

Computational Modeling of Solvent Effects on Protein-Ligand Interactions Using Fully Polarizable Continuum Model and Rational Drug Design

Fang Zheng and Chang-Guo Zhan*

Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, 789 South Limestone Street, Lexington, Kentucky 40536, USA.

Received 13 September 2011; Accepted (in revised version) 12 October 2011

Available online 12 June 2012

Abstract. This is a brief review of the computational modeling of protein-ligand interactions using a recently developed fully polarizable continuum model (FPCM) and rational drug design. Computational modeling has become a powerful tool in understanding detailed protein-ligand interactions at molecular level and in rational drug design. To study the binding of a protein with multiple molecular species of a ligand, one must accurately determine both the relative free energies of all of the molecular species in solution and the corresponding microscopic binding free energies for all of the molecular species binding with the protein. In this paper, we aim to provide a brief overview of the recent development in computational modeling of the solvent effects on the detailed protein-ligand interactions involving multiple molecular species of a ligand related to rational drug design. In particular, we first briefly discuss the main challenges in computational modeling of the detailed protein-ligand interactions involving the multiple molecular species and then focus on the FPCM model and its applications. The FPCM method allows accurate determination of the solvent effects in the first-principles quantum mechanism (QM) calculations on molecules in solution. The combined use of the FPCM-based QM calculations and other computational modeling and simulations enables us to accurately account for a protein binding with multiple molecular species of a ligand in solution. Based on the computational modeling of the detailed protein-ligand interactions, possible new drugs may be designed rationally as either small-molecule ligands of the protein or engineered proteins that bind/metabolize the ligand. The computational drug design has successfully led to discovery and development of promising drugs.

PACS: 31, 82, 87

Key words: Protein-ligand interaction, solvent effect, rational drug design, binding affinity.

*Corresponding author. *Email addresses:* fzhen2@email.uky.edu (F. Zheng), zhan@uky.edu (C.-G. Zhan)

1 Introduction

Structures and functions of biomolecular systems (such as protein, DNA, RNA, and their complexes with small-molecule ligands) are essential issues for understanding life processes at molecular level. Specially, when the biomolecule under consideration is a drug target, understanding the detailed structure and functions of the drug target at molecular level will provide a solid base for computational drug design. Information from experiments is always necessary, but often insufficient to achieve a complete understanding of the detailed structure and functions. Modern computational techniques of molecular modeling have been recognized to be a valuable complement to experiments, because an appropriate use of the state-of-the-art molecular modeling techniques can provide more detailed structural and mechanistic information that cannot be obtained from experiments alone, as demonstrated in many reports such as [1–8].

On the other hand, development of high-accuracy computational approaches to studying the structures and functions of biomolecules is particularly challenging. This is because many biomolecules is usually large in size and surrounded by a very complex chemical environment. The chemical environment surrounding a molecule in living system always includes a large number of solvent water molecules. Intermolecular interactions between a molecule under consideration (as the solute) and its solvent environment could dramatically change the structure and functions of the solute molecule. The experimental response of chemical, physical, and biochemical phenomena depends critically on the solvent effects. Thus, a reliable computational approach must appropriately account for the solvent effects in the practical computations.

A theoretically ideal computational approach would be to perform electronic structure calculations on the entire solvated biomolecular system, *i.e.* the entire biomolecule with its explicit chemical environment, at a sufficiently high-level *ab initio* quantum mechanical (QM) theory. This is a first-principles approach, which has been proven reliable in predicting the structures, properties, and chemical reactions of isolated small molecules (in vacuum, or in the gas phase). The reliability of the results calculated with this approach would not rely on any adjustable empirical parameters. Unfortunately, a high-level *ab initio* QM calculation on a biomolecule with its adequate chemical environment is impractical from a computational point of view [9], because the computing time required for a QM calculation will dramatically increase by adding additional atoms to the QM-treated system. For this reason, empirical molecular mechanics (MM) and related methods are currently very popular computational methodologies used in modeling and simulation of biomolecules.

A MM method simply considers all atoms to be classical particles with atomic forces determined by a set of parameterized interaction functions (force field), including bonded interactions (bonds, angles, and dihedral angles), non-bonded van der Waals interactions, and electrostatic interactions based on net atomic charges. By use of an empirical force field, a classical molecular dynamics (MD) simulation enables the study of time evolution of a large biomolecular system by taking many small successive time steps under

the force field. In addition to the all-atom models, united-atom models and simpler empirical models based on lattice or off-lattice representations with residue-level resolution have been developed to simplify the computations.

The MM-based methods are very useful in studying conformational structures and related properties of biomolecules with a given pattern of the covalent chemical bonds (first-order structure). A limitation of MM approach is that it cannot account for breaking or formation of a covalent chemical bond. To overcome the limitation, hybrid QM/MM methods [10–22] have been developed and used to study a variety of protein-ligand interactions and the mechanisms of enzymatic reactions [23–35]. A QM/MM method means to quantum mechanically treat some critically important part of a large biomolecular system under consideration and molecular mechanically treat the remaining part of the biomolecule and solvent. The QM-treated part of the biomolecular system may be regarded as a simplified model (such as the active site model of an enzyme), whereas the remaining part of the biomolecule and solvent are the MM-treated chemical environment of the model system. The QM/MM approach allows practical electronic structure calculations on the most important part of a large biomolecular structure and has been proven very useful [23–36].

