

# Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models

Junmei Wang\*, Tingjun Hou† and Xiaojie Xu\*

College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

**Abstract:** Recently, the combination of state-of-the-art molecular mechanical force fields with continuum solvation models enables us to make relatively accurate predictions of both structures and free energies for macromolecules from molecular dynamics trajectories. The first part of this review is focused on the history and basic theory of free energy calculations based on physically effective energy functions. The second part illustrates the applications of free energy calculations on many biological systems, including proteins, DNA, RNA, protein-ligand, protein-protein, protein-nucleic acid complexes, etc. Finally, the prospective and possible strategies to improve the techniques of MM-PBSA and MM-GBSA is discussed.

**Keywords:** Molecular mechanics, continuum solvation model, MM-PBSA, MM-GBSA, free energy, binding free energy.

## I. INTRODUCTION

Molecular mechanics (MM), though its theoretical background is not as solid as quantum mechanics (QM), has broad applications in studying biological systems for its simplicity and efficiency. The harmonic function form (Equation 1), which is widely-used in many popular molecular mechanical force fields, describes molecular energy using bond stretching, bond angle bending, torsional angle twisting as well as non-bonded electrostatic and van der Waals interactions. It is well-known that most biological procedures take place in aqueous solutions. Therefore, solvation effect cannot be neglected in studying the structures and interactions of biological systems. There are two basic ways to take the solvent effect into account: with either an explicit water model or an implicit water model. For the first approach, a biological system is usually immersed in a periodical water box; the second approach does not apply water molecules explicitly. Instead, solvation energy and force due to the solvent effect are calculated using some formulas or by solving some equations numerically or analytically. The widely used implicit solvation models include the generalized Born surface area (GBSA) [1-5] and the Poisson-Boltzmann surface area (PBSA) [6-15].

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

To study an ensemble, rather than a few conformations is important to guarantee the reliability and quality of calculation results. Molecular dynamics (MD) and Monte Carlo (MC) simulations are the two commonly used techniques for sampling conformations. Though MD and

MC have been proven to be successful in studying a variety of properties of small molecules for a long time, it is not until recent years that such techniques are useful in studying complex biological molecules. Since macromolecules have been studied with MD in the late seventies of the last century, there are three eras in the history of molecular simulations on macromolecules. In the first era (1976-1985), one could call the Dark Ages, molecular simulations were typically carried out without including explicit water or the aqueous environment around the macromolecules was represented in a primitive fashion. It was not a surprise if the crystal structures collapsed after MD or MC simulations in this era.

The main theme of the second era (1985-1998) is free energy calculations with free energy perturbation and thermodynamic integration. A typical scenario is like this: putting a water cap or a thin water sphere around the binding site of a biological system, and then mutating a residue or ligand bound to the macromolecule into another. The relative energies can be calculated and then compared to experimental findings. In the late nineties, with the implementation of computationally efficient particle mesh Ewald (PME) algorithm [16] in some molecular simulation

software packages, the electrostatic interaction can be calculated much more accurately and efficiently. It is since then that biological systems in aqueous environment can be modeled in a much more realistic fashion by using a periodic boundary condition.

In the third era (1998-), implicit solvent models, exemplified by GBSA [1-5] and PBSA [6-15], began to be applied in structure and free energy calculations. Although their theoretical background is not as rigorous as that of FEP and TI [17-23], MM-PBSA and MM-GBSA can be used to compute the structures and free energies of macromolecules

\*Address correspondence to this author at the College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China; E-mails: junmwang@pku.edu.cn; xiaojku@pku.edu.cn

†Current address: Department of Chemistry and Biochemistry, UCSD, E-mail: tihou@ucsd.edu

much faster. Moreover, the absolute binding free energy of a ligand or a substrate binding to a biological target can be computed with the two techniques. In the last few years, the two methods have been successfully applied to a wide variety of biological molecules and complexes. Several review papers have been published on both the methodology and applications of MM-GBSA [4] and MM-PBSA [6-9]. In this review, we will focus on the latest applications of the two techniques in studying the structures and free energies of biological systems. The limitations as well as the perspective of the two methods will also be discussed. First of all, the basic theory of MM-GBSA and MM-PBSA will be presented in the next section.

## II. THE MM-PBSA AND MM-GBSA APPROACHES

In the MM-PBSA and MM-GBSA theory, the energy of a molecule is made up of two parts, the gas phase MM energy and the solvation free energy (Equation 2).

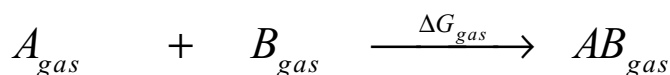
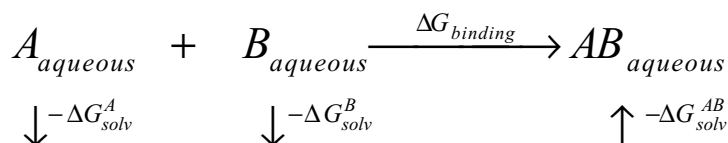
$$G = G_{gas} + \Delta G_{solv} \quad (2)$$

$$\begin{aligned} G_{gas} &= H_{gas} - TS_{MM} \\ &\approx E_{gas} - TS_{MM} \end{aligned} \quad (3)$$

MM-PBSA and MM-GBSA can be applied to calculate the relative free energy of a molecule between two

conformations directly by using Equations 2 and 3. However, to calculate the binding free energy of A + B ? AB, a thermodynamic cycle below must be utilized and the binding free energy is calculated with Equations 4-8.

$\Delta G_{binding}$ , the binding free energy, is made up of two parts, the gas phase molecular mechanical energy  $\Delta G_{gas}$  and the solvation free energy  $\Delta G_{solv}$ .  $\Delta G_{gas}$  is calculated with Equations 5 and 6.  $\Delta E_{inter}$  is typically neglected with an assumption that the intra-molecular energies of the ligand and receptor do not change significantly upon binding (the single trajectory protocol described below).  $\Delta G_{solv}$  is calculated with Equations 7 and 8. The solvation free energy of a molecule is further made up of two parts, the electrostatic interaction energy (polarization energy),  $\Delta G_{PB/\Delta GB}$ , and the non-electrostatic component (non-polar energy),  $\Delta G_{SA}$ . The polarization energy is calculated with either a GB or PB model, while the non-polar energy is simply estimated with solvent accessible surface area (SAS). If the solvation energies ( $\Delta G_{solv}^{AB}$ ,  $\Delta G_{solv}^A$  and  $\Delta G_{solv}^B$ ) and entropy ( $T\Delta S$ ) are omitted, Equation 4 becomes Equation 9. The scoring function described by Equation 9, which only considers the non-bonded intermolecular energy ( $\Delta E_{elec} + \Delta E_{vdw}$ ) between an inhibitor and a receptor, is widely used in some molecular docking programs because of its simplicity. However, in practice, a simple distance dependent dielectric constant instead of the gas dielectric constant is applied in the Coulombic term to roughly account for the screening effect of water molecules.



$$\Delta G_{binding} = G^{AB} - (G^A + G^B) = \Delta G_{gas} + \Delta G_{solv} \quad (4)$$

$$\Delta G_{gas} = \Delta H_{gas} - T\Delta S_{MM} \approx \Delta E_{gas} - T\Delta S_{MM} \quad (5)$$

$$\Delta E_{gas} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdw} \quad (6)$$

$$\Delta G_{solv} = \Delta G_{solv}^{AB} - (\Delta G_{solv}^A + \Delta G_{solv}^B) \quad (7)$$

$$\Delta G_{PBSA/GBSA} = \Delta G_{PB/GB} + \Delta G_{SA} \quad (8)$$

$$\Delta G_{binding}^{Dock} \approx \Delta E_{electrostatic} + \Delta E_{vdw} \quad (9)$$

$$G_{reaction} = -\frac{1}{2} \int \rho(r) \phi(r) dr \quad (10)$$

$$\nabla \cdot \epsilon(r) \nabla \phi(r) - \epsilon(r) \lambda(r) \kappa^2 \frac{k_B T}{q} \sinh \left[ \frac{q \phi(r)}{k_B T} \right] = -4\pi \rho(r) \quad (11)$$

$$G_{GB} = -\frac{1}{2} \left( 1 - \frac{1}{\epsilon} \right) \sum_{i,j}^{atoms} q_i q_j \gamma_{ij} \quad (12)$$

$$\gamma_{ij} = \left( r_{ij}^2 + \alpha_i \alpha_j e^{-r_{ij}^2/d_i \alpha_i \alpha_j} \right)^{1/2} \quad (13)$$

$$G_{SA} = \gamma A + b \quad (14)$$

The polarization energy by PB is the difference in the work of charging a molecule in the gas phase with a dielectric constant of unity and in solution with a dielectric constant of  $\epsilon$  (80 for water). The work of creating the charge distribution in a dielectric media is called the reaction field energy  $G_{reaction}$ , which can be determined with Equation 10. The electrostatic potential  $\phi(r)$  is calculated with Poisson-Boltzmann equation (Equation 11). Here  $\rho(r)$  is the free charge density;  $\Delta$  is the dielectric constant of the media;  $q$  represents the electronic charge;  $k_B$  is the Boltzmann constant;  $\chi(r)$  is a simple switching function, which is zero in regions inaccessible to the electrolyte and one otherwise; and, the Debye length, is a function of ionic strength of the electrolyte solution.

For a conducting sphere with a charge spreading out uniformly on the surface, the Poisson-Boltzmann equation has an analytical solution and the equation used to calculate the polarization energy of this ideal system is so-called the Born equation. The GB model is an extension of the Born model for a molecule with an arbitrary shape. The polarization energy of the GB model is calculated with Equation 12.  $\gamma_{ij}$  in this equation has a unit of inverse length and a widely-used functional form of it is given by Equation 13, where  $\gamma_{ij}$  is the inter-atomic distance;  $\alpha$  is effective Born radius; and  $d$  is a parameter.

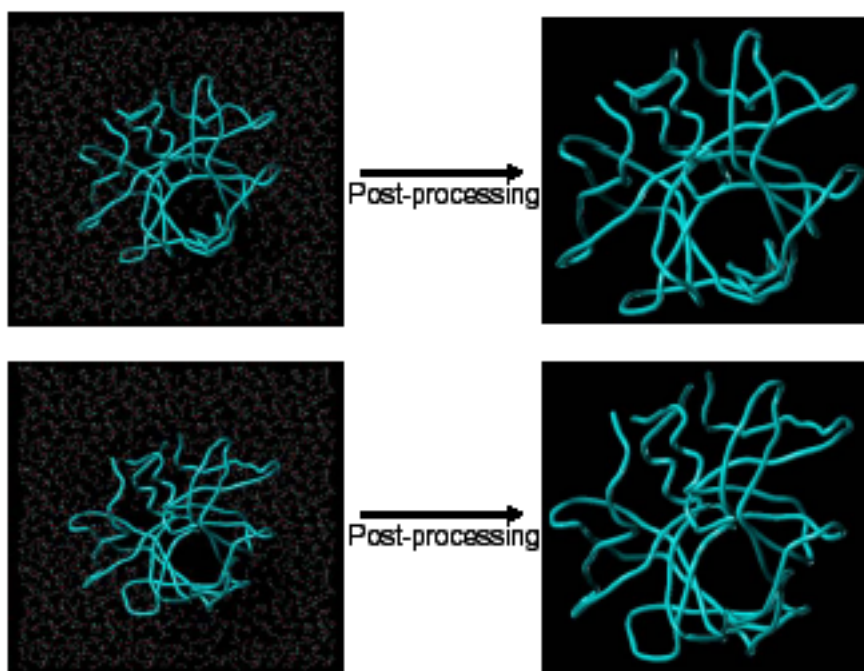
The non-polar energy in both models is the leftover excluding the polarization energy from the solvation free energy. It is comprised of the free energy required to form the solute cavity in a solvent that mainly accounts for the entropy penalty associated with the reorganization of solvent molecules around a solute, and the van der Waals interaction between the solute and solvent. The non-polar energy  $G_{SA}$  is simply estimated by multiplying a constant  $\gamma$  (the so-called surface tension) to the solvent accessible surface area ( $A$ ) of the solute (Equation 14) plus an intercept  $b$  [24-25].

By applying MM-PBSA and MM-GBSA, all the solvent coordinates are implicitly integrated out and the free energy between two “end points” can be calculated directly instead of calculating the relative free energies of a set of less interesting intermediate states along the mapping coordinate. This explains why MM-PBSA and MM-GBSA are much more efficient than FEP and TI. In contrast to LIE (linear interaction energy) [26-30], another popular method of free energy calculations, MM-PBSA/GBSA is more promising under some circumstances owing to the fact that it does not require a training set to derive empirical parameters for different biological systems in the first place, while LIE does. Therefore MM-GBSA and MM-PBSA are attractive methods for directly estimating binding free energies.

