Annu. Rev. Biophys. Biomol. Struct. 2001. 30:211–43

BIOMOLECULAR SIMULATIONS: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions

Wei Wang[1], Oreola Donini[2,3], Carolina M. Reyes[2], Peter A. Kollman[1,2]

[1]Graduate Group in Biophysics, [2]Department of Pharmaceutical Chemistry, University of California San Francisco, California 94143; e-mail: wwang@cgl.ucsf.edu; creyes@cgl.ucsf.edu; pak@cgl.ucsf.edu

[3]Kinetek Pharmaceuticals Inc., Vancouver, Canada V6P 6P2; e-mail: odonini@kintekpharm.com

Key Words molecular dynamics, molecular recognition, electronic structural calculations

■ **Abstract** Computer modeling has been developed and widely applied in studying molecules of biological interest. The force field is the cornerstone of computer simulations, and many force fields have been developed and successfully applied in these simulations. Two interesting areas are (*a*) studying enzyme catalytic mechanisms using a combination of quantum mechanics and molecular mechanics, and (*b*) studying macromolecular dynamics and interactions using molecular dynamics (MD) and free energy (FE) calculation methods. Enzyme catalysis involves forming and breaking of covalent bonds and requires the use of quantum mechanics. Noncovalent interactions appear ubiquitously in biology, but here we confine ourselves to review only noncovalent interactions between protein and protein, protein and ligand, and protein and nucleic acids.

CONTENTS

INTRODUCTION

This article reviews some applications of molecular modeling to biologically interesting molecules. The review is focused on three topics: (*a*) We review molecular mechanical (MM) force fields, which form the cornerstone of all simulations of complex chemical systems, including the ones considered in the other two topics; (*b*) we review the simultaneous use of both quantum mechanics and molecular mechanics calculations to study enzyme-catalyzed reactions; (*c*) we review simulation methods for studying noncovalent interactions involving biological molecules and applications to protein-ligand, protein-protein, and protein–nucleic acid interactions. Although these three areas are only a small fraction of the exciting work in modeling of biological molecules, each has been an exciting and vital research area in the last few years and thus is deserving of a review. We focus on papers that have been published in the last five years.

DEVELOPMENTS IN MOLECULAR MECHANICAL FORCE FIELD FOR BIOMOLECULAR SIMULATIONS

The cornerstone of all simulations is the MM force field. Such force fields have been developed for many types of molecules, but our review concentrates on those that have been extensively applied to biomolecules. A characteristic of all current force fields extensively applied to biopolymers to date is that they are two-body additive. This means that the potential energy function (Equation 1 as example) is a function of pairs of atoms.

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

In fact, such models allow one to include many body effects implicitly in the parameterization. Most biological molecules are surrounded by water, and

potential functions for water liquid have been under development for more than 30 years. The water models most currently used were developed in the 1980s and go by the names TIP3P, TIP4P (71), SPC, and SPC/E (16). The parameters in all four models are empirically adjusted so that they reproduce the enthalpy of vaporization and the density of water. They all do this very well, although the SPC/E model overshoots the enthalpy of vaporization deliberately and then empirically adds the energy needed to polarize the charge distribution from the gas phase into solution to make the net enthalpy of vaporization agree with experiment. None of those models does a good job of describing the temperature dependence of the density of water. Thus, Jorgensen recently developed a TIP5P model (97) that does this much better. All of these models have a dipole moment of approximately 2.3 Debye, compared to the gas phase water value of 1.85 Debye.

Continuing development of nonadditive models for water better reproduces the gas phase properties of the monomer and, by including polarization explicitly, attempts with a single model to reproduce the gas, liquid, and solid structure and thermodynamics of water (30, 133). A full analysis of such models is beyond the scope of this review, but there are as yet no protein or nucleic acid force fields fully developed with such an approach, although there will likely be some in the future. These nonadditive models are considerably (2–5 times) more computationally expensive than effective two-body additive models, and thus, given that the main bottleneck in application of simulations to biological molecules is usually the sampling problem, one usually does not want to accept the extra computational cost.

A number of force fields can simulate proteins. In alphabetical order, they include the force field developed within AMBER by Cornell et al (parm94) (27) and small modifications of the torsional energies [parm96 (74) and parm99 (164)]; the CHARMm force fields, the latest publication of which is by MacKerell and coworkers (95), the force field developed at BIOSYM (now MSI) called CVFF (46); the force field described by Daggett et al within Levitt's simulation program ENCAD (89); the GROMOS96 force field (153); and the OPLS-AA force field (72). In addition, the MMFF force field is aimed at simulating protein-ligand complexes (57–59). All but CVFF have a force field equation like that described in Equation 1. Subtle differences exist regarding how improper (out of plane) torsions are described, whether Urey-Bradley (1–3) interactions are included, or if scale factor(s) are used for nonbonded interactions between atoms separated by exactly 3 bonds. CVFF has a more complex functional form than that in Equation 1. All of these force fields use empirical fits to liquid- or solid-state small molecular systems in order to derive the van der Waals parameters. This ensures reasonable densities for the simulated systems. These force fields attempt to make their electrostatic interactions well-balanced with the water models noted above by using larger partial charges than would be found in a gas phase residue. All but Cornell et al attempt to do this empirically; Cornell et al use ab initio Hartree Fock calculations with a 6-31G* basis set to derive the electrostatic potential for residues or molecules and fit the partial charges to optimally reproduce the quantum mechanical (QM) potential. The reason for using this model is that it systematically overestimates

the polarity of the residues by a similar but somewhat smaller magnitude than do the standard water models.

The various models use some combination of empirical data and QM calculations to fit the torsional potentials in Equation 1. It is probably fair to say that various force fields are converging in their philosophy and parameters; twenty years ago, there were a number of proponents of deriving force field parameters strictly from QM calculations, but that was then and is even now quite impractical. This is because of the need for very high accuracy, inclusion of many body effects, and the fact that one has to go to a very accurate QM level to correctly represent dispersion attraction, a key attractive term in Equation 1.

A number of force fields have been developed that can be used for carbohydrates and membranes. The concepts/philosophy for the development are similar to those used in protein simulations, but often the details differ. In the area of carbohydrates, Brady and coworkers and Brickmann and coworkers have developed force fields within CHARMm (109, 128); Woods et al and Simmerling et al within AMBER (114, 144); and Damm et al within OPLS (29). For membranes, MacKerell et al have developed parameters within CHARMm (48); Berkowitz et al have modified parameters from AMBER models (147); and Berendsen and coworkers have developed models within GROMOS (158). Again, we emphasize that most of these parameter sets are quite transferable between the simulation programs, provided that in some cases the code is modified, because of the subtle differences in improper torsions, etc. mentioned for protein force fields above. Nonetheless, the force fields noted here for carbohydrates and membranes essentially use Equation 1 as a model.

A number of force fields have been developed for nucleic acids. The three major efforts are those by Cornell et al [parm94 (27) and parm98 (21)], MacKerell et al [CHARMm22 (96) and CHARMm27 (50)], and Langley (85) with a few applications using GROMOS (135) to study nucleic acids. These force fields have been reviewed by Cheatham & Kollman (22); they are not reviewed in detail here. In the case of nucleic acids, MD simulations with cutoff approaches have been shown to be unstable. Thus, one must use methods that more accurately treat the long-range electrostatic interactions, either Ewald-like methods [PME (45) or PPPM (13)] or simpler methods like force shift (20).

We close this section by briefly noting the efforts to develop nonadditive models for biomolecular systems. To our knowledge, such developments are being carried out by Friesner, Berne, and coworkers (152), CL Brooks (unpublished results), P Cieplak, J Caldwell, PA Kollman et al (unpublished results), and JW Ponder (unpublished results). There may be other efforts as well. It is clear that the development of the more physically correct nonadditive models is quite important because they should be more accurate at describing cation-pi interactions and processes at nonpolar-polar interfaces, whereas the additive models have to assume a charge distribution that is either water- or lipid-like. Nonetheless, additive models have proved surprisingly accurate for many phenomena, and since the sampling problem appears more daunting for simulations of biomolecular systems, one does

not want to spend the extra computer time required for a more accurate energy model. There are some systems such as simple liquids and rigid ionophores where the sampling problem is not so severe, and these have proven to be excellent testing grounds for nonadditive models.

MACROMOLECULAR MODELING OF REACTION MECHANISMS

Macromolecular modeling inevitably generates a requirement for model bond breaking and forming processes, especially when one considers the role of enzymes in catalysis. It is unfortunate that traditional QM methods, which excel at the treatment of such processes, are too computer intensive to accurately treat systems as large as proteins. It is only in the last two decades that methods have evolved to treat bond making and breaking processes at the QM level while still accounting for the environment of the protein. These methods, pioneered by Warshel & Levitt (170), have been continuously improving their ability to describe proteins as a result of both methodological advances and increasing computer power.

The underlying physical features of proteins that allow them to be such perfect catalysts have been the subject of both study and controversy. Unfortunately, while experimental methods can clearly delineate the efficiency of a given enzyme as a catalyst, it is much more difficult to pinpoint the exact physical cause of the efficiency (e.g., electrostatic pre-organization, transition state stabilization, etc.; see reference 36 for a recent discussion). Computer simulations, however, make it possible to dissociate the energetic contributions of residues, and even atoms, to the enzymatic reaction mechanism to a finer degree than even mutational analysis. Thus, as our power to investigate protein reactions has increased, so has our inherent understanding of protein mechanisms.

Theory

In order to adequately describe a macromolecular reaction, the simulation must encompass both QM descriptions of the bonds being broken and formed, and the effect of the protein environment. This requirement has led to the development of the so-called QM/MM methods. Unfortunately, this terminology has grown to encompass several methodologies and assumptions leading to some confusion in the literature. Here, we refer to the class of methodologies that inherently consider a protein reaction and its environment as QM-coupled-MM methods. These methodologies can be broadly classified into two major subdomains: those that structurally couple the QM and MM domains and those that thermodynamically couple them. A structural link indicates that the QM and MM energetic contributions are calculated simultaneously. That is, for every set of nuclear coordinates, $R_{QM} + R_{MM}$, the energies of each subset and their interaction are recalculated. In the thermodynamic linkage, the initial and final states of the reaction are

determined by quantum mechanics. The energetic difference between these states, due to the presence of the protein, is then considered.

The QM/MM methodology is a structural QM-coupled-MM technique by our terminology. The treatment of the boundary region between those atoms in the active site treated by QM methods and those outside the active site treated by classical MM methods has been a major methodological challenge in the structural QM-coupled-MM methods. QM calculations require a complete valence on each atom to calculate the electronic distribution. Protein active site residues are covalently bound to the rest of the protein. Thus, a boundary between X_{QM} and Y_{MM} spans a covalent bond. Initial approaches used "link atoms" (145) that were added to the QM calculation to satisfy the valence requirements. These link atoms could either be fixed relative to the active site (in which case they did not reflect any geometrical changes happening at the QM/MM interface) or free to move (in which case extra degrees of freedom were introduced into the simulation). The presence of the link atom also caused numerical problems in the calculations of the forces, and often electrostatic interactions across the boundary region were neglected because of the large forces incurred (129).

