Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Molecular Graphics and Modelling 24 (2006) 434–439
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Tools for building a comprehensive modeling system for virtual screening under real biological conditions: The Computational Titration algorithm

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

[a] *Department of Medicinal Chemistry & Institute for Structural Biology and Drug Discovery, School of Pharmacy, Virginia Commonwealth University, Box 980540, 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 7 September 2005
Available online 19 October 2005

Abstract

Computational tools utilizing a unique empirical modeling system based on the hydrophobic effect and the measurement of log $P_{o/w}$ (the partition coefficient for solvent transfer between 1-octanol and water) are described. The associated force field, Hydropathic INTeractions (HINT), contains much rich information about non-covalent interactions in the biological environment because of its basis in an experiment that measures interactions in solution. HINT is shown to be the core of an evolving virtual screening system that is capable of taking into account a number of factors often ignored such as entropy, effects of solvent molecules at the active site, and the ionization states of acidic and basic residues and ligand functional groups. The outline of a comprehensive modeling system for virtual screening that incorporates these features is described. In addition, a detailed description of the Computational Titration algorithm is provided. As an example, three complexes of dihydrofolate reductase (DHFR) are analyzed with our system and these results are compared with the experimental free energies of binding.

Keywords: Comprehensive modeling system; Virtual screening; Computational Titration algorithm; Hydropathic interactions; HINT

1. Introduction

Virtual screening is the computational analog of high-throughput assay screening (HTS) in that a library of putative ligands is docked and scored in silico to determine a binding efficacy that, in a perfect world, correlates with what would be determined in the actual assay. As the library itself can be ''virtual'', this can be a very valuable method for discovering new lead compounds or for setting synthesis priorities in a target-based drug discovery effort. The docking computational tools are relatively well-developed. Quite a few algorithms from various academic and commercial groups have been put forward and validated to place ligands in random, semi-random, or intelligently derived poses [1–6]. On the other hand, accurately scoring the poses, especially for comparing binding of multiple ligands, is less well-developed. Ideally, the score should represent the free energy of binding for the ligand. This is a quite difficult computational problem since almost always the only information available, even for a crystallographically determined structure, to say nothing of a docked pose, is the *end point* of binding, i.e., the equilibrium conformation of the protein–ligand complex. We know almost nothing about how the ligand arrived to its final position. Did the biomacromolecule change its conformation for the ligand to pass to the active site? How different from the bound form is the unbound form of the protein? What, and how many, water molecules at the active site were displaced? What happened to the waters that were solvating the ligand? There are other relevant questions—too numerous to list here. Detailed molecular dynamics simulations provide insight into many of these questions, but have yet to provide a complete picture of ligand binding.

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

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.09.001

However, the problem *is* important enough and the payoff potential *is* high enough that making these 'leaps of faith' in attempting to extrapolate the free energy of binding from static structures is a key area of computational chemistry practice. While free energy perturbation [7,8] and linear response methods [9,10] have been shown to be fairly accurate predictive tools for free energy of binding, both extensively utilize molecular dynamics simulations to recreate the fluxional nature of biological systems and are thus quite expensive and slow. The combination of these factors is the driving force for simpler scoring tools that retain as much of "reality" as possible, yet are robust and rapid enough to support real-time decision making.

Our scoring function, HINT (Hydropathic INTeractions), has been documented and reviewed previously [11–13]. This *natural* force field is particularly valuable for evaluation of non-covalent molecular interactions, and pays particular attention to entropically driven hydrophobic effects. An entire bimolecular interaction is scored by the following:

$$\sum_i \sum_j b_{ij} = \sum_i \sum_j (a_i S_i a_j S_j T_{ij} R_{ij} + r_{ij}), \quad (1)$$