Nevertheless, each free energy component may have intrinsically considerable uncertainties and the accumulated errors could be intolerable if the calculation protocol is not

well-designed. The widely-used procedure of calculating the relative free energy between two conformations of a molecule is as follows: for each conformation, MD simulations is typically carried out in a periodic box with water and counterions, the long-range electrostatic effects is correctly represented with PME, and a set of representative structures are collected after the system is well-equilibrated. Then post-process is carried out for these saved structures by removing any solvent and counterion molecules and then calculating the free energy according to Equations 2 and 3. As to the calculation of the binding free energy of  $A + B \rightarrow AB$ , a typical procedure is to run MD simulation for the complex in explicit water and to save a set of conformational snapshots; then the solvent and counterion molecules are stripped away and the free energies of the complex, the protein or DNA, and the substrate or the ligand, are calculated separately. This procedure, which may be called “single trajectory” protocol, assumes that there are no significant changes on the conformations of both A and B upon binding, and the gas phase molecular mechanical energy component,  $\Delta E_{gas}$  in Equations 5 and 6, is therefore the inter-molecular energy between A and B in C, the complex. When A and B undergo significant conformational change upon binding, the “single trajectory” protocol may cause substantial errors. If this is the case, one should apply an “individual trajectory” procedure, which samples A, B and C separately to calculate the binding free energy. Other simplified sampling protocols will be presented in the next section. It is critical that the same set of charges is used in both the gas phase molecular mechanical energy calculations and the solvation free energy calculations to ensure efficient cancellation of most calculation errors. The conformational entropy of a molecule,  $TS_{MM}$  in Equation 3, can be estimated by quasi harmonic analysis of the MD trajectories or by normal-mode analysis on selected snapshots [10]. The conformational entropy is likely to be much smaller than the other two terms in many applications of estimating relative free energies.  $TS_{MM}$  may be omitted if one does not need to calculate the absolute binding free energies.

The above procedures are illustrated in Fig. 1. It is demonstrated by many examples that the calculated  $\Delta G$  values are agreeable with experimental findings. In the following sections, the applications of MM-PBSA and MM-GBSA in free energy calculations for macromolecules are presented. We want to emphasize that we did not intent to collect all the papers on this topic, but some representatives to demonstrate what kind of problems this technique can address and for what kind of systems. Considering several reviews have been published on this topic in 2000-2002 [4,6-9], this review mainly focuses on the publications after the year of 2001. We also want to point out that there are many theoretical approaches for free energy calculations, which include free energy perturbation [17-23], linear interaction energy [26-30], hybrid-LIE/GB [31-32], hybrid-LIE/PB [33], -dynamics [34-35], the generalized Bjerrum approach [36] and ligand interaction scanning [37], double-coupling method [38], potential mean force based approach [39], wormhole Monte Carlo method [40], statistical mechanical method based on molecular correlation functions [41] and so on. The comparisons of these methods to MM-PBSA and MM-GBSA are beyond the scope of this paper.



**Fig. (1).** The protocol of calculating relative free energy between two conformations. The upper part was a crystal structure of a protein (PDB entry: 1aap) and the lower part was a molecular dynamic snapshot of the same molecule.

### III. APPLICATIONS OF MM-PBSA AND MM-GBSA

#### A. Relative Free Energies of Macromolecules

##### 1. Proteins

It is generally believed that the crystal structures are the global minima of macromolecules, such as proteins, DNA and RNA under most of circumstances. Theoretically, the crystal structures therefore have the lowest free energies with Equations 2 and 3 in a conformational decoy. Ideally, the smaller the RMS deviation of a conformation to the crystal ones, the lower its free energy is. If it is the case, one can reliably rank the predicted macromolecular structures and pick up the native one if it is among the decoy structures. In other words, MM-PBSA can be applied as a powerful tool in assessing the predicted structures and an essential technique in macromolecule modeling. If the hypothesis that the free energy landscape has a funnel shape [42] is true, the farther the structures lie from the native state, the less linear the relationship should be. Typically, the linear relationship is obvious only for structures that are within 5 Å of the native structure. In practice, the relationship between the calculated free energy and the RMSD may not be obvious even for conformations immediately surrounding the native states due to errors from all kinds of sources including the force field. The following is some successful stories.

Lee, Baker and Kollman calculated the relative free energy  $G$  for two small proteins, the 36-mer villin headpiece domain (HP-36) and the 65-mer structured region of ribosomal protein (S15) [11]. Starting from the native structure and a set of protein models by *ab initio* approach Rosetta [43], MD simulations were carried out followed by cluster analysis on the saved trajectories. Then the MM-PBSA free energy calculations were performed for the collected snapshots in each conformational cluster. The results were very encouraging: those conformational families

with the lowest average free energies also contained the best C $\alpha$  RMSD structures (1.4 Å for S15 and HP-36 core) and the lowest average C $\alpha$  RMSDs (1.8 Å for S15, 2.1 Å for HP-36 core); the ranking of the average free energies correlated well with the average C $\alpha$  RMSDs (the Spearman rank correlation coefficient is 0.77 for HP-36 and 0.83 for S15, respectively), whereas the Rosetta scores correlated poorly with the C $\alpha$  RMSDs.

In another work, they studied 12 small, single-domain proteins (four alpha (1gab, 1utg, 1uxd and 1pou), four beta (1sro, 1qyp, 1vif and 2cdx) and four mixed topologies (1leb, 2ptl, 5icb and 5znf)) by MD simulations and MM-PBSA free energy calculations [44]. They intended to answer the following questions: (1) Are the native proteins stable during the MD simulations? (2) What is the rank of the native structure in a conformational decoy? (3) How well do the MM-PBSA free energies correlate to native similarity measured by RMSD? (4) Can more native similarity be improved after the structural refinement by minimizations or simulations? The initial structures for MD simulations were generated by Rosetta. The MM-PBSA free energy of a protein model was calculated with a variety of parameter sets (force field parameters, the dielectric constant in  $\Delta E_{electrostatic}$  and  $\Delta G_{PB}$  calculations, and the surface tension  $\gamma$  in Equation 14 to estimate the non-polar contribution of solvation free energy). Here is what they found: (1) the native structures were reasonably stable along the 1 ns MD simulations and most of the C $\alpha$  RMSDs were under 2.0 Å; (2) With the “standard” parameter set, eight out of the twelve proteins, with exception of 1gab, 1uxd, 1pou and 2ptl, had their native structures ranked No 1 among a set of 36-77 decoy conformations based on the MM-PBSA energies. Moreover, the average RMSD values of the conformations having higher ranks than the natives were 2.08, 1.80, 3.70, and 11.01 Å for 1gab, 1uxd, 2ptl and

1pou, respectively. (3) For all the proteins except 1sro, 1pou and 5icb, a good linear correlation coefficient between C RMSDs and the MM-PBSA energies was achieved, but only for structural families that were less than 5 Å of the native state. This indicates that C $\alpha$  RMSD and an effective free energy such as MM-PBSA is only linear near the native state, and the relationship weakens dramatically beyond 5 Å C $\alpha$  RMSD. (4) In general, further relaxation of a protein structure can only slightly improve the native similarity.

Lee and Kollman applied the same protocol to study other 15 small single-domain proteins [14]. It was found that decoys were less energetically favorable than the native conformations for nine of the ten X-ray structures and none of the five NMR structures if only simple minimizations were conducted. However, all the 15 proteins had the lowest predicted free energies after short 150 ps MD simulations, which indicated that MD simulations is much more efficient in eliminating possible bad contact in NMR structures. Nicely, a strong correlation ( $r^2 = 0.86$ ) was found between the protein length, in term of the number of amino acid residues, and the predicted free energy of unfolding. The unfolding free energy is the difference between a protein's native state and its fully extended state, which is entirely alpha as suggested by Lee and Kollman. The unfolding free energy provides a useful criterion to evaluate how close a model protein represents its native even when the experimental structure is absent.

Loop structure prediction is one of the most important issues in protein modeling. Fogolari and Tosatto [45] recently applied MM-PBSA in combination with colony free energy proposed by Xiang and Honig *et al.* [46] to discriminate native or native-like loops from their decoys. Good correlations were found between the estimated free energies and the similarity to the native structures for their four test sets. The application of colony energy greatly hampered the strong dependence of MM-PBSA energy on minor conformational changes.

Santa *et al.* studied the  $\alpha$  r  $\beta$  and  $\beta\beta$  (Beta-Beta) conformations of tetrapeptides SALN and its mutants [47]. The  $\alpha$  r  $\beta$  turn was predicted to be slightly more stable than the conformation according to the MM-PBSA free energies. The authors suggested that r turn may be the most common turn type in peptides; it may be readily formed in aqueous solution and thereby plays important roles in the protein folding process. MM-PBSA has also been applied by Ma *et al.* to investigate the free energy landscapes of  $\beta$ -hairpin G peptide and its isomers [48]. They concluded that the changes in the sequence strongly modulated the relative stabilities of topologically similar regions in the energy landscape, rather than redefining the topology space.

## 2. Nucleic Acids

It is found that DNA is likely to adopt a right-hand double helix B form (B-DNA) under physiological conditions. However, repetitive sequences, such as poly(GC), poly(AT), are easy to take other DNA forms. The poly(AT) sequences are probably those exhibiting the widest range of accessible structures, including the antiparallel Hoogsteen duplex (apH). Recently, Cubero *et al.* [49] systematically studied a set of repetitive AT sequences, d(AT)n/2 for n = 4, 6, 8, 10, 12, 14 and 16, using MD

simulations and MM-PBSA/GBSA free energy calculations. They found that polyd(AT)n in both B and apH forms were stable along the nano-second MD simulations in aqueous solution (the RMSDs were smaller than 2.0 Å). Both their MM-PBSA and MM-GBSA results suggested that the relative free energies without counting the entropy contribution were negligible for the two forms. If entropy was considered, the B form was energetically more favorable than the apH helix and the difference slightly increased with the increase of the length of the oligonucleotide. As a conclusion, the population of the two helices in solution might depend on the existence of cofactors, specific hydration waters, and entropic considerations; the slight difference of structural, dynamical and energetic properties of the apH helices from those of the B form may be the basis for proteins and drugs to distinguish between these two helical structures.

Recently, Yan *et al.* carried out MD simulations on a pair of 11mer double-strand DNA that have an adenine residue covalently modified through reaction with mutagenic and tumorigenic metabolites of benzo[ $\alpha$ ]pyrene [50]. The two adducts, 10S(+) and 10R(-) trans-anti-[BP]-N2-dG, and 10S(+) and 10R(-) trans-anti-[BP]-N6-dA, are stereoisomers, which makes it suitable to calculate the relative free energy with Equations 2 and 3 using the sampling protocol in Fig. 1. The 10R(-) trans-anti-[BP]-N6-dA was 13 kcal/mol more stable than its stereoisomer according to the calculated MM-PBSA free energies. The enthalpy difference, which was about 10 kcal/mol, agreed quite well with observed differences in thermodynamic stability. In another work, Yan *et al.* studied a pair of guanine adducts, 10S(+) and 10R(-) trans-anti-[BP]-N2-dG, using a similar protocol [51]. The computed enthalpy difference ( $\Delta\Delta E_{\text{gas}} + \Delta\Delta G_{\text{PB}}$ ) between the guanine adducts (2.5 kcal/mol) was reasonably consistent with the experimental data based on DNA duplex formation; and the  $\Delta\Delta G$  was -0.9 kcal/mol, indicating that the two adducts had essentially equal stabilities. Another pair of DNA adducts, 1R(+) and 1S(-) trans-anti-B[c]Ph-N6-dA and 1R(+) trans-anti-B[c]Ph-N6-dA were studied by Wu *et al.* with a similar approach [52].

## B. Binding Free Energies

The combination of molecular docking, molecular dynamics simulations and MM-PBSA/GBSA free energy calculations enables us to address many problems in both structural biology and computer-aided drug design. With these techniques one can model complex structures of biological systems, calculate the binding free energies, elucidate the molecular interaction mechanism and identify the main factors and / or "hot spots" that make substantial contributions to the interaction by performing energetic component analysis. In the following sections, we will present some applications of MM-PBSA and MM-GBSA in this field for a variety of biological systems.