More recently, the link atom problem has been addressed by the frozen orbital (LSCF) approach (105, 155) and a related method known as generalized hybrid orbital (GHO) (52). In the LCSF method, the valence of the QM boundary atom is satisfied with a single strictly localized bond orbital calibrated from a small model system (105, 155, 170). GHO represents a further extension of the LSCF approach in which the boundary MM atom is described by four hybrid orbitals. The hybrid orbitals that would attach to other MM atoms are fixed and form an effective core potential. The remaining orbital, representing the bond to a QM atom, is optimized in the QM calculation. This additional flexibility removes the requirements for individual calibrations on small model systems once appropriate hybrid orbital parameters are developed (52). The link atom and LSCF approaches have been nicely summarized and tested by Karplus et al (129) and by Antes & Thiel (6).

At the other end of the spectrum, quantum mechanics/free energy (QM/FE) calculations represent the class of thermodynamic QM-coupled-MM methods. This method was inspired by the solution phase calculations of organic reactions by Jorgensen et al (105, 150). In this case, the active site residues are again isolated from the remainder of the protein, and broken valences are satisfied with a link hydrogen atom. Using the relative orientation of the residues from the active site, the QM atoms are optimized in the gas phase for both the reactants and the products. The QM atoms are then re-inserted into the protein, and the free energy difference between the reactant and the product (or transition state), due to the presence of the protein, is determined. The free energy of the reaction is therefore

$$\Delta G = \Delta E_{QM} + G_{MM} - T\Delta S_{QM}, \qquad (2)$$

where $T\Delta S_{QM}$ represents the entropy change in the QM atoms. The reference reaction for this calculation is the same reaction in water. This gives rise to two

predicted values: the reaction profile in water and in the protein. (See references 44 and 45 for recent reviews of this method.)

The empirical valence bond (EVB) method couples aspects of both structural and thermodynamic linkage. The EVB description of the reactants and products (or transition state) is calibrated to reproduce the known solution phase reaction energetics assuming a similar mechanism in water. These initial and final QM states then represent the end points of the free energy simulation of the system as a whole. At each geometrical configuration of the system, the QM portion of the energy is calculated as the off-diagonal terms of the Hamiltonians representing the given initial and final states (98, 168).

These methodologies have all been used successfully in the last few years and are giving rise to a consensus as to the role of proteins in enzymatic reactions. Nonetheless, there are a number of advantages and disadvantages to each of them. EVB methods are generally limited to a semi-empirical description of the QM states, particularly with the added use of MD in the free energy calculation, although there has been an attempt to calibrate the results with higher levels of QM theory (15).

QM/MM methods have traditionally been limited to minimizations, as opposed to dynamics and free energy calculations. Recently, there has been a QM/MM free energy calculation in which the points on the reaction path are calculated with minimizations and then, reminiscent of QM/FE approaches, the free energies between these points, due to the MM atoms, are calculated (176). Further developments of the QM/MM approach include the possibility of using Hartree-Fock or density functional theory levels of QM theory, given sufficient computational resources (93).

QM/FE calculations are explicitly used to calculate the free energy and can do so including the full range of MD simulation parameters [e.g., the Particle mesh Ewald (PME) method for long-range electrostatic treatment], which would currently be beyond the feasible computational limits of QM/MM, for example. However, the calculation is based on the assumption that the protein changes the relative stability of the reactant and product (or transition state) but not the identity of the transition state because these are determined in the gas phase QM calculation (75, 88). Presently, QM/FE calculations can incorporate a higher level of QM theory but have the disadvantage that they do not allow full coupling between the QM atoms and the environment.

Applications

Current opinions as to how enzymes catalyze reactions are changing. Thus, Warshel suggests that it is electrostatic effects that explain enzyme catalysis, particularly pre-organization of the enzyme to stabilize the transition state (169). The importance of electrostatic effects is becoming clear in the QM-coupled-MM applications to date although a complete understanding (or agreement) of protein catalysis has certainly not been attained.

In the past few years, there have been a number of macromolecular simulations involving enzyme catalysis and the use of quantum mechanics. In this review, we consider only calculations that are macromolecular (that is, we do not consider calculations on small model systems) and employ a coupling of QM and MM techniques (as opposed to those calculations that derive force field parameters solely from QM calculations).

The simplest (in terms of coupling) QM-coupled-MM calculations treat the MM part of the system as fixed atoms in the field of the QM atoms (7, 33, 102, 117, 139). Representative calculations involve phospholipase A2 (139), serine proteases (7, 33), glucose oxidase (102), and aspartylglucosaminidase (117). In all but one case (33), the transition state stabilization was due to electrostatic effects. In the remaining cases, Carr-Parinello MD was carried out on the QM atoms alone for very short time periods, and the catalytic effect was cited as due to a low-barrier hydrogen bond of a largely covalent nature.

Recent QM/MM calculations in the literature (3, 28, 92, 94, 107, 108, 117, 122) involve either minimization or minimization and dynamics techniques. Calculations on chorismate mutase (94), L-lactate dehydrogenase (122), and citrate synthase (107, 108) utilizing minimization only have been reported. The chorismate mutase system does not involve a covalent bond between the enzyme and substrate, although a range of simulations were carried out in which the QM portion varied from the substrate alone to the substrate and up to three neighboring residues. The catalytic effect was found to be due to both substrate strain and transition state stabilization as a result of electrostatic interactions with nearby arginine and glutamate residues. In the case of L-lactate dehydrogenase, the inclusion of the enzyme environment significantly changed the reaction profile as opposed to earlier QM supermolecule calculations (121, 122), although specific decomposition of the energetic contributions was not attempted. Finally, the citrate synthase calculation compared the energy of two competing intermediates (an enolate and an enol) traditionally suggested for the mechanism and found the enolate to be the most stabilized owing to the formation of local (not low-barrier) hydrogen bonds with an active site histidine and water (107, 108).

Systems on which more extensive MD simulations have been carried out with the QM/MM methodology include HIV protease (92), low molecular weight protein tyrosine phosphorylase (PTP) (3), dihydrofolate reductase (DHFR) (28), and orotidine monophosphate decarboxylase (171). In the earliest of these calculations involving HIV protease, limited 20 ps MD calculations were performed for a variety of points in the reaction profile yielding qualitative upper limits on the free energy differences. These simulations concluded that a general acid-base mechanism was reasonable and proposed an alternate position for the lytic water as opposed to that derived from inhibitor orientations in the crystal structures. No comment was made about the method of enzyme catalysis as compared to solution reactions (92). In the case of low molecular weight PTP, optimal alignment of both the reacting residues in the protein and dielectric screening by the protein promoted the attack of the anionic Cys12 in the protein reaction compared to the solution

one. (3). In the DHFR calculation, long-range electrostatics were not a major contributor to the enzymatic effect, but rather a local hydrogen bond was described as the major determinant (28). Finally, in the case of orotidine decarboxylase, free energy simulations resulted in quantitative agreement with experimental values for the solution and enzymatic reaction barriers. The catalytic advantage of the enzyme was attributed to electrostatic destabilization of the ground state as well as transition state stabilization. Nonetheless, electrostatic pre-organization of the active site was of paramount importance to the enzyme reaction (171).

EVB/FE calculations have also been applied to study a range of enzyme reactions including low molecular weight PTP (77, 78), glyoxylase I (47), and ribonuclease A (53). In each case, the catalytic effect was attributed to electrostatic pre-organization in the enzymatic reaction. Furthermore, mechanistic considerations were addressed for each system. For ribonuclease A, both a dianionic and a monoanionic intermediate were compared, but they could not be distinguished given the uncertainties in the calculated free energies. However, in the glyoxylase I system, the enolate was the preferred intermediate, whereas in the PTP mechanism, the catalytic cysteine residue was proposed to exist as the anionic form.

Finally, applications of QM/FE calculations can also be found. This method was introduced in 1998, with an application to trypsin (151). The major catalytic advantage of this enzyme over the solution reaction was the pre-organization free energy term (75, 88). That is, accurate results were obtained only when the free energy to assemble the reacting groups in solution was accounted for. A further example involving catechol O-methyl transferase similarly found pre-organization free energy terms to be important, whereas electrostatic terms were less so because the reaction resulted in the annihilation of charge in the transition state (83, 88). In fact for this system, both the enzyme and the solution reactions destabilized the gas phase reaction, although the enzyme did so to a lesser extent. Finally, calculations of the citrate synthase system comparing the relative stability of the enolate and enol intermediates have also been carried out (34, 75). In this case, the simulations were extremely sensitive to the treatment of the long-range electrostatic interactions, which suggests the importance of electrostatic interactions in the enzymatic mechanism. These results involving both the pre-organization energy term and the electrostatic dependence are reminiscent of Warshel's concept of electrostatic pre-organization (169).

Unfortunately, it is not always straightforward to compare methodologies because the same enzyme systems are not usually considered. To date, low molecular weight PTP is the only fully published example (3, 47, 78). Nonetheless, in the evaluation of any QM-coupled-MM methodology, it must be remembered that the traditional limitations of both QM and MM methods are still relevant. Thus, for any QM calculation, both the level of QM theory and the determination of the reaction mechanism in the gas phase can impact the quality of the results. On the MM side, consideration of the treatment of long-range electrostatics is, as always, of great importance (137). However, the advent of QM-coupled-MM techniques and their extension to free energy calculations make it possible to compare

theoretical results to experimental ones and to further consider the effects of some of these assumptions.

The exact role of the protein in a given enzymatic reaction is usually still a point of debate. Although the role of electrostatics is cited or apparent in many studies, it is not universally seen to be the only major contributor. This difference may be due to the particular enzyme or the particular method of calculation. Only further studies and comparisons between methods and with experiment will clarify these issues. Nonetheless, the emerging QM-coupled-MM methods are exciting tools with which to address the question of enzyme catalysis.

SIMULATIONS ON PROTEIN-PROTEIN, PROTEIN-LIGAND, PROTEIN-DNA, AND PROTEIN-RNA INTERACTIONS

Interactions between proteins and their substrates play central roles in many biological processes such as signal transduction, enzyme cooperativity, and metabolic reactions. With more and more complex structures solved, structure-based computational modeling has become a powerful tool to understand and predict binding. In this section, we focus on noncovalent binding and only discuss different methods to estimate absolute or relative binding free energies for protein-protein or protein-ligand interactions.

Molecular dynamics (MD) and Monte Carlo (MC) methods have provided dynamic and atomic insights to understand complicated biological systems. Free energy perturbation (FEP) and thermodynamic integration (TI) methods have been successfully applied to predict the binding strength of a complex (17, 73, 161). Nonetheless, these methods are computationally intensive. Thus, many techniques such as the λ-dynamics and the chemical Monte Carlo/molecular dynamics (CMC/MD) method have been developed to improve their efficiencies, and many other less rigorous methods have also been in development to estimate binding free energies quickly but with reasonable accuracy (2). Among them, we review the linear interaction energy (LIE) method, the molecular mechanics/Poisson Boltzmann surface area (MM/PBSA) method, the pictorial representation of free energy components (PROFEC), the one-window free energy grid (OWFEG) method, and their applications to study protein-protein or protein-ligand binding.

Free Energy Perturbation and Thermodynamic Integration Methods

Free energy perturbation (FEP) and thermodynamic integration (TI) methods are the most rigorous methods among those currently available for calculating free energies. In this section, we focus on the applications of these two methods to protein-ligand or protein-protein complexes.