where b_{ij} is the interaction score between atoms i and j, a the hydrophobic atom constant, S the atomic solvent accessible surface area, T_{ij} a logic function assuming +1 or −1 values based on the polar nature of interacting atoms i and j, and R_{ij} (usually a simple exponential function) and r_{ij} (an adaptation of the Lennard–Jones function) are functions of the distance between the atoms i and j. The key a parameters are calculated in a manner similar to the CLOG-P method of Hansch and Leo [14]. A positive b_{ij} value indicates hydropathically favorable contacts (hydrogen bonds, acid/base, and hydrophobic/hydrophobic), while negative b_{ij} values indicate unfavorable contacts (acid/acid, base/base, and hydrophobic/polar). Since the log $P_{o/w}$ (partition constant for 1-octanol/water) of a ligand or biological molecule is the sum of all hydrophobic atomic constants for that molecule, the a parameters are dimensionless thermodynamic quantities directly related to the free energy of atom transfer between 1-octanol and water. The HINT model relies on the assumption that each b_{ij} is related to a partial δg value for the i–j atom–atom interaction and the total HINT score (H_{TOTAL}) is directly comparable to the global $\Delta G_{interaction}$. It is important to emphasize that, because HINT is based on these free energy parameters that are obtained in solution, the HINT model implicitly includes some portion of entropy, solvation, and other effects otherwise difficult to estimate computationally.

We have shown HINT's applicability to predicting free energy of binding for protein–ligand systems [15–17], have illustrated a modeling process and tools we call "Computational Titration" by which the ionization states of active site residues and ligand functional groups can be explored [18,19], and have developed tools to build an understanding of the role of water molecules at the active site [20–22]. It is our current goal to incorporate this force field and set of tools into a comprehensive system for virtual screening. While this effort is on-going, we will report here the system design and, in particular, describe the Computational Titration tool in detail.

2. System design

2.1. General considerations

In our system, we describe a molecular model in three parts: the protein "receptor", the small molecule "ligand", and a "cofactor" that is for our purposes an array of solvent water molecules. Each of these components must be pre-processed to prepare for further calculations. Once atom types, etc., are validated, then all components need to be 'hydrogenated' with one of three levels: all, semi-essential, or essential hydrogens. Essential hydrogens are those attached to polar atoms, e.g., N, O, S, or P. Semi-essential hydrogens, a recent extension to the HINT method, also include all hydrogens attached to unsaturated carbons, and hydrogens attached to carbons that are α to heteroatoms. This latter class of hydrogen has been implicated in some recent reports of $–CH_n$ hydrogen bonding [23–25]. After hydrogens are placed, a molecular mechanics structure optimization is performed to idealize the positions and conformations of the hydrogen atoms.

This is the starting point for the scoring system: three molecular components (two if there is no cofactor) in which hydrogens or in rare cases some heavy atoms have been optimized with a molecular mechanics force field minimization. As this process can often be compromised by local minima involving the non-optimum rotation of $R–XH_n$ groups such as R–OH in tyrosines, serines, and threonines or R–SH in cysteines, an additional structure cleanup is applied as part of the scoring system (vide infra). Fig. 1 illustrates the following steps schematically.

Because of the vagaries of molecular modeling programs, force fields, and the partitioning and scoring methods of HINT [16,18,20], it is usually best to split the components into

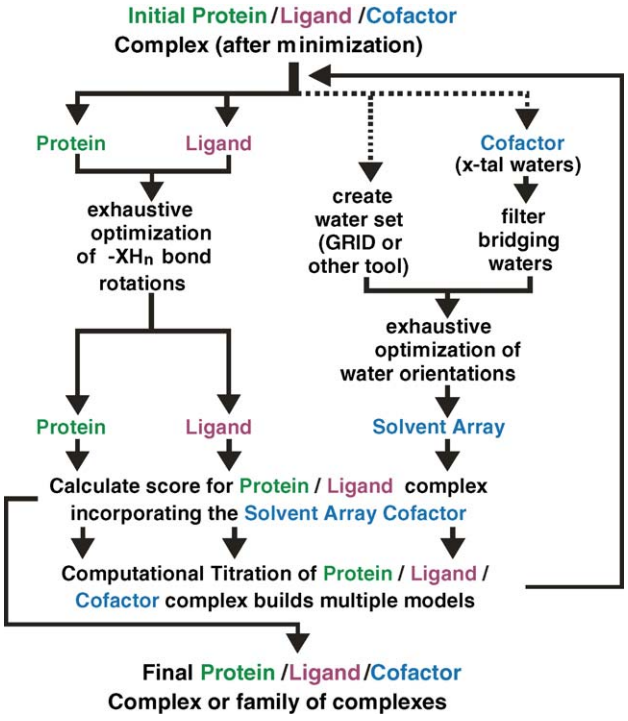

Fig. 1. Basic flowchart for scoring system.

