Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 22 (2004) 479–486
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Getting it right: modeling of pH, solvent and "nearly" everything else in virtual screening of biological targets

Glen E. Kellogg [a,*], Micaela Fornabaio [b,c], Francesca Spyrakis [b], Alessio Lodola [b], Pietro Cozzini [d], Andrea Mozzarelli [b,c,*], Donald J. Abraham [a]

[a] *Department of Medicinal Chemistry, School of Pharmacy, Institute for Structural Biology and Drug Discovery, Virginia Commonwealth University, Richmond, VA 23298-0540, USA*
[b] *Department of Biochemistry and Molecular Biology, University of Parma, 43100 Parma, Italy*
[c] *National Institute for the Physics of Matter, University of Parma, 43100 Parma, Italy*
[d] *Molecular Modeling Laboratory, Department of General and Inorganic Chemistry, University of Parma, 43100 Parma, Italy*

Accepted 4 March 2004

Available online 27 April 2004

Abstract

"Getting it right" refers to the careful modeling of all elements in the living system, i.e. biological macromolecules, ligands and water molecules. In addition, careful attention should be paid to the protonation state of ionizable functional groups on the ligands and residues at the active site. Computational technology based on the empirical HINT program is described to: (1) calculate free energy scores for ligand binding; (2) include the implicit and explicit effects of water in and around the ligand binding site; and (3) incorporate the effects of global and local pH in molecular models. This last point argues for the simultaneous consideration of a number of molecular models, each with different protonation profiles. Data from recent studies of protein–ligand systems (trypsin, thrombin, neuraminidase, HIV-1 protease and others) are used to illustrate the concepts in the paper. Also discussed are experimental factors related to accurate free energy predictions with this and other computational technologies.

Keywords: HINT; Hydropathy; Hydrophobic interactions; Free energy

1. Introduction

It is a critical goal of computational chemistry/biology to effectively predict free energy of binding for protein–ligand associations; in fact this goal was described as the "Holy Grail" by Pearlman and Kollman a number of years ago [1], and more recently by Gohlke and Klebe [2]. Molecular modeling of the biological milieu is a difficult task because there is just *so* much going on. In addition to modeling the protein, ligand(s) and cofactors, which is challenging enough, there is the additional "little" problem of solvent and its myriad of effects. Some of the solvent effects, however, can not be directly modeled, even with state-of-the-art in computational tools. For example, the hydrophobic effect is a consequence of the biomolecular ensemble exposing its more polar bits to the (water) solvent, where they can make hydrogen bonds. As this occurs, the hydrophobic portions of the ensemble are forced together, which gives the appearance of a "hydrophobic force" [3]. This, in turn, releases previously coordinated water to bulk, thus increasing the entropy of the ensemble. While no extant quantum mechanics calculation or "first principles" molecular mechanics force field can yet model the hydrophobic effect and/or its entropic consequences, complex and time-consuming molecular dynamics simulations with explicit water have been shown to correlate with free energy in a number of laboratories [4–7].

Our approach has been to apply an empirical "natural" model to understanding molecular associations in the biological environment. The interaction constants for the model are obtained from measurements of the partition coefficient for 1-octanol/water solubility, $P_{o/w}$, or as more commonly used, $\log P_{o/w}$. The partition coefficient is a thermodynamic quantity directly related to free energy for the transfer of a solute between the two solvents [8]. In our model we postulate that water and 1-octanol are representative of polar and

* Corresponding authors. Tel.: +1-804-828-6452/+39-0521-905138; fax: +1-804-827-3664/+39-0521-905151.
E-mail addresses: glen.kellogg@vcu.edu (G.E. Kellogg), biochim@unipr.it (A. Mozzarelli).

1093-3263/$ – see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2004.03.008

hydrophobic regions, respectively, of biomacromolecules [9]. Thus, the partition coefficient data for the two solvents should be applicable for general interactions in the biological environment. We call this model HINT for Hydropathic INTeractions [10]. Simply, we consider an interaction between two atoms to be the product of their hydrophobic atom constants (a_i, partial $\log P_{o/w}$) and solvent accessible surface areas (S_i) as follows:

$$b_{ij} = a_i S_i a_j S_j f(r_{ij}) \quad (1)$$

where $f(r_{ij})$ represents a function of the distance between the two atoms i and j [11]. The total interactions between two molecules is then the (double) sum over all atom–atom interactions, i.e.,

$$B = \sum\sum b_{ij} \quad (2)$$

Since $\log P_{o/w}$ for virtually any organic or drug-like molecule can be estimated with validated methods like CLOG-P [12], and extrapolated to atoms within proteins, and polynucleotides, the HINT model can be applied to many types of systems. We have, in fact, used HINT successfully in a number of studies, with a fairly wide swath of variety [10,13–17].

Because hydrogen bonds, Coulombic interactions, hydrophobic interactions, and solvation/desolvation all occur as part of the $\log P_{o/w}$ experiment, these effects are implicitly part of the HINT "forcefield". For example, the absolute value of fragment charge correlates well with the hydrophobic fragment constants for polar fragments [9], confirming that Coulombic effects are encoded in the hydrophobic parameters. Implicit in this is that the hard to characterize bulk effects of solvent are somehow encoded in the parameters. However, more specific solvent effects, such as those from highly constrained water molecules that are bridging between the interacting molecules, need to be explicitly considered.

In this report we describe how these solvation effects can be incorporated into molecular models and made part of a simple, relatively fast and intuitive free energy scoring technology based on HINT. The theme we are putting forth here is that careful consideration of a number of issues such as constrained water in the active site, the pH and ionization effects at the active site, can be combined with rapid scoring functions such as HINT to yield superior *and* intuitive binding models. This is what we mean by "getting it right".