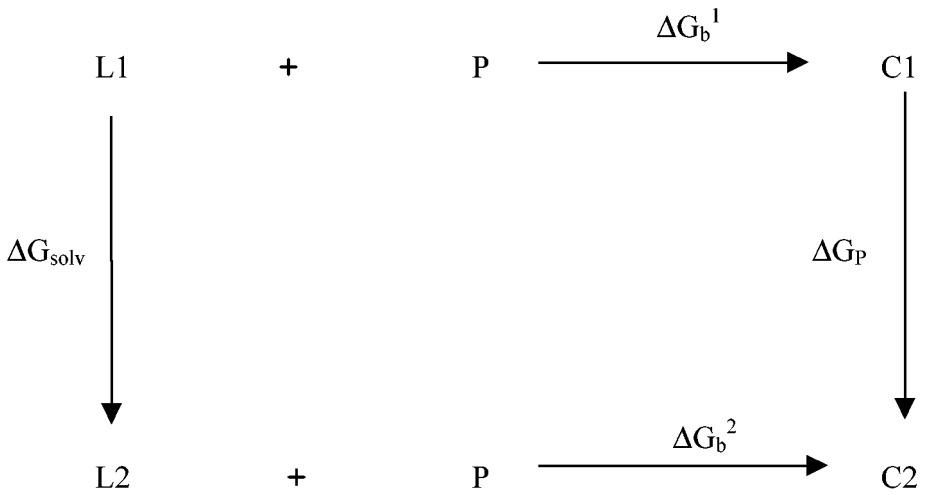

Figure 1 Thermodynamic cycle for calculating relative binding free energies between two ligands bound to the same protein.

Suppose that one wishes to calculate the binding free energy difference between two ligands bound to the same protein (the thermodynamic cycle is shown in Figure 1). Thus,

$$\Delta\Delta G = \Delta G_b^1 - \Delta G_b^2 = \Delta G_{solv} - \Delta G_P, \tag{3}$$

where ΔG_b^1 and ΔG_b^2 are binding free energies for ligand 1 and 2 respectively, and ΔG_{solv} and ΔG_P are nonphysical transmutation free energy from ligand 1 to ligand 2 in free and bound states. If ligand 1 and 2 are similar to each other, ΔG_{solv} and ΔG_P usually are easier to calculate than ΔG_b^1 and ΔG_b^2 because the mutation from ligand 1 to ligand 2 is assumed to cause only localized changes. FEP or TI is used to calculate ΔG_{solv} and ΔG_P. Equation 4 is used in FEP calculations.

$$\Delta G = -RT \sum_{i=1}^{N-1} \ln \left\langle \exp\left(-\frac{H(\lambda_{i+1}) - H(\lambda_i)}{RT} \right) \right\rangle_{\lambda_i}, \tag{4}$$

where ΔG is the free energy difference between two states, A and B. λ_i varies from 0 (state A) to 1 (state B), $H(\lambda_i)$ represents Hamiltonian of the system at λ_i, and $\langle\rangle_{\lambda_i}$ indicates an ensemble average. With the TI method, one calculates the average of derivatives of Hamiltonian at each λ, $H(\lambda)$, and then uses numerical integration over λ to calculate the free energy difference between two states (Equation 5), where λ has the same meaning as in FEP.

$$\Delta G = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle d\lambda. \tag{5}$$

In this section, we focus on the progress and applications in the last five years in the field of protein-ligand, protein-protein, and protein-nucleic acid interactions. Older papers can be found in previous reviews (70, 100).

Free energy calculations combined with MD or MC methods can provide rationale and insights for experimental observations and can suggest new experiments.

Jorgensen and coworkers have been applying MC and free energy calculations to study hydration free energies of organic molecules and binding free energies for ligand-protein complexes. Essex et al successfully applied MC and FEP to calculate accurate relative binding free energies for trypsin-benzamidine complexes (44). Their simulation was able to predict the strongest inhibitor among the four trypsin inhibitors. They also showed that changes of hydrogen bonding could not rationalize the calculated free energies. Instead, the relative binding affinities were justified in terms of bulk-solvation arguments whereby the more polar inhibitors had weaker binding affinities. This study is an excellent example of combining MC and FEP to study macromolecular systems.

Rastelli et al exploited FEP simulations to rationalize the binding differences between a benzocinnolinone carboxylic acid inhibitor of aldose reductase and its methoxylated analogs in four selected substitution sites (126). They were able to reproduce the experimental trend of binding. The four substitution sites were at the interface between protein residues and water. Thus, the perturbation involved only partial desolvation. This work sheds light on how to design new inhibitors targeting sites in the protein-water interface.

Fox et al (51) calculated the relative binding free energies between two transition state analog substrates of the catalytic antibody 17E8 using TI. The two substrates differ only in one side chain. The substitution of the -CH2- group to-S- leads to a 0.9 – 1.3 kcal/mol less favorable binding free energy. Their calculations showed that this preference for the -CH2- group over the-S- group was mainly due to the more favorable solvation free energy in the unbound form of the substrates. Free energy component analysis of the van der Waals and electrostatic contributions to the binding free energy indicated that these two terms contributed equally in solvent, whereas in the antibody, the van der Waals term clearly dominated. Several residues with large contributions to the binding were reported, and new site-specific mutagenesis experiments were suggested to test the calculated results.

Combined with MD and other simulation methods, free energy calculations can help determine some properties of the biological systems such as binding mode, protonation state of certain residues, and the flexibility of certain parts of the molecule. Several MD and FEP simulations have been performed to study carbohydrate-protein complexes (90, 114, 175). Zacharias et al (175) applied FEP simulations to study the differential binding between arabinose and fucose with arabinose binding protein (ABP). Liang et al (90) reported MD and FEP simulation results on binding of mannose versus galactose with a mannose binding protein (MBP). Recently, Pathiaseril & Woods (114) examined binding between analogs of the wild-type trisaccharide epitope of *Salmonella* serotype B and a fragment of the monoclonal anti-*Salmonella*, antibody Se155-4. All of these simulations obtained

free energy results consistent with experimental data. Pathiaseril & Woods (114) were able to reproduce the relative nuclear Overhauser effect (NOE) intensities for the wild-type ligand in solution and intermolecular hydrogen bond patterns in the complex from their MD trajectories. Their simulations showed that the free oligosaccharide oscillated around well-defined average glycosidic torsion angles, and the bound conformation was encompassed within those observed for the free ligand. Their free energy calculations also suggested that HIS-97 was diprotonated in the antibody. The insights obtained from these theoretical studies can be employed in the design of new ligands with higher binding affinity.

Sotriffer et al (148) showed a good example of how theoretical work could be carried out without direct experimental data, and they were able to provide useful information for designing new ligands. Starting from an antibody structure obtained by homology modeling (37), Sotriffer et al performed extensive docking searches to identify two pockets, S1 and S2, in the antibody IgE LB4 as the most probable binding site for three dinitrophenyl (DNP) amino acids (DNP-alanine, DNP-glycine, and DNP-serine). MD and FEP simulations were carried out for complexation on both pockets. A closed thermodynamic cycle was formed by transmutation between the three DNP amino acids (DNP-Ser $\rightarrow$ DNP-Ala $\rightarrow$ DNP-Gly $\rightarrow$ DNP-Ser), and the FEP calculations were validated by small closure errors of this cycle. The experimental free energy differences could only be reproduced for ligands binding to the S1 site. Analysis of the MD trajectory showed that the S1 complexes were characterized by a uniform binding mode, whereas ligand binding in the S2 site exhibited considerable variability. The authors concluded that the S1 site was expected to be the "real" binding site of those DNP amino acids. These theoretical predictions can be examined by crystallography or NMR experiments.

HIV-1 protease has been a therapeutic target for five FDA-approved AIDS drugs. Many free energy calculations have been performed on different inhibitors binding with the protease (25, 49, 123–125, 127, 160). Rao & Murcko have calculated relative binding free energies between HIV protease inhibitor A74704 and its diester analog (124). The diester analog inhibitor missed two hydrogen bonds with the protease active site, but its binding affinity was only tenfold weaker. They observed that Gly27 and Gly27′ loops were flexible, and thus, the hydrogen bonds between the inhibitor P1 and P1′ NH groups and the carbonyls of Gly27 and Gly27′ of the enzyme were weaker than those hydrogen bonds formed between the inhibitor and the flap water. Therefore, the net gain of binding due to hydrogen bond formation between the inhibitor and flexible parts of the enzyme was offset by the desolvation penalty of the polar hydrogen bonding groups and was unlikely to significantly increase binding strength. They pointed out that hydrophobic interactions with the enzyme and hydrogen bonding interactions with the two catalytic aspartates in the active site were crucial for potent inhibitors. Rick et al studied the drug resistant mutant I84V of the HIV-1 protease binding with three potent inhibitors, KNI-272, Indinavir, and Saquinavir (134). They applied TI to calculate relative binding free energies between the wild-type enzyme and the I84V mutant. Because

HIV protease is a homodimer, the perturbation involves I84V and I84′V mutations. They found that the free energy contribution from each side chain was correlated with the other side chain. The free energy from I84′V was more variable among the three inhibitors, due to the different P1′ group of the three inhibitors and therefore to different cavity sizes in the mutant complex. They observed that the cavity size, measured either in cavity volume or surface area, correlated very well with the measured free energy changes with slopes in the range of that found for protein stability. This stability is perhaps due to the peptidic nature of the inhibitors. McCarrick & Kollman carried out FEP simulations on haloperidol thioketal (THK) and three of its derivatives bound with HIV protease (101). Their simulations predicted tighter binding THK derivatives than the present THK compound.

FEP and TI have been widely exploited to calculate relative binding free energy for similar organic systems. Progress has been made in calculating absolute binding free energies for protein-ligand and DNA-ligand complexes as well (63, 104). Recently, Helms & Wade (63) reported the calculated absolute binding free energy for camphor binding to P450cam from *Pseudonomas putida*. By mutating the camphor into six water molecules in the binding site, they were able to reproduce the absolute binding free energy within 3 kJ/mol (<1.0 kcal/mol) of the experimental value.

It is well known that the most severe limitation in free energy calculations is sampling conformational space (12). It is not just a matter of sampling longer but also of sampling in the correct region of conformation space. In order to achieve good sampling, long-range electrostatic interactions and molecular polarization have to be treated appropriately. In their study of organic cations bound to a cyclophane host, Eriksson et al showed that using a nonadditive force field, which is necessary for considering polarization, and the PME method to consider long-range electrostatic interactions can improve the calculated relative free energy of association of an imminium (IM) and a guanidinium (GU) binding to the host from −2.3 kcal/mol to −4.0 kcal/mol compared to a measured value of −3.7 kcal/mol (38). Recently, Ota et al proposed a non-Boltzmann thermodynamic integration (NBTI) method, which is a combination of TI and umbrella sampling (umbrella sampling attempts to overcome the sampling problem by modifying the potential function so that the unfavorable states are sampled sufficiently) (5, 86), to enhance sampling conformational space for macromolecular systems (111, 112); they applied this method to calculate the relative binding free energy between benzamidine (BZD) and benzylamine (BZA) associated with trypsin. The calculated free energy value using NBTI (2.2 kcal/mol) was much closer to the measured value 2.6 kcal/mol than the value 0.8 kcal/mol obtained using conventional TI. This result is very encouraging.