separate molecules for the remainder of the process. When the cofactor is present and comprised of an array of water molecules, the array is first filtered to identify the water molecules that are potentially in ''bridging'' positions, i.e., able to interact with *both* the protein and ligand. In general, a water molecule is eliminated from the cofactor if its oxygen atom is not within a set distance, e.g., 5–6 Å, from any heavy atom in the protein *and* from any heavy atom in the ligand. The remaining water molecules are then exhaustively optimized as described previously [21]. If a water array is not available from the structure, it should be created using GRID [26] or another tool such as we have previously described [22], followed by optimization. The goal is to create a plausible set of bridging waters that would, optimally, match what would be experimentally found, if it were possible. However, any set of water molecules that provides *reasonable* water-mediated hydrogen bonds is of value for modeling protein–ligand interactions because at room (or physiological) temperature the water structure is very fluxional. If the cofactor is not a solvent array, but is instead, a protein or ligand-like small molecule, then it is processed as a member of those classes, and in our current system this would preclude having a solvent array in the model.

The key processing step for the protein and ligand models is to correct for local minima that would significantly impact the HINT energy score. In particular, we are interested in the rotation angle of R–XH_n bonds—these are exhaustively optimized for hydrogen bonding by forcing rotation through the entire 360°. Simply, this algorithm calculates the HINT score of each rotamer model for the –XH_n group as a ''ligand'' in a ''receptor'' comprised of all atoms within a range, e.g., 6 Å, of it. This algorithm is an adaptation of our earlier water optimization tool [21] that treated each water molecule as a ligand in a ''receptor'' of its neighborhood. Also, the orientation of histidines, asparagines, and glutamines are examined and rotated 180° if necessary.

At this point, a HINT score can be calculated for the protein–ligand interaction of the model. However, it is likely that there are a few to several ionizable residues and/or ligand functional groups in the active site, and that some of these may not be modeled in the most ideal ionization states for the interaction. It cannot be assumed that all functional groups adopt the ionization state that would be predicted by their pK_a and the solution pH. That said, the pH of the crystallization media and/or used in the solution binding measurements does have an influence on the ionization state of internal functional groups in a more subtle way. It would appear that the solution pH ''shifts'' the ionization states around a ''set point'' dictated by the local structure and interactions. This phenomenon, which we have been exploring with a modeling tool we call ''Computational Titration'', has been recently described [18,19]. In the case of the tripeptide Glu–Asp–Leu bound at HIV-1 protease, the model results showed a remarkable similarity to the experimental measurement of binding as a function of pH [27]. Thus, to build a molecular model that most accurately reflects the complex structure, the ionization states of residues and functional groups at the active site also need to be optimized. The next section of this report describes in some detail the algorithm of Computational Titration.

2.2. Computational Titration

The concept behind Computational Titration is that each of the many protonation models should be built and scored, and from that a ''titration curve'' can be constructed and a single or family of best models can be gleaned. First, analysis and enumeration of the actual protonation models must be made. Fig. 2 illustrates a protein with three potentially ionizable groups at the active site and a ligand with two ionizable sites. Our algorithm allows selection of the types of groups to be ''ionizable'', i.e., acid (Asp and Glu), bases (Lys, His, or Arg), Tyr, and Cys. We do not normally titrate tyrosine, but it is included in this example as an illustration. Note that carboxylic acids, whether a protein residue or on the ligand, have three possible states: deprotonated, protonation of $\mathbf{O_1}$, or protonation of $\mathbf{O_2}$. Histidine and Arginine have similar multiplicity, while all other titratable groups have two states: protonated or deprotonated. This is illustrated in Fig. 2 with two bins for protonation associated with the carboxylates, but only one bin with the amines and phenol. Note that only one of the carboxylate bins can be filled for any given state. Each of the filled circles in a bin represents a proton. Thus, there is only one state with zero protons (both carboxylates deprotonated, both amines in neutral state, and the phenol ionized). There are seven ways to place one proton as illustrated in the figure. The placement of two protons is only partially shown—there are, in total, 19 states. (There are 25, 16, and 4 states, respectively, for three, four, and five protons.) In this case, models for all of the 72 states would be built and each would be individually

