Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Structural Biology

Advances in all atom sampling methods for modeling protein–ligand binding affinities

Emilio Gallicchio and Ronald M Levy

Conformational dynamics plays a fundamental role in the regulation of molecular recognition processes. Conformational heterogeneity and entropy variations upon binding, although not always evident from the analysis of structural data, can substantially affect affinity and specificity. Computer modeling is able to provide some of the most direct insights into these aspects of molecular recognition. We review recent physics-based computational studies that employ advanced conformational sampling algorithms and effective potentials to model the three main classes of degrees of freedom relevant to the binding process: ligand positioning relative to the receptor, ligand and receptor internal reorganization, and hydration. Collectively these studies show that all of these elements are important for proper modeling of protein-ligand interactions.

Address
BioMaPS Institute for Quantitative Biology and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, United States

Corresponding author: Levy, Ronald M (ronlevy@lutece.rutgers.edu)

Current Opinion in Structural Biology 2011, **21**:161–166

This review comes from a themed issue on
Theory and simulation
Edited by Jeffrey Skolnick and Richard Friesner

Available online 19th February 2011

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2011.01.010

Introduction

Conformational dynamics is important aspect for both thermodynamic and kinetic aspects of ligand binding. The flexibility of both the ligand and the receptor, as well as the changes in flexibility upon binding, must be taken into account to correctly represent the thermodynamics of the binding reaction [1•]. Molecular recognition and biological activity is regulated by the conformational propensities of the binding partners in ways that are often not fully elucidated by the analysis of crystallographic structures alone [2]. Cytochrome P450 enzymes [3] represent a case in point in that their binding sites and substrates are very flexible, yet they act very specifically, and often crystal structures do not fully reflect the enzymatic mechanism [4].

In this review we focus on the role of conformational heterogeneity in binding free energy calculations[5]. We will focus in particular on approaches based on statistical mechanics [6•] that are limited, in principle, only by the accuracy of the force field and the extent of conformational sampling [7–9]. Approximate methods such as LIE and empirical scoring functions have been recently reviewed [10,11]. Ligand docking has been the focus of intense development and is nowadays a powerful tool for the search of candidate binding modes [12].

Conformational sampling algorithms applicable to binding free energy calculations must be able to not only find the relevant conformations of the system but also visit them with the correct probability. For example, it has been shown that strongly bound but rarely visited conformations of the complex are not necessarily the most relevant for the binding equilibrium [13••]. The need to consider both the bound and unbound states of the system further complicates the analysis [8]. Binding free energy protocols based on traditional sampling algorithms, such as Molecular Dynamics (MD) or Monte Carlo (MC), consider only a limited number of conformations and often suffer from poor reproducibility due presumably to slow convergence. Only relatively recently have enhanced sampling algorithms capable of equilibrating distinct conformational macrostates been applied to the protein–ligand binding problem.

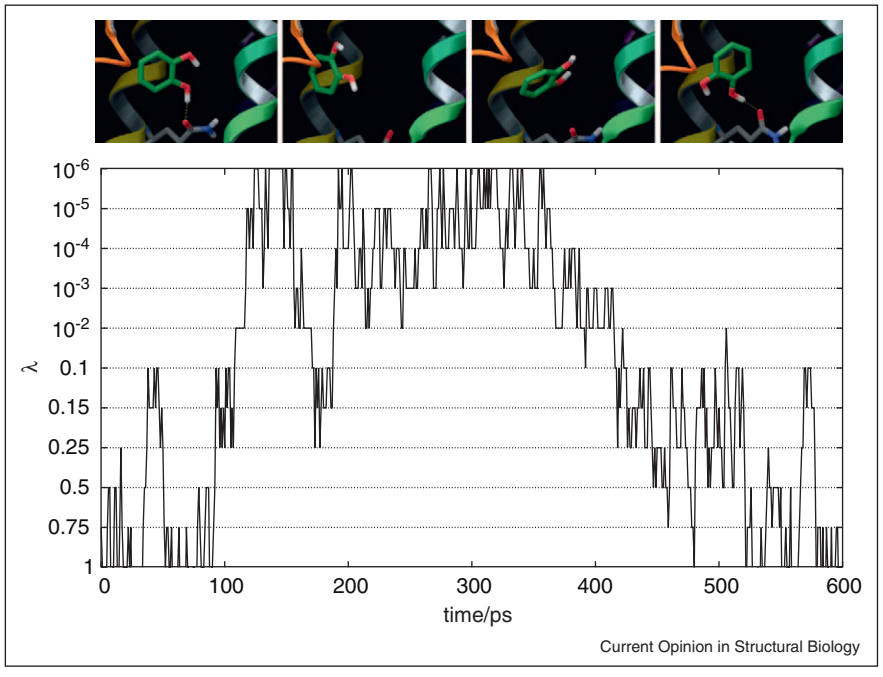
To simplify our view of the problem it is useful to consider the two main classes of degrees of freedom relevant to the molecular association process: the intermolecular motion of the ligand relative to the receptor, and the intramolecular dynamics of the ligand and the receptor [14]. In the following section we discuss enhanced free energy methods that attempt to address the former issue; recent studies of ligand and receptor reorganization, which pertain to changes in the distribution of intramolecular degrees of freedom induced by binding, are discussed next. Superimposed on these conformational degrees of freedom are the degrees of freedom of the solvent, some aspects of which will be discussed in closing. Unavoidably, any classification scheme such as this is imperfect. Most of the studies reviewed are not strictly limited to only one class of degrees of freedom. For example many of the free energy methods that we assign to the first class include some level of sampling of intramolecular coordinates. Nevertheless, in our view, the focus of the various methodologies is sufficiently distinct to make this classification useful.