Erion & Reddy have reported a new method that uses both QM and FEP methods for calculating relative changes in the hydration free energies between two similar molecules (41). Recently, they applied this method in designing inhibitors for adenosine deaminase and cytidine deaminase (42). They showed that heteroaromatic hydration was controlled by a multitude of molecular factors. Their

calculation of relative inhibitor potencies for adenosine deaminase agrees well with the experimental data (43).

As we mentioned above, FEP and TI are most rigorous methods and in principle can be used to calculate any free energy difference. Recent progress in developing and applying these methods to study complex macromolecular systems is promising. Combined with other simulation methods such as homology modeling and docking, FEP and TI will become more powerful tools in understanding biological problems.

"Multimolecule" Free Energy Calculation Methods

FEP and TI methods are intrinsically "pair-wise" methods, that is, each FEP/TI simulation has to be performed to obtain free energy difference between two states/molecules. It is more computationally efficient if the free energy differences between several states/molecules can be calculated in one simulation. Such "multimolecule" free energy calculation methods have been developed (55, 56, 79, 91, 118). They are specifically useful in calculating relative binding free energies for several similar ligands.

Brooks and coworkers have developed a new approach called λ-dynamics to evaluate relative hydration free energies or binding free energies between several molecules in a single run of simulation (55, 56, 79). In this λ-dynamics approach, they treated λ in Equation 3 as a set of variables λ_j {j = 1, n}, and each molecule was assigned a λ_j. {$\lambda_j = 0$; j = 1, n} and {$\lambda_j = 1$; j = 1, n} corresponded to start and end states respectively. An extended Hamiltonian of the whole system, $H_{extended}$ {λ_j, j = 1, n}, was a combination of the n molecules' Hamiltonians, a kinetic energy term associated with a set of fictitious masses and an umbrella potential (the potential function used in umbrella sampling). In order to optimize $H_{extended}$ {λ_j, j = 1, n} along a pathway from start to end state, the n molecules competed with each other. When the simulations reached equilibrium, different molecules had different λ values. The weighted histogram analysis method (WHAM) was then employed to generate free energy contours. This approach was successfully demonstrated in calculating the hydration free energies of several small organic molecules (CH_3CH_3, CH_3OH, CH_3SH, and CH_3CN) and identifying the best binder to trypsin among benzamidine and three of its paraderivatives. The results obtained from the λ-dynamics approach were consistent with experimental data and conventional FEP calculations (14, 40, 55, 56, 79, 118, 157).

Recently, Eriksson et al calculated binding free energies of TIBO-like HIV-1 reverse transcriptase (RT) inhibitors (40). In their study (40), the adaptive chemical Monte Carlo/molecular dynamics (CMC/MD), another "multimolecule" free energy calculation method, was exploited to rank 13 different TIBO derivatives with respect to their relative free energies. The CMC/MD method was developed by Pitera & Kollman and was able to rank binding affinities for several ligands in a single MD simulation (118). The MD was used to sample conformations

of each ligand, and the MC was used to sample "chemical space" of all ligands (14, 118, 157). A MD run started from the complex of the receptor and one of the several ligands. After a certain period of MD simulation, a mutation from the ligand to any ligand under consideration occurred. The Metropolis criteria were used to determine whether this mutation was accepted. At the end of the simulation, free energy differences between ligands could be obtained by analyzing the populations of each ligand, that is, ligands chosen more often by MC were assumed to bind more tightly to RT than those ligands chosen less often by MC in the whole simulation. The calculated values were consistent with measured ones, and some results were also confirmed by the Poisson-Boltzmann/solvent accessibility (PB/SA) method and FEP/TI methods (40). One new derivative, suggested by the program PROFEC (Pictorial Representation of Free Energy Components, see below; 120), was predicted to bind 1–2 kcal/mol better than the starting ligand, R86183 (8Cl-TIBO).

PROFEC and OWFEG

In drug design, the question often asked is, "What changes can be made to improve the binding constant?" Recently, two methods have been developed to suggest promising changes to improve the binding (87, 120). In their study of trypsin and its inhibitors (120), Radmer & Kollman have calculated the approximate free energy at each grid point (a probe was put at that point and a single window FEP was performed) surrounding an interesting region of one of the trypsin inhibitors, benzamidine. Free energies of all grid points were then displayed as contour surfaces around the inhibitor. This PROFEC method could quantitatively suggest relatively more favorable regions for molecular change and was shown promising to rank 9 trypsin inhibitors. Recently, Lee & Kollman (87) showed the strength of combining FEP and PROFEC methods to predict more potent inhibitors of thymidylate synthase (TS). TS is an enzyme that catalyzes dTMP synthesis for DNA synthesis. Inhibition of TS can block dTMP synthesis and therefore suggests chemotherapeutic use to combat cancer. Jones et al designed and synthesized 31 inhibitors of TS, most of which had low water solubility (68). Lee & Kollman predicted new, stronger inhibitors modified from one of the Jones et al inhibitors using PROFEC and confirmed the prediction by TI calculations. Their simulations provided guidelines for designing new potent inhibitors of TS with better solubility.

OWFEG (116) has made two modifications of PROFEC. First, each grid point underwent translation and rotation along with the atom of the ligand to which it was closest. Thus, flexible regions of the ligand could be explored. Second, three probes with neutral, positive, and negative charges were used instead of only a neutral probe to examine the desirability of introducing charged groups along the grid. This feature provided hints as to what type of charges should be placed at that gridpoint. In two test systems, quinoline and bis-pyrimidine, and FKBP-12·FK506 protein-ligand complex, the qualitative results derived from OWFEG showed excellent agreement with the standard TI simulations (116).

Linear Interaction Energy Method

The linear interaction energy (LIE) method was originally proposed by Åqvist et al to estimate the absolute binding free energies. The LIE method is based on linear response assumptions; that is, the solvent polarization responses to changes in the electrostatic field exerted by the solute is linear and characterized by a single dielectric constant (9). It divides the interaction between the ligand and its environment into electrostatic and van der Waals parts. The binding free energy is estimated as

$$\begin{aligned}\Delta G_{\text{bind}} &= \Delta G^{\text{el}}_{\text{bind}} + \Delta G^{\text{vdw}}_{\text{bind}} \\ &\approx \alpha\langle V^{\text{el}}_{\text{bound}} - V^{\text{el}}_{\text{free}}\rangle + \beta\langle V^{\text{vdw}}_{\text{bound}} - V^{\text{vdw}}_{\text{free}}\rangle, \end{aligned} \quad (6)$$

where $V^{\text{el}}_{\text{bound}}$ and $V^{\text{vdw}}_{\text{bound}}$ are the electrostatic and van der Waals interaction energies between the ligand and the solvated protein from a MD trajectory with ligand bound to protein; $V^{\text{el}}_{\text{free}}$ and $V^{\text{vdw}}_{\text{free}}$ are electrostatic and van der Waals interaction energies between the ligand and the water from an MD trajectory with the ligand in water; $\langle\rangle$ denotes an ensemble average, and α and β are two empirical parameters.

Åqvist and coworkers have applied this method to calculate absolute binding free energies of several protein-ligand complexes. They found that $\alpha = 0.5$ and $\beta = 0.16$ gave calculated binding free energies in good agreement with experimental data. In the calibration set, four inhibitors bound to endothiapepsin, this set of parameters gave a mean unsigned error of 0.39 kcal/mol and 0.59 kcal/mol for calculated absolute and relative binding free energies, respectively. The absolute binding free energy for the fifth inhibitor to endothiapepsin was predicted as -9.70 kcal/mol compared with the observed value -9.84 kcal/mol (9). This LIE method was also successfully applied to calculate absolute binding free energies of HIV protease inhibitors and two charged trypsin benzamidine inhibitors (8, 60). In these two studies, an additional correction term for long-range electrostatic contribution to the binding free energy was included. The calculated and observed absolute binding free energies agree well with each other using the same values of α and β. Paulsen & Ornstein, however, found that $\alpha = 0.5$ and $\beta = 1.043$ resulted in a good estimate of the binding free energies of 11 substrates binding to cytochrome P450 cam (115). The difference between the two sets of parameters was rationalized as perhaps owing to different force fields, GROMOS and CVFF respectively, used in the two studies (163). Wang et al (167) applied this method to calculate binding free energies of 14 compounds binding to avidin using the Cornell et al force field (27). Their results showed that $\alpha = 0.5$ and $\beta = 1.0$ gave reasonable estimates of the binding free energies with respect to the corresponding experimental results.

These studies raise an interesting question: Can one set of α and β be used in different protein-ligand complexes to give reasonable estimates of binding free energies? Although Wang et al used the Cornell et al force field (27), they found values of α and β similar to those of Åqvist et al for the trypsin-benzamidine

complex (163). This suggests that the use of different force fields cannot explain the difference in α and β found in different simulations. Wang et al (167) further examined this issue in seven different complex systems and found a relationship between the value of β and hydrophobicity of the ligand and the binding site of the receptor; that is, the more hydrophobic groups buried after binding, the more favorable the binding, and the larger the value of β. Different β values were determined for different inhibitors bound to avidin according to this relationship. Calculated absolute binding free energies were improved compared with those from using a fixed β value (167).

Jorgensen and coworkers extended this method further for calculating hydration and binding free energies. They added another term to Equation 6, which is proportional to solvent accessible surface area change upon binding. MC simulation was used to obtain the ensemble. The values of these coefficients were calibrated in a test set and then were transferred to predict other ligands bound with the same protein. They succeeded in calculating binding free energies for sulfonamide inhibitors with human thrombin (69) and FKBP12 inhibitors (84).

The LIE method is useful for estimating absolute binding free energies for protein-ligand systems. This method is more computationally efficient than the FEP/TI method.

MM/PBSA

Recent computational advances in parallel computing, force fields, and more accurate treatment of electrostatic interactions have enabled multinanosecond MD simulations of highly charged macromolecules such as nucleic acids (20, 31). Analyses of dynamics alone, however, do not sufficiently describe macromolecular recognition and complex formation. Conventional free energy calculations, as described above, have also been applied to protein–nucleic acid complexes. More recently, a hybrid method combining molecular mechanics and continuum solvent calculations has increased in popularity to analyze the free energies of binding and relative free energies of different conformations (76, 146, 149, 172–174).

The method takes solute configurations, or snapshots, from a MD trajectory with explicit solvent. The solvent molecules are removed to obtain the molecular mechanics energy (E_{MM}) of the solute. This is computed for each snapshot with the same molecular mechanics potential as in the simulation but with no cut-offs to incorporate all of the nonbonded interactions. The conformational entropy of the solute, $T\Delta S$, including rotational and vibrational contributions, is estimated from normal mode analyses:

$$\Delta G = E_{MM} = T\Delta S + \Delta G_{solvation}, \quad \text{and} \tag{7}$$

$$\Delta G_{solvation} = \Delta G_{PB} + \Delta G_{nonpolar}. \tag{8}$$

The free energy of solvation, $\Delta G_{solvation}$, is approximated as the sum of electrostatic and nonpolar contributions. The electrostatic solvation term is calculated with the PB approach, whereas the nonpolar term as a surface area (SA)–dependent term, hence the name MM-PBSA.

