitz, Eds.; Wiley-VHS,* **2001**; Vol. *17*; pp. 1-60.

[45] Verlinde, C.L.M.J.; Hol, W.G.J. *Structure* **1994**, *2*, 577-587.

[46] Gane, P.J.; Dean, P.M. *Curr. Op. Struct. Biol.* **2000**, *10*, 401-404.

[47] Bamborough, P.; Cohen, F.E. *Curr. Opin. Struct. Biol*. **1996**, *6*, 236-241.

[48] Charifson, P.S.; Walters, W.P. J. *Comp. Aided Mol. Des.* **2002**, *16*, 311-323.

[49] Bohm, H.J.; Stahl, M. *Rev. Comput. Chem*. **2002**, *18*, 41-87.

[50] Gohlke, H.; Klebe, G. *Angew. Chem. Int. Ed*. **2002**, *41*, 2644-2676.

[51] Lazaridis, T. *Curr. Org. Chem.* **2002**, *6*, 1319-1332.

[52] Wang, W.; Donini, O.; Reyes, C.M.; Kollman, P.A. *Annu. Rev. Biophys. Biomol. Struct*. **2001**, *30*, 211-243.

[53] Pattabiraman, N. *Curr. Med. Chem.* **2002**, *9*, 609-621.

[54] Ladbury, J.E. *Chem. Biol.* **1996**, *3*, 973-980.
[55] Connelly, P.R. In *Structure-based drug design. Thermodynamics modeling and strategy*; J. E. Ladbury, P. R. Connelly, Eds.; Springer-Verlag and Landes Bioscience: Berlin, **1997**; pp. 143-159.
[56] Reinhardt, W.P.; Miller, M.A.; Amon, L.M. *Acc. Chem. Res.* **2001**, *34*, 607-614.
[57] Renzoni, D.A.; Zvelebil, M.J.J.M.; Lundback, T.; Ladbury, J.E. In *Structure-based drug design. Thermodynamics modeling and strategy*; J. E. Ladbury, P. R. Connelly, Eds.; Springer-Verlag and Landes Bioscience: Berlin, **1997**; pp. 162-180.
[58] Westhof, E. *Water and Biological Macromolecules;* CRC Press, Boca Raton, **1993**.
[59] Aldhous, P. *Nature* **2003**, *422*, 251.
[60] Rupley, J.A.; Careri, G. *Adv. Protein Chem.* **1991**, *41*, 37-172.
[61] Gregory, R.B. *Protein- solvent interactions*; Marcel Dekker, Inc. New York, **1995**.
[62] Barron, L.D.; Hecht, L.; Wilson, G. *Biochemistry* **1997**, *36*, 13143-13147.
[63] Karplus, P.A.; Faerman, C. *Curr. Opin. Struct. Biol.* **1994**, *4*, 770-776.
[64] Israelachvili, J.; Wennerstrom, H. *Nature* **1996**, *379*, 219-225.
[65] Levitt, M.; Park, B.H. *Structure* **1993**, *1*, 223-226.
[66] Carugo, O.; Bordo, D. *Acta Crystal. D* **1999**, *55 (Pt. 2)*, 479-483.
[67] Timasheff, S.N. *Biochemistry* **2002**, *41*, 13473-13482.
[68] Eisenberg, D.; McLachlan, A.D. *Nature* **1986**, *319*, 199-203.
[69] Vedani, A.; Huntha, D. *J. Am. Chem. Soc.* **1991**, *113*, 5860-5862.
[70] Mezei, M.; Beveridge, D.L. *Meth. Enzymol.* **1986**, *127*, 21-47.
[71] Lounnas, V.; Pettitt, B.M.; Phillips, G.N. Jr. *Biophys. J.* **1994**, *66*, 601-614.
[72] Pettitt, B.M.; Makarov, V.A.; Andrews, B.K. *Curr. Opin. Struct. Biol.* **1998**, *8*, 218-221.
[73] Makarov, V.; Pettitt, B.M.; Feig, M. *Acc. Chem. Res.* **2002**, *35*, 376-384.
[74] Otting, G.; Liepinsh, E.; Wuthrich, K. *Science* **1991**, *254*, 974-80.
[75] Makarov, V.A.; Andrews, B.K.; Smith, P.E.; Pettitt, B.M. *Biophys. J.* **2000**, *79*, 2966-2974.