Generally speaking, the above-mentioned three types of computational approaches (QM, MM, and QM/MM) complement each other. Different types of structural/mechanistic problems may be solved by using different types of computational approaches. However, all of these types of computational approaches are valuable in the state-of-the-art computational design. The present review concerns the recent development of the computational strategies and methods related to computational drug design. As discussed below, it is crucial for a reliable computational drug design to accurately evaluate solvent effects on protein-ligand interactions. Generally speaking, the solvent effects may be accounted for either explicitly (*i.e.* including actual solvent molecules in the computation) or implicitly (*e.g.*, using a continuum solvent model based on the Poisson-Boltzmann equation [37], Generalized Born model [38], or conductor-like screening model [39]). Thus, in this paper, we will first briefly discuss the solvent effects on protein-ligand interactions in Section 2 and review the recent development of a unique solvation model in Section 3. Based on the computational methodology development, we will further review recent development of computational design strategies, along with specific examples in practical studies of protein-ligand interactions and computational drug design, in Section 4. Concluding remarks are provided in Section 5.

2 Solvent effects on protein-ligand interactions

2.1 Protein binding with a given ligand species

In practical drug design, most drug targets are proteins. Thus, for convenience of discussion, we will focus on proteins below; the general concept of computational methods discussed here is also applicable to the DNA and RNA targets. A ligand of protein can be

a naturally occurring biomolecule, such as substrate of an enzyme, or a drug candidate like inhibitor of an enzyme or agonist/antagonist of a receptor protein. The Gibbs free energy of binding (ΔG_{bind}) between a protein and a ligand is defined as the Gibbs free energy change from the free protein plus the free ligand ($G_P + G_L$) to the protein-ligand complex (G_{P-L}):

$$\Delta G_{\text{bind}} = G_{P-L} - (G_P + G_L). \quad (2.1)$$

Eq. (2.1) applies to the binding of a protein with a given ligand structure. ΔG_{bind} can be evaluated by using the thermodynamic cycle shown in Fig. 1. Solvent effects on the binding free energy, ΔG_{bind} , can be accounted for either explicitly or implicitly. For example, within the explicit solvation method, one may explicitly include solvent molecules in the solvated system and directly simulate the binding process using a biasing force and determine a potential of mean force (PMF) corresponding to the free energy change during the binding process [40, 41]. However, only a limited number of solvent molecules can actually be included in the practical simulation and, thus, the bulk solvent effects may not be accounted for very well.

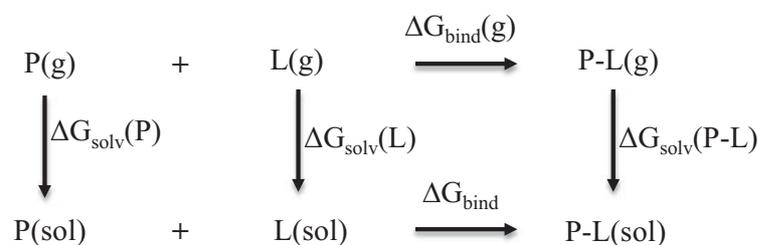


Figure 1: Thermodynamic cycle for calculating protein-ligand binding free energy (ΔG_{bind}) in solution. $\Delta G_{\text{solv}}(P)$, $\Delta G_{\text{solv}}(L)$, and $\Delta G_{\text{solv}}(P-L)$ refer to the solvent shifts of the Gibbs free energies of the protein (P), ligand (L), and complex (P-L), respectively.

A more efficient and more popularly used computational method is to implicitly account for the solvent effects on the binding free energy in the final energy calculations, even though the protein-ligand binding structure used in the energy calculations may be determined by using an explicit solvation method (*i.e.* MD simulation and/or energy minimization on the explicitly solvated system). Specifically, one may first evaluate the binding free energy in the gas phase by using an MM- or QM/MM-based method ignoring the solvent effects on the energies. Then, the solvent shift of the binding free energy may be calculated by using an implicit solvation method, such as the popularly used Poisson-Boltzmann surface area (PBSA) or Generalized Born Surface Area (GBSA) [42–44], as demonstrated in reported computational studies on protein-ligand interactions [7, 45–58].

2.2 Protein binding with multiple molecular species of a ligand

It should be pointed out that all of the computational methods mentioned here, including the PMF, MM-PBSA, QM/MM-PBSA, MM-GBSA, and QM/MM-GBSA, can be used to

determine binding free energy for a protein binding with a given ligand species. However, in practical studies on protein-ligand interactions related to computational drug design, a ligand could have multiple molecular species (including protonated and deprotonated states) that coexist in solution. For example, the widely abused nicotine, *i.e.* (S)-(-)-nicotine, is a well known agonist for nicotinic acetylcholine receptors (nAChRs). It has been known [59] that (S)-(-)-nicotine has a total of six molecular species (two neutral species, two cations, and two dications) that coexist in solution, and that the most favorable molecular species in solution is not necessarily the most favorable molecular species in a protein binding site. As nAChRs are recognized as important drug targets in various therapeutic areas, a larger number of small-molecule ligands have been identified as the agonists, antagonists, or partial agonists of nAChRs. Majority of the reported ligands contain at least one protonable amine group so that each nAChR ligand can have multiple protonation states [60].