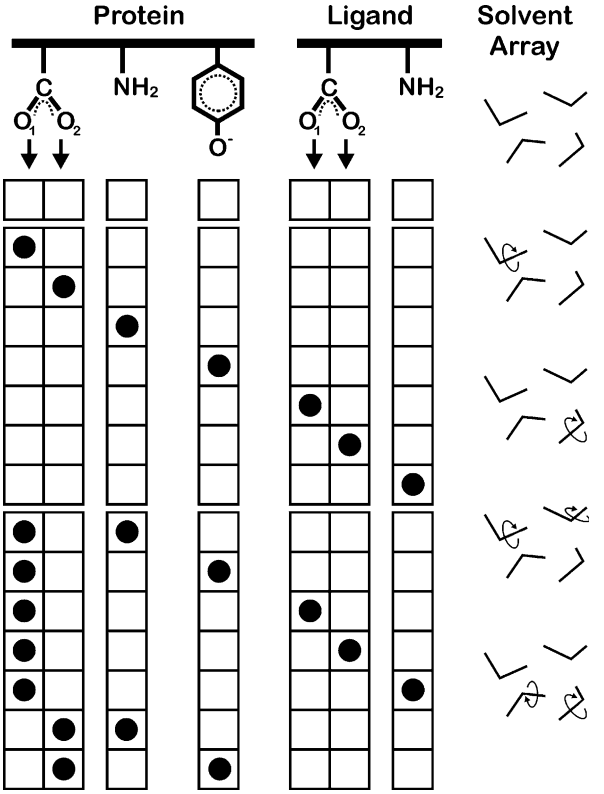

Fig. 2. Schematic illustration of Computational Titration algorithm. Each filled circle represents a protonation of that functional group.

optimized as described above, i.e., all rotatable R–XH_n bonds would be examined (including those newly created in protonated carboxylates, etc.), all asparagines, glutamines, etc., would be re-examined, and all of the water molecules at the active site would be re-optimized. This latter step may be necessary because the hydrogen bonding characteristics of an atom will change with its ionization state and water molecules interacting with that atom may need to rearrange so that they can present a different face to it. To account for the possibility that the order of these optimizations may influence the results, we iterate each model four cycles, which is enough to converge models built from different starting points to the same, i.e., within ±1% in HINT score, end point. This four-cycle iteration is the most CPU-intensive portion of the procedure.

As they are calculated, the results are tabulated: HINT score as a function of site charge. In some cases [16,19], we have calibration equations relating binding free energy to HINT score for the specific protein system; in other cases, we must use the reported [16] general equation. This conversion to an actual energy is an important step as it allows us to apply a statistical thermodynamics energy partition function to calculate Boltzmann-weighted binding free energy averages. For the Computational Titration, we calculate this Boltzmann average for *each* site charge and assume that these site charges and solution pH are correlated. ''Titration'' plots of this type will be presented in the next section for a handful of cases.

3. Results

3.1. Analysis of DHFR complexes

As an illustration of our score system, we have examined three crystallographically characterized complexes of dihydrofolate reductase (DHFR): biopterin (**1**), folate (**2**), and methotrexate (**3**) (see Fig. 3 and Table 1). These are the three DHFR–ligand complexes of the X-Score database of Wang et al. [28,29], so they can be assumed to be validated and self-consistent. All models were initially prepared with default protonation states for the ligand functional groups and protein residues at the active site (defined to be within 6 Å of the ligand). These are termed the ''pH 7'' models; the calculated binding energies for these models are indicated in Table 1. The raw HINT scores were converted to free energies using the general equation:

$$\Delta G = -0.00195 H_{\mathrm{TOTAL}} - 5.543, \quad (2)$$

where H_{TOTAL} is the total HINT score as described earlier [16]. These values were calculated for the three protein–ligand complexes, first without including the effects of active site waters, and then including the water–ligand HINT scores [19] for the pre-selected (as described above) bridging water molecules. In the former case, the average error of prediction is 3.38 kcal mol^{-1}, while in the latter this error is 2.15 kcal mol^{-1}, thus underscoring the value of including water in protein–ligand models and in binding energy predictions.