[76] Likic, V.A.; Prendergast, F.G. *Proteins* **2001**, *43*, 65-72.
[77] Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 7557-7561.
[78] Knapp, E.W.; Muegge, I. *J. Phys. Chem.* **1993**, *97*, 11339-11343.
[79] Muegge, I.; Knapp, E.W. *J. Phys. Chem.* **1995**, *99*, 1371-1374.
[80] Brunne, R.M.; Liepinsh, E.; Otting, G.; Wuthrich, K.; van Gunsteren, W.F. *J. Mol. Biol.* **1993**, *231*, 1040-1048.
[81] Ahlstrom, P.; Teleman, O.; Jonsson, B. *J. Am. Chem. Soc.* **1988**, *110*, 4198-4203.
[82] Henchman, R.H.; McCammon, J.A. *Protein Sci.* **2002**, *11*, 2080-2090.
[83] Wang, H.; Ben-Naim, A. *J. Med. Chem.* **1996**, *39*, 1531-1539.
[84] Poornima, C.S.; Dean, P.M. *J. Comput. Aided Mol. Des.* **1995**, *9*, 500-512.
[85] Poornima, C.S.; Dean, P.M. *J. Comput. Aided Mol. Des.* **1995**, *9*, 513-520.
[86] Poornima, C.S.; Dean, P.M. *J. Comput. Aided Mol. Des.* **1995**, *9*, 521-531.
[87] 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*, 1089-1093.
[88] Covell, D.G.; Wallqvist, A. *J. Mol. Biol.* **1997**, *269*, 281-297.
[89] Tame, J.R.; Sleigh, S.H.; Wilkinson, A.J.; Ladbury, J.E. *Nature Struct. Biol.* **1996**, *3*, 998-1001.
[90] 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*, 321-329.
[91] Lemieux, R. *Acc. Chem. Res*. **1996**, *29*, 373-380.
[92] Dunitz, J.D. *Science* **1994**, *264*, 670.
[93] Lam, P.Y.S.; Jadhav, P.K.; Eyermann, C.J.; Hodge, C.N.; Ru, Y.; Bacheler, 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.; Ericksonviitanen, S. *Science* **1994**, *263*, 380-384.
[94] Connelly, P.R.; Aldape, R.A.; Bruzzese, F.J.; Chambers, S.P.; Fitzgibbon, M.J.; Fleming, M.A.; Itoh, S.; Livingston, D.J.; Navia, M.A.; Thomson, J.A.; Wilson, K.P. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1964-1968.
[95] Watson, K.A.; Mitchell, E.P.; Johnson, L.N.; Son, J.C.; Bichard, C.J.F.; Orchard, M.G.; Fleet, G.W.J.; Oikonomakos, N.G.; Leonidas, D.D.; Kontou, M.; Papageorgioui, A. *Biochemistry* **1994**, *33*, 5745-5758.
[96] Ringe, D. *Curr. Opin. Struct. Biol.* **1995**, *5*, 825-829.
[97] Goodford, P.J. *J. Med. Chem.* **1985**, *28*, 849-857.
[98] Pastor, M.; Cruciani, G.; Watson, K.A. *J. Med. Chem.* **1997**, *40*, 4089-4102.
[99] Kastenholz, M.A.; Pastor, M.; Cruciani, G.; Haaksma, E.E.; Fox, T. *J. Med. Chem.* **2000**, *43*, 3033-3044.
[100] Wade, R.C.; Goodford, P.J. *J. Med. Chem.* **1993**, *36*, 148-156.
[101] Wade, R.C.; Clark, K.J.; Goodford, P.J. *J. Med. Chem*. **1993**, *36*, 140-147.
[102] Boobbyer, D.N.; Goodford, P.J.; McWhinnie, P.M.; Wade, R.C. *J. Med. Chem.* **1989**, *32*, 1083-1094.
[103] Rarey, M.; Kramer, B.; Lengauer, T. *Proteins* **1999**, *34*, 17-28.
[104] Pitt, W.R.; Murrayrust, J.; Godfellow, J.M. *J. Comput. Chem*. **1993,** *14*, 1007-1018.
[105] Roe, S.M.; Teeter, M.M. *J. Mol. Biol.* **1993**, *229*, 419-427.