2. Results

2.1. Intuitive free energy scores

The program HINT calculates a "score" for each atom–atom interaction in a biomolecular association as in Eqs. (1) and (2). The basis of the HINT model is that quantitatively significant data of biomolecular association are encoded in the experimental determination of hydrophobicity, particularly from the water–octanol system. Each atom–atom score is a partial δg that has a character representing the type of interaction (hydrogen bond, hydrophobic, acid–base, base–base, acid–acid or hydrophobic–polar) and a magnitude of the interaction [18]. The sum of atom–atom scores for an association represents the total strength of the interaction. We have shown that these total HINT scores can be correlated with $\Delta G_{interaction}$ for a set of 76 experimentally characterized protein–ligand complexes, structurally determined at resolution better than 3.2 Å (Fig. 1a). The corresponding equation is

$$\Delta G = -0.0019 H_{total} - 3.927 \quad (3)$$

where H_{total} is the total HINT score. This correlation exhibits $r^2 = 0.48$ and standard error of $\pm 2.4\,kcal\,mol^{-1}$. When the HINT analysis was carried out on a subset of 56 complexes structurally determined at a resolution better than 2.5 Å the correlation significantly improved ($r^2 = 0.72$, standard error $= \pm 1.8\,kcal\,mol^{-1}$) (Fig. 1b). Within a series of related biomolecule–ligand complexes the correlation between ΔG and HINT score was found to be very good (Fig. 1c). In Fig. 1c, the following equations apply:

$$\Delta G = -0.0028 H_{total} - 0.588 \quad (4)$$

$$\Delta G = -0.0023 H_{total} - 4.544 \quad (5)$$

for bovine trypsin (eight complexes, resolution better than 2.0 Å, $r^2 = 0.76$, standard error of $\pm 1.2\,kcal\,mol^{-1}$) and human and bovine thrombin (nine complexes with five at a resolution between 2.5 and 3.2 Å, $r^2 = 0.51$, standard error of $\pm 1.8\,kcal\,mol^{-1}$), respectively [13]. It is interesting that the slopes of the two lines are quite consistent and well within the regression uncertainty while there is considerably more variation in the ΔG intercept. We are currently exploring the factors contributing to both the slope and y-intercept parameters (vide infra). It would thus seem that deriving a unique relationship for each family of structurally and thermodynamically characterized complexes would provide improved accuracy and precision of ΔG predictions for complexes. Even though the data in Fig. 1c were assembled from crystallographic and solution binding data collected in multiple laboratories [13], this suggests that accuracies in ΔG predictions on the order of $\pm 1\,kcal\,mol^{-1}$ should be routinely achievable with the HINT free energy scoring methodology applied to a single system. On the other hand, the experimental uncertainties in experimental binding energy measurements, coupled with uncertainties in structural measurements, indicate that we should not really expect to predict ΔG with greater than $\pm 1\,kcal\,mol^{-1}$ accuracy (see Section 3). Lastly, Fig. 1d, an example of virtual screening [14], illustrates the correlation between calculated score and experimental free energy for 26 cyclin-dependent kinase inhibitor complexes generated by docking. The regression of ΔG versus H_{total} for these data is