The three complexes were next subjected to Computational Titration analysis in order to optimize the ionization states of functional groups on the ligand and on active site amino acid residues. Parameters and results for the Computational Titration analyses are illustrated in Table 2 and Fig. 4. Because ligand **1** itself has but one potentially ionizable group and interacts with only one ionizable residue (Glu30), there are only a total of 6 models for **1**. Ligand **2** has four ionizable groups and interacts with two ionizable residues (Asp21 and Glu30) yielding a total of 412 models. Ligand **3** has five ionizable groups, interacts with only Asp27, and has 216 potential ionization models. We think of the Boltzmann-weighted energy predictions as graphed in Fig. 4 as the ''titration'' curve and choose as the ''optimal'' calculated binding energy the minimum (lowest energy) point on the titration curve. These

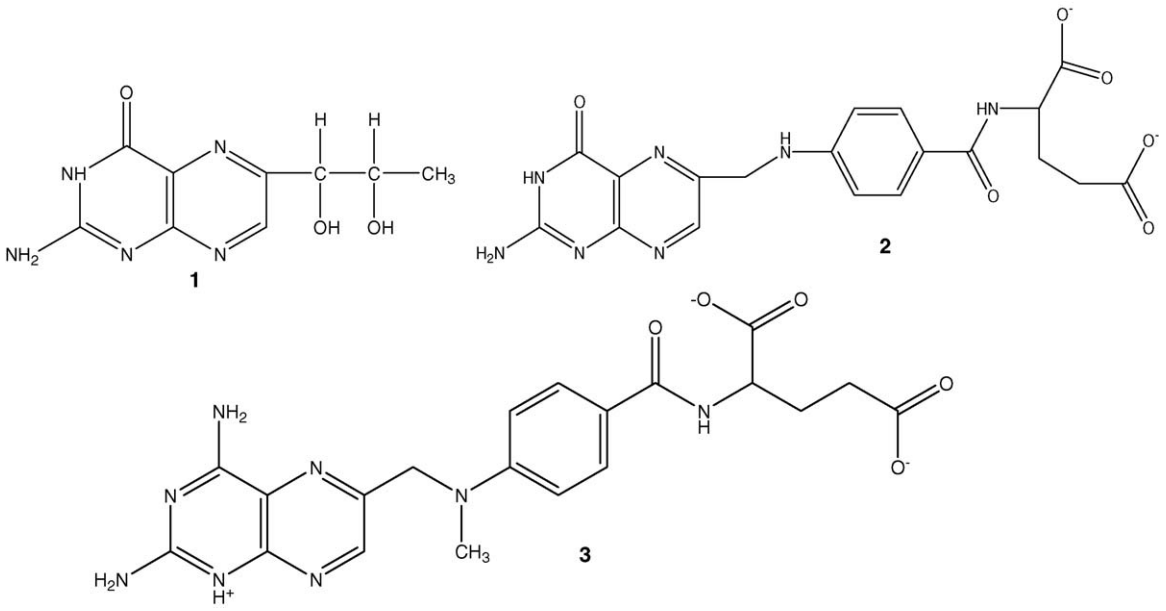

Fig. 3. Dihydrofolate reductase ligands.

Table 1
DHFR ligands with calculated and experimental binding data

Ligand	PDB code	No. waters	"pH 7" ΔG (kcal mol^{-1})		Optimal (calculated) ΔG (kcal mol^{-1})	Experimental ΔG (kcal mol^{-1})	References
			No water	With water			
1	1dr1	4	−4.37	−4.35	−5.44	−7.60	[30,28]
2	1drf	9	−7.26	−11.23	−9.12	−10.15	[31,28]
3	4dfr	12	−9.22	−11.12	−13.76	−13.23	[32,33]

Table 2
Computational Titration results for DHFR complexes (energies in kcal mol^{-1})

Ligand	Site charge	No. of models	Normal average	Maximum	Minimum	Boltzmann average
1	−1	1	−4.60	−4.60	−4.60	−4.60
	0	3	−3.94	−1.95	−6.08	−5.44
	1	2	−4.51	−4.22	−4.81	−4.58
2	−4	1	−6.60	−6.60	−6.60	−6.60
	−3	10	−6.70	−5.17	−9.83	−8.49
	−2	41	−6.83	−3.98	−10.44	−8.89
	−1	88	−6.97	−3.24	−10.74	−9.12
	0	104	−7.15	−3.26	−10.79	−9.08
	1	64	−7.30	−3.75	−10.65	−8.91
	2	16	−7.55	−6.39	−8.71	−8.03
3	−3	1	−9.55	−9.55	−9.55	−9.55
	−2	9	−10.40	−8.28	−12.77	−11.95
	−1	33	−10.84	−7.31	−14.96	−13.10
	0	63	−11.47	−7.31	−15.59	−13.62
	1	66	−11.98	−8.43	−15.44	−13.72
	2	36	−12.50	−9.59	−15.37	−13.76
	3	8	−13.00	−11.59	−14.41	−13.69

are also reported in Table 1; the average error of prediction in this case is a very respectable 1.24 kcal mol^{-1}.