A finite difference solution to the PB equation is calculated using the Delphi II program (142, 143):

$$\nabla\varepsilon(\mathrm{r})\nabla\phi(\mathrm{r}) - \kappa'\phi(\mathrm{r}) = -4\pi\rho(\mathrm{r}), \tag{9}$$

where ϕ(r) is the electrostatic potential, ε(r) is the dielectric function, ρ(r) is the charge density, and κ' is related to the Debye-Huckel inverse length. In the Delphi program, the solute is mapped onto a cubic lattice grid. Values for the electrostatic potential, charge density, dielectric constant, and ionic strength are assigned to each grid point (86). The derivatives of the PB equation are calculated with a finite difference formula and iteratively computed to convergence. The electrostatic component of the solvation free energy is the change of electrostatic energy from transferring the solute from a low dielectric (vacuum) to high dielectric medium using the same grid and solute dielectric.

$$\Delta G_{PB} = \frac{1}{2}\sum_{i} q_i\left(\phi_i^{80} - \phi_i^{1}\right). \tag{10}$$

The nonpolar solvation term is approximated as linearly dependent on the solvent accessible surface area:

$$\Delta G_{nonpolar} = \gamma(\mathrm{SASA}) + \beta, \tag{11}$$

where $\gamma = 0.00542$ and $\beta = 0.92$ kcal/mol (149). The surface area is computed with Sanner's MSMS software (138) using a water-sized probe. The MM energies and solvation free energies are computed for each snapshot of the solute and then averaged to compute the difference in free energies. The free energy difference can be computed for absolute binding or the relative binding of different mutants.

Chong et al (23) applied this method to study dianionic hapten binding to a germ line and mature forms of the 48G7 antibody Fab fragments. Reasonable absolute and good relative binding free energies compared with experimental data were obtained. Their calculations indicated that van der Waals interaction energies and nonpolar contributions to solvation energy were almost identical for both antigens. The >10,000-fold tighter binding of the matured antibody than that of the germ line was due to the gain of more favorable electrostatic interactions over the desolvation penalty through optimizing the binding site geometry. This work sheds light on understanding the process of antibody maturation.

Contrary to the electrostatic discrimination between two antigens, Wang et al (166) observed that in the complexes of Sem-5 SH3 domain and its ligands, van der Waals interactions were primarily responsible for significant binding affinity differences between Cα- and N-substituted ligands. This shows that binding is

dominated by different interactions in different complex systems. In their study, they were also able to identify several critical residues for binding by considering van der Waals energy and conservation of each residue (166).

Another application of this MM/PBSA method to study biotin and its derivatives binding with avidin/streptavidin was carried out by Kuhn & Kollman (82). They were able to reproduce relative binding free energies of 9-methylbiotin compounds with a very good correlation to experimental values. A so-called "computational fluorine scanning" technique—that is, substituting hydrogen by fluorine at different sites of the biotin in a single trajectory obtained from one compound, and then calculating the binding free energy for the substituent upon the "substituted trajectory"—was shown to work well for ranking the nine compounds. This makes free energy calculations more efficient.

Donini et al (35) have used a single trajectory of a ligand binding to a matrix metalloprotease to calculate the binding free energy of five other analog inhibitors. The relative binding free energy of the neutral inhibitors and charged inhibitors were correctly ranked within their series, but the neutral inhibitors were calculated to bind more strongly than the charged inhibitors relative to experiment.

The precedent of the computational fluorine scanning is the "computational alanine scanning" technique used by Massova & Kollman (99) in their study of p53-MDM2 interactions. Mutating 20 amino acids to alanine in the native trajectory allowed them to quickly compare calculated binding free energy with measured ones. Further examination of W23A mutation by PROFEC led to the conclusion that an additional methyl group on the aromatic rings of W23 might substantially improve binding.

The above work was followed by Huo et al (66), who compared results obtained from the computational and from the experimental alanine scanning on human growth hormone/human growth hormone receptor (HGH/HGHr). Twelve residues were mutated; in all cases but two (R43A and R216A), the calculated $\Delta\Delta G_{bind}$ were in reasonable agreement with the experimental ones. Significant conformational change of mutating Arg43 or Arg216 to Ala caused the overestimation of the $\Delta\Delta G_{bind}$.

Wang & Kollman (165) also applied this MM/PBSA method in studying dimer stability of the HIV protease. They were able to reproduce the relative ranking order of different dimers in agreement with experiments. A rapid screening method, which identified cavities on the dimer interface and suggested that favorable van der Waals contacts would be created if the cavity was filled with a larger side chain, was exploited to suggest new possible mutations that might enhance binding. Conformational search and minimization were performed for mutation to larger side chains on the dimer interface. Several new stronger associated heterodimers were suggested in their study by this so-called "virtual mutagenesis" method.

Recently, Kuhn & Kollman (81) compared MM/PBSA and LIE methods in calculating binding free energies for diverse avidin and streptavidin ligands. Their calculations were able to reproduce experimental ΔG_{bind} with a correlation

coefficient of $r^2 = 0.92$, which was much better that the results obtained from the LIE method with fixed parameters ($\alpha = 0.5$ and $\beta = 1$) and whose r^2 is 0.55. Although the β value can be adjusted based on hydrophobicity of the binding site (see above), the MM/PBSA method does not introduce any empirical parameters on a protein-by-protein basis.

Protein-DNA Complexes

Several structures of DNA-protein complexes have been elucidated and studied by MD (110). Eriksson & Nilsson studied the binding of estrogen receptor DNA-binding domain (ERDBD) to DNA both as a dimer and as a monomer (39). Their comparative MD study found that the monomer ERDBD is more flexible than the dimer and suggested that dimerization of ERDBD orders the ZnII region and facilitates DNA binding. Schulten and coworkers also performed shorter MD simulations of ERDBD bound to DNA and focused on the hydration on the binding interface and local structural changes of the DNA (80). Harris et al modeled the ERDBD in complex with a nonconsensus estrogen response element DNA sequence and detailed intermolecular H-bonding interactions from the 300 ps simulation (61).

Roxstrom et al simulated a zinc-finger protein bound to DNA in their nanosecond trajectory of Zif268-DNA complex (136). They found comparable fluctuation patterns to experimental B-factors. Tang & Nilsson performed MD simulations on the human sex-determining region Y (hSRY) protein-DNA complex and highlighted the dominance of hydrophobic interactions. They also modeled the free monomers and predicted that both hSRY and DNA would undergo conformational change upon binding.

There have also been several computational studies of repressor protein–DNA complexes. Harris et al simulated the bacteriophage 434 cI repressor protein homodimer in complex with its cognate operator DNA sequence (62). They ran a nanosecond trajectory of the complex with a 10 Å water shell and observed direct and water-mediated H-bonding in the interface. Thomasson et al computed the free energy profile of nonspecific binding of Cro repressor protein to DNA (156). They employed Brownian dynamics method to generate a Boltzmann population of protein conformations docked to the DNA. The statistical ensemble of docked complexes was used to obtain a potential of mean force (PMF), or free energy of interaction between the protein and DNA. They computed comparable ΔG of association to the experimental value. Moreover, from the PMF calculations, they estimated the lifetime of nonspecific docking of DNA to the Cro repressor protein to be 5 microseconds. Another repressor protein–DNA complex was studied by Misra et al with the nonlinear PB method to compute the electrostatic contribution to the binding free energy of the λcI repressor to DNA (103). They calculated an overall unfavorable electrostatic contribution to the binding free energy. This was based on highly unfavorable desolvation of the monomers as charged, and polar groups were removed from solvent upon binding.

The EcoRI-DNA complex was a subject of several computational studies. Jayaram et al computed its binding free energy with an MM-continuum solvent hybrid method (67). The binding free energy was decomposed into 24 components, including structural adaptation, electrostatic and van der Waals terms, hydrophobic contributions, and ion effects. Their binding free energy analysis resulted in reasonable agreement with experimental binding. Sen & Nilsson performed MD simulations and FEP calculations on the EcoRI-DNA complexes modifying functional groups in Ade^5, Ade^6, and Gua^4, finding good qualitative agreement with experiment (140, 141).

RNA-Protein Complexes

MD simulations and free energy calculations have long been applied to DNA-protein complexes. The theoretical study of RNA-protein complexes, however, is a fairly new and emerging field concurrent with the explosion of RNA-protein structures determined by both crystallography and NMR (32, 36, 65, 162). Often, the assembly of RNA-protein complexes requires extensive conformational rearrangement or induced-fit. One of the best-studied RNA-protein complexes, the paradigm of ribonucleoprotein complex interactions, is U1A-RNA.

The solution of the crystal structure of the N-terminal region of the U1A protein bound to its cognate hairpin RNA at 1.9 Å resolution provided structural data on one of the best-characterized RBDs (113). The U1A protein is a component of the U1 small nuclear ribonucleoprotein (snRNP), which binds to the 5′ end of the primary transcript and initiates spliceosomal assembly. The crystal structure of U1A-RNA is a paradigm for RBD containing proteins and shows an intricate web of interactions on the binding interface. Splayed-out RNA loop bases stack with hydrophobic residues on the β-sheet surface of the protein. Moreover, the protein's $\beta 2$–$\beta 3$ loop protrudes through the RNA loop and contacts the loop-closing base pair. In 1996, the NMR structure of U1A bound to an internal loop RNA was published. The structure showed very similar interactions to the U1A-hairpin complex (4). A particularly interesting point is that U1A recognizes these two structurally distinct RNAs with similarly high (nanomolar) affinity. Moreover, a comparison with the structures of the unbound monomers (10, 54) indicates large conformational rearrangements must occur upon binding. Extensive structural and biochemical data on U1A-RNA has catalyzed a number of computational studies focused on investigating the structure, thermodynamics, and conformational reorganization of U1A-RNA binding.

Hermann & Westhof described the MD simulations of the U1A protein bound to the hairpin RNA at low and high salt concentration in explicit solvent (64). At 1M NaCl concentration, they observed increased fluctuations at the binding interface. They suggested that the increased fluctuations at high ionic strength represent the microscopic basis of salt-induced complex dissociation. Tang & Nilsson reported a 0.6 ns MD simulation of the U1A-hairpin complex and a 1–2 ns simulation of the free RNA hairpin (154). They observed a more flexible RNA loop in the

free hairpin simulation compared to the complex simulation in agreement with experimental observation.

Reyes & Kollman compared the dynamics of the U1A-hairpin and U1A–internal loop complexes and observed a global hinge motion in the internal loop complex simulation that was not observed in the hairpin complex simulation (130). This suggests that although both complexes have equivalent nanomolar binding affinities, they do not necessarily have the same thermodynamic driving forces.

The thermodynamics of U1A-RNA binding was the subject of another study by Reyes & Kollman (132). Using MD and free energy analyses with MM-PBSA, Reyes & Kollman estimated the energetic cost of conformational reorganization upon U1A-RNA binding. They also computed the absolute binding free energies of both complexes and found good agreement with experimentally measured binding affinities. In another paper, the relative binding of different U1A mutants to the hairpin RNA was computed using a computationally inexpensive method for investigating and predicting the effects of site-specific mutations (131). Reyes & Kollman obtained good agreement with experimental mutagenesis and verified specific mutations that either abolish or improve binding.