$$\Delta G = -0.0029 H_{total} - 3.438 \quad (6)$$

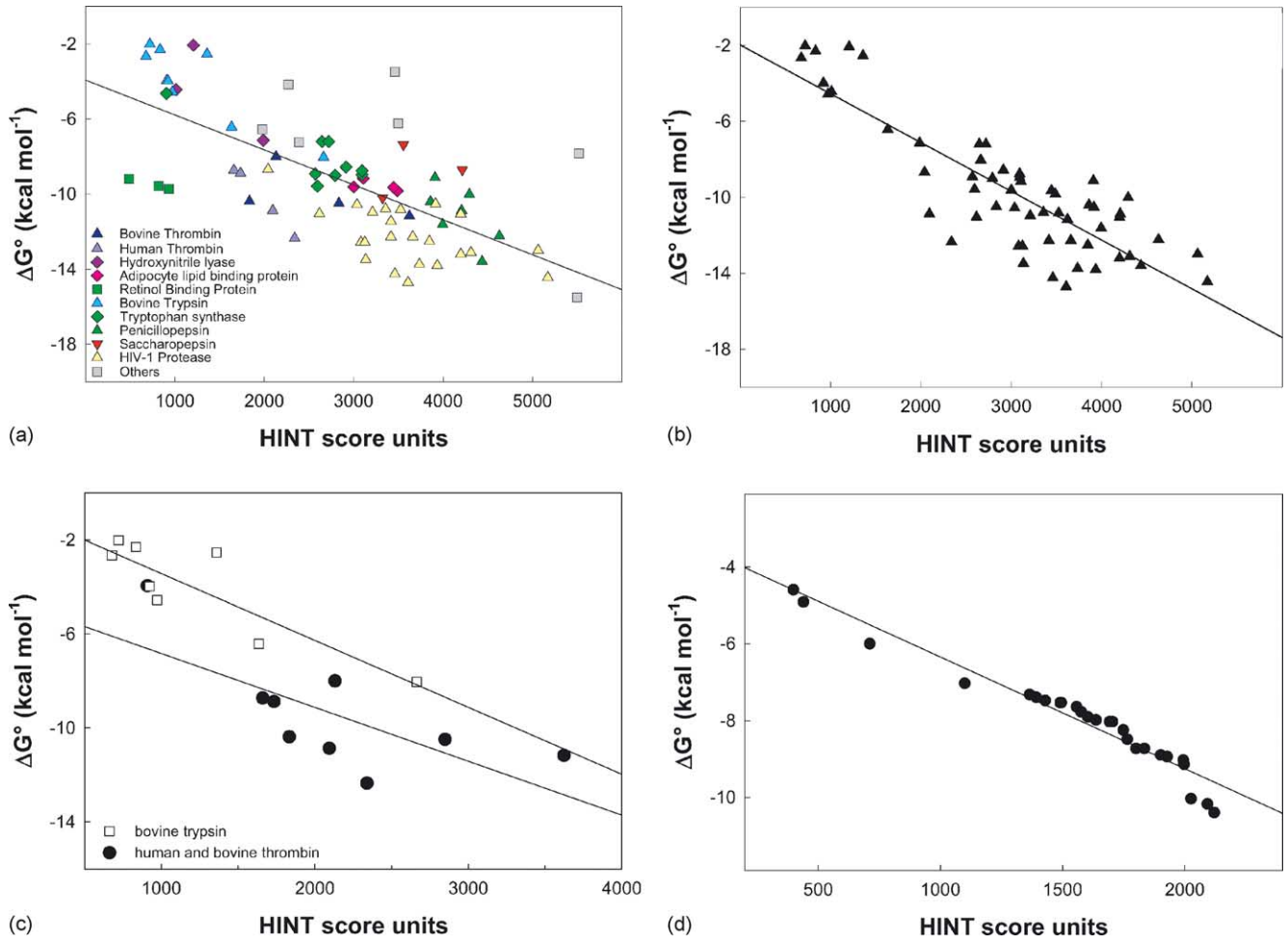

Fig. 1. (a) Correlation between calculated HINT scores and measured free energy of binding for 76 diverse protein–ligand complexes structurally determined at a resolution better than 3.2 Å; (b) correlation between calculated HINT scores and measured free energy of binding for a subset of 56 diverse protein–ligand complexes structurally determined at a resolution better than 2.5 Å; (c) free energy of binding as a function of calculated HINT score for bovine trypsin and human and bovine thrombin; (d) free energy of binding as a function of HINT score for docked models of a series of paullone cyclin-dependent kinase inhibitors [14].

where the standard error is ± 0.3 kcal mol^{-1} and r^2 is 0.94. Because all of the molecular models in this case were constructed from a single homology model structure [19], and all of the inhibition measurements were made in the same laboratory [20], the uncertainties due to these factors are constant across the series, thus yielding the deceivingly impressive statistics of Eq. (6).

2.2. Contribution from bound water

It is important to emphasize that the HINT model specifically and explicitly accounts for the hydrophobic effect, and that this is a unique capability of HINT. The hydrophobic effect is manifested in two primary ways: first, hydrophobic–hydrophobic interactions are rewarded with favorable energy scores; second, desolvation, an energy cost, is scored with (unfavorable) hydrophobic–polar interactions [18]. These latter terms are observed when polar groups of one interacting entity are forced into a hydrophobic region in the other entity that is hydrophobic, i.e., the first entity is desolvated. These are both due, indirectly, to the presence of water in the biological environment and also in the experimental conditions where $\log P_{o/w}$ was measured. Water molecules have more *direct* consequences that are also modeled by HINT. A number of water molecules are involved in bridging "water-mediated" hydrogen bonds, i.e., as below for the association of molecules A and B:

$$H_{\text{total}} = H_{\text{A–B}} + H_{\text{A–water}} + H_{\text{B–water}} \quad (7)$$

where H_{total} is the total score (free energy) for the association, $H_{\text{A–B}}$ is the score for the A–B interaction and $H_{\text{A–water}}$ and $H_{\text{B–water}}$ are the interactions between the individual components and bridging water. We have shown the energetic effect of these water molecules in native and mutant hemoglobins [15] and also in a recent study of HIV-1 protease–inhibitor complexes [21] (see Table 1). The primary water of interest, labeled 301 in the crystallographic models of HIV-protease, lies between the flaps and the lig-

Table 1
Measured and HINT predicted free energy of binding for HIV-1 protease inhibitors

PDB code for complex	ΔG_{exp} (kcal mol^{-1})	$\Delta G_{predicted}$ (kcal mol^{-1}) ligand–protein only	$\Delta G_{predicted}$ (kcal mol^{-1}) with water correction
1hbv	−8.68	−10.35	−9.58
1ajv[a]	−10.52	−12.60	−11.45
1sbg	−10.56	−11.55	−11.87
1ajx[a]	−10.79	−11.93	−10.50
1g2k[a]	−10.82	−12.13	−10.78
1hih	−10.97	−11.76	−12.08
1htf	−11.04	−11.07	−10.51
1g35[a]	−11.06	−12.94	−11.93
1aaq	−11.45	−12.00	−11.67
1hvl	−12.27	−12.00	−12.73
1hiv	−12.27	−12.30	−13.27
4phv	−12.51	−12.62	−12.82
1hpv	−12.57	−11.60	−11.82
1hps	−12.57	−11.65	−11.51
1dmp[a]	−12.99	−13.89	−13.27
7hvp	−13.11	−13.08	−14.21
1htg	−13.20	−12.97	−14.14
1hxb	−13.49	−11.67	−11.90
1hvi	−13.74	−12.38	−13.20
1hvk	−13.80	−12.63	−13.29
1hvj	−14.25	−12.06	−12.72
1qbt[a]	−14.44	−14.11	−13.58
1hxw	−14.71	−12.23	−13.39

[a] Ligands designed to displace water 301 [23,24].