Intermolecular degrees of freedom: positioning the ligand relative to the receptor

In this section we review binding free energy computational methods mainly focused on enhancing the sampling of the positioning of the ligand relative to the receptor. A majority of these protocols employ alchemical methods based on Molecular Dynamics (MD) or Monte Carlo (MC) sampling. Relative free energies are commonly computed by Free Energy Perturbation (FEP) and Thermodynamic Integration (TI) formulas [9], while double decoupling (DDM) [15], Potential of Mean Force (PMF) [16], and Binding Energy Distribution Analysis (BEDAM) [13••] methods are applicable to absolute binding free energies. Both classes of protocols are based on a thermodynamic path along a progress parameter λ which modulates ligand–environment interactions. Following recent progress in the development of efficient free energy estimators [17], the field is turning its attention to the conformational sampling problem [8] (while the assessment of potential function models necessarily depends on the binding free energy estimates being truly converged). Most studies to date are characterized by uncertain coverage of conformational space. These limitations are generally reflected in poor rates of convergence and hysteresis effects.

Among the algorithms applied to remedy these problems, some of the most promising are those based on generalized ensembles and parallel replica exchange (RE) protocols. Methods such as RETI [18], λ-dynamics [19•], FEP/REMD [20,21], and BEDAM[13••] have been shown to yield superior conformational sampling and more rapid convergence rates by allowing the simulation threads to migrate from one λ-window to another. Unlike traditional approaches, 'λ-hopping' moves improve mixing by allowing conformational transitions to occur at the value of λ at which they are most likely to occur and to be then propagated to other windows (see Figure 1). The use of reference states based on soft-core potentials has also been shown to enhance sampling of alternative binding modes [22,23]. Integration over part approaches [24,25,26••], which express the absolute binding free energy from an appropriate combination of contributions from each binding mode, is also attractive because it is far easier to perform complete sampling within a local macrostate than achieving equilibration between distinct binding modes. The challenge is to identify the collection of modes that contribute the most to the total binding free energy. Misidentification of the highest contributing mode can introduce major errors, while neglecting secondary modes can also affect accuracy [13••,25]. Improved convergence

Figure 1



Illustrative example of the sampling enhancement achieved by parallel RE sampling in λ space. The plot shows the value of λ assumed by a particular replica of the system as a function of time. In this example (taken from the BEDAM trajectory data of reference [13••] for catechol binding to L99A/M102Q T4-lysozyme) λ = 1 corresponds to full ligand–receptor interactions and λ = 0 to no ligand–receptor interactions. Above the plot, the structure frames corresponding to 0, 200, 400, and 600 ps simulation times are shown. The trajectory starts near λ = 1 with catechol forming a hydrogen bond (yellow line) with Gln102 of T4-lysozyme (first frame). Then the replica progressively assumes smaller values of λ and concurrently the ligand partially dissociates from the receptor (second and third frames). Subsequently the ligand reassociates with Gln102 in a different hydrogen bonding configuration when λ increases again (fourth frame).