Summary

In this section, we reviewed recent studies of protein-ligand, protein-protein, protein-DNA, and protein-RNA interactions using different free energy calculation methods. The same methods have also been applied to study hydration and interactions involving organic molecules, which we have not addressed here.

FEP and TI methods are the most rigorous but require extensive computer resources. With the rapid increase of computer power, we can expect wider application of these methods in the future. Multimolecule free energy calculation methods are promising based on recent studies. They are certainly worth further investigation. The LIE method has a unique advantage because it allows the calculation of absolute binding free energies. With appropriate empirical parameters, this method is useful for studying specific complex systems. Replacing explicit water molecules with a solvent continuum can accelerate MD simulations and enable binding free energies to be calculated directly. Thus, MM/PBSA is a promising direction for evaluating binding affinities. Combined with other modeling tools, free energy calculation methods will be used in a broader range of research, from evaluating stability of folding structures to the design of new drugs.

Perhaps the most exciting recent development in RNA-protein structure determination is the determination of the three-dimensional structure of the ribosome. The ribosome is a ribonucleoprotein particle that carries out the translation of mRNA to proteins with the help of amino-acid-charged tRNAs and accessory proteins. The molecular weight of ribosomal particles ranges from 2.5 MD (typical of a bacterial ribosome) to 4.5 MD (in eukaryotic cytoplasm) (106). The bacterial ribosome, as the smallest ribosome, has been the focus of many structural efforts. The low-resolution structure and now the atomic-resolution structure of the 30S

and 50S subunits, as well as the structure of 70S ribosome, have recently appeared (1, 11, 19, 26, 106, 159).

CONCLUSIONS AND PERSPECTIVES

We have presented a review of three areas of biomolecular simulation: force fields, combined quantum/molecular methods to study enzyme catalysis, and free energy calculations on noncovalent interactions. Our general view is that these are all exciting and vital areas where much more research needs to be done. It is probably worth noting what has not been covered here, including enhanced sampling techniques in MD, MC, and other approaches to improve the conformational sampling of complex biological systems, DOCKing approaches to screen many molecules at a time in ligand design, the use of quantitative structure activity relationships (QSAR) studies when a receptor structure is not available, and a variety of lower resolutions than all atom approaches to protein folding. Obviously, each of these topics is probably deserving of a review by itself.

LITERATURE CITED

1. Agalarov SC, Prasad GS, Funke PM, Stout CD, Williamson JR. 2000. Structure of the S15,S18-rRNA complex: assembly of the 30S ribosome central domain. *Science* 288:107–12
2. Ajay, Murcko MA. 1995. Computational methods to predict binding free energy in ligand-receptor complexes. *J. Med. Chem.* 38:4953–67
3. Alhambra C, Wu L, Zhang ZY, Gao JL. 1998. Walden-inversion-enforced transition-state stabilization in a protein tyrosine phosphatase. *J. Am. Chem. Soc.* 120:3858–66
4. Allain FHT, Gubser CC, Howe PWA, Nagai K, Neuhaus D, Varani G. 1996. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature* 380:646–50
5. Allen MP, Tildesley DJ. 1987. *Computer Simulation of Liquids*, pp. 213–18. New York: University Press
6. Antes I, Thiel W. 1998. On the treatment of linkatoms in hybrid methods. *ACS Symp. Ser.* 712:50
7. Antonczak S, Monard G, Ruiz-Lopez MF, Rivail JL. 1998. Modeling of peptide hydrolysis by thermolysin. A semiempirical and QM/MM study. *J. Am. Chem. Soc.* 120:8825–33
8. Åqvist J. 1996. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comp. Chem.* 17:1587–97
9. Åqvist J, Medina C, Samuelsson JE. 1994. New method for predicting binding affinity in computer-aided drug design. *Protein Eng.* 7:385–91
10. Avis JM, Allain FHT, Howe PWA, Varani G, Nagai K, Neuhaus D. 1996. Solution structure of the N-terminal Rnp domain of U1A protein—the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.* 257:398–411
11. Ban N, Nissen P, Hansen J, Capel M, Moore PB, Steitz TA. 1999. Placement

of protein and RNA structures into a 5 angstrom-resolution map of the 50S ribosomal subunit. *Nature* 400:841–47
12. Bash PA, Singh UC, Langridge R, Kollman PA. 1987. Free energy calculations by computer simulation. *Science* 236:564–68
13. Beckers JVL, Lowe CP, DeLeeuw SW. 1998. An iterative PPPM method for simulating Coulombic systems on distributed memory parallel computers. *Mol. Simul.* 20:369–83
14. Bennett CH. 1976. Efficient estimation of free energy differences from Monte Carlo data. *J. Comp. Phys.* 22:245–68
15. Bentzien J, Muller RP, Florian J, Warshel A. 1998. Hybrid ab initio quantum mechanics molecular mechanics calculations of free energy surfaces for enzymatic reactions: the nucleophilic attack in subtilisin. *J. Phys. Chem. B* 102:2293–301
16. Berendsen HJC, Grigera JR, Straatsma TP. 1987. The missing term in effective pair potentials. *J. Phys. Chem.* 91:6269–74
17. Beveridge DL, Dicapua FM. 1989. Free energy via molecular simulations: application to chemical and biochemical system. *Annu. Rev. Biophys. Biophys. Chem.* 18:431–92
18. Deleted in proof
19. Cate JH, Yusupov MM, Yusupova GZ, Earnest TN, Noller HF. 1999. X-ray crystal structures of 70S ribosome functional complexes. *Science* 285:2095–104
20. Cheatham TE, Brooks BR. 1998. Recent advances in molecular dynamics simulation towards the realistic representation of biomolecules in solution. *Theor. Chem. Acc.* 99:279–88
21. Cheatham TE, Cieplak P, Kollman PA. 1999. A modified version of the Cornell et al force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct.* 16:845–62
22. Cheatham TE, Kollman PA. 2000. Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.* 51:435–71
23. Chong LT, Duan Y, Wang L, Massova I, Kollman PA. 1999. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Sci. USA* 96:14330–35
24. Cieplak P, Caldwell J, Kollman PA. 2000. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and n-methyl acetimide. Nucleic acid base and amide hydrogen bonding and chloroform/waste coefficients of the nucleic acid bases. *J. Comp. Chem.*
25. Cieplak P, Kollman PA. 1993. Peptide mimetics as enzyme inhibitors—use of free energy perturbation calculations to evaluate isosteric replacement for amide bonds in a potent HIV protease inhibitor. *J. Comput. Aided Mol. Des.* 7:291–304
26. Clemons WM, May JLC, Wimberly BT, McCutcheon JP, Capel MS, Ramakrishnan V. 1999. Structure of a bacterial 30S ribosomal subunit at 5.5 angstrom resolution. *Nature* 400:833–40
27. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, et al. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–97
28. Cummins PL, Gready JE. 1998. Molecular dynamics and free energy perturbation study of hydride-ion transfer step in dihydrofolate reductase using combined quantum and molecular mechanical model. *J. Comp. Chem.* 19:977–88
29. Damm W, Frontera A, Tirado-Rives J, Jorgensen WL. 1997. OPLS all-atom force field for carbohydrates. *J. Comp. Chem.* 18:1955–70
30. Dang LX, Kollman PA. 1990. Free energy of association of the 18-Crown-6-K+ complex in water—a molecular dynamics simulation. *J. Am. Chem. Soc.* 112:5716–20
31. Darden TA, Toukmaji A, Pedersen LG. 1997. Long-range electrostatic effects in biomolecular simulations. *J. Chim. Phys. Phys.-Chim. Biol.* 94:1346–64

32. De Guzman RN, Turner RB, Summers MF. 1998. Protein-RNA recognition. *Biopolymers* 48:181–95
33. De Santis L, Carloni P. 1999. Serine proteases: an ab initio molecular dynamics study. *Proteins: Struct. Funct. Genet.* 37:611–18
34. Donini O, Darden T, Kollman PA. 2000. QM-FE calculations of alphatic hydrogen abstruction in citrate synthase and solution: reproduction of the effect of enzyme catalysis and demonstration that an enolate rather than an enol is formed. *J. Am. Chem. Soc.* In press
35. Donini O, Kollman PA. 2000. Calculation and prediction of binding free energies for the matix metalloproteins. *J. Med. Chem.* 43:4180–88
36. Draper DE. 1995. Protein-RNA recognition. *Annu. Rev. Biochem.* 64:593–620
37. Droupadi PR, Varga JM, Linthicum DS. 1994. Mechanism of allergenic cross-reactions. V. Evidence for participation of aromatic residues in the ligand binding site of two multi-specific Ige monoclonal antibodies. *Mol. Immunol.* 31:537–48
38. Eriksson MAL, Morgantini PY, Kollman PA. 1999. Binding of organic cations to a cyclophane host as studied with molecular dynamics simulations and free energy calculations. *J. Phys. Chem. B* 103:4474–80
39. Eriksson MAL, Nilsson L. 1999. Structural and dynamic differences of the estrogen receptor DNA binding domain, binding as a dimer and as a monomer to DNA: molecular dynamics simulation studies. *Eur. Biophys. J.* 28:102–11
40. Eriksson MAL, Pitera J, Kollman PA. 1999. Prediction of the binding free energies of new TIBO-like HIV-1 reverse transcriptase inhibitors using a combination of PROFEC, PB/SA, CMC/MD, and free energy calculations. *J. Med. Chem.* 42:868–81
41. Erion MD, Reddy MR. 1995. Calculation of relative free energy differences for the covalent hydration of organic compounds—a combined quantum mechanical and free energy perturbation study. *J. Comp. Chem.* 16:1513–21
42. Erion MD, Reddy MR. 1998. Calculation of relative hydration free energy differences for heteroaromatic compounds: use in the design of adenosine deaminase and cytidine deaminase inhibitors. *J. Am. Chem. Soc.* 120:3295–304
43. Erion MD, van Poelje PD, Reddy MR. 2000. Computer-assisted scanning of ligand interactions: analysis of the fructose 1,6-bisphosphatase-AMP complex using free energy calculations. *J. Am. Chem. Soc.* 122:6114–15
44. Essex JW, Severance DL, Tirado-Rives J, Jorgensen WL. 1997. Monte Carlo simulations for proteins: binding affinities for trypsin-benzamidine complexes via free-energy perturbations. *J. Phys. Chem. B* 101:9663–69
45. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–93
46. Ewig CS, Thacher TS, Hagler AT. 1999. Derivation of class II force fields. VII. Nonbonded force field parameters for organic compounds. *J. Phys. Chem. B* 103:6998–7014
47. Feierberg I, Cameron AD, Åqvist J. 1999. Energetics of the proposed rate-determining step of the glyoxalase I reaction. *FEBS Lett.* 453:90–94
48. Feller SE, Yin DX, Pastor RW, MacKerell AD. 1997. Molecular dynamics simulation of unsaturated lipid bilayers at low hydration: parameterization and comparison with diffraction studies. *Biophys. J.* 73:2269–79
49. Ferguson DM, Radmer RJ, Kollman PA. 1991. Determination of the relative binding free energies of peptide inhibitors to the HIV-1 protease. *J. Med. Chem.* 34:2654–59
50. Foloppe N, MacKerell AD. 2000. All-atom empirical force field for nucleic acids. I. Parameter optimization based on small