and and generally forms four hydrogen bonds—two that accept from protein backbone nitrogens and two that donate into carbonyl oxygens of the ligands. This pattern is illustrated in Fig. 2 where HINT interaction maps [10] show the water to ligand interactions and the protein to ligand interactions for HIV-1 protease inhibitor CGP 53820 (pdb code 1hih) [22]. The favorable contribution of the water 301 is indicated by the strong blue contour surfaces between water 301 and the two carbonyls of the ligand. We have found that inclusion of the energetic contribution of this tightly constrained water 301 significantly improves the correlation between HINT scores and experimental free energy, and that cyclic urea (and cyclic sulfamide) based inhibitors designed to displace water 301 [23,24] only gain the energy of the abolished water to ligand interactions. In Fig. 3 the dashed line indicates the correlation for the six water-displacing compounds, which quite nearly superimposes on the correlation line for the entire series post-correction for water 301. These studies clearly indicate the value of explicitly modeling water at molecular interfaces. There are a number of well-validated computational tools such as GRID [25] that can determine the locations of probable water molecules in biomolecular systems when the experimental data are weak (e.g., low crystallographic resolution) or when building models for cases where experimental data do not exist.

Therefore it is a surprise that so few virtual screening protocols pay any heed to the solvent as virtual (and real) ligands are rapidly docked and scored to make binary decisions regarding potential lead compounds. Clearly, knowledge of the role of water in the active sites of the targets would impact some of these choices [26]. It appears that speed is the overarching goal in many implementations of virtual screening even though "getting it right" would argue for including the energetic contributions of structural water in virtual screening models.

2.3. pH and ionization states

Another important aspect for accurately modeling biomolecular complexes is the dependence of the free energy of association on the ionization state(s) of residues and/or ligand functional groups involved in binding. In addition to global pH effects, the local effective "pH" in specific loci can be perturbed enough to protonate or deprotonate acidic/basic residues, or, in other words, the pK_a of ionizable residues can be affected by environment. This is a significant issue in many drug design projects because virtual screening experiments of databases against protein targets are almost exclusively carried out with no or very limited attempts to optimize the ionization state of residues surrounding the binding site and/or functional groups on the ligands themselves. It is not difficult to envision cases where the protonation of a single acid residue or ligand functional group would make the difference between a favorable and unfavorable binding event. "Getting it right" argues for consideration of pH effects in virtual screening protocols.

We have developed a new modeling tool, called computational titration, which includes a careful modeling of the ionization states and resonance forms for the ligands and protein residues at the binding site and the evaluation of the free energy of binding for modeled complexes by HINT. This procedure was recently applied to a collection of influenza virus neuraminidase–inhibitors [27] that had previously been difficult to model because of the observed wide range in enzyme inhibition with seemingly minor changes in ligand structure [28,29]. With computational titration each binding site can be "titrated" to reveal the optimum protonation condition for the association (see Fig. 4). The curve in Fig. 4b indicates the HINT score as a function of added (titrated) protons to the complex. The peak at two protons is suggestive of the optimum protonation state for the complex. After correction with the computational titration algorithm of the neuraminidase–inhibitor complex models for pH effects and likely ionization changes, the regression of H_{total} versus ΔG according to equation

$$\Delta G = -0.0020 H_{total} + 4.052 \quad (8)$$

with $r^2 = 0.65$ and standard error of ± 2.1 kcal mol^{-1}, was obtained [27].

One of the consequences of this approach is that it forces us to recognize that there is not a *single* global model for each protein–ligand system, i.e., a well-determined ionized state for each defined residue, because protons are not static and the ionization state of residues is a group function. In