### 1. Protein-Ligand Interactions

Recently, we successfully applied molecular docking combined with MM-PBSA to determine the binding mode of HIV-1 RT/efavirenz [13]. In this blind test, not only the calculated binding free energy was in good agreement with the experiment, but also the crystal structure, which was

released after our manuscript had been submitted, was well-predicted by the combination of molecular docking and molecular dynamic simulations (the RMSD of the non-nucleoside reverse transcriptase inhibitor and 54  $\alpha$ -carbons of the key residues around the binding site was 1.1 Å). Our strategy of modeling protein-ligand complex is described as follows: docking the ligand into the receptor; performing the cluster analysis and selecting a representative docking pose from each cluster (up to 10), then running MD simulations for each docking pose followed by MM-PBSA analysis. The pose that has the most favorable binding free energy is the most reasonable complex structure. The same target was also studied by Weininger *et al.* and a very good correlation was achieved between the MM-PBSA binding free energies and binding affinities estimated by IC<sub>50</sub> for efavirenz and a set of benzoxazinone derivatives [53].

Huo *et al.* studied a set of cathepsin D (CatD) complexes by molecular docking, MD simulations and MM-PBSA calculations [54]. They were able to reproduce the experimental binding affinities for seven inhibitors of CatD with an average error of 1.0 kcal/mol and a correlation coefficient of 0.98, in contrast to the correlation coefficient of 0.2 of the docking scores. The ligand conformation that was found in an X-ray structure of the peptide (pepstatin)/cathepsin complex was successfully identified in the MD simulations of the CatD inhibitors. Kuhn *et al.* reported MM-PBSA analysis on nine diverse biotin derivatives in complex with avidin [55]. Although the absolute binding free energies were 3.3 kcal/mol away from the experiment in average, a nice correlation coefficient ( $r^2 = 0.92$ ) was achieved between the calculated and the experimental binding free energies.

Wang and Kollman also studied HIV-1 protease's drug resistance using MM-PBSA [56]. First they calculated the binding free energies of five marketed drugs and one substrate. Then the free energy contribution of each residue was analyzed. Finally, they suggested a mechanism for drug resistance: if the large free energy contribution came from a not well conserved residue, in another word, this residue was unimportant or the substrate could tolerate the mutations for viral activity, the mutations at that site would not affect the function of the protease but would be able to significantly reduce the inhibition of drugs; and as a result, the mutations would cause drug resistance. In another study, they applied MM-PBSA in analyzing the interactions between the Sem-5 SH3 domain and its ligands, a set of *N*-substituted peptoids at site -1, 0 and 2 [57]. They found that the calculated relative binding free energies (without the contribution of entropy) correlated well with the experimental data. They also examined the effect of different dielectric constants and different ligand charge methods on MM-PBSA binding free energies. For the first molecule set, which included the wild type and seven mutants for which mutations occurred at different sites, the correlation coefficient squares between the calculated and experimental were similar for  $n = 1$  and  $n = 4$  ( $r^2 = 0.88$ ), whereas  $n = 4$  gave significantly better results than  $n = 1$  ( $r^2 = 0.78$  versus  $r^2 = 0.21$ ) for nine mutants substituted at site -1. They found that AM1-BCC [58-59], which was a much efficient charge method and could be calculated with the antechamber module in the AMBER packages for arbitrary organic molecules, gave comparable results to those applying the RESP [60-61] charges in MM-PBSA

calculations. Finally, Wang *et al.* found that the binding free energies calculated from computational mutagenesis using the wild-type peptide trajectory correlated poorly with the experimental data ( $r^2 = 0.34$ ).

Recently, Swanson and McCammon [62] took the FK506 binding protein (FBP-12) and the ligand 4-hydroxy-2-butanone as an example to calibrate the MM-PBSA method for end-point free energy calculations. Instead of applying normal model analysis, the entropic component of the binding free energy was calculated with quasiharmonic analysis. The calculated binding free energy was in reasonable agreement with experiment (-7.4 versus -4.5 kcal/mol).

Hou and Xu *et al.* applied both MM-PBSA and LIE to calculate the binding free energies of eight hydroxamate inhibitors of gelatinase-A [63]. A good correlation was achieved between the predicted binding free energies and the experimental data ( $r = 0.84$  and  $q = 0.78$ ). The absolute binding free energies were reasonably predicted with an average unsigned error of 2.9 kcal/mol. In contrast, the best LIE model achieved a  $q^2$  of 0.83. However, the LIE model had three fitting parameters and may not be transferable to other systems. Díaz *et al.* recently studied the molecular interactions between TEM-1  $\beta$ -Lactamase and cephalothin (CEP) and benzylpenicillin (BP) by MD simulations and MM-PBSA calculations [64]. The initial structure of lactamase/CEP was suggested by AutoDock. The predicted relative binding free energies between the two ligands ranged from 1.8 to 5.7 kcal/mol for different computational protocols, favoring benzylpenicillin. Interestingly, the result of the standard MM-PBSA was consistent with that of the semi-empirical quantum chemical PBSA. The absolute MM-PBSA binding free energies were more than 10 kcal/mol negative than the experiment, perhaps because of the inaccurate solvation free energies of ligands.

Recently, Brigo *et al.* [65] carried molecular docking, explicit solvent MD simulations and MM-PBSA calculations for HIV-1 integrase in complex with L-731988, one of the most active molecules of the class of  $\beta$ -diketo acids. To study the molecular mechanism of drug resistance of T66I/M154I to the inhibitor, they docked the ligand to two protein conformations, which were chosen from prior MD trajectories and orientated differently. Then they performed MD simulations on the wild type and the mutants of HIV-1 integrase in complex with L-731988. Significant differences were observed in the mobility of HIV-1 integrase catalytic loop (residues 138-149). They also identified Gln62 as a hot spot that played an important role in the interactions between the inhibitor and the protein.

von Langen *et al.* studied the binding of five steroids to human glucocorticoid receptor (hGR) through homology modeling, docking, MD simulations, and free energy calculations [66]. They found the binding free energies with both MM/PBSA and FlexX could discriminate strongly and weakly binding ligands. Both methods recognized cortisol, which had a nearly perfect steric and electrostatic complementarity with the hGR binding pocket as the endogenous ligand of the hGR *in silico*. Schwarzl *et al.* [67] recently docked six benzamidine-like ligands to trypsin, and then calculated the binding free energies with a scoring function described as Equation 2 except that the van der

Waals energy  $E_{vdw}$  was scaled down by 85 %. The RMS difference between the calculated and the experiment was 0.55 kcal/mol. The binding free energies of seven aliphatic cyclic ureas to HIV-1 protease were calculated using a predominant states method and an MM-PBSA by Mardis *et al.* [68]. They found that the MM-PBSA binding free energies could reproduce the observed U-shaped trend of binding free energy as a function of aliphatic chain length. However the GBSA, which yielded a much smaller change in solvation free energy with chain length trends, could not reproduce the experimental binding affinity trend. Peräkylä *et al.* performed MD simulations and MM-PBSA analysis on anti-progesterone antibody DB3 in complex with two steroids [69]. The relative binding free energy of the two steroids, PRG and 5AD, was in fine agreement with the experimental energy, 1.29 kcal/mol.

Site-directed mutagenesis has been widely used in the study of protein functions. It can be used to identify the “hot spots” that make substantial contributions to receptor/ligand binding. Computational mutagenesis is attractive due to its efficiency both timely and financially. Recently, a computational scanning mutagenesis method has been developed in Kollman’s group [70-71]. The basic idea is to firstly run MD simulations on a wild type complex, then map and alert the coordinates of the new residues, and finally perform MM-PBSA analysis on the new species. This protocol, in principal, can work at least as long as the mutation does not cause significant conformational change in the binding interface and/or destroy the important interaction network with other residues, such as salt bridges or disulfide bonds. Because alanine is the second smallest natural amino acid residue and not as flexible as glycine, it is not a surprise that residues are typically mutated to alanine in mutagenesis studies. The technique of computational scanning mutagenesis can be applied as a tool to optimize the interacting species for the binding, or as a ranking tool in high throughput screening of peptide/protein design. Computational alanine scanning has been successfully used in many protein/ligand systems, which include P53-binding domain of oncoprotein Mdm2 in complex with 12-residue N-terminal stretch of tumor suppressor protein p53 (Massova *et al.*) [70] and human growth hormone complexed with its receptor (Huo *et al.*) [71]. In the later case, Huo *et al.* could reproduce the experimental  $\Delta\Delta G_{\text{binding}}$  with an average unsigned error of  $\sim 1$  kcal/mol for the alanine mutations of hydrophobic residues and polar/charged residues without buried salt bridge by using a single MD simulation trajectory. They also found that the minimization protocol described below was not able to discriminate the “hot spots” of binding free energy from the non-“hot spot”.

The dynamics of buried water molecules in the cavities of apolipoprotein E were studied by MD simulations and MM-PBSA free energy calculations [72]. The calculated electrostatic component of the binding free energy of the five crystal buried water molecules that exchanged in the course of the simulations ranged from -4.8 to -1.4 kcal/mol.

## 2. Protein-Protein, Protein-Peptide Interactions

What is the overall guideline to use single and separate trajectory sampling protocols in protein-protein interaction studies? Noskov *et al.* [73] designed three protocols to

calculate the MM-GBSA binding free energies for two systems: trypsin complexed with bovine pancreatic trypsin inhibitor (BPTI), and the fragment variable (Fv) region of mouse monoclonal antibody, D1.3, bound to hen egg-white lysozyme (HEL). For both systems, the crystal structures of bound and two unbound proteins were available. In protocol 1, unbound receptor and ligand structures were minimized prior to MM-PBSA analysis; in protocol 2 and protocol 3, the bound receptor and ligand were minimized without and with constraint on all heavy atoms prior to MM-PBSA analysis, respectively. The result showed that the GGB values with protocol 1 were in excellent agreement with the experiments (-11.4 versus -11.4 kcal/mol for D1.3/Hel and -18.6 versus -18.1 for BPTI/trypsin). The protocol 3 was superior to protocol 2 in reproducing the experimental absolute binding free energies, but inferior to protocol 1. This indicated that simple minimizations could not bring the bound ligand or receptor to their global minima; what is more, the error cancellation became less efficient.

Recently, Gohlke and Case applied both MM-GBSA and MM-PBSA to calculate the protein-protein interaction energies of H-Ras/C-Raf1 and H-Ras/RalGDS [74-75]. With the separate trajectory protocol, the calculated binding free energies were in fair agreement with the experimentally determined values (-15.0 (modified GB model described by Jayaram [5]) versus -9.6 kcal/mol for Ras-Raf; -19.5 (GB) versus -8.4 kcal/mol for Ras/RalGDS). The experimental and calculated relative binding free energies between the “hot spot” residues and their alanine mutants yielded an obvious correlation with  $r^2$  of 0.55 and 0.46 for Ras-Raf and Ras/RaGDS, respectively. They also found that different PB and GB models/protocols could produce substantially different values for the absolute binding free energies. Thus, a delicate computational protocol that balances the different energetic and entropic contributions to maximize error cancellation is critical in absolute binding free energy calculations.

Wang *et al.* recently studied HIV protease dimerization using MM-PBSA [76]. They firstly calculated the binding free energies between the wild-type HIV protease with the two catalytic aspartic acid residues at different protonation states. The finding that the double ionic state had the most favorable binding free energy was consistent with the experiment. They also developed a qualitative geometrical criterion to seek mutations that could affect dimerization free energy and then used a rapid, minimization-based method to evaluate their MM-PBSA dimerization free energies. Several new mutants that might further stabilize heterodimer stability were identified. With a similar approach described in Ref.13, Wu *et al.* modeled the complex of scorpion toxin ScyTx, a 31-residue protein bound to a small conductance calcium-activated potassium channel rsk2 [77]. The NMR structures of ScyTx were docked to a homology structure of rsk2 with ZDOCK, followed by MD simulations on four binding poses. The best pose that had the most favorable MM-PBSA binding free energies was then applied to perform computational alanine-scanning. The mutagenesis result was consistent with the experimental findings.

MD simulations and MM-PBSA analysis were carried out by Suenaga *et al.* for SH2 domain of Grb2 and ErbB phosphotyrosyl peptides [78]. The calculated binding free

energies of nine peptides were in excellent agreement with the surface plasmon resonance (SPR) experimental data ( $r = 0.92$ ). Component analysis of the calculated binding free energies revealed that van der Waals interaction between the Grb2 and the phosphotyrosyl peptide was the dominant factor for specificity and binding affinity. Recently, they successfully conducted MD simulations followed by MM-PBSA free energy calculations to identify the binding mode of eight ErbB3 receptor-derived phosphotyrosyl peptides in complex with the SH2 domain of the p85 subunit of phosphatidylinositol 3-kinase [79]. They found that some peptides favored the N1 binding site in the N-terminal region, while the others favored the N2 binding site in the C-terminal region as indicated by the MM-PBSA binding free energies at both sites. An excellent agreement between the calculated and the experimental binding free energies was achieved with a correlation coefficient of 0.91.