can also be achieved by restricting the position of the ligand to a single macrostate [27]. Approaches of this kind are formally correct as the imposition of conformational restraints is matched by their release at a later stage of the calculation. The effects of multiple binding modes, however, are shifted to the restraining free energy component which, since it is computed with full receptor–ligand interactions, it is likely difficult to fully converge unless accelerated conformational sampling strategies are again employed to sample all important conformations. The role of the binding site volume, which is required by the theory [28], is, we think, another related aspect of some absolute binding free energy implementations that deserves further investigation.

The mining minima (MM) method [29] is an example of a binding free energy protocol which, not relying on MD or MC sampling, does not suffer from slow rates of transition typical of importance sampling algorithms. This advantage is, however, partially counterbalanced by the challenge of performing complete enumeration of all important stable minima of large receptor–ligand complexes.

Intramolecular degrees of freedom: ligand and receptor reorganization

Often substantial reorganization of the conformational ensembles of the binding partners occurs upon binding. This phenomenon, referred in the literature as conformational reorganization, induced fit, or conformational selection [30], corresponds to the free energy penalty associated with the unfavorable work for restraining ligand and receptor to their bound conformational states. For the purpose of this discussion reorganization refers to large conformational changes, such as changes in rotameric states and opening and closing of the receptor, rather than small readjustments that can be modeled with traditional sampling methods. Reorganization is nowadays recognized as a key factor to understand protein–protein recognition, as well as drug potency, specificity and resistance. Some protein–ligand binding free energy models include reorganization, explicitly or implicitly, as

$$\Delta F_{\mathrm{bind}} = \Delta F_{\mathrm{inter}} + \Delta F_{\mathrm{reorg}}, \quad (1)$$

where the second term represents the reorganization free energy and $\Delta F_{\mathrm{inter}}$ represents the term obtainable by the free energy methods described in the previous section. While most efforts have focused on the latter, the reorganization component can play an important role by, for example, regulating binding specificity. For congeneric series of complexes sharing the same binding mode it is often reasonable to assume that reorganization is not discriminating. However notable exceptions exist. For example, reorganization has been successfully used as the design principle for the optimization of the presentation of HIV epitopes for vaccine development [31]. Here, as well as for any other system where the nature of the binding interface is biologically constrained, preorganizing the ligand to the bound conformation is the only viable route for optimizing the binding affinity. In another recent example [32] optimization of a class of inhibitors was achieved by rigidification. In this case structural analysis indicated that enhanced binding was indeed solely due to smaller reorganization penalties.

The separation between energetic and reorganization terms is most evident in empirical docking scores [33] including ligand entropic reorganization contributions based, for example, on the number of ligand rotatable bonds. Binding free energy decompositions as in Eqn (1) are also the basis for rigorous free energy models [1•,8,29,34•]. These, however, are still in their infancy and are not widely employed in structure based drug design enterprises. Lack of progress in this area is due to the complexities of reorganization effects, which are inherently dynamical phenomena [14] requiring the knowledge of both a range of conformational states and their probability of occurrence in solution. Nevertheless, because these aspects of binding are very difficult to study experimentally, it is expected that computer modeling will eventually provide some of the most useful insights into the problem.

The definition of the reorganization free energy is not unique. As with any free energy decomposition scheme, individual contributions depend on the specific model under consideration. For example, the term corresponding to $\Delta F_{\mathrm{inter}}$ in Eqn (1) for dual trajectory MM-GB/PBSA models includes in principle enthalpic changes due to intramolecular interactions, while entropic reorganization effects are modeled separately [35,36]. The corresponding $\Delta F_{\mathrm{inter}}$ term for single trajectory MM-GB/PBSA schemes, however, does not include the intramolecular reorganization enthalpy and, consequently, both entropic and enthalpic terms are sometimes included in the $\Delta F_{\mathrm{reorg}}$ term [37]. Similarly, while it is often interpreted as either being only entropic or enthalpic in nature, in general reorganization causes a free energy change, with both enthalpic and entropic signatures [32].