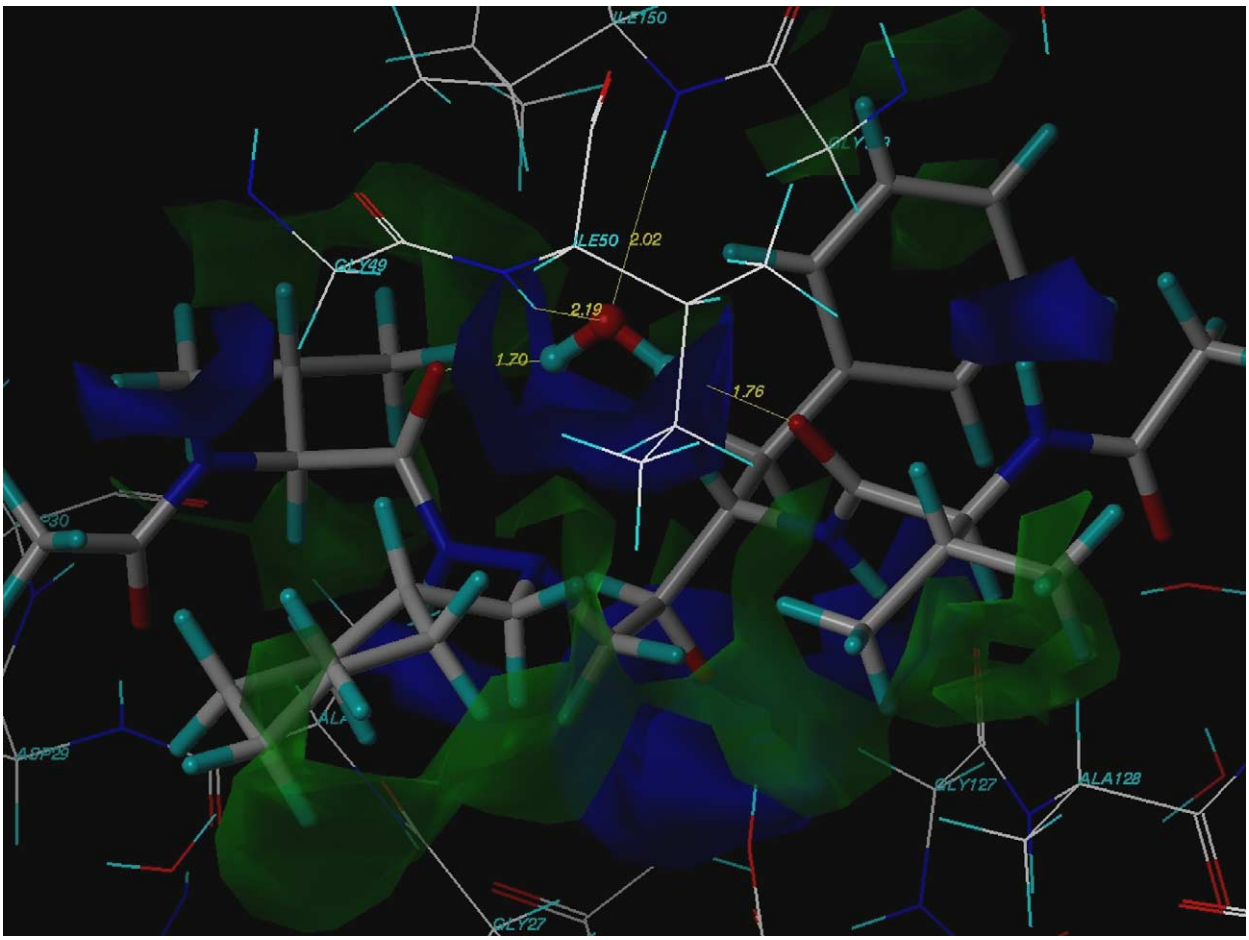

Fig. 2. HINT interaction maps for HIV-1 protease complex with inhibitor CGP 53820 (pdb code 1hih) [22]. The inhibitor is shown in rendered stick form, the water 301 in ball and stick, and the active site residues of the protein in line display. The HINT interaction map illustrates the through-space interactions as contours with character and magnitude (volume): blue contour surfaces represent favorable polar interactions, which are largely hydrogen bonds between the protein (or water 301) and the ligand; green contour surfaces (translucent) represent favorable hydrophobic–hydrophobic interactions. The hydrogen bonds to water 301 and their distances are indicated.

other words, there may be multiple energetically accessible states for each complex. Proton transfer between molecules and, in particular, proton migration across hydrogen bonds, has been identified as one of the fundamental mechanisms for biological processes [30]. Thus, modeling protein–ligand complexes of this type as an ensemble of multiple ionization models is a more biologically reasonable approach.

3. Discussion

The association of two molecular entities in the biological environment is a process governed by free energy, that is to say that entropy is important. Especially significant is the role of solvent, i.e., water. The displacement of water molecules when two biological molecules associate is clearly a major source of entropy. This is manifested by the hydrophobic effect and desolvation of functional groups buried by the association. Solvent effects also modulate the ionization states of acidic and basic functional groups on the protein side chains (or ligand functional groups) and thereby biomolecular associations. These properties have been little studied and are poorly understood but must be computationally accounted for in any realistic model. It is well known that “local” pH at specific sites within proteins can be considerably different than the global solution pH, and this, in turn, affects the ionization state of these residues [31–36].

Computational approaches to target the above factors (entropy, hydrophobicity, solvation/desolvation, and pH) in the biological environment are currently at the cutting edge of simulations of molecular interactions. There are, loosely, three tactics applied in modeling complex systems such as those involving protein–ligand interactions. The first is based on extensive molecular dynamics simulations of the complete system in a box including explicit water molecules with boundary conditions. By configuration of these simulations for thermodynamic integration (TI) or free energy perturbation (FEP), the free energy for specific events, e.g., ligand binding, can be estimated [4–7]. Alternatively, the linear response method analyzes the states of multiple dynamics

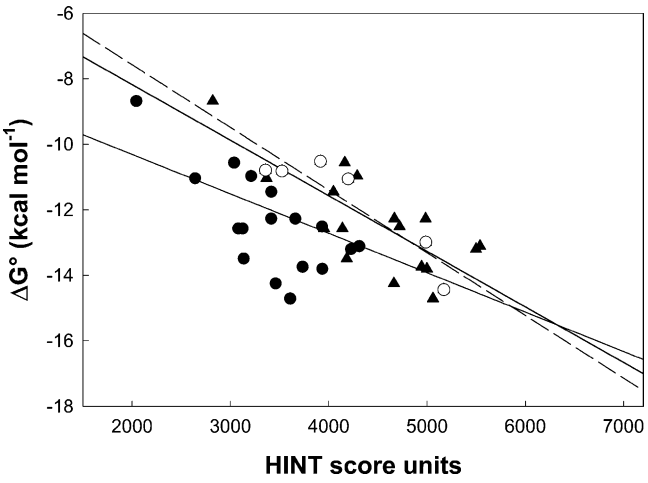

Fig. 3. Correlation of HINT score with free energy of binding for HIV-1 protease inhibitors. Circles: score for inhibitors when the tight-binding water 301 is not included (correlation: $r^2 = 0.30$, standard error $= \pm 1.30\,\text{kcal mol}^{-1}$). Open circles are six inhibitors designed to displace 301 and dashed line is fit for these. Triangles: water 301 score contribution included for the 17 inhibitors where 301 was present (correlation including the six inhibitors without water 301: $r^2 = 0.63$, standard error $= \pm 0.95\,\text{kcal mol}^{-1}$).

simulations to obtain a semi-empirical estimation of free energy [37,38]. These methods are computationally expensive to perform and are tied to the quality of the underlying Newtonian molecular mechanics (MM) forcefields. Thus, only interaction types specifically programmed into the forcefield will be modeled and observed. In particular, MM forcefields do not include specific terms for the hydrophobic effect and hydrophobic–hydrophobic interactions are usually indicated to be energetically unfavorable.