[106] Raymer, M.L.; Sanschagrin, P.C.; Punch, W.F.; Venkataraman, S.; Goodman, E.D.; Kuhn, L.A. *J. Mol. Biol.* **1997**, *265*, 445-464.
[107] Kollman, P.A. *Chem. Rev.* **1993**, *93*, 2395-2417.
[108] Dunitz, J.D. *Chem. Biol.* **1995**, *2,* 709-712.
[109] Shoichet, B.K.; Leach, A.R.; Kuntz, I.D. *Proteins* **1999**, *34*, 4-16.
[110] Chothia, C.; Janin, J. *Nature* **1975**, *256*, 705-708.
[111] Lee, B.; Richards, F.M. J. *Mol. Biol.* **1971**, *55*, 379-400.
[112] Chothia, C. *Nature* **1974**, *248*, 338-339.
[113] Tanford, C. *Thermochemical Kinetics: Methods for Estimation of thermochemical data and rate parameters*; 2nd ed.; John Wiley & Sons, Inc., New York, **1980**.
[114] Sharp, K.A.; Nicholls, A.; Fine, R.F.; Honig, B. *Science* **1991**, *252*, 106-9.
[115] Hine, J.; Mookerjee, P.K. *J. Phys. Rev.* **1975**, *40*, 292-298.
[116] Leo, A.; Jow, P.Y.; Silipo, C.; Hansch, C. *J. Med. Chem*. **1975**, *18*, 865-868.
[117] Wang, J.; Wang, W.; Huo, S.; Lee, B.; Kollman, P. *J. Phys. Chem. B* **2001**, *105*, 5055-5067.
[118] Tawa, G.J.; Topol, I.A.; Burt, S.K.; Erickson, J.W. *J. Am. Chem. Soc.* **1998**, *120*, 8856-8863.
[119] Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. *J. Mol. Biol.* **1982**, *161*, 269-288.
[120] Miller, M.D.; Kearsley, S.K.; Underwood, D.J.; Sheridan, R.P. *J. Comput. Aided Mol. Des.* **1994**, *8*, 153-174.
[121] Toney, J.H.; Fitzgerald, P.M.; Grover-Sharma, N.; Olson, S.H.; May, W.J.; Sundelof, J.G.; Vanderwall, D.E.; Cleary, K.A.; Grant, S.K.; Wu, J.K.; Kozarich, J.W.; Pompliano, D.L.; Hammond, G.G. *Chem. Biol.* **1998**, *5*, 185-196.
[122] Miller, M.D.; Sheridan, R.P.; Kearsley, S.K.; Underwood, D.J. *Meth. Enzymol.* **1994**, *241*, 354-370.
[123] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; R., H.; Hart, W.E.; Belew, R.K.; Olson, A.J. *J. Comput. Chem.* **1998**, *19*, 1639-1662.
[124] Osterberg, F.; Morris, G.M.; Sanner, M.F.; Olson, A.J.; Goodsell, D.S. *Proteins* **2002**, *46*, 34-40.
[125] Dill, K.A. *J. Biol. Chem.* **1997**, *272*, 701-704.
[126] Pearlman, D.A.; Case, D.A.; Caldwell, J.W.; Ross, W.S.; Cheatham III, T.E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P.A. *Comput. Phys. Comm.* **1995**, *91*, 1-41.
[127] Cornell, W.E.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.S.; Fox, T.; Caldwell, J.W.; Kollman, P.A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
[128] Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *2,* 187-217.
[129] Koheler, J.E.H.; Saenger, W.; van Gunsteren, W.F. *Eur. J. Biophys.* **1987**, *15*, 197-210.
[130] Kamsinski, G.; Jorgensen, W.L. *J. Phys. Chem.* **1996**, *100*, 18010-18013.
[131] Dinur, U.; Hagler, A.T. In *Rev. Comput. Chem.* K. B. Lipkowitz, D. B. Boyd, Eds.; VCH Publishers,: New York, **1991**; Vol. *2*; pp. 99-164.
[132] SYBYL; 6.5 ed.; Tripos Inc.: St. Louis, MO, **1995**.
[133] Halgren, T.A. *J. Comput. Chem*. **1996**, *17*, 490-519.
[134] Allinger, N.L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.