Recently, the bindings between the Abl SH3 domain and 35 peptide ligands (10 binders and 20 non-binders) were analyzed using MD simulations and MM-PBSA calculations by Hou *et al.* [80]. The calculated binding free energies correlated well with the rank order of the binding peptides and clearly distinguished binders and non-binders. Free energy component analysis revealed that the van der Waals interactions dictated the binding strength of peptides while the binding specificity was determined by the electrostatic interaction and the polar contribution of desolvation. The binding motif of the Abl SH3 domain was then determined by a virtual mutagenesis (VM) method, which mutates the residue at each position of the template peptide to all other 19 amino acids and calculates the binding free energy difference between the template and the mutated peptides using MM/PBSA. A single position mutation free energy profile (SPMFEP) was thus established and used as a scoring matrix to search peptides recognized by the Abl SH3 domain in the human genome. Ten out of the thirteen experimentally-determined binding partners of the Abl SH3 domain were identified from ~ 6.2 10<sup>7</sup> decapeptides in the SWISS-PROT database. This application demonstrated that the combination of MD simulations, MM-PBSA calculations and virtual mutagenesis would be a powerful tool to identify possible binding partners of the modular protein domains.

Recently, Myshkin *et al.* [81] studied the protein-protein interactions between plastocyanine (Pc)/photosystem I (PSI) using a set of docking programs, including GRAMM, FTDOCK, DOT and AUTODOCK. Then MM-PBSA free energy calculations were performed on the docked complexes that were best consistent with the available biological information. The free energy rank of the wild-type Pc, as well as the hydrophobic patch Tyr<sub>12</sub>Gly and Pro<sub>14</sub>Leu Pc mutants was in agreement with the experimental mutagenesis result. The neuregulin/ErbB system is a growth factor/receptor cascade that has been proven to be essential in the development of the heart and the sympathetic nervous system. Recently, Luo *et al.* [82] carried out MM-PBSA analysis for two complexes: NRG-1 $\beta$ /ErbB3 and NRG-1 $\beta$ /ErbB4, which was constructed against the homologous proteins. The specificity of ligand-receptor recognition mechanism was also elucidated by computational alanine scanning mutagenesis in the binding site of NRG-1 $\beta$ . Some

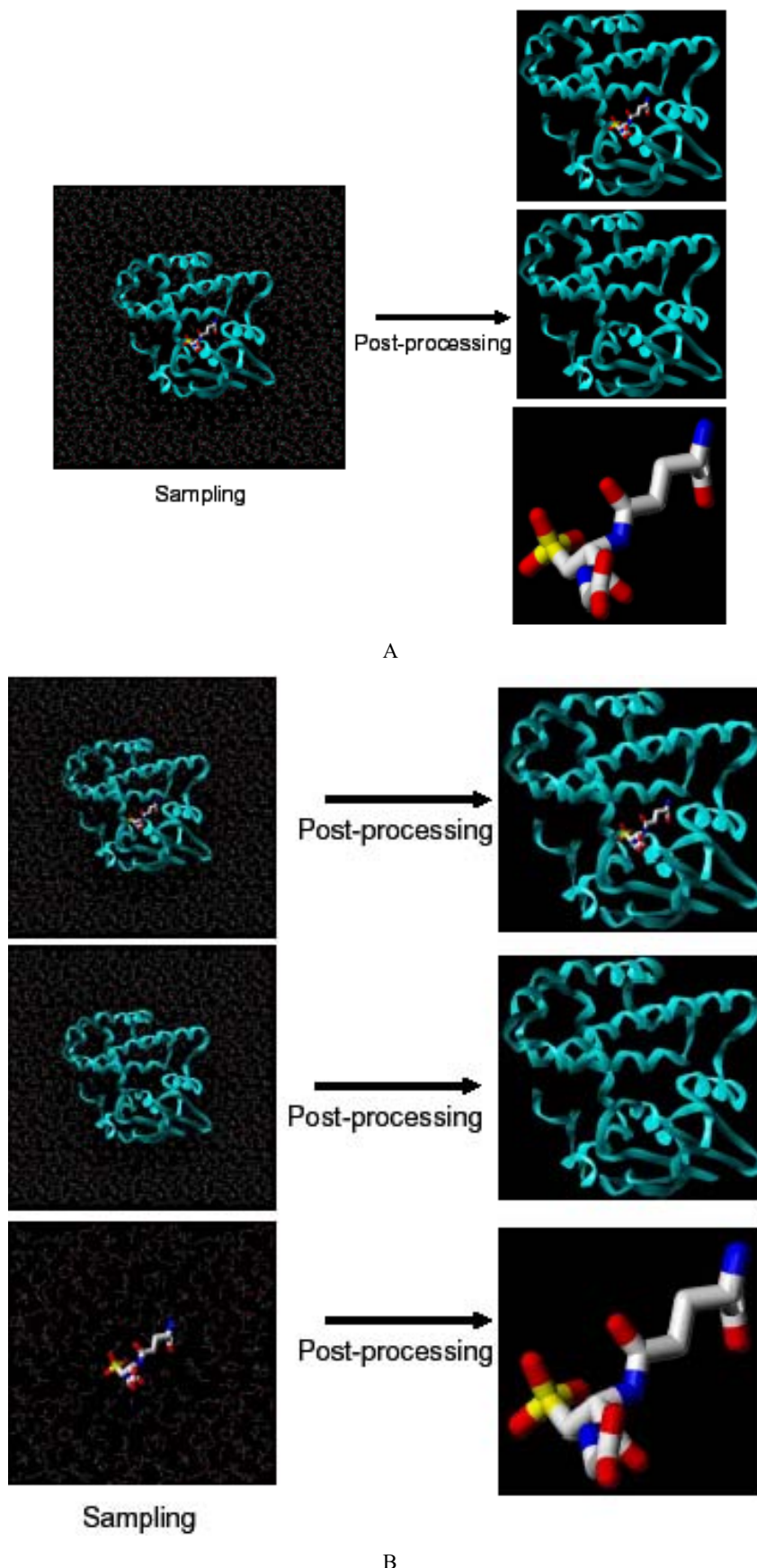
“hot spots” were identified and the MM-PBSA binding free energies of NRG-1 $\beta$  mutants binding to ErbB3 and ErbB4 were in agreement with the experimental data. The computational mutagenesis result was useful in designing mutagenesis experiment to further improve the binding affinity and optimize the specificity of NRG-1 $\beta$  binding to ErbB3 and ErbB4.

A coiled-coil protein composed of two oligo-peptides was studied by MM-PBSA [83]. The calculated binding free energy, the difference between the energies of coiled-coil complex and two  $\alpha$ -helices, was -87.0 kcal/mol, i.e. about -1.2 kcal/mol per amino acid residue. Politicelli *et al.* studied Cruzain S2 in complex with small peptides by MD simulations followed MM-PBSA analysis [84]. The calculated absolute binding free energies were consistently overestimated for about 10.0 kcal/mol. MD and MM-PBSA free energy calculations were applied to study the formation of amyloid  $\beta$  dimer by Urbanc *et al.* [85].

### 3. Protein-Nucleic Acid Interactions

MD simulations were carried out on the bovine immunodeficiency virus BIV Tat-TAR complex by Reyes and Kollman [86]. They ran MD simulations on the native complex and then calculated the binding free energies of a set of mutants with two simple post-processing protocols. In the first protocol, a representative MD structure, such as the last snapshot, was taken to calculate the binding free energy of mutants after a series of minimizations with a distance-dependant dielectric. The second was a generalized alanine-scanning procedure to generate mutant structures directly from the wild-type MD trajectory by altering the corresponding coordinates. Then MM-PBSA post-processing was performed on both the native and the mutated topologies without further minimizations. This second protocol should work, in principle, at least as long as the involved mutant topology does not make the residue larger. Seven mutations on the Tat peptide were carried out with the first protocol and the calculated relative binding free energies  $\Delta\Delta G$  were in reasonable agreement with experiment. Three among the seven computational mutations were also conducted with the second protocol and the calculated relative free energies were similar to those obtained by using the first protocol. In another work, Reyes and Kollman carried out MD simulations on the spliceosomal protein U1A that bound to both an internal loop (IL) and a hairpin loop (HL) of a comparable sequence [87]. In many cases, RNA-protein complexes are formed by an “adaptive binding” mechanism, wherein both molecules undergo significant conformational changes upon binding. Thus, the commonly used protocol of only sampling complex conformations is not adequate. Instead, one needs to run MD simulations for A, B and C in Fig. 2 separately, and apply Equations 2 and 3 to calculate the free energy for each species with its own trajectory. The adaptive free energy of A or B, which is the free energy of conformational change that accompanies complex formation, can also be calculated:  $\Delta\Delta G_{adaptive} = \Delta G_{bound} - \Delta G_{free}$ , where  $\Delta G_{free}$  is the free energy in free bound state and  $\Delta G_{bound}$ , the free energy in bound state, is calculated with the monomer structures taken from the complex trajectory. As to the system of U1A-RNA, the adaptive free energies of U1A were 12.62 and 11.28





**Fig. (2).** The protocols of calculating binding free energies based on (a) single trajectory and (b) individual trajectory. The complex, the protein and the ligand were taken from the protein data bank (PDB entry:1glp).

kcal/mol for IL and HL binding, respectively; whereas the adaptive free energies of IL and HL RNA were 7.93 and 8.73 kcal/mol, respectively. The calculated absolute binding free energies of both U1A-hairpin RNA and U1A-internal loop RNA ( $\Delta G \sim -5$  kcal/mol) were close to those found experimentally ( $\Delta G \sim -12$  to  $-14$  kcal/mol). In contrast, the binding free energies calculated with single trajectories of complexes were more than 10 kcal/mol more negative than experiment. Computational mutagenesis [88] were carried out with a set of hierarchical sampling approaches (from a simple minimization on the mutant built from the experimental structure, to mutation on a set of snapshots from the wild-type trajectory by atomic coordinate removal, to the “standard” sampling protocol of running separated trajectory for each mutant) to calculate the relative binding free energies of different U1A-hairpin RNA mutants. The calculated relative binding free energies were in good agreement with experimental studies and the mutations that abolished and improved binding were verified.

The electrostatic and non-electrostatic effect of protein-DNA recognition was studied for 20 complexes that had crystal structures available by using MM-PBSA analysis [89]. It was found that the desolvation penalty showed some degree of correlation with the buried SAS; both favorable and unfavorable salt-independent electrostatic free energies were observed in the different protein-DNA complexes. The association free energy from the hydrophobic effect was on the same magnitude as the experimental estimates. The number of heavy atoms that contributed to the buried SAS varied between 68 and 180 atoms and the binding free energy per heavy atoms varied from  $-0.07$  to  $-0.14$  kcal/mol/atom.

#### 4. Protein-Carbohydrate Interactions

Recently, Shilov and Kurnikova studied the interaction between a transmembrane pore protein  $\alpha$ -hemolysine ( $\alpha$ -HL) and a cyclic oligosaccharide  $\alpha$ -cyclodextrin ( $\alpha$ -CD) theoretically [90]. A set of starting structures (10) were constructed by orientating  $\alpha$ -CD towards the *cis* or *trans* side of the channel of  $\alpha$ -HL and placing the center of mass of  $\alpha$ -CD at  $z = 30, 35, 40, 45$  and  $50$  Å. To mimic the lipid bilayer environment, a slab of heavy dummy atoms with Lennard-Jones potential, were added around the stem of  $\alpha$ -HL. They found that the equilibrated configurations with  $\alpha$ -CD residing in the vicinity of Met113 residue of  $\alpha$ -HL protein and with wider rim oriented toward the *trans*-side of the membrane were the most favorable in terms of both interaction and MM-PBSA binding free energies. This result was consistent with the experimental observations.

#### 5. Nucleic Acid-Ligand Interactions

Gouda *et al.* applied MM-PBSA to study the interaction of theophylline and its derivatives with an RNA aptamer [91]. Although the rank of the relative binding free energies of the five theophylline analogs was the same as that of experiment with one exception, no good correlation between the calculated and the experimental binding free energies could be found. In contrast, the relative binding free energies by thermodynamic integration, a much more expensive method, were well predicted with an average unsigned error of 0.4 kcal/mol. The binding free energies of three anticancer

compounds (Mit, PyrI, PyrII) to DNA were calculated by Baginski *et al* using MM-PBSA [92]. For each molecule, several topologically different modes were first constructed manually. Then MM-PBSA analysis was carried out on the minimized complexes. The most favorable MM-PBSA binding free energies were in fair agreement to experiment ( $-16.8$  versus  $-10.3$  kcal/mol for Mit,  $-14.1$  versus  $-8.9$  kcal/mol for PyrI and  $-4.7$  versus  $-6.6$  kcal/mol for PyrII).