3.2. Summary and outlook

The one question we cannot easily answer, however, is "which *specific* model is correct?" This is an exceedingly interesting question because it requires us to think about molecular modeling and free energy prediction in new (and perhaps uncomfortable) ways. For starters, there cannot be only *one* correct model when so many states are energetically similar. In fact, at room temperature many of these states are populated. This leads to another issue: how valid an approach is attempting to model room temperature assay data using crystal structures obtained at very low temperature? The tight electron density envelopes that lead to such beautiful three-dimensional structures are potentially very misleading when modeling events associated with ligand binding. Molecular dynamics is a partial cure, but fluidity in protons must also be considered. All of the models examined in Computational Titration differ only in the location of protons—no heavy atoms are moved. In a sense, we can say that these models are "isocrystallographic" in that all of them would fit in the experimental electron density envelopes.

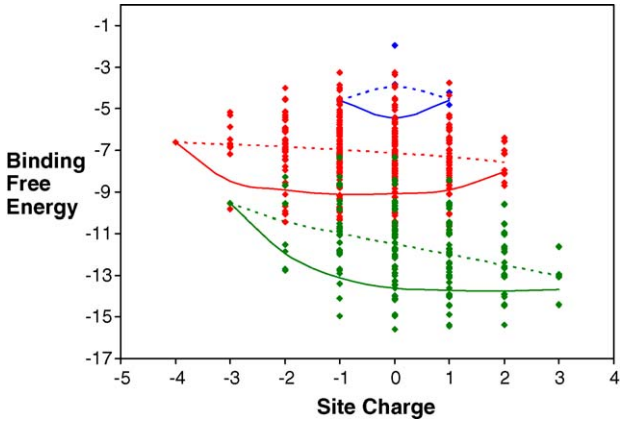

Fig. 4. Computational Titration results for ligands **1** (blue), **2** (red) and **3** (green). Each data point represents a protonation model for the DHFR complex. The dashed lines are the normal averages of energy as a function of site charge, while the solid lines are the Boltzmann averages.

As it now exists, the Computational Titration calculation can be very computationally expensive. The number of discrete states increases exponentially with the number of ionizable groups at the active site. The examination of a large data set with this tool is a daunting task: some protein–ligand complexes have as many as 12 ionizable groups—yielding on the order of 10^6–10^7 models to be examined for just that 1 complex. While each state model is built, optimized and analyzed on the order of 1–10 s (longer times for more complex state models), this is clearly a bottleneck in our system. Addressing this is a long-term goal of our research program. We are implementing multi-processing versions of the Computational Titration tool and are also exploring other strategies such as genetic algorithms for this purpose.

Developing a pragmatic, but more or less complete, understanding of the multiple roles of water in the biological environment, particularly with respect to ligand binding, is another difficult task [34]. That may, in fact, be one of the next major challenges of computational chemistry.

Acknowledgements

We gratefully acknowledge the support of the National Institutes of Health (Grant 5R01GM071894, G.E.K.), Virginia Commonwealth University, and the Italian Instruction, University and Research Ministry Grant PRIN2003, FIRB 2003, and the National Institute for the Physics of Matter (A.M.) for partial support of this research. Technical assistance from Austin R. Hess and helpful discussions with J. Neel Scarsdale are also acknowledged.

References

[1] R.L. DesJarlais, R.P. Sheridan, G.L. Seibel, J.S. Dixon, I.D. Kuntz, R. Venkataraghavan, Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure, J. Med. Chem. 31 (1988) 722–729.

[2] E.C. Meng, B.K. Schoichet, I.D. Kuntz, Automated docking with grid-based energy evaluation, J. Comp. Chem. 13 (1993) 505–524.