[135] Allinger, N.L.; Yuh, Y.H.; J.-H., L. *J. Am. Chem. Soc.* **1989**, *111*, 8551.
[136] Bowen, J.P.; Allinger, N.L. In *Rev. Comput. Chem*. K. B. Lipkowitz, D. B. Boyd, Eds.; VCH Publishers: New York, **1991**; Vol. *2;* pp. 81-97.
[137] Reddy, M.R.; Erion, M.D.; Agarwal, A. *Rev. Comput. Chem.* **2000**, *16*, 217-304.

[138] Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144-1149.
[139] Cramer, C.J.; Truhlar, D.G. *Chem. Rev*. **1999**, *99*, 2161-2200.
[140] Gilson, M.K.; Honig, B. *Proteins* **1988**, *4*, 7-18.
[141] Cramer, C.J.; Truhlar, D.G. *Science* **1992**, *256*, 213-217.
[142] Lazaridis, T.; Karplus, M. *J. Mol. Biol.* **1999**, *288*, 477-487.
[143] Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.
[144] Grant, J.A.; Pichup, B.T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608-640.
[145] Lamm, E. *Rev. Comput. Chem.* **2003**, *19*, 147-365.
[146] Jorgensen, W.L.; Chandrasenkhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. *J. Chem. Phys.* **1983**, *79*, 926-935.
[147] Berendsen, H.J.C.; Postma, J.P., M.; Van Gunsteren, W.F.; Hermans, J. In *Intermolecular Forces*; B. Pullman, Ed.; Reidel: Dordrecht, Holland, **1981**; pp. 331-342.
[148] Berendsen, H.J.C.; Grigera, J.R.; Straatsma, P. *J. Chem. Phys.* **1987**, *91*, 6269-6271.
[149] Mahoney, M.W.; Jorgensen, W.L. *J. Chem. Phys.* **2000**, *112*, 8910-8922.
[150] Mitchell, M.J.; McCammon, J.A. *J. Comput. Chem.* **1991**, *12*, 271-275.
[151] Pearlman, D.A.; Kollman, P.A. *J. Chem. Phys.* **1991**, *94*, 4532-4545.
[152] Chipot, M.; Kollman, P.A.; Pearlman, D.A. *J. Comput. Chem.* **1996**, *17*, 1112-1131.
[153] Beveridge, D.L.; DiCapua, F.M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431-492.
[154] Jorgensen, W.L. *Acc. Chem. Res.* **1989**, *22*, 184-189.
[155] Aqvist, J.; Luzhkov, V.B.; Brandsal, B.O. *Acc. Chem. Res.* **2002**, *35*, 358-365.
[156] Ajay; Murcko, M.A. *J. Med. Chem*. **1995**, *38*, 4953-4967.
[157] Tame, J.R. *J. Comput. Aided Mol. Des.* **1999**, *13*, 99-108.
[158] Straatsma, T.P. In *Rev. Comput. Chem*. K. B. Lipkowitz, D. B. Boyd, Eds.; VCH Publishers: New York, **1996**; Vol. *9*; pp. 81-127.
[159] van Gunsteren, W.F. *Angew. Chem. Int. Ed.* **1990**, *29*, 992-1023.
[160] Aqvist, J.; Medina, C.; Samuelsson, J.-E. *Protein Eng.* **1994**, *7*, 385-391.
[161] Wong, C.F.; McCammon, J.A. *J. Am. Chem. Soc.* **1986**, *108*, 3830-3832.
[162] Bash, P.A.; Singh, U.C.; Langridge, R.; Kollman, P.A. *Science* **1987**, *236*, 3830-3832.
[163] Merz, K.M.; Kollman, P.A. *J. Am. Chem. Soc.* **1989**, *111*, 5649-5658.
[164] Morgan, B.P.; Scholtz, J.M.; Ballinger, M.D.; Zipkin, I.D.; Bartlett, P.A. *J. Am. Chem. Soc*. **1991**, *113*, 297-307.
[165] Graffner-Nordberg, M.; Marelius, J.; Ohlsson, S.; Persson, A.; Swedberg, G.; Andersson, P.; Andersson, S.E.; Aqvist, J.; Hallberg, A. *J. Med. Chem.* **2000**, *43*, 3852-3861.