In general, to appropriately account for a protein target binding with multiple molecular species of a ligand, one must first reliably determine the relative free energies of all molecular species of the free ligand in solution, and then calculate the microscopic binding free energy for the protein binding with each molecular species of the ligand in solution [59]. With all of these two types of energetic data available, one can evaluate the distribution of various microscopic protein-ligand structures and the macroscopic/phenomenological binding affinity [59]. The distribution of various microscopic protein-ligand structures will also reveal the primary microscopic binding structure for the protein-ligand binding [59–61].

Within the two types of energetic data required to determine the phenomenological binding affinity, the microscopic binding free energy for protein interacting with each molecular species of the ligand can be determined by using Eq. (2.1) as discussed above. It is particularly challenging to accurately determine the relative free energies of different molecular species of the free ligand in solution because the structural differences between different molecular species are usually associated with the differences in covalent bonds. Thus, it would be unreasonable to determine the relative free energies of different molecular species of the ligand by simply using an MM-based force field.

To determine the relative free energies of all molecular species of the free ligand in solution, the structures and free energies of all molecular species of the ligand must be examined at a sufficiently high level of theory which accurately accounts for the changes in covalent bonds and the solvent effects. It is desirable to determine the relative free energies of the free ligand in solution by using a truly reliable QM-based approach. The primary challenge of the QM calculations for this purpose is the determination of the solvent shifts of the Gibbs free energies, as discussed below. Thus, in the rest part of this section and in the next section, we will mainly discuss how to accurately account for the solvent effects in the QM calculations.

2.3 First-principles QM approach for determining the solvent effects on the relative free energies of multiple molecular species of a ligand in solution

First-principles QM approach has become a very powerful tool for studying a variety of scientific problems [62–65]. As well known, *ab initio* QM calculations on an isolated molecular species can readily achieve the desired chemical accuracy (*i.e.* the error is smaller than 1 kcal/mol) so long as one uses a sufficiently large basis set, considers the electron correlation at a sufficiently high level, and appropriately accounts for the relativistic effects in the calculation. By performing practical QM calculations at various levels, in principle one can always try to get results converged to the desired “exact” solution of a QM equation, *e.g.*, Schrödinger equation (if the relativistic effects can be neglected) or the Dirac equation (if the relativistic effects must be included), for a given molecular system existing in vacuum or the gas phase.

On the other hand, development of first-principle QM approach for molecules existing in solution is much more challenging and, thus, is now far behind that for molecules in vacuum. Thus, molecular properties in solution should not be expected to predict with chemical accuracy by performing QM calculations in vacuum without properly accounting for the solvent effects, even if one could perform the calculations at the highest level of QM theory, *i.e.* full configuration-interaction (full-CI) with a complete basis set (CBS), provided that the relativistic effects are negligible (or otherwise are also included). It is highly desirable that the properties of molecules in solution be described with a similar level of accuracy that can be achieved for molecules in the gas phase. A major roadblock is that directly performing *ab initio* QM calculation on the solute-solvent system including even a modest number of solvent molecules is impractical from a computational point of view. Therefore, a realistic treatment of the electronic structure of molecules in solution is one of the most important issues in theoretical and computational chemistry. Below, we will focus on the methods for effective treatment of the solvent effects on solute in the QM calculation.

The main, basic strategies to describe the solvent polarizing effect on solute in QM calculations include the supermolecule approach [66], QM/MM strategy [11, 13, 16, 67–71], and self-consistent reaction field (SCRF) methods [72, 73] although there have been efforts to describe the solvent through different ways such as the statistical mechanical reference interaction site model (RISM) [74–76] effective fragment potential (EFP) [77] and ONIOM-XS methods [78, 79]. All of these solvation approaches treat the solute quantum mechanically, and the differences lie only in description of the solvent. The first two are explicit solvation models. In the supermolecule approach, the solvent molecules are also described at the QM level, but only a limited number of solvent molecules can be explicitly included due to the expense of the QM calculations. Thus, the effect of the bulk solvent cannot be dealt with properly. In the combined QM/MM strategy, the solvent molecules are treated as classical entities, whose charge distribution is described by partial atomic charges or point dipoles. The SCRF methods consider the solvent to be a dielectric continuum medium that can be polarized by the solute leading to a reaction

field that in turn polarizes the solute itself. The reaction field is obtained from classical electrostatics by solving the requisite Poisson's equation, or the more general Poisson-Boltzmann equation (which is necessary to account for ionic strength of ionic solution), to obtain a mutual potential energy of solute-solvent interaction that is included in the solute Hamiltonian. Solutions to these electrostatic equations can be conveniently represented by certain apparent polarization charge density distributions.

There are, at least, two principal advantages of continuum models over explicit solvation models in QM calculations [73]. The first is a significant reduction in the system's number of degrees of freedom. For example, if one explicitly treats 1,000 solvent water molecules, this adds 9,000 degrees of freedom. Observable structural and dynamical properties of a solute must be averaged over these degrees of freedom, typically by Monte Carlo or MD simulations. However, if one can treat the solvent as a continuous medium bathing the solute, the averaging becomes implicit in the properties attributed to the bath. The second advantage is that SCRF theory based on continuum model provides a very convenient and accurate way to treat the strong, long-range electrostatic solute-solvent interactions that dominate many solvation phenomena. These are why the SCRF methods based on the continuum theory have been being widely employed to investigate solvent effects on molecular structures, properties, and chemical reactions in solution.