molecule and condensed phase macromolecular target data. *J. Comp. Chem.* 21:86–120
51. Fox T, Scanlan TS, Kollman PA. 1997. Ligand binding in the catalytic antibody 17E8. A free energy perturbation calculation study. *J. Am. Chem. Soc.* 119:11571–77
52. Gao JL, Amara P, Alhambra C, Field MJ. 1998. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A* 102:4714–21
53. Glennon TM, Warshel A. 1998. Energetics of the catalytic reaction of ribonuclease A: a computational study of alternative mechanisms. *J. Am. Chem. Soc.* 120:10234–47
54. Gubser CC, Varani G. 1996. Structure of the polyadenylation regulatory element of the human U1A pre-MRNA 3′-untranslated region and interaction with the U1A protein. *Biochemistry* 35:2253–67
55. Guo Z, Brooks CL, Kong X. 1998. Efficient and flexible algorithm for free energy calculations using the lambda-dynamics approach. *J. Phys. Chem. B* 102:2032–36
56. Guo ZY, Brooks CL. 1998. Rapid screening of binding affinities: application of the lambda-dynamics method to a trypsin-inhibitor system. *J. Am. Chem. Soc.* 120:1920–21
57. Halgren TA. 1996. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of Mmff94. *J. Comp. Chem.* 17:490–519
58. Halgren TA. 1996. Merck molecular force field. II. Mmff94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comp. Chem.* 17:520–52
59. Halgren TA. 1996. Merck molecular force field. III. Molecular geometries and vibrational frequencies for Mmff94. *J. Comp. Chem.* 17:553–86
60. Hansson T, Åqvist J. 1995. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* 8:1137–44
61. Harris LF, Sullivan MR, Popken-Harris PD. 1997. Molecular dynamics simulation in solvent of the estrogen receptor protein DNA binding domain in complex with a non-consensus estrogen response element DNA sequence. *J. Biomol. Struct.* 15:407–30
62. Harris LF, Sullivan MR, Popken-Harris PD. 1999. Molecular dynamics simulation in solvent of the bacteriophage 434 cI repressor protein DNA binding domain amino acids (R1-69) in complex with its cognate operator (OR1) DNA sequence. *J. Biomol. Struct.* 17:1–17
63. Helms V, Wade RC. 1998. Computational alchemy to calculate absolute protein-ligand binding free energy. *J. Am. Chem. Soc.* 120:2710–13
64. Hermann T, Westhof E. 1999. Simulations of the dynamics at an RNA-protein interface. *Nat. Struct. Biol.* 6:540–44
65. Holbrook SR. 1998. RNA crystallography. In *RNA Structure and Function*, ed. RW Simons, M Grunberg-Manago, pp. 147. Cold Spring Harbor, NY: Cold Spring Harbor Lab. Press
66. Huo S, Massova I, Kollman PA. 2000. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J. Comput. Chem.* In press
67. Jayaram B, McConnell KJ, Dixit SB, Beveridge DL. 1999. Free energy analysis of protein-DNA binding: the EcoRI endonuclease-DNA complex. *J. Comp. Phys.* 151:333–57
68. Jones TR, Varney MD, Webber SE, Lewis KK, Marzoni GP, et al. 1996. Structure-based design of lipophilic quinazoline inhibitors of thymidylate synthase. *J. Med. Chem.* 39:904–17
69. Jones-Hertzog DK, Jorgensen WL. 1997. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *J. Med. Chem.* 40:1539–49
70. Jorgensen WL. 1989. Free-energy calculations: a breakthrough for modeling

organic chemistry in solution. *Acc. Chem. Res.* 22:184–89
71. Jorgensen WL, Chandrasekhar J, Madura J, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–35
72. Kaminski G, Jorgensen WL. 1996. Performance of the Amber94, Mmff94, and Opls-Aa force fields for modeling organic liquids. *J. Phys. Chem.* 100:18010–13
73. Kollman P. 1993. Free energy calculations—applications to chemical and biochemical phenomena. *Chem. Rev.* 93:2395–17
74. Kollman PA. 1997. *Computer Simulations of Biomolecular Systems*, ed. WF van Gunsteren, PK Weiner, A Wilkinson. Leiden: ESCOM
75. Kollman PA, Kuhn B, Donini O, Perakyla M, Stanton RV, Bakowies D. 2000. *Acc. Chem. Res.* In press
76. Kollman PA, Massova I, Reyes CM, Kuhn B, Huo S, et al. 2001. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33:889–97
77. Kolmodin K, Åqvist J. 1999. Computational modeling of the rate limiting step in low molecular weight protein tyrosine phosphatase. *FEBS Lett.* 456:301–5
78. Kolmodin K, Nordlund P, Åqvist J. 1999. Mechanism of substrate dephosphorylation in low M-r protein tyrosine phosphatase. *Proteins: Struct. Funct. Genet.* 36:370–79
79. Kong XJ, Brooks CL. 1996. Lambda-dynamics—a new approach to free energy calculations. *J. Chem. Phys.* 105:2414–23
80. Kosztin D, Bishop TC, Schulten K. 1997. Binding of the estrogen receptor to DNA. The role of waters. *Biophys. J.* 73:557–70
81. Kuhn B, Kollman PA. 2000. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanisms and continuum solvation models. *J. Med. Chem.* 43:3786–91
82. Kuhn B, Kollman PA. 2000. A ligand that is predicted to bind better to avidin than biotin: insights from computational fluorine scanning. *J. Am. Chem. Soc.* 122:3909–16
83. Kuhn B, Kollman PA. 2000. QM-FE and molecular dynamics calculations on catechol O-methyltransferase: free energy of activation in the enzyme and in aqueous solution and regioselectivity of the enzyme-catalyzed reaction. *J. Am. Chem. Soc.* 122:2586–96
84. Lamb ML, Tirado-Rives J, Jorgensen WL. 1999. Estimation of the binding affinities of FKBP12 inhibitors using a linear response method. *Bioorg. Chem.* 7:851–60
85. Langley DR. 1998. Molecular dynamic simulations of environment and sequence dependent DNA conformations: the development of the BMS nucleic acid force field and comparison with experimental results. *J. Biomol. Struct.* 16:487–509
86. Leach AR. 1996. *Molecular Modeling: Principles and Applications*. Mountain View, CA: Addison-Wesley
87. Lee TS, Kollman PA. 2000. Theoretical studies suggest a new antifolate as a more potent inhibitor of thymidylate synthase. *J. Am. Chem. Soc.* 122:4385–93
88. Lee TS, Massova I, Kuhn B, Kollman PA. 2000. QM and QM-FE simulations on reactions of relevance to enzyme catalysis: trypsin, catechol O-methyltransferase, beta-lactamase and pseudouridine synthase. *J. Chem. Soc. Perkin Trans.* 2:409–15
89. Levitt M, Hirshberg M, Sharon R, Daggett V. 1995. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Comput. Phys. Commun.* 91:215–31
90. Liang G, Schmidt RK, Yu HA, Cumming DA, Brady JW. 1996. Free energy simulation studies of the binding specificity of

mannose-binding protein. *J. Phys. Chem.* 100:2528–34
91. Liu HY, Mark AE, Van Gunsteren WF. 1996. Estimating the relative free energy of different molecular states with respect to a single reference state. *J. Phys. Chem.* 100:9485–94
92. Liu HY, Mullerplathe F, Van Gunsteren WF. 1996. A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease. *J. Mol. Biol.* 261:454–69
93. Lyne PD, Hodoscek M, Karplus M. 1999. A hybrid QM-MM potential employing Hartree-Fock or density functional methods in the quantum region. *J. Phys. Chem. A* 103:3462–71
94. Lyne PD, Mulholland AJ, Richards WG. 1995. Insights into chorismate mutase catalysis from a combined QM/MM simulation of the enzyme feaction. *J. Am. Chem. Soc.* 117:11345–50
95. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102:3586–616
96. MacKerell AD, Wiorkiewiczkuczera J, Karplus M. 1995. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* 117:11946–75
97. Mahoney MW, Jorgensen WL. 2000. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* 112:8910–22
98. Marelius J, Kolmodin K, Åqvist J. 1998. Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J. Mol. Graph. Mod.* 16:213–25
99. Massova I, Kollman PA. 1999. Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* 121:8133–43
100. McCammon JA. 1991. Free energy from simulations. *Curr. Opin. Struct. Biol.* 1:196–200
101. McCarrick MA, Kollman PA. 1999. Predicting relative binding affinities of non-peptide HIV protease inhibitors with free energy perturbation calculations. *J. Comput. Aided Mol. Des.* 13:109–21
102. Meyer M, Wohlfahrt G, Knablein J, Schomburg D. 1998. Aspects of the mechanism of catalysis of glucose oxidase: a docking, molecular mechanics and quantum chemical study. *J. Comput. Aided Mol. Des.* 12:425–40
103. Misra VK, Hecht JL, Yang AS, Honig B. 1998. Electrostatic contributions to the binding free energy of the lambda cl repressor to DNA. *Biophys. J.* 75:2262–73
104. Miyamoto S, Kollman PA. 1993. What determines the strength of noncovalent association of ligands to proteins in aqueous solution? *Proc. Natl. Acad. Sci. USA* 90:8402–6
105. Monard G, Loos M, Thery V, Baka K, Rivail JL. 1996. Hybrid classical quantum force field for modeling very large molecules. *Int. J. Quantum Chem.* 58:153–59
106. Moore PB. 1998. The three-dimensional structure of the ribosome and its components. *Annu. Rev. Biophys. Biomol. Struct.* 27:35–58
107. Mulholland AJ, Lyne PD, Karplus M. 2000. Ab initio QM/MM study of the citrate synthase mechanism. A low-barrier hydrogen bond is not involved. *J. Am. Chem. Soc.* 122:534–35
108. Mulholland AJ, Richards WG. 1997. Acetyl-CoA enolization in citrate synthase: a quantum mechanical molecular mechanical (QM/MM) study. *Proteins: Struct. Funct. Genet.* 27:9–25
109. Naidoo KJ, Brady JW. 1999. Calculation of the Ramachandran potential of mean force for a disaccharide in aqueous solution. *J. Am. Chem. Soc.* 121:2244–52