[166] Ota, N.; Stroupe, C.; Ferreira-da-Silva, J.M.; Shah, S.A.; Mares-Guia, M.; Brunger, A.T. *Proteins* **1999**, *37*, 641-53.
[167] Hansson, T.; Aqvist, J. *Protein Eng.* **1995**, *8*, 1137-1144.
[168] Aqvist, J. *J. Comput. Chem.* **1996**, *17*, 1587-1597.
[169] King, G.; Warshel, A. *J. Chem. Phys*. **1989**, *91*, 3647-3661.
[170] Hansson, T.; Marelius, J.; Aqvist, J. *J. Comput. Aided Mol. Des.* **1998**, 12, *27*-35.
[171] Marelius, J.; Graffner-Nordberg, M.; Hansson, T.; Hallberg, A.; Aqvist, J. *J. Comput. Aided Mol. Des.* **1998**, *12*, 119-131.
[172] Ljungberg, K.B.; Marelius, J.; Musil, D.; Svensson, P.; Norden, B.; Aqvist, J. *Eur. J. Pharm. Sci.* **2001**, *12*, 441-446.
[173] Wang, W.; Wang, J.; Kollman, P.A. *Proteins* **1999**, *34*, 395-402.
[174] Srinivasan, J.; Cheatham, T.E.; Cieplak, P.; Kollman, P.A.; Case, D.A. *J. Am. Chem. Soc.* **1998**, *120*, 9401-9409.
[175] Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham, T.E., *3rd Acc. Chem. Res.* **2000**, *33*, 889-897.
[176] Smith, K.C.; Honig, B. *Proteins* **1994**, *18*, 119-132.
[177] Kuhn, B.; Kollman, P.A. *J. Med. Chem*. **2000**, *43*, 3786-3791.
[178] Wang, W.; Kollman, P.A. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14937-14942.
[179] Wang, W.; Lim, W.A.; Jakalian, A.; Wang, J.; Wang, J.; Luo, R.; Bayly, C.I.; Kollman, P.A. *J. Am. Chem. Soc.* **2001**, *123*, 3986-3994.
[180] Sharp, K.A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301-332.
[181] Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem. B* **1993**, *97*, 1101-1109.
[182] Sippl, M.J. *Curr. Opin. Struct. Biol.* **1995**, *5*, 229-235.
[183] Jernigan, R.L.; Bahar, I. *Curr. Op. Struct. Biol.* **1996**, *6*, 195-209.
[184] Thomas, P.D.; Dill, K.A. *J. Mol. Biol.* **1996**, *257*, 457-469.
[185] Moult, J. *Curr. Op. Struct. Biol.* **1997**, *7*, 194-199.
[186] Verkhivker, G.; Appelt, K.; Freer, S.T.; Villafranca, J.E. *Protein Eng.* **1995**, *8,* 677-691.
[187] Sharp, K.A.; Nicholls, A.; Friedman, R.; Honig, B. *Biochemistry* **1991**, *30*, 9686-9697.
[188] Muegge, I.; Martin, Y.C. *J. Med. Chem*. **1999**, *42*, 791-804.
[189] Mitchell, J.B.O.; Laskowski, R.A.; Alex, A.; Forster, M.J.; Thornton, J.M. *J. Comput. Chem.* **1999**, *20*, 1177-1185.
[190] Mitchell, J.B.O.; Laskowski, R.A.; Alex, A.; Thornton, J.M. *J. Comput. Chem*. **1999**, *20*, 1165-1176.
[191] Goodfellow, J.M.; Pitt, W.R.; Smart, O.S.; Williams, M.A. *Comput. Phys. Comm.* **1995**, *91*, 321-329.
[192] Muegge, I. *J. Comput. Chem.* **2001**, *22*, 418-425.
[193] Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol.* **2000**, *295*, 337-356.
[194] Bohm, H.J. *J. Comput. Aided Mol. Des.* **1994**, *8*, 623-632.
[195] Gohlke, H.; Hendlich, M.; Klebe, G. *Perspectives Drug Discov. Des.* **2000**, *20*, 115-144.
[196] Sotriffer, C.A.; Gohlke, H.; Klebe, G. *J. Med. Chem.* **2002**, *45*, 1967-1970.
[197] Gohlke, H.; Klebe, G. *J. Med. Chem*. **2002**, *45*, 4153-4170.