A number of modeling studies have focused on ligand reorganization, which is simpler to model than receptor reorganization. A comprehensive analysis [38] had concluded that the majority of ligand structures in complexes deposited in the Protein Data Bank correspond to energetically strained ligand conformers in solution. Yang *et al.* [37] obtained significantly better correlation with experimental affinities when GBSA scores were augmented with a free energy term accounting for ligand reorganization. A thorough study by Gao *et al.* [34•] on 233 protein–ligand complexes confirmed the usefulness of including a model for the ligand reorganization free energy for binding affinity prediction. One issue apparent in this and other studies is the difficulty to clearly judge the benefits

of including ligand reorganization effects by comparing model predictions with experimental affinity data, given that its contribution is partly obfuscated by other approximations in the model, such as the omission of receptor reorganization effects. Given the relatively small size of conformational space, rigorous statistical mechanic approaches exist to model the ligand reorganization free energy. Approaches based on the mining minima (MM) framework [29,34•] assemble the reorganization free energy by directly computing the configurational partition functions of a set of low energy conformational states. It has been recently confirmed [39] that MD sampling aided by temperature RE can also be used to accurately compute conformational populations of ligand conformational states. Generalized-ensemble MD sampling is a promising approach given its generality and scaling properties.

Binding modeling studies explicitly incorporating receptor reorganization effects are beginning to appear. Major challenges exist due to the size of conformational space and the rarity of conformational transitions. Some recent studies have focused on the role of protein sidechain motion. Mobley *et al.* [40] have introduced a confine and release method to model the free energy associated with the conformation variability of a selected set of sidechains in the binding site region. In a number of cases it was shown that including these terms improved the accuracy of binding affinity predictions [26••]. Similarly, a two-dimensional Hamiltonian RE free energy perturbation approach has been proposed to soften sidechain torsional barriers [41].

Solvent interactions

Water plays a fundamental role in molecular association phenomena of biological interest. Water molecules often mediate protein–ligand interactions [42] and understanding water dynamics is important for rationalizing the physics of binding in solution [43]. While both explicit and implicit models of hydration have been employed in computational models of protein–ligand binding [7,13••,44], explicit models are considered better able to capture the detailed behavior of water molecules in binding sites. Modeling has revealed the complexity of the thermodynamics of water molecules in protein cavities [45,46]. Recent studies have focused on the free energy gain or loss corresponding to the work of displacing water molecules to accommodate the ligand [47], and the use of this knowledge in rational ligand design. Michel *et al.* [48•] have recently employed an algorithm [49] for water placement and scoring based on a double decoupling scheme [50] to compute the free energy for displacing key water molecules in the binding sites of scytalone dehydratase, p38-α MAP kinase, and EGFR. They found that water displacement can both favor or disfavor binding depending on the details of the structural and energetic properties of the ligand and the receptor. Young *et al.* [51] have recently completed a modeling survey of 'dry' active sites in proteins that, because they do not desolvate upon binding, are excellent receptors for hydrophobic ligands.

Significant advances have been recently documented in the application of the inhomogeneous fluid solvation theory developed by Themis Lazaridis to understand the thermodynamic contributions of bound water molecules [52]. The role of water molecule clusters in protein–ligand complexes with cyclophilin A [52] and Factor Xa [53••] has been analyzed. The theory has elucidated the origin of the extremely high binding affinity between streptavidin and avidin [54] in terms of a cluster of low entropy water molecules displaced upon association.

Slow diffusion and kinetic trapping of water molecules near the binding site lead to systematic errors and slow convergence of binding free energy calculations. To remedy this problem some recent studies [55•,56] employed a Grand Canonical Monte Carlo (GCMC) algorithm which explicitly promotes exchanges between bound water molecules and the bulk.

Concluding remarks

The mechanisms of molecular recognition are still poorly understood at a quantitative level. A companion review in this volume [57] summarizes the many challenges for robust predictions of protein–ligand binding affinities. In particular the field has developed an appreciation, which we have attempted to capture in this review, for the important role of conformational heterogeneity and conformational sampling. Although much remains to be done, the past few years have witnessed tremendous progress in the incorporation of multiple binding modes, and reorganization and solvation effects in binding free energy models.

Acknowledgements

This work has been supported in part by a research grant from the National Institute of Health (GM30580).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Baron R, McCammon JA: **(Thermo)dynamic role of receptor flexibility, entropy, and motional correlation in protein–ligand binding**. *ChemPhysChem* 2008, **9**:983-988.