In addition, the SCRF calculation can be coupled with an appropriate use of a supermolecule model or a combined QM/MM approach or other explicit/discrete solvation model to further consider some important specific solute-solvent interactions and to develop semi-continuum approach. The combined use of an SCRF procedure and a supermolecule model may also be called hybrid supermolecule-continuum approach [80–85], in which the solute and a few solvent molecules are considered as a supermolecule treated quantum mechanically and the bulk solvent is considered as a dielectric continuum. Semi-continuum approach, or called combined discrete/continuum approach, also includes the combination of an SCRF method with a combined QM/MM method or other discrete solvation model, in which part of the solvent is explicitly accounted for with certain approximation, such as the point charge representation, dipole polarizability representation, and effective fragment potential (EFP) representation, and the remaining bulk solvent is modeled as a dielectric continuum medium.

3 Fully polarizable continuum model (FPCM)

3.1 Basic FPCM method

The continuum approach is currently the most popular choice for describing solvent in the QM-based electronic structure calculations, due to the reasons mentioned above. Despite the advantages of continuum models, previously employed SCRF methods based on continuum models have limitations that compromise their accuracy in determination of the effective polarization charge distribution that represents the solvent reaction field. Many practical SCRF implementations for general molecular-shape cavities describe the

solvent polarization either explicitly or implicitly in terms of an apparent surface charge distribution spread over the boundary of the solute cavity. However, unconstrained QM calculation of the solute electronic structure usually leads to a tail of the wave function penetrating outside the solute cavity, thereby producing an additional volume polarization [86] in the reaction field that has been rarely recognized or treated. One of the major problems existing in most of the previous SCRF implementations is the neglect of volume polarization or the inaccuracy of its treatment. The effects of volume polarization can be represented by an apparent volume charge distribution spread throughout the entire dielectric medium. Neglecting charge penetration while retaining the simple dielectric continuum model leads to inconsistencies during the solution of Poisson's equation which several groups of researchers have attempted to rectify in cursory fashion through various *ad hoc* charge renormalization schemes, such as those used in the polarizable continuum model (PCM) [87], isodensity surface polarizable continuum model (IPCM) [88], integral equation formulation of polarizable continuum model (IEFPCM) [89, 90], and conductor-like screening model (COSMO, or CPCM) [91, 92] implemented in the popularly employed *Gaussian* program.

In order to accurately treat the long-range electrostatic and associated with them inductive interactions between solute and solvent, an efficient implementation of the surface and volume polarization for electrostatic interaction (SVPE) [93–96] has been developed based on the continuum model. The SVPE method is also known as the fully polarizable continuum model (FPCM) [4, 97–109], because it fully accounts for both surface and volume polarization in the SCRF calculation. According to the FPCM method [93], the effective Schrödinger equation of a solute in a solvent environment (with a dielectric constant ϵ) can be expressed as

$$(H^{vac} + V^{pol})|\psi\rangle = E|\psi\rangle. \quad (3.1)$$

In this equation, $|\psi\rangle$ is the wavefunction of the solute in solution and E is the corresponding energy. The effective Hamiltonian (H) of solute in solution is described as sum of the solute Hamiltonian H^{vac} in vacuum and long-ranged solute-solvent electrostatic interaction energy V^{pol} : $H = H^{vac} + V^{pol}$. According to the FPCM method [93], the potential energy V^{pol} due to the solute-solvent interaction can concisely and conveniently be represented by effective solvent polarization charge distributions $\sigma(\mathbf{s})$ on the solute cavity surface Γ (*i.e.* surface polarization charge distribution) and $\beta(\mathbf{r})$ outside the cavity surface (*i.e.* volume polarization charge distribution). Both the surface and volume polarization charge distributions, *i.e.* $\sigma(\mathbf{s})$ and $\beta(\mathbf{r})$, are determined by solving the Poisson's equation for a given solute charge density $\rho(\mathbf{r})$,

$$\nabla^2\Phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \begin{cases} 1, & \mathbf{r} \subseteq V^{int}, \\ 1/\epsilon, & \mathbf{r} \subseteq V^{ext}. \end{cases} \quad (3.2)$$

Here solute stays within the inner region V^{int} inside the solute cavity and solvent resides in the outer region V^{ext} outside the cavity. Such volume separation naturally leads to the

separation of the electrostatic potential $\Phi(\mathbf{r})$ into two parts — vacuum potential $\Phi^{vac}(\mathbf{r})$ and reaction potential $\Phi^{pol}(\mathbf{r})$. Eq. (3.2) is solved numerically for a general irregularly-shaped solute cavity in the FPCM method, without neglecting $\rho(\mathbf{r})$ when $\mathbf{r} \subseteq V^{ext}$. Many other SCRF methods, such as PCM, IPCM, and self-consistent IPCM (SCIPCM) implemented in the *Gaussian* program, neglect the volume polarization. In other words, when $\mathbf{r} \subseteq V^{ext}$, $\rho(\mathbf{r})$ is considered to be zero in the PCM, IPCM, and SCIPCM methods such that the Poisson's equation is simplified as the Laplace's equation. So, those SCRF methods approximate the reaction potential by accounting only for the surface polarization, arising from the discontinuity of the dielectric constant across the cavity surface. The surface polarization can be described as a charge distribution $\sigma(\mathbf{s})$ on the cavity surface Γ , and it is proportional to the normal component of the electric field, $E_n(\mathbf{s})$, on the cavity surface,

$$\sigma(\mathbf{s}) = -\frac{\varepsilon-1}{\varepsilon+1} \frac{1}{2\pi} E_n(\mathbf{s}). \quad (3.3)$$