Both MM-PBSA and MM-GBSA were applied by Burkhardt *et al.* to study monovalent and divalent cations to two adenine-adenine platform structures from the *Tetrahymena* group I intron ribozyme [93]. Qualitative agreement between the calculated and experimental ion placements and binding selectivity was obtained. The inclusion of solvation effects turned out to be important to obtain the low energy structures and ion binding placements in agreement with the experiment. The calculated alkali ion binding selectivity by PB for both platforms followed the order  $K^+ > Na^+ > Rb^+ > Cs^+ > Li^+$  in case of allowing RNA conformational relaxation during docking. The GB result was similar to that of PB except that the binding free energy of  $Na^+$  was marginally more favorable than  $K^+$  in the first AA platform. However, if the RNA conformations were rigid during the docking, the binding free energies of  $Na^+$  were higher than that of  $Rb^+$  in all the circumstances.

MD simulations and MM-PBSA analysis were performed for DNA-DAP (4',6-Diamidino-2-phenylindole) binding at the minor groove by Spačková *et al.* [94]. DNA sequences including two (AATT and ATTG) for which the binding modes were observed experimentally and two (AATT and ATTC) with alternative shifted binding modes were investigated. The absolute MM-PBSA binding energies of both single and separate trajectories were significantly underestimated ( $\sim 0$  versus  $-9$  to  $-12$  kcal/mol in experiment). The relative binding free energies, which should be meaningful, suggested that the AATT site was weakly favored in both binding sites. DAP binding to seven dodecanucleotides including the two DNA sequences mentioned above were studied by de Castro and Zacharias using a combination of docking, GB-minimization and PBSA calculations [95]. Qualitative agreement were obtained between the results of GB and PB approaches as well as between the calculated and experimentally observed trends regarding the sequence specificity of DAPI binding to B-DNA. Cieplak [96] also performed MM-PBSA analysis for long MD trajectories to calculate the binding free energies of some DNA-drug complexes and got encouraging results.

#### 6. Other Systems

Some small guest-host systems were also studied by MM-PBSA. For example, the experimental result of enantiodifferentiation observed in the complexation of cizolirtine and its parent carbinol with  $\beta$ -cyclodextrin ( $\beta$ -CD) was rationalized and interpreted by using MD simulations and MM-PBSA free energy calculations [97]. The chiral discrimination of N-acetylphenyl-alanine enantiomers by  $\beta$ -CD was studied by Choi *et al.* [98]. The calculated relative binding free energy was in fine agreement with the experimental determined value ( $-0.381$  kJ/mol vs.  $-0.26$  kJ/mol).

### C. Ligand Design

In this section, we present the application of MM-PBSA and MM-GBSA in combination with other approaches to design new and optimize known ligands. We have shown several examples [13,77] above that MM-PBSA and MM-GBSA together with molecular docking and MD simulations can reliably model a protein and DNA complex *a priori*. Evidently, this strategy provides a reliable means to evaluate a protein model only if the binding affinities of a set of inhibitors are known. A “successful” protein model should give considerably accurate MM-PBSA binding free energies compared to experiment. This strategy could have great applications in GPCR (the G-protein coupled receptors) modeling for which very few crystal structures are available.

MM-PBSA and MM-GBSA in drug discovery can also be used as a promising filter in virtual screening. We have successfully applied a set of hierarchical filters that include a pharmacophore model, multiple-conformation rigid docking, solvation docking and MM-PBSA in exploring promising inhibitors for HIV-1 RT [99]. The basic idea of the virtual screening strategy is to first employ rapid but less accurate methods, such as docking, to screen out less interesting compounds and then to apply MD techniques and MM-PBSA to perform more accurate calculations on the most promising ligands. 15,000 compounds in a refined available chemical directory (ACD) database were subjected to the four filters and 3360 compounds survived the first three filters. In a control test, 22 out of 37 known HIV-1 RT inhibitors survived the first three filters and 16 known ligands had the calculated MM-PBSA binding free energies better than -6.8 kcal/mol. Overall the enrichment factor for the first three filters was 25-fold and the hit rate for all the four filters was predicted to be 41%. We also pointed out that although MD simulations followed by MM-PBSA analysis required massive amount of computational resource, it was expected that one could screen as many as 3,500 compounds with this filter within a reasonable timeframe (less than a week) for a pharmaceutical company. It is worth mentioning that to apply MM-PBSA in database searching, one needs a general or universal force field that is consistent to the force field used by protein and nucleic acids for organic molecules. We have developed and will continue to improve a general AMBER force field [100] for this purpose.

Hou and Xu *et al.* recently studied a set of quinazoline-like inhibitors of epidermal growth factor receptor with 3D QSAR, molecular docking and MM-PBSA [101]. Following the same approach described above, they used MM-PBSA to determine the most favorable binding mode among those suggested by docking. The proposed protein-ligand complex, which had a MM-PBSA binding free energy 10 kcal/mol more favorable than the second best one, could explain the SAR data and was in good agreement with the contour maps of the comparative molecular field analysis (CoMFA) model ( $q^2 = 0.6$ ,  $F = 124.51$ ). Furthermore, the predicted complexed structure was quite close to the crystal structure, which was released after the manuscript had been submitted. The cross-validation between different methods provides a useful way to evaluate a computer model in complement to experiment.

Recently, Kuhn *et al.* further evaluated MM-PBSA as a tool for drug discovery [12]. Their studies covered a variety

of drug design approaches, which included virtual screening and *de novo* design, and eight proteins and a large number of inhibitors were involved. They concluded that MM-PBSA was valuable as a post-docking filter in further enriching virtual screening hits and was helpful in prioritizing *de novo* design solutions, and could distinguish between good and weak binders.

It is believed that the receptor conformation that is adequate for a ligand to bind to occurs infrequently in the unliganded receptor. Therefore, the common docking protocol, flexible-ligand and rigid-receptor may fail to find correct ligand-receptor binding modes. A receptor must be flexible to adopt an appropriate conformation to accommodate a specific ligand. Lin and McCammon *et al.* [102] proposed a relaxed receptor method in ligand design. They first docked ligands to multiple snapshots collected from a long MD simulation of the receptor. Then MM-PBSA was employed to re-score the docking poses. They found that the average distribution of the docking free energies was about 2-3 kcal/mol for FK506 binding protein FKBP-12 binding to a set of small molecules. They concluded that by using the MM/PBSA protocol the binding modes that were in agreement with the x-ray studies were consistently ranked as the most stable complexes.

GBSA was applied by Taylor *et al.* in a two-stage docking studies [103]. The first stage, the geometry-based docking had four steps: clique detection, clique filtering, clique embedding and clique clustering. In the second stage, Monto Carlo simulations were used to further optimize the poses produced in the first stage. A soft-core interaction function and a GBSA model were employed in the molecular mechanical energy-based scoring function. 13 out of the 15 test protein complexes were able to find the experimental binding mode in the rigid-protein, flexible-ligand docking; in contrast to 11 out of the 15 for the both flexible protein and ligand docking.

The solvation energy has also been applied as a descriptor in QSAR studies. Nair *et al.* [104] found that polarization energy calculated by GB, the polar, non-polar and total surface areas in addition to the 3D-QSAR descriptors could substantially improve the performance of QSAR models for the inhibitors of HIV-1 protease.

### D. Other Applications

With a linear response of the dielectric assumption, Sulea *et al.* [105] showed that the total reaction field was the superposition of all the individual reaction fields of the charges in the cavity and the reaction field energy had a quadratic function form. The reaction field energy could be rapidly calculated for an arbitrary value of  $q$  at atom center  $i$  if the coefficients were pre-calculated. Therefore, for a given binding site and a given spatial arrangement of atoms in a ligand, there existed an optimal set of partial charges at the atom centers that would optimize the net electrostatic binding free energy of the ligand. This idea could have a great use in molecular docking and *de novo* ligand design.

MM-PBSA scoring function was applied by Silberstein *et al.* [106] to identify the substrate binding sites of enzymes. The algorithm placed a set of small molecules or functional groups on a protein surface, and found the region

that had the most favorable binding free energy. Their method could find the consensus site that bound the highest number of different probes. They also successfully mapped thermolysin, for which experimental mapping results were available, and six other enzymes that had no experimental mapping data, but whose binding site were well characterized.

## E. Summary

In Table 1, we list the protocols and some key parameters of MM-PBSA and MM-GBSA analysis employed in 54 papers. The sampling protocols include single trajectory, individual trajectory, docking and minimization; the parameters considered include the charge method, radius parameter set and the solute dielectric

**Table 1. A Summary of MM-PBSA and MM-GBSA Free Energy Calculations for 54 Papers**

No.	Ref.	Sys Type1	Calc Type2	Wat Mod3	Sampl Protoc4	Free Energy Type5	PB Prog and GB Protocol 6	Radius Param Set7	Solute Dielec Const8	Surface Tension9	Entropy Method10	Performance / Results11
1	11	1	1	1	1	1	1	1	1	1	4	3 and 4
2	12	3	2	1	2	1	1	1	1	1	1	1 and 2
3	13	3	2	2	2	1	1	1	1	1	1	1
4	14	1	1 and 1a	1	2	1	1	1				3
5	44	1	1	1	1	1	1	1	4 and 1	3 and 8	4	3,4
6	45	1c	1	3	3	1	2		1	5	5	3
7	47	1a	1	1	1	1	3	2	1	11	1	4
8	48	1b	5	1	1	1	1		1	7		4
9	49	2	1	1	1	3	5		1	N/A	1	4
10	50	2	1	1	1	1	1	1	1	1	1	3,4
11	51	2	1	1	1	1	1	1	1	1	1	3,4
12	52	2	1	1	1	1	1	1		1	1	4
13	54	3	2	2	2	1	1	1	1	1	1	1,3
14	55	3	2	2	2	1	1	1	1	1	1	1,3
15	56	3	2	1	2	1	1	1	1	1		1,3
16	57	3	3	1	1 and 3 after 2a	1	1	1	4 and 1	1	4	2,3
17	62	3	2	1	1 and 2	1	4		1	1	2	1
18	63	3	2	2	2	1	1	1	1	1	1	3
19	64	3	2	1	2	1	1	1	1	1	1	2
20	66	3	2	1	2	1						2
21	67	3	2		4	1	2		4	6	1	3
22	68	3	2	GB	MD	3				10	4	
23	69	3a	3	2	2	1	1	1	1	1	4	2
24	70	3	2 and 3	1	2	1		1	1	1	1	2
25	71	3	2 and 3	1	2	1	1	1	1	1	1	2,3
26	72	3b	4	1	2	1	2					4
27	73	3	2		3	3	1 and 9	2	1	2	2	1
28	74	3	2 and 3	1	1 and 2	3	10 and 11	1	2 for PB and 1 for GB	1	1	2,3
29	75	3	2 and 3	1	1	2	10		1	2	1	2,3

(Table 1)contd....

No.	Ref.	Sys Type1	Calc Type2	Wat Mod3	Sampl Protoc4	Free Energy Type5	PB Prog and GB Protocol 6	Radius Param Set7	Solute Dielec Const8	Surface Tension9	Entropy Method10	Performance / Results11
30	77	4	2	GB	GB-MD	1						4
31	78	4a	2	2	2	1	5	1	4	1	1	3
32	79	4a	2	2	1	1	1	1	1	1	1	3
33	81	4	2	1	2	1	1	1	1	1	4	4
34	83	4	2	1	1	1				1	3	4
35	84	4a	2	4	2 after 3 after 4	1	1	1	2	9	3	4
36	85	4b	2	1	1	1					4	4
37	86	5a	2	1	2	1	1	1	1	1	4	1
38	87	5a	2	1	1	1	1	1	1	1	1	1
39	88	5a	3	1	3 after 4a	1	1	1	1	1	4	2
40	89	5	2		3	1	1		2	7	3	4
41	90	6	2	1a	1	1	6	5	2	4	4	
42	91	7	2	1	2	1	1	1	1	1	4	2
43	92	7	2	5	3 after 4	1	2		4	7		1
44	93	7a	2	GB	3 after 4	3	2 and 8	5 for GB and 4 for PB	1	4		4
45	94	7	2	1	1 and 2a	1	1	1	1	1	4	2
46	95	7	2		4	3	2 and 8	1	1	2	4	4
47	97	8	2		2	1						4
48	98	8	3		2	1	3		1	1	4	2
49	99	3	2	2	2	1	1	1	1	1	1	1
50	101	3	1	2	1	1	1	1	1	1	1	3
51	102	3	2	1	4	1	4		2	6	1	2
52	103	3 and 4	2		4	2	8			2	4	2,4
53	106	1	1		N/A	1	7		4	N/A	4	4
54	122	4	2		4	1a	2	3	1	N/A	4	4