The second class of tactic for simulating the biological environment is based on the electrostatic properties of the molecules and their response to the dielectric variations at the molecule/solvent interface. Several implementations of solutions to the Poisson–Boltzmann equation (PBE) have been used in the biological environment [31,36,39]. While PBE approaches are very robust in terms of representing electrostatic effects, particularly at molecular surfaces, they do not have terms to represent other non-covalent forces so can be of limited utility in cases that are not dominated by electrostatic effects. For this reason, some of the more recent simulations have utilized methods combining PBE (or generalized Born, a faster, less robust electrostatic method) with FEP or TI calculations [40–42].

The third tactic for understanding the biological environment is the application of empirical free energy scoring algorithms. These tools have often been developed for feedback in docking or de novo molecule design programs, and thus have been optimized for speed. In general, simple metrics such as partial charge and solvent accessible surface area, combined with structural features as extracted from experimental measurements, produce a "score" that can often be correlated with free energy [43,44]. For example, the hydrophobic effect is usually represented by apolar surface contact area. While useful within their native context, i.e., docking or de novo building, these algorithms are often invalid outside their training and validation environment. The HINT forcefield described in this work is empirical, but differs from other scoring functions in that all key parameters are derived from an experiment that uniquely measures intermolecular interactions in the biological environment.

3.1. The relationship between HINT score and free energy

We have shown previously that, because $\log P_{o/w}$ is a thermodynamic quantity representing the free energy of solvent transfer for a solute, the HINT score is also a free energy-like parameter [9]. In Eqs. (3)–(6) and (8) and in other publications [15,45] we have indicated a linear relationship between H_{total} and ΔG. Somewhat surprisingly the slopes of these lines have been rather consistent, similar to 500 ± 100 HINT units per kcal mol^{-1}. Less of a surprise, perhaps, is that the

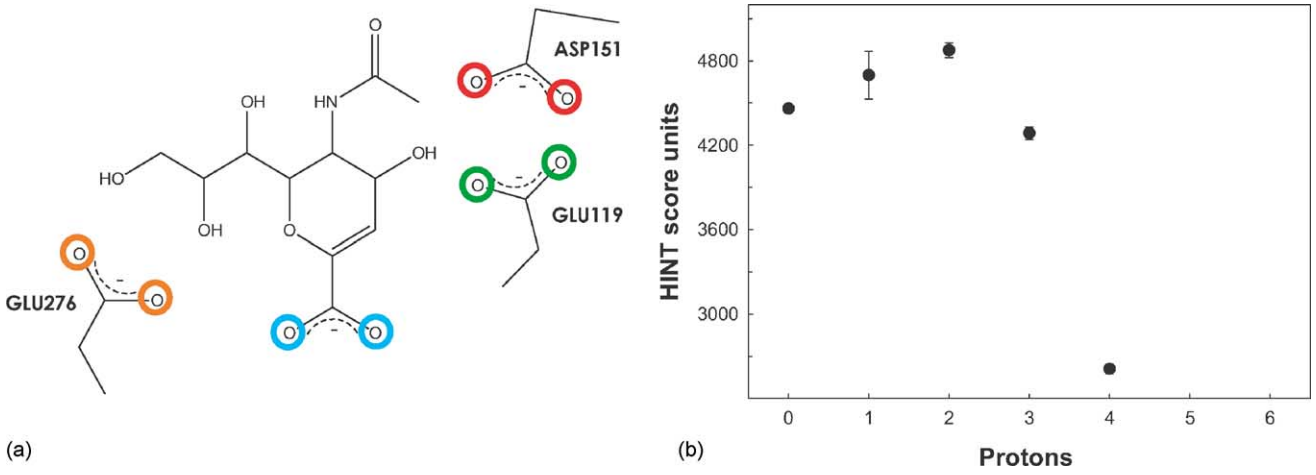

Fig. 4. (a) Influenza virus neuraminidase inhibitor DANA in binding site indicating ionizable residues and functional groups; (b) computational titration of DANA where the addition of two protons to the model at left generates the optimum binding model.

y-intercepts of these lines have had a much larger variance, even changing sign. We interpret the consistency in slope as an indication that $\Delta\Delta G$ is modeled well by HINT, and that this follows from the thermodynamic basis of $\log P_{o/w}$. The *y*-intercept term is much harder to rationalize. It must be due to a collection of hard-to-quantify factors such as internal energy of the macromolecule and ligands themselves, protein conformational entropy, as well as other factors such as experimental uncertainty from the structural and solution measurements.

Experimental uncertainties and errors are, in some ways, the untold story of molecular modeling and computational predictions of binding. Typical reproducibility in inhibition, etc. measurements between laboratories is around an order of magnitude, or around 1 kcal mol^{-1}. Biomacromolecular crystallography is also subject to a number of reproducibility issues, from the possibility that different crystal forms may be analyzed, with different packing forces (which would be particularly significant for surface residues), to the actual uncertainties and errors made via subjective and potentially biased hand-fitting of electron density. As shown in Fig. 1a and b, an increase in the quality of the structural data causes a significant improvement in the prediction. Also, in the limit of low resolution where the electron density is less well-defined, refinement can be driven more by molecular mechanics forcefields than by the experimental electron density, which can introduce a different set of biases into "experimental" structure data. Considering all of these factors, an uncertainty of prediction of ± 1 kcal mol^{-1} is probably conservative.