The assumption, in which the total electric field is the sum of electric fields produced by the solute charge distribution and by the surface polarization, *i.e.*

$$E_n(\mathbf{s}) = E_n^{vac}(\mathbf{s}) + E_n^\sigma(\mathbf{s}), \quad (3.4)$$

is true only if there is no solute charge outside the cavity. In Eq. (3.4), $E_n^{vac}(\mathbf{s})$ and $E_n^\sigma(\mathbf{s})$ refer to the contributions from the solute charge distribution and by the surface polarization, respectively. As discussed in the previous studies [93], the solute charge penetration through the cavity surface into the external region has two distinct contributions to the reaction potential $\Phi^{pol}(\mathbf{r})$. It contributes directly to the reaction potential through the volume polarization charge distribution $\beta(\mathbf{r})$ as given by

$$\Phi^\beta(\mathbf{s}) = \int_V \frac{\beta(\mathbf{r})}{|\mathbf{s}-\mathbf{r}|} d^3\mathbf{r}. \quad (3.5)$$

Here the volume polarization charge $\beta(\mathbf{r})$ is zero inside the cavity (internal region V^{int}) and is equal to $-((\varepsilon-1)/\varepsilon)\rho(\mathbf{r})$ outside the cavity (external region V^{ext}). Furthermore, it contributes indirectly to the reaction potential through the electric field $E_n^\beta(\mathbf{s})$ generated on the cavity surface Γ ,

$$E_n^\beta(\mathbf{s}) = -\frac{\varepsilon-1}{\varepsilon} \int_{V^{ext}} \rho(\mathbf{r}) \frac{n(\mathbf{s})(\mathbf{s}-\mathbf{r})}{|\mathbf{s}-\mathbf{r}|^3} d^3\mathbf{r} \quad (3.6)$$

which affects the surface polarization charge distribution when solving the following equation to self-consistency:

$$\sigma(\mathbf{s}) = \frac{1}{2\pi} \frac{\varepsilon-1}{\varepsilon+1} \left[E_n^{vac}(\mathbf{s}) + E_n^\sigma(\mathbf{s}) + E_n^\beta(\mathbf{s}) \right]. \quad (3.7)$$

The free energy of the solute in solution is written as

$$G = E^{\text{int}} + \frac{1}{2} \langle V^{\text{pol}} \rangle = E^{\text{int}} + \frac{1}{2} \langle V_{\sigma}^{\text{pol}} \rangle + \frac{1}{2} \langle V_{\beta}^{\text{pol}} \rangle = E^{\text{int}} + \frac{1}{2} \sum_i \sigma_i \Phi_i + \frac{1}{2} \sum_{ik} \beta_{ik} \Phi_{ik}, \quad (3.8)$$

where $E^{\text{int}} = \langle \Psi | H^{\text{vac}} | \Psi \rangle$ is the internal energy of solute and is evaluated by using the Hamiltonian in vacuum. The electrostatic potential (Φ) includes both electronic and nuclear contributions. The discrete point charges β_{ik} (volume polarization charge at angular ray i and volume layer k) are calculated by using multiple volume layers and reproduce the exact charge sum rule determined by the well-known Gauss Law:

$$\sigma + \beta = -\rho(\epsilon - 1)/\epsilon \quad (3.9)$$

in which σ , β , and ρ refer to the total surface polarization charge, total volume polarization charge, and total (net) charge of the solute, respectively.

The aforementioned discussion shows that the FPCM method is capable of determining volume polarization effects for a general irregularly-shaped solute cavity in addition to the more commonly treated surface polarization. In other SCRF implementations, volume polarization effects are ignored or approximately modeled by modifying the surface polarization charge distribution through the surface charge simulation [110–113] or/and charge renormalization, or the solute charge distribution is simply represented by a set of point charges at the solute nuclei.

The accuracy of results obtained from the numerical FPCM computation is related to the number (N) of surface nodes used to describe the surface polarization charge distribution on the cavity surface and to the number (M) of layers used to describe the volume polarization charge distribution outside the cavity [93]. According to the FPCM implementation, if one could employ infinite number of surface nodes and infinite number of layers outside the cavity, then the calculated numerical results would be exactly the same as those obtained from the exact analytic solutions of the Poisson's equation, provided that both are employed with a same QM approximation level. Practically, the FPCM calculations converge very quickly with increasing number of surface nodes and with increasing number of layers. Compared to the exact volume polarization correction determined by the FPCM calculation, all the charge renormalization schemes examined can sometimes lead to energy corrections of the wrong sign [93]. So, an SCRF calculation with a charge renormalization scheme could give even worse results than the corresponding SCRF calculation in which volume polarization is completely ignored.