[198] Wallqvist, A.; Jernigan, R.L.; Covell, D.G. *Protein Sci*. **1995**, *4*, 1881-1903.
[199] Privalov, P.L.; Makhatadze, G.I. *J. Mol. Biol.* **1990**, *213*, 385-391.
[200] Makhatadze, G.I.; Privalov, P.L. *J. Mol. Biol.* **1990**, *213*, 375-384.
[201] Privalov, P.L.; Makhatadze, G.I. J. *Mol. Biol*. **1992**, *224*, 715-723.
[202] DeWitte, R.S.; Shakhnovich, E.I. *J. Am. Chem. Soc.* **1996**, *118*, 11733-11744.
[203] Bohm, H.J.; Stahl, M. *Med. Chem. Res.* **1999**, *9*, 445-462.
[204] Bohm, H.J. *J. Comput. Aided Mol. Des.* **1998**, *12*, 309-323.
[205] Head, R.D.; Smythe, M.L.; Oprea, T.I.; Waller, C.L.; Green, S.M.; Marshall, G.R. *J. Med. Chem.* **1996**, *118*, 3959-3969.
[206] Eldridge, M.D.; Murray, C.W.; Auton, T.R.; Paolini, G.V.; Mee, R.P. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425-445.
[207] Murray, C.W.; Auton, T.R.; Eldridge, M.D. *J. Comput. Aided Mol. Des.* **1998**, *12*, 503-519.
[208] Jain, A.N. *J. Comput. Aided Mol. Des.* **1996**, *10*, 427-440.
[209] Wang, R.; Liu, L.; Tang, Y. *J. Mol. Model.* **1998**, *4*, 379-394.
[210] Takamatsu, Y.; Itai, A. *Proteins* **1998**, *33*, 62-73.
[211] Venkatarangan, P.; Hopfinger, A.J. *J. Med. Chem*. **1999**, *42*, 2169-2179.
[212] Rognan, D.; Lauemoller, S.L.; Holm, A.; Buus, S.; Tschinke, V. *J. Med. Chem*. **1999**, *42*, 4650-4658.
[213] Bohacek, R.S.; McMartin, C. *J. Am. Chem. Soc.* **1994**, *116*, 5560-5571.
[214] Kasper, P.; Christen, P.; Gehring, H. *Proteins* **2000**, *40*, 185-92.
[215] Hansch, C.; Leo, A.J. *Substituent constants for correlation analysis in chemistry and biology*; John Wiley and Sons: New York, **1979**.
[216] Kellogg, G.E.; Semus, S.F.; Abraham, D.J. *J. Comput. Aided Mol. Des.* **1991**, *5*, 545-552.
[217] Kellogg, G.E.; Abraham, D.J. *Eur. J. Med. Chem.* **2000**, *35*, 651-661.
[218] Kellogg, G.E.; Burnett, J.C.; Abraham, D.J. *J. Comput. Aided Mol. Des.* **2001**, *15*, 381-393.
[219] Kellogg, G.E.; Joshi, G.S.; Abraham, D.J. *Med. Chem. Res.* **1992**, *1*, 444-453.
[220] Wireko, F.C.; Kellogg, G.E.; Abraham, D.J. *J. Med. Chem.* **1991**, *34*, 758-767.
[221] Kellogg, G.E.; Abraham, D.J. J. *Mol. Graph.* **1992**, *10*, 212-217.
[222] Mannhold, R.; van de Waterbeemd, H. *J. Compu. Aided Mol. Des.* **2001**, *15*, 337-354.
[223] Mannhold, R.; Rekker, R.F. Perspectives *Drug Discov. Des.* **2000**, *18,* 1-18.
[224] Abraham, D.J.; Leo, A.J. *Proteins* **1987**, *2*, 130-152.
[225] Eros, D.; Örfi, K.L.; Takács-Novák, K.; Acsády, G.; Kéri, G. *Curr. Med. Chem.* **2002**, *9*, 1819-1829.
[226] Hansch, C.; Leo, A. *Exploring QSAR. Fundamentals and applications in chemistry and biology*; American Chemical Society: Washington, **1995**.
[227] Burnett, J.C.; Kellogg, G.E.; Abraham, D.J. *Biochemistry* **2000**, *39*, 1622-1633.