4. Conclusions

Of course we desire accuracy in our models and predictions of binding. Remember the Holy Grail? However, if we are attempting to screen thousands or millions of candidate ligands as potential lead compounds we also desire speed. Optimally we are thus looking for a compromise between accuracy and speed. We believe that the computational protocols we have been developing for empirical scoring, treatment of water and modeling of pH and ionization are such a compromise. We are able to achieve respectable correlations between the HINT scores and experimental free energy measurements with simple scoring functions and associated consideration of implicit and explicit solvent effects. However, these (or any) approaches are certainly not universally "getting it right". That goal is a long way off!

Acknowledgements

We gratefully acknowledge the support of the National Institutes of Health (DJA, Grant 5R01HL32793-15), Virginia Commonwealth University and the Italian Instruction, University and Research Ministry Grant PRIN01, FIRB 2003 and the National Institute for the Physics of Matter (AM) for partial support of this research. Helpful discussions with J. Neel Scarsdale are also acknowledged.

References

[1] D.A. Pearlman, P.A. Kollman, The calculated free energy effects of 5-methyl cytosine on the B to Z transition in DNA, Biopolymers 29 (1990) 1193–1209.
[2] H. Gohlke, G. Klebe, Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, Angew. Chem. Int. Ed. Engl. 41 (2002) 2644–2676.
[3] C. Tanford, Amphiphile orientation: physical chemistry and biological function, Biochem. Soc. Trans. 15 (1987) 1S–7S.
[4] P.A. Kollman, Free energy calculations: applications to chemical and biochemical phenomena, Chem. Rev. 93 (1993) 2395–2417.
[5] M.R. Reddy, M.D. Erion, A. Agarwal, Free energy calculations: use and limitations in predicting binding affinities, Rev. Comput. Chem. 16 (2000) 217–304.
[6] T.P. Straatsma, Free energy by molecular simulation, Rev. Comput. Chem. 9 (1996) 81–127.
[7] N. Ota, C. Stroupe, J.M. Ferreira da Silva, S.A. Shah, M. Mares-Guia, A.T. Brunger, Non-Boltzmann thermodynamic integration (NBTI) for macromolecular systems: relative free energy of binding of trypsin to benzamidine and benxylamine, Proteins: Struct. Funct. Genet. 37 (1999) 641–653.
[8] D.J. Abraham, A.J. Leo, Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients, Proteins: Struct. Funct. Genet. 2 (1987) 130–152.
[9] G.E. Kellogg, D.J. Abraham, Hydrophobicity. Is $\log P_{o/w}$ more than the sum of its parts? Eur. J. Med. Chem. 35 (2000) 651–661.
[10] F.C. Wireko, G.E. Kellogg, D.J. Abraham, Allosteric modifiers of hemoglobin. 2. Crystallographically determined binding sites and hydrophobic binding/interaction analysis of novel hemoglobin oxygen effectors, J. Med. Chem. 34 (1991) 758–767.
[11] G.E. Kellogg, G.S. Joshi, D.J. Abraham, New tools for modeling and understanding hydrophobicity and hydrophobic interactions, Med. Chem. Res. 1 (1992) 444–453.
[12] C. Hansch, A.J. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, 1979.
[13] P. Cozzini, M. Fornabaio, A. Marabotti, D.J. Abraham, G.E. Kellogg, A. Mozzarelli, Simple, intuitive calculations of free energy of binding for protein–ligand complexes. 1. Models without explicit constrained water, J. Med. Chem. 45 (2002) 2469–2483.
[14] R. Gussio, D.W. Zaharevitz, C.F. McGrath, N. Pattabiraman, G.E. Kellogg, C. Schultz, A. Link, C. Kunick, M. Leost, L. Meijer, E.A. Sausville, Structure-based design modifications of the paullone molecular scaffold for cyclin-dependent kinase inhibition, Anti-Cancer Drug Des. 15 (2000) 53–66.
[15] J.C. Burnett, P. Botti, D.J. Abraham, G.E. Kellogg, Computationally accessible method for estimating free energy changes resulting from site-specific mutations of biomolecules: systematic model building and structural/hydropathic analysis of deoxy and oxy hemoglobins, Proteins: Struct. Funct. Genet. 42 (2001) 355–377.
[16] D.J. Cashman, J.N. Scarsdale, G.E. Kellogg, Hydropathic analysis of the free energy differences in anthracycline antibiotic binding to DNA, Nucl. Acids Res. 31 (2003) 4410–4416.
[17] C.S. Wright, G.E. Kellogg, Differences in hydropathic properties of ligand binding at four independent sites in wheat germ agglutinin–oligosaccharide crystal complexes, Protein Sci. 5 (1996) 1466–1476.
[18] G.E. Kellogg, J.C. Burnett, D.J. Abraham, Very empirical treatment of solvation and entropy: a force field derived from $\log P_{o/w}$, J. Comput.-Aided Mol. Des. 15 (2001) 381–393.