110. Nilsson L. 1998. Protein-nucleic acid interaction. In *Encyclopedia of Computational Chemsitry*, ed. PVR Schleyer, pp. 2220–29. New York: Wiley
111. Ota N, Brunger AT. 1997. Overcoming barriers in macromolecular simulations: non-Boltzmann thermodynamic integration. *Theor. Chem. Acc.* 98:171–81
112. Ota N, Stroupe C, Ferreira-da-Silva JMS, Shah SA, Mares-Guia M, Brunger AT. 1999. Non-Boltzmann thermodynamic integration (NBTI) for macromolecular systems: relative free energy of binding of trypsin to benzamidine and benzylamine. *Proteins: Struct. Funct. Genet.* 37:641–53
113. Oubridge C, Ito H, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1. 92 angstrom resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 372:432–38
114. Pathiaseril A, Woods RJ. 2000. Relative energies of binding for antibody-carbohydrate-antigen complexes computed from free-energy simulations. *J. Am. Chem. Soc.* 122:331–38
115. Paulsen MD, Ornstein RL. 1996. Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.* 9:567–71
116. Pearlman DA. 1999. Free energy grids: a practical qualitative application of free energy perturbation to ligand design using the OWFEG method. *J. Med. Chem.* 42:4313
117. Perakyla M, Kollman PA. 1997. A simulation of the catalytic mechanism of aspartylglucosaminidase using ab initio quantum mechanics and molecular dynamics. *J. Am. Chem. Soc.* 119:1189–96
118. Pitera J, Kollman P. 1998. Designing an optimum guest for a host using multimolecule free energy calculations: predicting the best ligand for Rebek's "tennis ball." *J. Am. Chem. Soc.* 120:7557–67
119. Deleted in proof
120. Radmer RJ, Kollman PA. 1998. The application of three approximate free energy calculations methods to structure based ligand design: trypsin and its complex with inhibitors. *J. Comput. Aided Mol. Des.* 12:215–27
121. Ranganathan S, Gready JE. 1994. Mechanistic aspects of biological redox reactions involving NADH. V. Am1 transition-state studies for the pyruvate L-lactate interconversion in L-lactate dehydrogenase. *J. Chem. Soc. Faraday Trans.* 90:2047–56
122. Ranganathan S, Gready JE. 1997. Hybrid quantum and molecular mechanical (QM/MM) studies on the pyruvate to L-lactate interconversion in L-lactate dehydrogenase. *J. Phys. Chem. B* 101:5614–18
123. Rao BG, Murcko MA. 1994. Reversed stereochemical preference in binding of Ro 31-8959 to HIV-1 proteinase—a free energy perturbation analysis. *J. Comp. Chem.* 15:1241–53
124. Rao BG, Murcko MA. 1996. Free energy perturbation studies on binding of a-74704 and its diester analog to HIV-1 protease. *Protein Eng.* 9:767–71
125. Rao BG, Tilton RF, Singh UC. 1992. Free energy perturbation studies on inhibitor binding to HIV-1 proteinase. *J. Am. Chem. Soc.* 114:4447–52
126. Rastelli G, Costantino L, Vianello P, Barlocco D. 1998. Free energy perturbation studies on binding of the inhibitor 5, 6-dihydrobenzo[h] cinnolin-3(2H)one-2 acetic acid and its methoxylated analogs to aldose reductase. *Tetrahedron* 54: 9415–28
127. Reddy MR, Viswanadhan VN, Weinstein JN. 1991. Relative differences in the binding free energies of human immunodeficiency virus-1 protease inhibitors—a thermodynamic cycle-perturbation approach. *Proc. Natl. Acad. Sci. USA* 88:10287–91
128. Reiling S, Schlenkrich M, Brickmann J.

1996. Force field parameters for carbohydrates. *J. Comp. Chem.* 17:450–68
129. Reuter N, Dejaegere A, Maigret B, Karplus M. 2000. Frontier bonds in QM/-MM methods: a comparison of different approaches. *J. Phys. Chem. A* 104:1720–35
130. Reyes CM, Kollman PA. 1999. Molecular dynamics studies of U1A-RNA complexes. *RNA Publ. RNA Soc.* 5:235–44
131. Reyes CM, Kollman PA. 2000. Investigating the binding specificity of U1A-RNA by computational mutagenesis. *J. Mol. Biol.* 295:1–6
132. Reyes CM, Kollman PA. 2000. Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J. Mol. Biol.* 297:1145–58
133. Rick SW, Stuart SJ, Berne BJ. 1994. Dynamical fluctuating charge force fields—application to liquid water. *J. Chem. Phys.* 101:6141–56
134. Rick SW, Topol IA, Erickson JW, Burt SK. 1998. Molecular mechanisms of resistance: free energy calculations of mutation effects on inhibitor binding to HIV-1 protease. *Protein Sci.* 7:750–56
135. Roxstrom G, Velazquez I, Paulino M, Tapia O. 1998. DNA structure and fluctuations sensed from a 1.1ns molecular dynamics trajectory of a fully charged Zif268-DNA complex in water. *J. Biomol. Struct.* 16:301–12
136. Roxstrom G, Velazquez I, Paulino M, Tapia O. 1998. Molecular dynamics simulation of a Zif268-DNA complex in water. Spatial patterns and fluctuations sensed from a nanosecond trajectory. *J. Phys. Chem. B* 102:1828–32
137. Sagui C, Darden TA. 1999. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.* 28:155–79
138. Sanner MF, Olson AJ, Spehner JC. 1996. Reduced surface—an efficient way to compute molecular surfaces. *Biopolymers* 38:305–20
139. Schurer G, Lanig H, Clark T. 2000. The mode of action of phospholipase A(2): semiempirical MO calculations including the protein environment. *J. Phys. Chem. B* 104:1349–61
140. Sen S, Nilsson L. 1999. Free energy calculations and molecular dynamics simulations of wild-type and variants of the DNA-EcoRI complex. *Biophys. J.* 77:1801–10
141. Sen S, Nilsson L. 1999. Structure, interaction, dynamics, and solvent effects on the DNA-EcoRI complex in aqueous solution from molecular dynamics simulation. *Biophys. J.* 77:1782–1800
142. Sharp KA, Honig B. 1990. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem.* 94:7684–92
143. Sharp KA, Honig B. 1990. Electrostatic interactions in macromolecules—theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* 19:301–32
144. Simmerling C, Fox T, Kollman PA. 1998. Use of locally enhanced sampling in free energy calculations: testing and application to the alpha→beta anomerization of glucose. *J. Am. Chem. Soc.* 120:5771–82
145. Singh UC, Kollman PA. 1986. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the CH/sub 3/Cl+Cl/sup -/ exchange reaction and gas phase protonation of polyethers. *J. Comp. Chem.* 7:718–30
146. Smith KC, Honig B. 1994. Evaluation of the conformational free energies of loops in proteins. *Proteins: Struct. Funct. Genet.* 18:119–32
147. Smondyrev AM, Berkowitz ML. 2000. Molecular dynamics simulation of dipalmitoylphosphatidylcholine membrane with cholesterol sulfate. *Biophys. J.* 78:1672–80

148. Sotriffer CA, Flader W, Cooper A, Rode BM, Linthicum DS, et al. 1999. Ligand binding by antibody IgE lb4: assessment of binding site preferences using microcalorimetry, docking, and free energy simulations. *Biophys. J.* 76:2966–77
149. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. 1998. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate—DNA helices. *J. Am. Chem. Soc.* 120:9401–9
150. Stanton RV, Hartsough DS, Merz KM. 1995. An examination of a density functional molecular mechanical coupled potential. *J. Comp. Chem.* 16:113–28
151. Stanton RV, Perakyla M, Bakowies D, Kollman PA. 1998. Combined ab initio and free energy calculations to study reactions in enzymes and solution: amide hydrolysis in trypsin and aqueous solution. *J. Am. Chem. Soc.* 120:3448–57
152. Stern HA, Kaminski GA, Banks JL, Zhou RH, Berne BJ, Friesner RA. 1999. Fluctuating charge, polarizable dipole, and combined models: parameterization from ab initio quantum chemistry. *J. Phys. Chem. B* 103:4730–37
153. Stocker U, van Gunsteren WF. 2000. Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMOS96 force field against nuclear magnetic resonance data. *Proteins: Struct. Funct. Genet.* 40:145–53
154. Tang Y, Nilsson L. 1999. Molecular dynamics simulations of the complex between human U1A protein and hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys. J.* 77:1284–1305
155. Thery V, Rinaldi D, Rivail JL, Maigret B, Ferenczy GG. 1994. Quantum mechanical computations on very large molecular systems—the local self-consistent field method. *J. Comp. Chem.* 15:269–82
156. Thomasson KA, Ouporov IV, Baumgartner T, Czlapinski J, Kaldor T, Northrup SH. 1997. Free energy of nonspecific binding of Cro repressor protein to DNA. *J. Phys. Chem. B* 101:9127–36
157. Tidor B. 1993. Simulated annealing on free energy surfaces by a combined molecular dynamics and Monte-Carlo approach. *J. Phys. Chem.* 97:1069–73
158. Tieleman DP, Forrest LR, Sansom MSP, Berendsen HJC. 1998. Lipid properties and the orientation of aromatic residues in OmpF, influenza M2, and alamethicin systems: molecular dynamics simulations. *Biochemistry* 37:17554–61
159. Tocilj A, Schlunzen F, Janell D, Gluhmann M, Hansen HAS, et al. 1999. The small ribosomal subunit from *Thermus thermophilus* at 4.5 angstrom resolution: pattern fittings and the identification of a functional site. *Proc. Natl. Acad. Sci. USA* 96:14252–57
160. Tropsha A, Hermans J. 1992. Application of free energy simulations to the binding of a transition-state-analogue inhibitor to HIV protease. *Protein Eng.* 5:29–33
161. van Gunsteren WF. 1989. Methods for calculation of free energies and binding constants: successes and problems. In *Computer Simulation of Biomolecular Systems*, ed. WF van Gunsteren, PK Weiner, pp. 27–59. Leiden: ESCOM
162. Varani G. 1997. RNA-protein intermolecular recognition. *Acc. Chem. Res.* 30:189–95
163. Wang J, Dixon R, Kollman PA. 1999. Ranking ligand binding affinities with avidin: a molecular dynamics-based interaction energy study. *Proteins: Struct. Funct. Genet.* 34:69–81
164. Wang JM, Cieplak P, Kollman PA. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comp. Chem.* 21:1049–74
165. Wang W, Kollman PA. 2000. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and

continuum solvent model. *J. Mol. Biol.* 303:567–82

166. Wang W, Lim WA, Jakalian A, Wang J, Wang JM, et al. 2000. An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations and sequence analysis. *J. Am. Chem. Soc.* In press
167. Wang W, Wang J, Kollman PA. 1999. What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins: Struct. Funct. Genet.* 34:395–402
168. Warshel A. 1991. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. New York: Wiley
169. Warshel A. 1998. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem.* 273:27035–38
170. Warshel A, Levitt M. 1976. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103:227–49
171. Wu N, Mo YR, Gao JL, Pai EF. 2000. Electrostatic stress in catalysis: structure and mechanism of the enzyme orotidine monophosphate decarboxylase. *Proc. Natl. Acad. Sci. USA* 97:2017–22
172. Yang AS, Hitz B, Honig B. 1996. Free energy determinants of secondary structure formation. III. Beta-turns and their role in protein folding. *J. Mol. Biol.* 259:873–82
173. Yang AS, Honig B. 1995. Free energy determinants of secondary structure formation. I. Alpha-helices. *J. Mol. Biol.* 252:351–65
174. Yang AS, Honig B. 1995. Free energy determinants of secondary structure formation. II. Antiparallel beta-sheets. *J. Mol. Biol.* 252:366–76
175. Zacharias M, Straatsma TP, McCammon JA, Quiocho FA. 1993. Inversion of receptor binding preferences by mutagenesis—free energy thermodynamic integration studies on sugar binding to L-arabinose binding proteins. *Biochemistry* 32:7428–34
176. Zhang YK, Liu HY, Yang WT. 2000. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.* 112:3483–92