Long molecular dynamic simulations of ligated and unligated forms of cytochrome C peroxidase are analyzed to investigate the receptor reorganization entropy contributions to the binding free energy. It is found that upon binding the protein suffers substantial conformational entropy loss due mostly to increased correlation among degrees of freedom.

2. Shi Y, Liu H, Gai D, Ma J, Chen XS: **A computational analysis of ATP binding of SV40 large tumor antigen helicase motor**. *PLoS Comput Biol* 2009, **5**:e1000514.

3. Cojocaru V, Winn PJ, Wade RC: **The ins and outs of cytochrome P450s**. *Biochim Biophys Acta* 2007, **1770**:390-401.

4. Ravindranathan KP, Gallicchio E, Friesner RA, McDermott AE, Levy RM: **Conformational equilibrium of cytochrome p450 bm-3 complexed with n-palmitoylglycine: a replica exchange molecular dynamics study**. *J Am Chem Soc* 2006, **128**:5786-5791.

5. Shirts MR, Mobley DL, Scott P: **Free-energy calculations in structure-based drug design**. *Brown Drug Design-Structure- and Ligand-Based Approaches*, Cambridge University Press; 2010:61–86.

6. • Zhou H-X, Gilson MK: **Theory of free energy and entropy in noncovalent binding**. *Chem Rev* 2009, **109**:4092-4107.
A comprehensive yet pedagogical review of the statistical thermodynamics of binding. The role of entropy and conformational decomposition is discussed in particular detail. An interesting critical analysis of traditional concepts such as functional group decomposition and ligand efficiency is provided.

7. Jorgensen WL: **Efficient drug lead discovery and optimization**. *Acc Chem Res* 2009, **42**:724-733.

8. Mobley DL, Dill KA: **Binding of small-molecule ligands to proteins: ''what you see'' is not always ''what you get''**. *Structure* 2009, **17**:489-498.

9. Michel J, Essex JW: **Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations**. *J Comput Aided Mol Des* 2010, **24**:639-658.

10. Gilson MK, Zhou H-X: **Calculation of protein–ligand binding affinities**. *Annu Rev Biophys Biomol Struct* 2007, **36**:21-42.

11. Guvench O, MacKerell AD: **Computational evaluation of protein–small molecule binding**. *Curr Opin Struct Biol* 2009, **19**:56-61.

12. Zhou Z, Felts AK, Friesner RA, Levy RM: **Comparative performance of several flexible docking programs and scoring functions: enrichment studies for a diverse set of pharmaceutically relevant targets**. *J Chem Inf Model* 2007, **47**:1599-1608.

13. •• Gallicchio E, Lapelosa M, Levy RM: **Binding energy distribution analysis method (BEDAM) for estimation of protein–ligand binding affinities**. *J Chem Theor Comput* 2010, **6**:2961-2977.
An absolute binding free energy protocol based on the analysis of probability distributions of the binding energy computed with implicit solvation, λ-hopping replica exchange MD sampling, and histogram reweighting. Results for complexes of T4-lysozyme indicate that conformations with a wide range of binding energies contribute significantly to the binding affinity depending on their probability of occupancy. The method takes fully into account the entropic cost of restricting the ligand within the binding site region.

14. Karplus M: **Dynamical aspects of molecular recognition**. *J Mol Recognit* 2010, **23**:102-104.

15. Gilson MK, Given JA, Bush BL, McCammon JA: **The statistical-thermodynamic basis for computation of binding affinities: a critical review**. *Biophys J* 1997, **72**:1047-1069.

16. Woo H-J, Roux B: **Calculation of absolute protein–ligand binding free energy from computer simulations**. *Proc Natl Acad Sci U S A* 2005, **102**:6825-6830.

17. Shirts MR, Chodera JD: **Statistically optimal analysis of samples from multiple equilibrium states**. *J Chem Phys* 2008, **129**:124105.

18. Woods CJ, Essex JW, King MA: **Enhanced configurational sampling in binding free-energy calculations**. *J Phys Chem B* 2003, **107**:13711-13718.

19. • Knight JL, Brooks CL: **Lambda-dynamics free energy simulation methods**. *J Comput Chem* 2009, **30**:1692-1700.
A comprehensive review of the λ-dynamics free energy methodology and its applications to date. In this extended Lagrangian method λ is a dynamical variable that simultaneously connects multiple ligand states. The dual topology version of the method is capable of effectively sampling multiple binding modes.