- System types: 1~proteins, 1a~peptide, 1b~protein and isomers, 1c~protein loops, 2~DNA, 3~protein-ligand, 3a~antibody-ligand, 3b~protein-water, 4~protein-protein, 4a~protein-peptide, 4b~peptide-peptide, 5~protein-DNA, 5a~protein-RNA, 6~protein-carbohydrate, 7~DNA-ligand, 7a~RNA-ion, 8~guest-host.
- Calculation Type: 1~relative free energy, 1a~unfolding free energy, 2~binding free energy, 3~relative binding free energy, 4~relative electrostatic binding free energy, 5~energy landscape.
- Water Model: 1~water box, 1a~water box, dummy atoms with LJ parameters that mimic the interior of lipid bilayer, 2~Water cap or water sphere, 3~GB; 4~explicit water, 5~vacuum.
- Sampling Protocol: 1~individual trajectory, 2~single trajectory, 2a~single trajectory mutation, 3~minimization, 4~docking.
- Free Energy Type: 1~MM-PBSA, 1a~MM-PB, 2~MM-GBSA, 3~MM-PBSA and MM-GBSA.
- PB Program and GB Protocol: 1~Delphi, 2-UHBD, 3~PBEQ in CHARMM, 4~APBS of Baker et al., 5~MEAD, 6~PNP, 7~CONGEN, 8~Hawkins et al. (GBSA) 9~Qiu et al. (GBSA), 10~Jayaram et al. (GBSA), 11~Other GBSA models.
- Radius Parameter Set: 1~Parse, 2~CHARMM, 3~AMBER, 4~BONDI. 5~(C: 1.7, H:1.0, N:1.5, O:1.6, S:1.9, P:2.0Å)
- Dielectric constant of solute in PB or GB calculations
- Surface Tension: 1~0.0054 + 0.92, 2~0.0072, 3~0.00542, 4~0.0055, 5~0.02, 6~0.025, 7~0.05, 8~0.054, 9~0.05818, 10~0.006 kcal/mol/ Å<sup>2</sup>, 11~ Kyte and Doolittle's (*J. Mol. Biol.*, **1982**, *157*, 105-132).
- Entropy Method: 1~normal mode analysis, 2~quasiharmonic, 3~empirical approaches, 4~Not calculated, 5~implicitly considered in the colony free energy calculations.
- 1~absolute values of free energies are well predicted, 2~relative values of free energies are well predicted, 3~good correlations between the calculated and experiment are achieved, 4~can explain other experimental data.

constant for PB or GB calculations, the surface tension for estimating the non-polar solvation energy, and the entropy calculation approaches. The following is a summary: most studies performed MD simulations on complexes in explicit water to get trajectories for single trajectory analysis; the parse [25] parameter set was commonly used and the solute dielectric constant was usually set to 1; dielectric constants of 2 and 4 were also tried in some studies; a set of diverse surface tensions, including 0.0025, 0.0054, 0.006, 0.0072, 0.025, 0.05, 0.05818 kcal/mol/Å<sup>2</sup>, were used; however, most applications used 0.0054 with the constant  $b$  in Equation 14 ( $b$  was set to 0.92 kcal/mol) [24-25]; about half of the studies estimated the entropy contribution and normal mode analysis was used in most cases; quasiharmonic analysis and empirical methods were also utilized occasionally by some users.

In terms of the performance of MM-PBSA and MM-GBSA, in general, relative free energies and relative binding free energies could be calculated with a considerable accuracy (<2.0 kcal/mol), while the absolute binding free energies, although in some studies good results were achieved, could have a large error compared to experiment (up to 10.0 kcal/mol). Fortunately, the calculated free energies usually could explain some experimental data and phenomena even when large calculation errors happened. As to for what systems MM-PBSA/GBSA tend to work well and for what systems not, in general, for systems that have substantial contributions to the free energy from buried charges, MM-PBSA/GBSA may not perform well; on the other hand, good results can be expected for the hydrophobic interaction dominated systems. Yet, this is not always the case and good results of both relative and absolute free energies could be observed for almost all kinds of systems, including proteins [14,44], nucleic acids [50-51], protein-ligand [13,54,63], and protein-protein [76-79]. It is emphasized that the above summary may be biased to some degree since we only selected some representative papers published recently in this field.

The absolute and relative MM-PB/GBSA free energies are affected by many adjustable parameters that include the van der Waals radii used in PB or GB solvation models, the exterior and interior dielectric constants, and even the force field parameters, etc. Although one could make the calculated free energies reproduce experimental data well by adjusting some parameters for his/her systems, the modified parameters may not be transferred to other systems without a problem. This free energy calculation approach should be improved in a systematic and physical way. In the next section, we will first discuss the perspective of the method followed by discussion on how to improve the methodologies from two aspects, both in accuracy and efficiency.

#### IV. PERSPECTIVE AND CONCLUSION

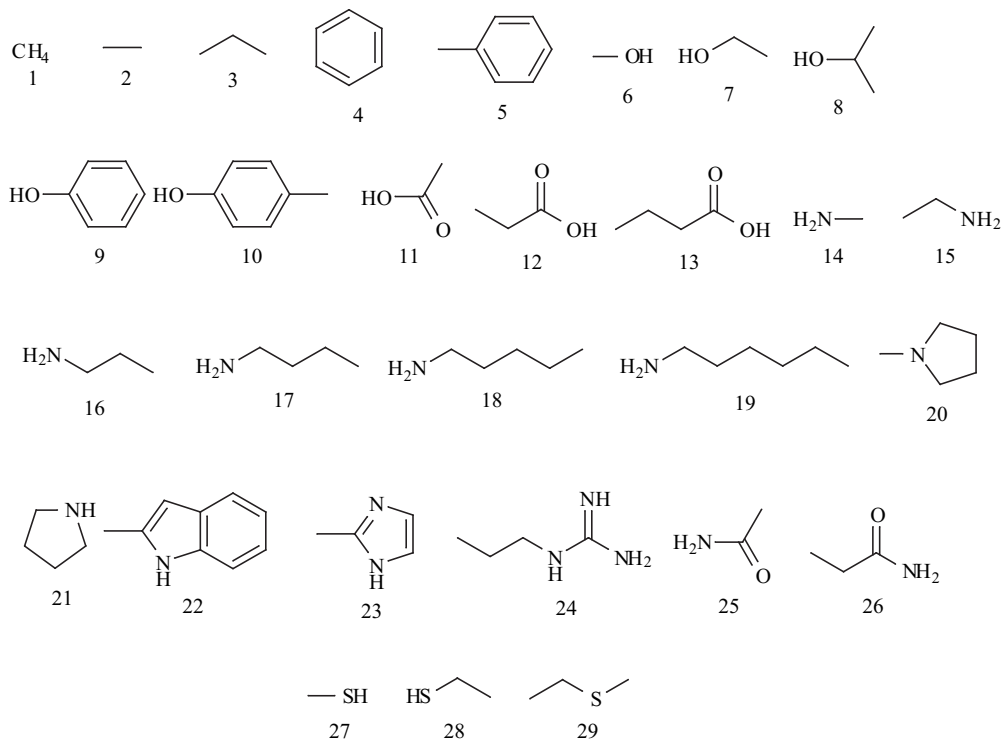
In the above section, we have shown that MM-PBSA and MM-GBSA have considerable promise in calculating free energies for a wide variety of biological systems. Although MM-PBSA and MM-GBSA do not have as a solid theoretical basis as FEP and TI, it is computationally more efficient. Moreover, compared to other free energy

calculation methods including LIE and solvation docking, MM-PBSA is more promising under some circumstances owing to the fact that it does not require a training set to derive empirical parameters in the first place, while LIE and solvation docking do. As a newborn technique of free energy calculations, MM-PBSA and MM-GBSA may have limitations in many situations, for example, when explicit water molecules form critical hydrogen-bond interactions and some binding that involves divalent ions such as Mg<sup>2+</sup> and Zn<sup>2+</sup>. Furthermore, it is difficult to determine which solute dielectric constant should be used *a priori*, although the overall guideline is to use 1 for non-polar binding sites and 2 for polar binding sites and 4 if charged residues exist in the binding sites. Experience on how to choose a reasonable dielectric constant may be obtained by studying a set of variant protein or DNA complexes (up to 100) that have not only high-quality experimental structures, but also experimental binding free energies. In the following section, we will discuss how to improve MM-PBSA and MM-GBSA.

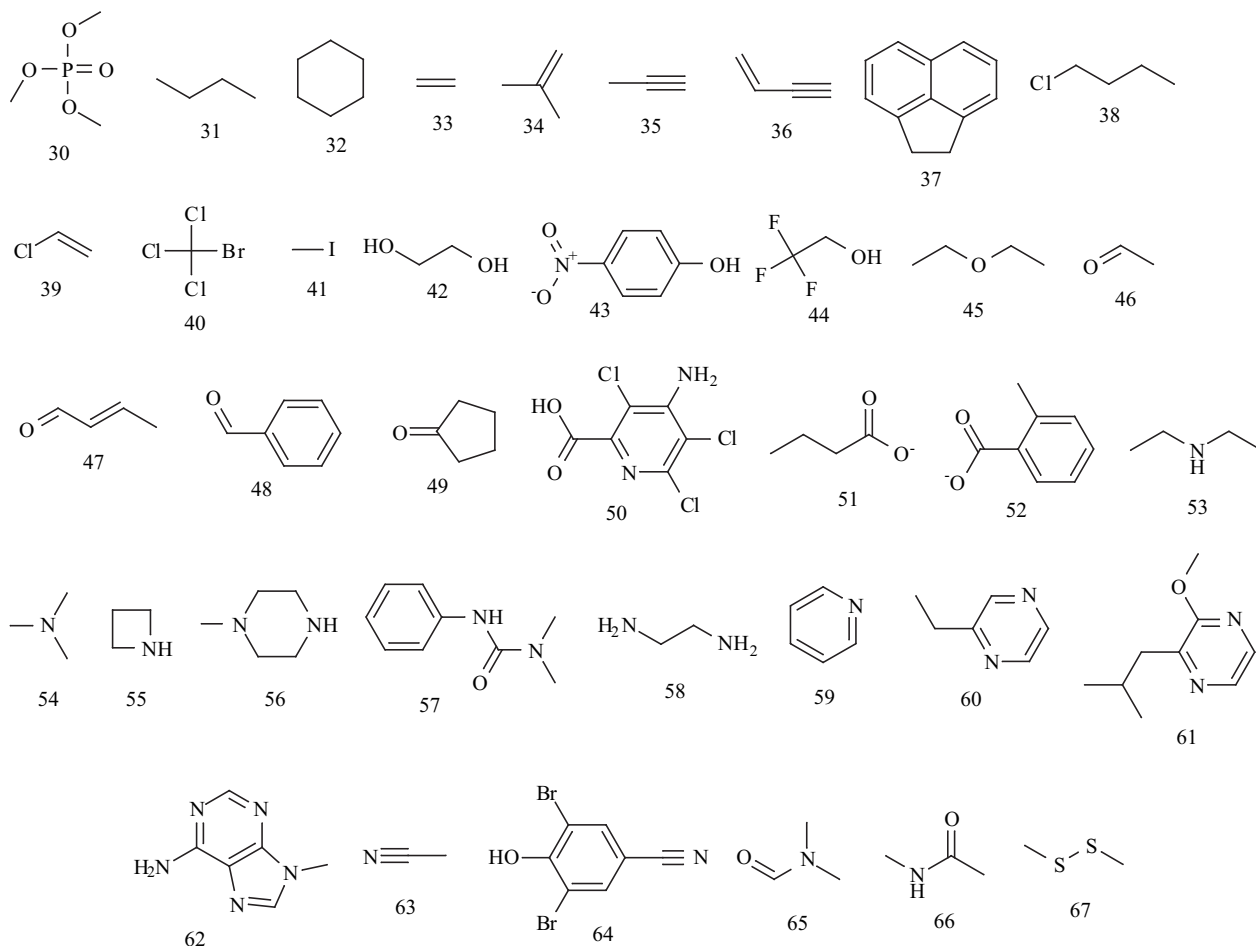
This technology can be improved in two aspects, both in accuracy and efficiency. The errors of MM-PBSA and MM-GBSA may come from several sources: the molecular mechanical force field error, the solvation free energy error, the entropy error, and the error due to inadequate sampling, etc. One of the most important issues in continuum solvation models is how to define the dielectric boundary. The dielectric medium is usually set to begin at the van der Waals surface of the solute molecule. However, the van der Waals radius parameters for PB and GB calculations are charge method dependent and may not be necessarily the same set used in the molecular mechanical force field. We have argued that one needs to apply the same charges in both MM and solvation free energy calculations in order to make errors cancelled to a great degree. Unfortunately, there are no radius-parameter sets developed for most molecular mechanical force fields. The Parse parameter set, which theoretically should be used together with the Parse charges, is widely-used in combination with other charge methods, including the HF/6-31G\* RESP charges. The performance of PBSA solvation free energies of a set of small molecules by using RESP charges and Parse radius parameters are listed in Table 2. The first 29 molecules are amino acid side chain analogs, while the other 38 small molecules are chosen to cover most of the functional groups. The structures of these molecules are shown in Fig. 3. The average unsigned errors (AUE) and the root-mean-square errors (RMSE) are listed in Table 3. One could see that the errors are very large for all three solute dielectric constants ( $\epsilon = 1$ ,  $\epsilon = 2$ ,  $\epsilon = 4$ ). The smallest AUE and RMSE were achieved when dielectric constant was 2, which were 1.55 and 2.04 kcal/mol, respectively. It should be pointed out that the error of calculated MM-PBSA binding free energy by using the combination of RESP charge and Parse radii could be unacceptably large for a small or medium-sized molecule, especially in a scenario that the ligand molecule is well-buried in the binding pocket and the errors can not be well cancelled.