[228] Burnett, J.C.; Botti, P.; Abraham, D.J.; Kellogg, G.E. *Proteins* **2001**, *42*, 355-377.

[229] Kellogg, G.E.; Scarsdale, J.N.; Fornari, F.A. Jr. *Nucleic Acids Res.* **1998**, *26*, 4721-4732.
[230] Wright, C.S.; Kellogg, G.E. *Protein Sci.* **1996**, *5*, 1466-1476.
[231] Cashman, D.J.; Rife, J.P.; Kellogg, G.E. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 119-122.
[232] Gussio, R.; Pattabiraman, N.; Zaharevitz, D.W.; Kellogg, G.E.; Topol, I.A.; Rice, W.G.; Schaeffer, C.A.; Erickson, J.W.; Burt, S.K. *J. Med. Chem.* **1996**, *39*, 1645-1650.
[233] Gussio, R.; Pattabiraman, N.; Kellogg, G.E.; Zaharevitz, D.W. *Methods* **1998**, *14*, 255-263.
[234] Gussio, R.; Zaharevitz, D.W.; McGrath, C.F.; Pattabiraman, N.; Kellogg, G.E.; Schultz, C.; Link, A.; Kunick, C.; Leost, M.; Meijer, L.; Sausville, E.A. *Anticancer Drug Des.* ***2000****, 15, 53-66.*
[235] Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D.J.; Kellogg, G.E.; Mozzarelli, A. *J. Med. Chem.* **2002**, *45*, 2469-2483.
[236] Kellogg, G.; Fornabaio, M.; Spyrakis, F.; Lodola, A.; Cozzini, P.; Mozzarelli, A.; Abraham, D.J. *J. Mol. Graph. Model.* **2004**, in press.
[237] Abraham, D.J.; Kellogg, G.E.; Holt, J.M.; Ackers, G.K. *J. Mol. Biol.* **1997**, *272*, 613-632.
[238] Fornabaio, M.; Cozzini, P.; Mozzarelli, A.; Abraham, D.J.; Kellogg, G.E. *J. Med. Chem.* **2003**, submitted.
[239] Davies, D.R. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 189-215.
[240] Navia, M.A.; Fitzgerald, P.M.; McKeever, B.M.; Leu, C.T.; Heimbach, J.C.; Herber, W.K.; Sigal, I.S.; Darke, P.L.; Springer, J.P. *Nature* **1989**, *337*, 615-20.
[241] Abdel-Rahman, H.M.; Al-karamany, G.S.; El-Koussi, N.A.; Youssef, A.F.; Kiso, Y. *Curr. Med. Chem.* **2002**, *9*, 1905-1922.
[242] Venkatesan, N.; Kim, B.H. *Curr. Med. Chem.* **2003**, *9*, 2243-2270.
[243] Wlodawer, A.; Vondrasek, J. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249-284.
[244] Priestle, J.P.; Fassler, A.; Rosel, J.; Tintelnot-Blomley, M.; Strop, P.; Grutter, M.G. *Structure* **1995**, *3*, 381-389.
[245] Pawson, T.; Nash, P. *Science* **2003**, *300*, 445-52.
[246] Royer, W.E., Jr.; Pardanani, A.; Gibson, Q.H.; Peterson, E.S.; Friedman, J.M. *Proc Natl Acad Sci* USA **1996**, *93*, 14526-31.
[247] Hosur, M.V.; Bhat, T.N.; Kempf, D.J.; Baldwin, E.T.; Liu, B.; Gulnik, S.; Wideburg, N.E.; Norbeck, D.W.; Appelt, K.; Erickson, J.W. *J. Am. Chem. Soc.* **1994**, *116*, 847-855.
[248] Dreyer, G.B.; Lambert, D.M.; Meek, T.D.; Carr, T.J.; Tomaszek, T.A.; Fernandez, A.V.; Bartus, H.; Cacciavillani, E.; Hassell, A.M.; Minnich, M.; Petteway, S.R.; Metcalf, B.W. *Biochemistry* **1992**, *31*, 6646-6659.
[249] Kim, E.E.; Baker, C.T.; Dwyer, M.D.; Murcko, M.A.; Rao, B.G.; Tung, R.D.; Navia, M.A. *J. Am. Chem. Soc.* **1995**, *117*, 1181-1182.