20. Jiang W, Hodoscek M, Roux B: **Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics**. *J Chem Theor Comput* 2009, **5**:2583-2588.

21. Wu Y, Lin T, Shelley JC, Sherman W: **Lambda hopping: an efficient replica exchange-based sampling method for relative free energy perturbation calculations**. In preparation.

22. Hritz J, Oostenbrink C: **Efficient free energy calculations for compounds with multiple stable conformations separated by high energy barriers**. *J Phys Chem B* 2009, **113**:12711-12720.

23. Christ CD, Mark AE, van Gunsteren WF: **Basic ingredients of free energy calculations: a review**. *J Comput Chem* 2010, **31**:1569-1582.

24. Jayachandran G, Shirts MR, Park S, Pande VS: **Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics**. *J Chem Phys* 2006, **125**:084901.

25. Mobley DL, Chodera JD, Dill KA: **On the use of orientational restraints and symmetry corrections in alchemical free energy calculations**. *J Chem Phys* 2006, **125**:084902.

26. •• Boyce SE, Mobley DL, Rocklin GJ, Graves AP, Dill KA, Shoichet BK: **Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site**. *J Mol Biol* 2009, **394**:747-763.
An extensive binding free energy study in which the accuracy of numerical prediction are tested experimentally in terms of both thermodynamic and structural properties. This study gives an unbiased perspective on the merits and challenges of some of the best available modeling techniques. It is shown that all of the main aspects of conformational sampling discussed in this review (ligand positioning, reorganization, and hydration) are important to achieve sufficient predictive ability.

27. Deng Y, Roux B: **Computations of standard binding free energies with molecular dynamics simulations**. *J Phys Chem B* 2009, **113**:2234-2246.

28. Mihailescu M, Gilson MK: **On the theory of noncovalent binding**. *Biophys J* 2004, **87**:23-36.

29. Chang CA, Chen W, Gilson MK: **Ligand configurational entropy and protein binding**. *Proc Natl Acad Sci U S A* 2007, **104**:1534-1539.

30. Csermely P, Palotai R, Nussinov R: **Induced fit, conformational selection and independent dynamic segments: an extended view of binding events**. *Trends Biochem Sci* 2010, **35**:539–546.

31. Lapelosa M, Arnold GF, Gallicchio E, Arnold E, Levy RM: **Antigenic characteristics of rhinovirus chimeras designed in silico for enhanced presentation of HIV-1 gp41 epitopes**. *J Mol Biol* 2010, **397**:752-766.

32. DeLorbe JE, Clements JH, Teresk MG, Benfield AP, Plake HR, Millspaugh LE, Martin SF: **Thermodynamic and structural effects of conformational constraints in protein–ligand interactions. Entropic paradoxy associated with ligand preorganization**. *J Am Chem Soc* 2009, **131**:16758-16770.

33. McInnes C: **Virtual screening strategies in drug discovery**. *Curr Opin Chem Biol* 2007, **11**:494-502.

34. • Gao C, Park M-S, Stern HA: **Accounting for ligand conformational restriction in calculations of protein–ligand binding affinities**. *Biophys J* 2010, **98**:901-910.
A GBSA scoring function for binding is augmented with a model based on quasi-harmonic/free energy perturbation analysis to account for the ligand reorganization free energy. The model is tested on 233 complexes from the PDBBind database resulting in slightly better correlation to measured binding affinities than two empirical scoring functions. A similar model accounting for only ligand reorganization enthalpy is shown to be less effective in this respect.

35. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W *et al.*: **Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models**. *Acc Chem Res* 2000, **33**:889-897.

36. Foloppe N, Hubbard R: **Towards predictive ligand design with free-energy based computational methods?** *Curr Med Chem* 2006, **13**:3583-3608.

37. Yang C-Y, Sun H, Chen J, Nikolovska-Coleska Z, Wang S: **Importance of ligand reorganization free energy in protein–ligand binding-affinity prediction**. *J Am Chem Soc* 2009, **131**:13709-13721.

38. Perola E, Charifson PS: **Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding**. *J Med Chem.* 2004, **47**:2499-2510.