[3] B. Kramer, G. Metz, M. Rarey, T. Lengauer, Ligand docking and screening with FlexX, Med. Chem. Res. 9 (1999) 463–478.

[4] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function, J. Comp. Chem. 19 (1998) 1639–1662.

[5] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.

[6] M.R. McGann, H.R. Almond, A. Nicholls, F.K. Brown, Gaussian docking functions, Biopolymers 68 (2000) 76–90.

[7] P.A. Kollman, Free energy calculations: applications to chemical and biochemical phenomena, Chem. Rev. 93 (1993) 2395–2417.

[8] 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.

[9] J. Åqvist, V.B. Luzhkov, B.O. Brandsal, Ligand binding affinities from MD simulations, Acc. Chem. Res. 35 (2002) 358–365.

[10] W.L. Jorgensen, Free energy calculations: a breakthrough for modeling organic chemistry in solution, Acc. Chem. Res. 22 (1989) 184–189.

[11] G.E. Kellogg, M. Fornabaio, F. Spyrakis, A. Lodola, P. Cozzini, A. Mozzarelli, D.J. Abraham, Getting it right: modeling of pH, solvent and "nearly" everything else in virtual screening of biological targets, J. Mol. Graph. Model. 22 (2004) 479–496.

[12] 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.

[13] 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.

[14] C. Hansch, A.J. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, John Wiley and Sons, New York, 1979.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] F. Spyrakis, M. Fornabaio, P. Cozzini, A. Mozzarelli, D.J. Abraham, G.E. Kellogg, Computational Titration analysis of a multiprotic HIV-1 protease–ligand complex, J. Am. Chem. Soc. 126 (2004) 11764–11765.

[20] 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 in HIV-1 protease complexes, J. Med. Chem. 47 (2004) 4507–4516.

[21] G.E. Kellogg, D.L. Chen, The importance of being exhaustive. Optimization of bridging structural water molecules and water networks in models of biological systems, Chem. Biodivers. 1 (2004) 98–105.

[22] G.E. Kellogg, M. Fornabaio, D.L. Chen, D.J. Abraham, New application design for a 3D hydropathic map-based search for potential water molecules bridging between protein and ligand, Internet Elec. J. Mol. Des. 4 (2005) 194–209.

[23] Z.S. Derewenda, L. Lee, U. Derewenda, The occurrence of C–H···O hydrogen bonds in proteins, J. Mol. Biol. 252 (1995) 248–262.

[24] M.C. Wahl, M. Sundaralingam, C–H···O hydrogen bonding in biology, TIBS 22 (1997) 97–102.

[25] L. Jiang, L. Lai, CH···O hydrogen bonds at protein–protein interfaces, J. Biol. Chem. 277 (2002) 37732–37740.

[26] P.J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, J. Med. Chem. 28 (1985) 849–857.

[27] J.M. Louis, F. Dyda, N.T. Nashed, A.R. Kimmel, D.R. Davies, Hydrophilic peptides derived from the transframe region of Gag–Pol inhibit the HIV-1 protease, Biochemistry 37 (1998) 2105–2110.

[28] R. Wang, L. Lai, S. Wang, Further development and validation of empirical scoring functions for structure-based binding affinity prediction, J. Comput. Aided Mol. Des. 16 (2002) 11–26.

[29] R. Wang, Y. Lu, S. Wang, Comparative evaluation of 11 scoring functions for molecular docking, J. Med. Chem. 46 (2003) 2287–2303.

[30] M.A. McTigue, J.F. Davies II, B.T. Kaufman, J. Kraut, Crystal structure of chicken liver dihydrofolate reductase complexed with NADP+ and biopterin, Biochemistry 31 (1992) 7264–7273.

[31] C. Oefner, A. D'Arcy, F.K. Winkler, Crystal structure of human dihydrofolate reductase complexed with folate, Eur. J. Biochem. 174 (1988) 377–385.

[32] J.T. Bolin, D.J. Filman, D.A. Matthews, R.C. Hamlin, J. Kraut, Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution. I. General features and binding of methotrexate, J. Biol. Chem. 257 (1982) 13650–13662.

[33] H.-J. Bohm, Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs, J. Comput. Aided Mol. Des. 12 (1998) 309–324.

[34] 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. 11 (2004) 3093–3118.