[250] Swain, A.L.; Miller, M.M.; Green, J.; Rich, D.H.; Schneider, J.; Kent, S.B.H.; Wlodawer, A. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8805-8809.
[251] Jhoti, H.; Singh, O.M.P.; Weir, M.P.; Cooke, R.; Murray-Rust, P.; Wonacott, A. *Biochemistry* **1994**, *33*, 8417-8427.
[252] Abdel-Meguid, S.S.; Metcalf, B.W.; Carr, T.J.; Demarsh, P.; DesJarlais, R.L.; Fisher, S.; Green, D.W.; Ivanoff, L.; Lambert, D.M.; Murthy, K.H.M.; Petteway, S.R.; Pitts, W.J.; Tomaszek, T.A.; Winborne, E.; Zhao, B.; Dreyer, G.B.; Meek, T.D. *Biochemistry* **1994**, *33*, 11671-11677.
[253] Bone, R.; Vacca, J.P.; Anderson, P.S.; Holloway, M.K. *J. Am. Chem. Soc.* **1991**, *113*, 9382-9384.
[254] Thompson, S.K.; Murthy, K.H.M.; B., Z.; Winborne, E.; Green, D.W.; Fisher, S.M.; DesJarlais, R.L.; Tomaszek, T.A.; Meek, T.D.; Gleason, J.G.; Abdel-Meguid, S.S. *J. Med. Chem.* **1994**, *37*, 3100-3107.
[255] Kempf, D.J.; Marsh, K.C.; Denissen, J.F.; McDonald, E.; Vasavanonda, S.; Flentge, C.A.; Green, B.E.; Fino, L.; Park, C.H.; Kong, X.P.; Wideburg, N.E.; Saldivar, A.; Ruiz, L.; Kati, W.M.; Sham, H.L.; Robins, T.; Stewart, K.D.; Hsu, A.; Plattner, J.J.; Leonard, J.M.; Norbeck, D.W. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2484-2488.
[256] Krohn, A.; Redshaw, S.; Ritchie, J.C.; Graves, B.J.; Hatada, M.H. *J. Med. Chem.* **1991**, *34*, 3340-3342.
[257] Thanki, N.; Rao, J.K.M.; Foundling, S.I.; Howe, W.J.; Moon, J.B.; Hui, J.O.; Tomasselli, A.G.; Heinrikson, R.L.; Thaisrivongs, S.; Wlodawer, A. *Protein Sci.* **1992**, *1*, 1061-1072.
[258] Bäckbro, K.; Löwgren, S.; Österlund, K.; Atepo, J.; Unge, T. *J. Med. Chem.* **1997**, *40*, 898-902.
[259] Jadhav, P.K.; Ala, P.; Woerner, F.J.; Chang, C.H.; Garber, S.S.; Anton, E.D.; Bacheler, L.T. *J. Med. Chem.* **1997**, *40*, 181-191.
[260] Hodge, C.N.; Aldrich, P.E.; Bacheler, L.T.; Chang, C.H.; Eyermann, C.J.; Garber, S.; Grubb, M.; Jackson, D.A.; Jadhav, P.K.; Korant, B.; lam, P.Y.S.; Maurin, M.B.; Meek, J.L.; Otto, M.J.; Rayner, M.M.; Reid, C.; Sharpem, T.S.; L., S.; Winslowm, D.L.; Erickson-Viitanen, S. *Chem. Biol.* **1996**, *3*, 301-314.
[261] Schaal, W.; Karlsson, A.; Ahlsén, G.; Lindberg, J.; Andersson, H.O.; Danielson, U.H.; Classon, B.; Unge, T.; Samuelsson, B.; Hultén, J.; Hallberg, A.; Karlén, A. *J. Med. Chem.* **2001**, *44*, 155-169.
[262] Hoog, S.S.; Zhao, B.; Winborne, E.; Fisher, S.; Green, D.W.; DesJarlais, R.L.; Newlander, K.A.; Callahan, J.F.; Moore, M.L.; Huffman, W.F.; Abdel-Meguid, S.S. *J. Med. Chem.* **1995**, *38*, 3246-3252.
[263] Sachpatzidis, A.; Dealwis, C.; Lubetsky, J.B.; Liang, P.H.; Anderson, K.S.; Lolis, E. *Biochemistry* **1999**, *38*, 12665-12674.