39. Okumura H, Gallicchio E, Levy RM: **Conformational populations of ligand-sized molecules by replica exchange molecular dynamics and temperature reweighting**. *J Comput Chem* 2010, **31**:1357-1367.

40. Mobley DL, Chodera JD, Dill KA: **The confine-and-release method: obtaining correct binding free energies in the presence of protein conformational change**. *J Chem Theor Comput* 2007, **3**:1231-1235.

41. Jiang W, Roux B: **Free energy perturbation hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations**. *J Chem Theor Comput* 2010, **6**:2559-2565.

42. Lu Y, Wang R, Yang C-Y, Wang S: **Analysis of ligand-bound water molecules in high-resolution crystal structures of protein–ligand complexes**. *J Chem Inf Model* 2007, **47**:668-675.

43. Li Z, Lazaridis T: **Water at biomolecular binding interfaces**. *Phys Chem Chem Phys* 2007, **9**:573-581.

44. Michel, Julien ML, Verdonk JW, Essex: **Protein–ligand binding affinity predictions by implicit solvent simulations: a tool for lead optimization?** *J Med Chem* 2006, **49**:7427-7439.

45. Yu H, Rick SW: **Free energies and entropies of water molecules at the inhibitor–protein interface of dna gyrase**. *J Am Chem Soc* 2009, **131**:6608-6613.

46. Baron R, Piotr Setny J, McCammon, Andrew: **Water in cavity–ligand recognition**. *J Am Chem Soc* 2010, **132**:12091-12097.

47. Barillari C, Taylor J, Viner R, Essex JW: **Classification of water molecules in protein binding sites**. *J Am Chem Soc* 2007, **129**:2577-2587.

48. • Michel J, Tirado-Rives J, Jorgensen WL: **Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization**. *J Am Chem Soc* 2009, **131**:15403-15411.
A thermodynamic cycle is employed to study the thermodynamic effect of displacing water molecules from the binding site. The procedure explicitly included the alchemical processes of transforming one ligand into another with or without a key water molecule present and, in addition, the process of transfer of the water molecule from the binding site to the bulk by double decoupling. While simple steric consideration rationalize some of the results, it is shown that often water displacement favors or disfavors binding depending on the subtleties of specific molecular interactions.

49. Michel J, Tirado-Rives J, Jorgensen WL: **Prediction of the water content in protein binding sites**. *J Phys Chem B* 2009, **113**:13337-13346.

50. Hamelberg D, McCammon AJ: **Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method**. *J Am Chem Soc* 2004, **126**:7683-7689.

51. Young T, Hua L, Huang X, Abel R, Friesner R, Berne BJ: **Dewetting transitions in protein cavities**. *Proteins* 2010, **78**:1856-1869.

52. Li Z, Lazaridis T: **Thermodynamics of buried water clusters at a protein–ligand binding interface**. *J Phys Chem B* 2006, **110**:1464-1475.

53. •• Robert A, Young T, Farid R, Berne BJ, Friesner RA: **Role of the active-site solvent in the thermodynamics of factor Xa ligand binding**. *J Am Chem Soc* 2008, **130**:2817-2831.
A map of water occupancy and chemical potential for the fXa active site, obtained by means of the inhomogeneous theory of solvation, is used to rationalize the relative binding affinities of a congeneric series of ligands whose binding is believed to differ mainly on the nature of the bound water molecules they displace. The study highlights the influence of water displacement and release on the thermodynamics of binding.

54. Young T, Abel R, Kim B, Berne BJ, Friesner RA: **Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding**. *Proc Natl Acad Sci U S A* 2007, **104**:808-813.

55. • Deng Y, Roux B: **Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations**. *J Chem Phys* 2008, **128**:115103.
This study introduces the use of Grand Canonical Monte Carlo (GCMC) in absolute binding free energies calculations to enhance the equilibration of water molecules between the binding site and the bulk. The methodology is illustrated for camphor binding to a P450 enzyme, where a number of deeply bound water molecules is expelled upon binding.

56. Ge X, Roux B: **Absolute binding free energy calculations of sparsomycin analogs to the bacterial ribosome**. *J Phys Chem B* 2010, **114**:9525-9539.

57. Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS: **Alchemical free energy methods for drug discovery: progress and challenges**. *Curr Opin Struct Biol*, (2011), doi:10.1016/j.sbi.2011.01.011.