3.2 Cavity size for the FPCM calculations

Since the solute cavity surface is defined as a solute electronic charge isodensity contour determined self-consistently during the FPCM iteration process, the FPCM results, converged to the exact solution of Poisson's equation, rather than the Laplace's equation, with a given numerical tolerance [93], depend only on the contour value at a given dielectric constant and on the QM level of theory used. This single parameter value has

been determined to be 0.001 a.u. based on an extensive calibration study [94] seeking the best overall agreement with experimental conformational free energy differences (62 experimental observations) in various polar solutes existing in various solvents. Based on the fitting process employed in the calibration, the root-mean-squares (rms) deviation of the 62 experimental values from the results calculated by FPCM method using the 0.001 a.u. contour is 0.096 kcal/mol [94]. The calibration studies [94] also revealed that the 0.001 a.u. contour is not affected significantly by the volume polarization. For this reason, the calibrated 0.001 a.u. contour may be used also for the SCRf calculations using other simplified continuum solvation models in which the volume polarization is neglected.

3.3 Effects of volume polarization effects

Practical FPCM calculations were carried out to determine solvation effects on a variety of conformational free energy differences [93], NMR chemical shifts [95, 109], pK_a [4, 100, 102, 107], and (free) energy barriers for chemical reactions [97, 101, 103–105, 114]. The calculated results indicate that the SCRf calculations can consistently reach high accuracy only when the volume polarization is accurately determined. For example, with the calibrated 0.001 a.u. contour the solvent shifts of nitrogen chemical shift in CH₃CN determined by the FPCM calculations are in good agreement with the available experimental data, whereas the solvent shifts determined by the corresponding SCRf calculations neglecting volume polarization are ~ 12 times larger than the FPCM results [95].

3.4 Available software with the FPCM functionality

The basic FPCM method was initially implemented in a local version [93], and then the publicly available version, of the GAMESS program [115]. More recently, the FPCM method has been implemented in a local version [116] of Gaussian program [117]. In the newest local version (M. J. Vilkas and C.-G. Zhan, unpublished results) of the Gaussian program, the analytical first energy derivatives have been developed for the FPCM method so that the FPCM method can be used for the geometry optimizations and also for the vibrational frequency calculations by using numerical second energy derivatives based on the analytical first energy derivatives.

3.5 Non-electrostatic solute-solvent interactions

It should be pointed out that the SVPE or FPCM method itself only accurately evaluates the dominant electrostatic part of the solvation free energy. The relatively less-important non-electrostatic solute-solvent interactions, including the cavitation, dispersion, and Pauli repulsion, should be estimated otherwise or after the SVPE or FPCM calculation itself. The non-electrostatic interactions are usually short-range, whereas the electrostatic interactions are usually long-range. For most practical chemical applications, accurate

determination of the electrostatic part of the solvation free energy is good enough, as demonstrated in the previous computational studies [96–98, 103–108, 110, 114]. This is because the non-electrostatic contributions can be cancelled out for many chemical problems [104], such as the calculated free energy barriers and reaction free energies. Usually, the non-electrostatic contributions to the total solvation free energy of a reaction system usually do not change significantly during the reaction process. For example, the estimated non-electrostatic contributions to the total solvation free energies of transition state structures are very close to those of the corresponding reactants [104].

When it is necessary to determine the absolute free energy of solvation, the non-electrostatic contributions must be estimated. The non-electrostatic contributions may be estimated either empirically [118] or by using a hybrid supermolecule-continuum approach (see below). For the empirical approach, a new method, denoted by SMVLE (which represents the Solvation Model including surface, Volume, and Local Electrostatic effects and atomic surface tensions) [119], has been developed recently. The SMVLE method, which has been implemented in the GAMESS program, can accurately predict absolute aqueous free energies of solvation by combining (1) the SVPE method [116], (2) semiempirical atomic surface tensions as used in the SM6 model [120], and (3) a new functional form that explicitly accounts for the local electrostatic effect. Specifically, the free energy of solvation is a sum of three terms:

$$\Delta G_{\text{solv}} = \Delta G_{\text{SVPE}} + G_{\text{CDS}} + G_{\text{L}}, \quad (3.10)$$

where ΔG_{solv} is the absolute solvation free energy, ΔG_{SVPE} is the bulk electrostatic portion calculated by SVPE method, G_{CDS} is the semiempirical term based on atomic surface tensions, and G_{L} is the semiempirical electric-field-dependent term; see ref. [119] for the detailed equations. The parameters for SMVLE have been calibrated against a broad range of solutes, including 272 neutrals and 143 ions [119]. The predicted aqueous solvation free energies by the parameterized SMVLE method correlate very well with experiment and have a value of the square of the correlation coefficient equal to 0.9945 and a slope of 0.9847. Comparisons with previous SMx solvation models show that the SMVLE model not only has comparable accuracy for neutrals but that it also impressively increases the predictive accuracy for ions. The semiempirical terms derived from the electric field have been found to be primarily responsible for the increase in predictive accuracy for ions. The outward-directed normal electric fields that make the most important contributions account for strong interactions between the ionic solute and the nearby solvent, which makes the addition of explicit water molecules unnecessary. These encouraging results [119] demonstrate that the parameterized SMVLE is accurate and effective in predicting absolute solvation free energies not only for neutral molecules, but also for ions exhibiting strong solute-solvent interactions.

3.6 FPCM-based hybrid supermolecule-continuum approach

Concerning the short-range non-electrostatic interactions between the solute and solvent, pure dielectric continuum solvation models completely ignore the solvent structure, and,

therefore, might not account for some important effects caused by specific solute-solvent interactions, especially for chemical reactions assisted directly by solvent molecules. The pure reaction field calculation can be improved by coupling with a supermolecule model that includes solute and a few solvent molecules interacting with the solute.