Recently, we have developed a set of radius parameters that are consistent to the HF/6-31G\* charges. The 404 training set molecules are from our previous work.<sup>107</sup> The PBSA free energies with the newly-developed radius sets are

A



B



**Fig. (3).** Structures of the molecules listed in Table 2. (a) amino acid side chain analogs; (d) divers molecules with different functional groups.

also listed in Table 1 for comparison purpose. The AUE and RMSE, which are only marginally larger than the experimental errors, are significantly smaller than those of the RESP/Parse protocols. The parameterization for GBSA was also reported recently. Zhang *et al.* [108] found that the GB radii, which were obtained by timing 1.1 of the corresponding OPLS Lennard-Jones radii ( $\sigma/2$ ) plus 0.05, could make the GB solvation energies well reproduce those calculated by free energy perturbation for 38 small molecules (the AUE is 0.6 kcal/mol). A popular analytical GB solvation model was modified by Onufriev and Case *et al.* [109] to improve its accuracy in calculating the solvent polarization part of free energy changes in large-scale conformational transitions, such as protein folding. The new algorithm was implemented in AMBER7 and AMBER8.

It is believed that the repulsive cavity term and attractive van der Waals solute-solvent interaction term of non-polar solvation energy can be modeled independently. The cavity term, which mainly accounts for the entropy change of solvent reorganization when a solute immerses in solvent, is proportional to the solvent accessible surface area. However, the solute-solvent dispersion interaction energy component depends on the atomic composition of the solute. Gakkucgui *et al.* [110] and Levy *et al.* [111] tried to decompose the non-polar component and to model the solute-solvent van der Waals energy separately. Gakkucgui *et al.* designed seven fitting functions to optimize the surface tensions and constants of SAS classified by atom types to reproduce experimental solvation energies of a set of small organic molecules. Unfortunately, the parameters may not be used for macromolecules without modification, since buried atoms in macromolecules may also contribute to the solute-solvent van der Waals energy. Levy *et al.* found that the linear correlation between SAS and the solute-solvent van der Waals energies calculated through explicit MD simulations could not be well transferred from the training set to the test set.

To improve the efficiency of MM-PBSA and MM-GBSA free energy calculations, actions may be taken in the following two aspects. First of all, one may run MD simulations using implicit water models, such as in GBSA and PBSA, instead of an explicit water model. PB-MD [112] and GB-MD have been available in Macromodel [113] and the AMBER software packages. Although the conformations sampled by PB-MD and GB-MD for post energetic analysis may not be as good as those sampled by MD in explicit solvent, PB-MD and GB-MD can save a lot of CPU hours not only in the sampling stage, but also in the post-analysis stage since many solvation energy terms are calculated during the simulations. The simplified sampling protocol used by Reyes [86-88] can also be applied under certain circumstances.

Secondly, the entropies (translational, rotational and vibrational) of each species need to be calculated if one wants to know the absolute binding free energies or the values are substantially different for two conformations in relative energy comparisons. Normal mode analysis (NMA) is the widely approach although it is not perfect due to its not considering anharmonic effect. NMA, which is computationally expensive, especially for macromolecules of more than 5000 atoms, is one of the bottlenecks of MM-

PBSA and MM-GBSA. Another approach, so called quasiharmonic analysis, is clearly sensitive to simulation length and inadequate or unconverged sampling may cause substantial errors [62]. A much simpler approach of estimating the conformational entropy is to perform side-chain rotational analysis [114]. An empirical scale, RA, is defined as the calculated accessible surface area of the side chain divided by the surface area of the side chain in the extended state. If RA is greater than 60%, the side chain is assumed to rotate freely, and if not, it is regarded as a buried side chain and just one rotamer. Recently, Gohlke *et al.* [115] proposed a rapid method to estimate vibrational entropy changes upon macromolecular complex formation by using the result of network analysis [116]. For a data set of 10 protein-protein complexes with widely varying properties, total vibrational entropy changes determined by their method correlated well ( $r^2 = 0.84$ ) with those obtained from NMA but it only required a fraction of computational time of NMA. A new algorithm, which is based on solvent-accessible surface area, is being developed to calculate the entropy accurately and efficiently by ourselves.

Thirdly, one could develop a set of MM-PBSA-like approaches by replacing PBSA and GBSA with simpler models, such as solvation models based on weighted solvent accessible surface area classified by atom types [107,117-119], and atomic contact energy (ACE) approaches [120-121].

Recently, Cerutti and McCammon *et al.* developed an efficient algorithm to calculate the electrostatic and non-polar solvation energies of protein-protein interaction. Their method, so-called ELSCA (Energy by Linear Superposition of corrections Approximation), was a correction protocol based on a distance-dependent dielectric, a scaleable function describing the buried surface area between two interacting spheres, and a set of potentials of mean force between distinct types of atoms. ELSCA was trained by a linear least-squares fit on more than 39,000 putative complexes and tested against over 8000 non-native complexes. A good correlation of 0.962 was achieved between ELSCA and the MM-PBSA energies of the non-native complexes. When applied to native complexes (45 protein systems), ELSCA reproduced PBSA results with a lower correlation of 0.787. As a very fast method, ELSCA is useful in macromolecular docking and protein association simulations [122].

With these improvements, we expect that MM-PBSA and MM-GBSA could help to make molecular mechanics approaches more useful in the "end games" of protein structure prediction and drug design [123].

## REFERENCES

- [1] Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T. *J. Am. Chem. Soc.*, **1990**, *112*, 6127- 6129.
- [2] Qiu, D.; Shenkin, P.S.; Hollinger, F.P.; Still, W.C. *J. Phys. Chem.*, **1997**, *101*, 3005-3014.
- [3] Hawkins, G.D.; Cramer, C.J.; Truhlar, D.G. *Chem. Phys. Lett.*, **1995**, *246*, 122-129.
- [4] Bashford, D.; Case, D.A. *Annu. Rev. Phys. Chem.*, **2000**, *51*, 129-152.
- [5] Jayaram, B.; Sprous, D.; Beveridge, D.L. *J. Phys. Chem. B*, **1998**, *102*, 9571-9576.
- [6] Kollman, P.A.; Massova, I.; Reyes, C.M.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.;



- Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham III, T.E. *Acc. Chem. Res.*, **2000**, *33*, 889-897.
- [7] Cieplak, P. *Molecular Simulation*, **2002**, *28*, 173 – 186.
- [8] Wang, W.; Donini, O.; Reyes, C.M.; Kollman, P.A. *Annu. Rev. Biophys. Biomol. Struct.*, **2001**, *30*, 211-43.
- [9] Simonson, T. *Curr. Opin. Struct. Biol.*, **2001**, *11*, 243-252.
- [10] Srinivasan, J.; Cheatham, T.E.; Cieplak, P.; Kollman, P.A.; Case, D.A. *J. Am. Chem. Soc.*, **1998**, *120*, 9401-9409.
- [11] Lee, M.R.; Baker, D.; Kollman, P.A. *J. Am. Chem. Soc.*, **2001**, *123*, 1040-1046.
- [12] Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. *J. Med. Chem.*, **2005**, *48*, 4040-4048.
- [13] Wang, J.; Morin, P.; Wang, W.; Kollman, P.A. *J. Am. Chem. Soc.*, **2001**, *123*, 5221-5230.
- [14] Lee, M.R.; Kollman, P.A. *Structure*, **2001**, *9*, 905-916.
- [15] Fogolari, F.; Esposito, G.; Viglino, P.; Molinari, H. *J. Comput. Chem.*, **2001**, *22*, 1830-1842.
- [16] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.*, **1993**, *98*, 10089-10092.
- [17] Michielin, O.; Karplus, M. *J. Mol. Biol.*, **2002**, *324*, 547-569.
- [18] Dixon, R.W.; Radmer, R.J.; Kuhn, B.; Kollman, P.A. *J. Org. Chem.*, **2002**, *67*, 1827-1837.
- [19] Price, D.J.; Jorgensen, W.L. *J. Comput.-Aided Mol. Des.*, **2001**, *15*, 681-695.
- [20] Reddy, M.R.; Erion, M.D. *J. Am. Chem. Soc.*, **2001**, *123*, 6246-6252.
- [21] Reddy, M.R.; Singh, U.C.; Erion, M.D. *J. Am. Chem. Soc.*, **2004**, *126*, 6224-6225.
- [22] Florián, J.; Goodman, M.F.; Warshel, A. *J. Phys. Chem. B*, **2000**, *104*, 10092-10099.
- [23] Loeffler, H.H.; Sotriffer, C.A.; Winger, R.H.; Liedl, K.R.; Rode, B.M. *J. Comput. Chem.*, **2001**, *22*, 846-860.
- [24] Sanner, M.F.; Olson, A.J.; Spehner, J.C. *Biopolymers*, **1996**, *38*, 305-320.
- [25] Sitkoff, D.; Sharp, K.A.; Honig, B. *J. Phys. Chem. B*, **1998**, *98*, 1978-1983.
- [26] Åqvist, J.; Medina, C.; Samuelsson, J.E. *Protein Eng.* **1994**, *7*, 385-391.
- [27] Almlöf M.; Brandsdal, B.O.; Åqvist, J. *J. Comput. Chem.*, **2004**, *25*, 1242-1254.
- [28] Rizzo, R.C.; Tirado-Rives, J.; Jorgensen, W.L. *J. Med. Chem.*, **2001**, *44*, 145-154.
- [29] Huang, D.; Cafilisch, A. *J. Med. Chem.*, **2004**, *47*, 5791-5797.
- [30] Singh, P.; Mhaka, A.M.; Christensen, S.B.; Gray, J.J.; Denmeade, S.R.; Isaacs, J.T. *J. Med. Chem.*, **2005**, *48*, 3005-3014.
- [31] Zou, X.Q.; Sun, Y.; Kuntz, I.D. *J. Am. Chem. Soc.*, **1999**, *121*, 8033-8043.
- [32] Liu, H.-Y.; Kuntz, I.D.; Zou, X. *J. Phys. Chem. B*, **2004**, *108*, 5453-5462.
- [33] Zoete, V.; Michielin, O.; Karplus, M. *J. Comput.-Aided Mol. Des.*, **2003**, *17*, 861-880.
- [34] Banba, S.; Guo, Z.; Brooks, C.L., III. In *Free Energy Calculations in Rational Drug Design*, Kluwer Academic/Plenum Publishers: New York, **2001**; pp. 195-223.
- [35] Banba, S.; Guo, Z.; Brooks III, C.L. *J. Phys. Chem. B*, **2000**, *104*, 6903-6910.
- [36] Luo, H.; Sharp, K. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 10399-10404.
- [37] Erion, M.D.; Reddy, M.R. In *Free Energy Calculations in Rational Drug Design*, Kluwer Academic/Plenum Publishers: New York, **2001**; pp. 225-241.
- [38] Hamelberg, D.; McCammon, J.A. *J. Am. Chem. Soc.*, **2004**, *126*, 7683-7689.
- [39] Woo, H.-J.; Roux, B. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 6825-6830.
- [40] Karney, C.F.F.; Ferrara, J.E.; Brunner, S. *J. Comput. Chem.*, **2005**, *26*, 243-251.
- [41] Li, Z.; Lazaridis, T. *J. Phys. Chem. B*, **2005**, *109*, 662-670.
- [42] Dill, K.A.; Chan, H.S. *Nat. Struct. Biol.*, **1997**, *4*, 10-19.
- [43] Simons, K.T.; Bonneau, R.; Ruczinski, I.; Baker, D. *Proteins*, **1999**, Suppl. 3, 171-176.
- [44] Lee, M.R.; Jerry Tsai, J.; David Baker, D.; Kollman, P.A. *J. Mol. Biol.*, **2001**, *303*, 417-430.
- [45] Fogolari, F.; Tosatto, S.C.E. *Prot. Sci.*, **2005**, *14*, 889-901.
- [46] Xiang, Z.; Soto, S.C.; Honig, B. *Proc. Natl. Acad. Sci.*, **2002**, *99*, 7432-7437.
- [47] Santa, H.; Ylisirniö, M.; Hassinen, T.; Laatikainen, R.; Peräkylä, M. *Prot. Engin.*, **2002**, *15*, 651-657.
- [48] Ma, B.; Nussinov, R. *Prot. Sci.*, **2003**, *12*, 1882-1893.
- [49] Cubero, E.; Abrescia, N.G.A.; Subirana, J.A.; Luque, F.J.; Orozco, M. *J. Am. Chem. Soc.*, **2003**, *125*, 14603-14612.
- [50] Yan, S.; Shapiro, R.; Geacintov, N.E.; Broyde, S. *J. Am. Chem. Soc.*, **2001**, *123*, 7054-7066.
- [51] Yan, S.; Wu, M.; Patel, D.J.; Geacintov, N.E.; Broyde, S. *Biophys. J.*, **2003**, *84*, 2137-2148.
- [52] Wu, M.; Yan, S.; Patel, D.J.; Geacintov, N.E.; Broyde, S. *Nucleic Acids Res.*, **2002**, *30*, 3422-3432.
- [53] Weinzinger, P.; Hannongbua, S.; Wolschann, P. *J. Enzyme Inhibition Med. Chem.*, **2005**, *20*, 129-134.
- [54] Huo, S.; Wang, J.; Cieplak, P.; Kollman, P.A.; Kuntz, I.D. *J. Med. Chem.*, **2002**, *45*, 1412-1419.
- [55] Kuhn, B.; Kollman, P.A. *J. Med. Chem.*, **2000**, *43*, 3786-3791.
- [56] Wang, W.; Kollman, P. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 14937-14942.
- [57] Wang, W.; Lim, W.A.; Jakalian, A.; Wang, J.; Wang, J.; Luo, R.; Bayly, C.I.; Kollman, P.A. *J. Am. Chem. Soc.*, **2001**, *123*, 3986-3994.
- [58] Jakalian, A.; Bush, B.L.; Jack, D.B.; Bayly, C.I. *J. Comput. Chem.*, **2000**, *21*, 132-146.
- [59] Jakalian, A.; Jack, D.B.; Bayly, C.I. *J. Comput. Chem.*, **2002**, *23*, 1623-1641.
- [60] Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P.A. *J. Phys. Chem. B*, **1993**, *97*, 10269-10280.
- [61] Cieplak, P.; Cornell, W.D.; Bayly, C.; Kollman, P.A. *J. Comput. Chem.*, **1995**, *16*, 1357-1376.
- [62] Swanson, J.M.J.; Henchman, R.H.; McCammon, J.A. *Biophys. J.*, **2004**, *86*, 67-74.
- [63] Hou, T.; Guo, S.; Xu, X. *J. Phys. Chem. B*, **2002**, *106*, 5527-5535.
- [64] Diaz, N.; Suárez, D.; Merz Jr., K.M.; Sordo, T.L. *J. Med. Chem.*, **2005**, *48*, 780-791.
- [65] Brigo, A.; Lee, K.W.; Fogolari, F.; Mustata, G.I.; Briggs, J.M. *Prot. Struct. Funct. Bioinf.*, **2004**, *59*, 723-741.
- [66] von Langen, J.; Fritzsche, K.-H.; Diekmann, S.; Hillisch, A. *Chem. Bio. Chem.*, **2005**, *6*, 1110 – 1118.
- [67] Schwarzl, S.M.; Tschopp, T.B.; Smith, J.C.; Fischer, S. *J. Comput. Chem.*, **2002**, *23*, 1143-1149.
- [68] Mardis, K.L.; Luo, R.; Gilson, M.K. *J. Mol. Biol.*, **2001**, *309*, 507-517.
- [69] Peräkylä, M.; Nordman, N. *Prot. Engin.*, **2001**, *14*, 753-758.
- [70] Massova, I.; Kollman, P.A. *Perspectives In Drug Discovery And Design*, **2000**, *18*, 113-135.
- [71] Huo, S.; Massova, I.; Kollman, P.A. *J. Comput. Chem.*, **2002**, *23*, 15-27.
- [72] Prévost, M. *Biopolymers*, **2004**, *75*, 196-207.
- [73] Noskov, S.Y.; Lim, C. *Biophys. J.*, **2001**, *81*, 737-750.
- [74] Gohlke, H.; Case, D.A. *J. Comput. Chem.*, **2004**, *25*, 238-250.
- [75] Gohlke, H.; Kiel, C.; Case, D.A. *J. Mol. Biol.*, **2003**, *330*, 891-913.
- [76] Wang, W.; Kollman, P.A. *J. Mol. Biol.*, **2000**, *303*, 567-582.
- [77] Wu, Y.; Cao, Z.; Yi, H.; Jiang, D.; Mao, X.; Liu, H.; Li, W. *Biophys. J.*, **2004**, *87*, 105-112.
- [78] Suenaga, A.; Hatakeyama, M.; Ichikawa, M.; Yu, X.; Futatsugi, N.; Narumi, T.; Fukui, K.; Terada, T.; Taiji, M.; Shirouzu, M.; Yokoyama, S.; Konagaya, A. *Biochemistry*, **2003**, *42*, 5195-5200.
- [79] Suenaga, A.; Takada, N.; Hatakeyama, M.; Ichikawa, M.; Yu, X.; Tomii, K.; Okimoto, N.; Futatsugi, N.; Narumi, T.; Shirouzu, M.; Yokoyama, S.; Konagaya, A.; Taiji, M. *J. Biol. Chem.*, **2005**, *280*, 1321-1326.
- [80] Hou, T.J.; Chen, K.; McLaughlin, W.A.; Lu, B.Z.; Wang, W. *PLoS Comput. Biol.*, in press.
- [81] Myshkin, E.; Leontis, N.B.; Bullerjahn, G.S. *Biophys. J.*, **2002**, *82*, 3305-3313.
- [82] Luo, C.; Xu, L.; Zheng, S.; Luo, X.; Shen, J.; Jiang, H.; Liu, X.; Zhou, M. *Prot. Struct. Funct. And Bioinf.*, **2004**, *59*, 742-756.
- [83] Orzechowski, M.; Cieplak, P.; Piela, L. *J. Comput. Chem.*, **2002**, *23*, 106-110.
- [84] Polticelli, F.; Zaini, G.; Bolli, A.; Antonini, G.; Gradoni, L.; Ascenzi, P. *Biochemistry*, **2005**, *44*, 2781-2789.
- [85] Urbanc, B.; Cruz, L.; Ding, F.; Sammond, D.; Khare, S.; Buldyrev, S.V.; Stanley, H.E.; Dokholyan, N.V. *Biophys. J.*, **2004**, *87*, 2310-2321.
- [86] Reyes, C.M.; Nifosi, R.; Frankel, A.D.; Kollman, P.A. *Biophys. J.*, **2001**, *80*, 2833-2842.
- [87] Reyes, C.M.; Kollman, P.A. *J. Mol. Biol.*, **2000**, *297*, 1145-1158.

- [88] Reyes, C.M.; Kollman, P.A. *J. Mol. Biol.*, **2000**, *295*, 1-6.
- [89] Norberg, J. *Arch. Biochem. Biophys.*, **2002**, *410*, 48-68.
- [90] Shilov, I.Y.; Kurnikova M.G. *J. Phys. Chem. B*, **2003**, *107*, 7189-7201.
- [91] Gouda, H.; Kuntz, I.D.; Case, D.A.; Kollman, P.A. *Biopolymer*, **2003**, *68*, 16-34.
- [92] Baginski, M.; Polucci, P.; Antonini, I.; Martelli, S. *J. Mol. Model.*, **2002**, *8*, 24-32.
- [93] Burkhardt, C.; Zacharias, M. *Nucleic Acids Res.*, **2001**, *29*, 3910-3918.
- [94] ?paèková, N.; Cheatham III, T.E.; Ryjáèek, F.; Lankas, F.; van Meervelt, L.; Hobza, P.; ?poner, J. *J. Am. Chem. Soc.*, **2003**, *125*, 1759-1769.
- [95] de Castro, L.F.P.; Zacharias, M. *J. Mol. Recognit.*, **2002**, *15*, 209-220.
- [96] Cieplak, P. *Molecular Simulation*, **2002**, *28*, 173-186.
- [97] Beà, I.; Jaime, C.; Redondo, J.; Bonnet, P.; Torrens, A.; Frigola, J. *Supramol. Chem.*, **2002**, *14*, 33-39.
- [98] Choi Y.; Jung S. *Carbohydrate Research*, **2004**, *339*, 1961-1966.
- [99] Wang, J.; Kang, X.; Kuntz, I.D.; Kollman, P.A. *J. Med. Chem.*, **2005**, *48*, 2432-2444.
- [100] Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. *J. Comput. Chem.*, **2004**, *25*, 1157-1174.
- [101] Hou, T.; Zhu, L.; Chen, L.; Xu, X. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 273-287.
- [102] Lin, J.-H.; Perryman, A.L.; Schames, J.R.; McCammon, J.A. *Biopolymers*, **2003**, *68*, 47-62.
- [103] Taylor, R.D.; Jewsbury, F.J.; Essex, J.W. *J. Comput. Chem.*, **2003**, *24*, 1637-1656.
- [104] Nair, A.C.; Jayatilleke, P.; Wang, X.; Miertus, S.; Welsh, W.J. *J. Med. Chem.*, **2002**, *45*, 973-983.
- [105] Sulea, T.; Purisima, E.O. *J. Phys. Chem. B*, **2001**, *105*, 889-899.
- [106] Silberstein, M.; Dennis, S.; Brown III, L.; Kortvelyesi, T.; Clodfelter, K.; Vajda, S. *J. Mol. Biol.*, **2003**, *332*, 1095-1113.
- [107] Wang, J.; Wang, W.; Huo, S.; Lee, M.; Kollman, P.A. *J. Phys. Chem. B*, **2001**, *105*, 5055-5067.
- [108] Zhang, L.Y.; Gallicchio, E.; Friesner, R.A.; Levy, R.M. *J. Comput. Chem.*, **2001**, *22*, 591-607.
- [109] Onufriev, A.; Bashford, D.; Case, D.A. *Proteins*, **2004**, *55*, 383-394.
- [110] Gallicchio, E.; Zhang, L.Y.; Levy, R.M. *J. Comput. Chem.*, **2002**, *23*, 517-529.
- [111] Levy, R.M.; Zhang, L.Y.; Gallicchio, E.; Felts, A.K.; *J. Am. Chem. Soc.*, **2003**, *125*, 9523-9530.
- [112] Fogolari, F.; Brigo, A.; Molinari, H. *Biophys. J.*, **2003**, *85*, 159-166.
- [113] Guvench, O.; Weiser, J.; Shenkin, P.; Kolossváry, I.; Still, W.C. *J. Comput. Chem.*, **2002**, *23*, 214-221.
- [114] Pickett, S.D.; Sternberg, M.J.E. *J. Mol. Biol.*, **1993**, *231*, 825-839.
- [115] Gohlke, H. Gordon Research Conference on Computer-aided Drug Design, **2005**.
- [116] Jacobs, D.J.; Rader, A.J.; Kuhn, L.A.; Thorpe, M.F. *Proteins*, **2002**, *44*, 150-165.
- [117] Hou, T.; Qiao, X.; Zhang, W.; Xu, X. *J. Phys. Chem. B*, **2002**, *106*, 11295-11304.
- [118] Filikov, A.V.; Mohan, V.; Vickers, T.A.; Griffey, R.H.; Cook, P.D.; Abagyan, R.A.; James, T.L. *J. Comput-Aided Mol. Des.*, **2000**, *14*, 593-610.
- [119] Liu, Z.; Dominy, B.N.; Shakhnovich, E.I. *J. Am. Chem. Soc.*, **2004**, *126*, 8515-8528.
- [120] Kimura, S.R.; Brower, R.C.; Vajda, S.; Camacho, C.J. *Biophys. J.*, **2001**, *80*, 635-642.
- [121] Zhang, C.; Vasmatzis, G.; Cornette, J. and DeLisi, C. *J. Mol. Biol.*, **1997**, *267*, 707-726.
- [122] Cerutti, D.S.; Eyck, L.F.T.; McCammon, J.A. *J. Chem. Theo. Comput.*, **2005**, *1*, 143-152.
- [123] Schueler-Furman, O.; Wang, C.; Bradley, P.; Misura, K.; Baker, D. *Science*, **2005**, *310*, 638-642.