[19] C.F. McGrath, R. Gussio, N. Pattabiraman, G.E. Kellogg, C. Schultz, A. Link, M. Leost, L. Meijer, C. Kunick, E.A. Sausville, D.W. Zaharevitz, Unpublished data.
[20] L. Meijer, A. Borgne, O. Mulner, J.P. Chong, J.J. Blow, N. Inagaki, M. Inagaki, J.G. Delcros, J.P. Moulinoux, Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of cyclin-dependent kinases cdc2 cdk2 and cdk5, Eur. J. Biochem. 243 (1997) 527–536.
[21] M. Fornabaio, F. Spyrakis, A. Mozzarelli, P. Cozzini, D.J. Abraham, G.E. Kellogg, Simple, intuitive calculations of free energy of binding for protein–ligand complexes. 3. Including the free energy contribution of structural water molecules, J. Med. Chem., Unpublished data.
[22] J.P. Priestle, A. Fässler, J. Rösel, M. Tintelnot-Blomley, P. Strop, M.G. Grütter, Comparative analysis of the X-ray structures of HIV-1 and HIV-2 proteases in complex with CGP 53820, a novel pseudosymmetric inhibitor, Structure 3 (1995) 381–389.
[23] P.Y.S. Lam, P.K. Jadhav, C.J. Eyermann, C.N. Hodge, Y. Ru, L.T. Bacheler, J.L. Meek, M.J. Otto, M.M. Rayner, Y.N. Wong, C.-H. Chang, P.C. Weber, D.A. Jackson, T.R. Sharpe, S. Erickson-Viitanen, Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors, Science 263 (1994) 380–384.
[24] W. Schaal, A. Karlsson, G. Ahlsén, J. Lindberg, H.O. Andersson, U.H. Danielson, B. Classon, T. Unge, B. Samuelsson, J. Hultén, A. Hallberg, A. Karlén, Synthesis and comparative molecular field analysis (CoMFA) of symmetric and nonsymmetric cyclic sulfamide HIV-1 protease inhibitors, J. Med. Chem. 44 (2001) 155–169.
[25] P.J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, J. Med. Chem. 28 (1985) 849–857.
[26] P. Cozzini, M. Fornabaio, A. Marabotti, D.J. Abraham, G.E. Kellogg, A. Mozzarelli, Free energy of ligand binding to protein: evaluation of the contribution of water molecules by computational methods, Curr. Med. Chem., in press.
[27] M. Fornabaio, P. Cozzini, A. Mozzarelli, D.J. Abraham, G.E. Kellogg, Simple, intuitive calculations of free energy of binding for protein–ligand complexes. 2. Computational titration and pH effects in molecular models of neuraminidase–inhibitor complexes, J. Med. Chem. 46 (2003) 4487–4500.
[28] B.J. Smith, P.M. Colman, M. Von Itzstein, B. Danylec, J.N. Varghese, Analysis of inhibitor binding in influenza virus neuraminidase, Protein Sci. 10 (2001) 689–696.
[29] I. Muegge, The effect of small changes in protein structure on predicted binding modes of known inhibitors of influenza virus neuraminidase: PMF-scoring in DOCK4, Med. Chem. Res. 9 (1999) 490–500.
[30] T. Bountis (Ed.), Proton Transfer in Hydrogen-Bonded Systems, NATO ASI Series B: Physics, vol. 291, Plenum Press, New York, 1992, pp. 1–355.
[31] M.K. Gilson, B.H. Honig, Calculation of electrostatic potentials in an enzyme active site, Nature 330 (1987) 84–86.
[32] D. Bashford, M. Karplus, pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model, Biochemistry 29 (1990) 10219–10225.
[33] S.J. Dixon, P.C. Jurs, Estimation of pK_a for organic oxyacids using calculated atomic charges, J. Comput. Chem. 14 (1993) 1460–1467.
[34] Y.Y. Sham, Z.T. Chu, A. Warshel, Consistent calculations of pK_a's of ionizable residues in proteins: semi-microscopic and microscopic approaches, J. Phys. Chem. B 101 (1997) 4458–4472.
[35] G.M. Ullmann, E.W. Knapp, Electrostatic models for computing protonation and redox equilibria in proteins, Eur. Biophys. J. 28 (1999) 533–551.
[36] J.E. Nielsen, J.A. McCammon, On the evaluation and optimization of protein X-ray structures for pK_a calculations, Protein Sci. 12 (2003) 313–326.
[37] J. Åqvist, V.B. Luzhkov, B.O. Brandsal, Ligand binding affinities from MD simulations, Acc. Chem. Res. 35 (2002) 358–365.
[38] W.L. Jorgensen, Free energy calculations: a breakthrough for modeling organic chemistry in solution, Acc. Chem. Res. 22 (1989) 184–189.
[39] J.A. Grant, B.T. Pickup, A. Nicholls, A smooth permittivity function for Poisson–Boltzmann solvation methods, J. Comput. Chem. 22 (2001) 608–640.
[40] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham III, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, Acc. Chem. Res. 33 (2000) 889–897.
[41] F. Fogolari, A. Brigo, H. Molinari, Protocol for MM/PBSA molecular dynamics simulations of proteins, Biophys. J. 85 (2003) 159–166.
[42] H. Gohlke, C. Kiel, D.A. Case, Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RaIGDS complexes, J. Mol. Biol. 330 (2003) 891–913.
[43] H.-J. Böhm, M. Stahl, Rapid empirical scoring functions in virtual screening applications, Med. Chem. Res. 9 (1999) 445–462.
[44] I. Muegge, Y.C. Martin, A general and fast scoring function for protein–ligand interactions: a simplified potential approach, J. Med. Chem. 42 (1999) 791–804.
[45] D.J. Cashman, J.P. Rife, G.E. Kellogg, Which aminoglycoside ring is most important for binding? A hydropathic analysis of gentamicin, paromomycin, and analogues, Bioorg. Med. Chem. Lett. 11 (2001) 119–122.