An FPCM-based hybrid supermolecule-continuum approach [80–85] has been developed to predict free energies of solvation. In the FPCM-based hybrid supermolecule-continuum approach, the solute and part of the solvent surrounding the solute is treated quantum mechanically and the remaining bulk solvent is considered to be a dielectric continuum medium accounted for by using the FPCM method. According to this approach, the calculated results can systematically be improved by increasing the number of quantum mechanically treated explicit solvent molecules. It has been shown [81–85] that the FPCM-based hybrid supermolecule-continuum calculations can quickly converge to the infinite number of explicit solvent molecules. For example, the free energy barriers calculated for the base-catalyzed hydrolysis of amides at the CCSD(T)/aug-cc-pVDZ level are converged at $n=5$ [85], and the absolute hydration free energy of the proton calculated at high levels are converged at $n=4$ [81]. The FPCM and the FPCM-based hybrid supermolecule-continuum approach have been used to solve a variety of crucial scientific problems, including those that cannot be solved by experiment alone [80–85].

4 Determination of protein-ligand interaction and computational drug design

The basic computational methods discussed above provide a foundation for practical computational studies on a variety of protein-ligand interactions and, thus, for rational drug design. Below we will briefly discuss some representative computational studies and rational drug design efforts, illustrating how one can employ the state-of-the-art computational approaches to study of protein-ligand interactions and perform rational drug design.

4.1 Determine the most favorable molecular species of ligand interacting with a protein

For a protein binding with multiple molecular species of a ligand, each molecular species may form a microscopic binding complex with the protein. Thus, one may have multiple microscopic binding complexes for a pair of protein and ligand. The computational methods mentioned above allow us to predict the relative free energies of various possible molecular species of the free ligand in solution and to calculate the microscopic binding free energy of each species with the protein. With all of these energetic data, one can determine the statistical distribution of the various molecular species in the protein binding site [59], because the relative free energies associated with the determined

microscopic protein-ligand binding complexes can be evaluated as

$$\Delta G_{P-L}(i) = \Delta G_L(i) + \Delta G_{\text{bind}}(i), \quad (4.1)$$

where $\Delta G_L(i)$ is the relative free energy of the i th molecular species of the free ligand in solution, $\Delta G_{\text{bind}}(i)$ represents the microscopic binding free energy for the protein with the i th molecular species of the ligand, and $\Delta G_{P-L}(i)$ is the relative free energy of the i th microscopic binding complex between the protein and the i th molecular species. The most favorable microscopic binding complex, or the most favorable molecular species in the protein binding site, is associated with the lowest $\Delta G_{P-L}(i)$ value.

According to Eq. (4.1), the lowest $\Delta G_{P-L}(i)$ value is not necessarily associated with the lowest $\Delta G_L(i)$ value. In other words, the most favorable molecular species of a ligand in the protein binding site is not necessarily the most favorable molecular species of the free ligand in solution. For example, (S)-(-)-nicotine can have three types of protonation states: the free base (deprotonated state, neutral species), singly protonated state (cation), and doubly protonated state (dication). Both computational and experimental studies have consistently demonstrated that the dominant molecular species of the free (S)-(-)-nicotine in a neutral (aqueous) solution or an aqueous solution with the physiologic pH (pH 7.4) is a singly protonated state. However, the combined MM-PBSA and FPCM-based QM calculations [121] have revealed that the most favorable molecular species of (S)-(-)-nicotine in the active site of cytochrome P450 2A6 (CYP2A6) is the free base (the deprotonated state). CYP2A6 is a crucial enzyme responsible for nicotine metabolism in the body. The finding of the most favorable molecular species of nicotine from the combined MM-PBSA and FPCM-based QM calculations provides an essentially important starting point for further computational studies on the detailed metabolic pathway of nicotine at the molecular level. In fact, following the combined MM-PBSA and FPCM-based QM calculations, further first-principles QM/MM-free energy (QM/MM-FE) calculations have been performed to uncover the detailed reaction pathways and the corresponding free energy profiles for CYP2A6-catalyzed metabolic reactions of nicotine [121]. The computational results [121] are consistent with available experimental data and provide a solid base for future rational design of novel drugs that aim to control the nicotine metabolism.

It should also be pointed out that the most favorable molecular species of a ligand in the binding site of a protein is not necessarily the same as that of the same ligand in the binding site of another protein. For example, whereas the most favorable molecular species of (S)-(-)-nicotine in the active site of CYP2A6 is the free base [121], the most favorable molecular species of (S)-(-)-nicotine in the binding sites of nAChRs is always a singly protonated state according to the FPCM-based QM calculations in combination with the microscopic binding free energy calculations [59–61].

These examples demonstrate that, for a reliable computational determination of the most favorable molecular species of a ligand in the binding site of a protein, one must accurately determine both the relative free energies of all molecular species of the free ligand in solution and the corresponding microscopic binding free energies.

In addition, based on the FPCM-based QM calculations along with other computational modeling and simulations, a reliable computational strategy [1, 2] has been developed to study the structural identity of a catalytic ligand bridging metal ions in the active sites of metalloenzymes. The computational strategy has been employed to successfully determine the active site structures and catalytic mechanisms of phosphotriesterase (PTE) and phosphodiesterase (PDE) [1–6]. These computational studies on the detailed protein structures demonstrate how computational modeling and simulations can be carried out to assess the protein structure questions not resolvable from the X-ray diffraction techniques.

4.2 From microscopic binding to phenomenological binding affinity

The FPCM-based QM calculations in combination with the microscopic binding free energy calculations can also be performed to determine not only the most favorable molecular species of a ligand in the binding site of a protein, but also the phenomenological binding affinity associated with all molecular species of a ligand binding with a protein. In particular, the relative $\Delta G_{P-L}(i)$ values calculated by using Eq. (4.1) can be used to determine the Boltzmann distribution of all molecular species of the ligand in the binding site. Based on the determined Boltzmann distribution of the all molecular species, one can conveniently evaluate the phenomenological binding affinity which is experimentally observable, as demonstrated in a computational study [59].

In particular, the FPCM-based QM calculations in combination with the microscopic binding free energy calculations were carried out to study how the $\alpha 4\beta 2$ nAChR binds with various molecular species of two typical agonists, (*S*)-(-)-nicotine and (*R*)-(-)-deschloroepibatidine [59], each of which are distinguished by different free bases and protonation states. Based on the computational results, predictions were made regarding the corresponding microscopic binding free energies. Hydrogen bonding and cation- π interactions between the receptor and the respective ligands were found to be the dominant factors differentiating the binding strengths of different microscopic binding species. The calculated results and analyses demonstrate that for each agonist, all the species are interchangeable and can quickly achieve a thermodynamic equilibration in solution and at the nAChR binding site. This allows us to evaluate the equilibrium concentration distributions of the free ligand species and the corresponding microscopic ligand-receptor binding species. The calculated equilibrium concentration distributions of the ligand species clearly show their pH-dependence and provide the microscopic information required for further determination of the phenomenological binding affinity of the ligand with the $\alpha 4\beta 2$ nAChR [59]. The predicted equilibrium concentration distributions, pK_a values, absolute phenomenological binding affinities of the ligand species and their pH-dependence are all in good agreement with available experimental data, which suggests that the computational strategy of studying interactions of ligands with receptors from their microscopic binding species and affinities to the phenomenological binding affinity is reliable for studying protein-ligand binding, and thus, should be a valuable approach

for future rational design of drugs targeting the $\alpha 4\beta 2$ nAChR [59].

Besides the studies on $\alpha 4\beta 2$ nAChR binding with the above two agonists, the same computational approach has also been employed to study how $\alpha 4\beta 2$ nAChR and other nAChR subtypes bind with their agonists and antagonists, leading to the detailed understanding of the observed relative binding affinities and the subtype selectivity of the ligands [60,61]. It is essential for rational drug design to achieve a detailed understanding of the observed relative binding affinities and the subtype selectivity of the ligands.

The general strategy of the "from-microscopic-binding-to-phenomenological-binding" approach [59] could also be useful in future studies of other types of ligand-protein interactions involving multiple molecular species of a ligand and in other related rational drug design endeavors.

4.3 Determine whether a ligand of a receptor protein should be an agonist or antagonist

A receptor protein, such as a nAChR (ligand-gated sodium channel), may have two different states: open- and closed-channel states. For convenience, here we discuss nAChR as an example. In theory, to computationally predict whether a nAChR ligand should be an agonist or antagonist, one may first determine how the ligand binds with the closed-channel state (the rest state) of the nAChR and then carry out a sufficiently long MD simulation on the determined nAChR-ligand binding structure in a reasonable model of the physiological environment. If the ligand is an agonist, then the channel should eventually open during the MD simulation. If the ligand is an antagonist, then the channel should not open during the MD simulation. Practically, this theoretically "reasonable" approach does not work for a nAChR. This is because the average time required to open nAChR channels is in milliseconds (ms), *e.g.*, ~ 59 ms for $\alpha 4\beta 2$ nAChR [122, 123] and, therefore, the MD simulation on a nAChR-ligand complex must be performed for at least many milliseconds to be really meaningful. Such a time scale is insurmountable for a fully relaxed (real-time) MD simulation (with a usual time step of 1 or 2 fs) of a protein system as large as a nAChR on any supercomputer in the World at this point of time. Currently, a meaningful MD simulation (with a usual time step of 1 or 2 fs) on a fully solvated nAChR system can only be performed for nanoseconds by using supercomputing time in days.

In fact, targeted MD simulations (*i.e.* the MD simulations with certain artificial forces that accelerate the change of nAChR structure from the starting closed/open-channel state to the targeted open/closed-channel state) were performed on nAChRs to study some major molecular motions related to the opening and closing of nAChR channels [124]. As expected, the artificial forces speeded up the opening/closing of the channel so that the channel opening/closing could be simulated in only nanoseconds. However, when the targeted MD simulation is performed on a nAChR-ligand complex, the channel opening/closing can always be observed due to the use of the artificial forces no matter whether the ligand is an agonist or antagonist. So, the targeted MD simulations cannot

be used to predict whether a nAChR ligand is agonist or antagonist.

In a recently reported computational study [125], a practical and fast computational approach was developed to predict whether a nAChR ligand is an agonist or antagonist by calculating the microscopic binding free energies for both the open and closed states of $\alpha 4\beta 2$ nAChR interacting with the protonated and deprotonated forms of 27 representative ligands (agonists and antagonists) along with the FPCM-based QM calculations. The FPCM-based QM calculations were carried out to determine the relative free energies of the protonated and deprotonated forms and, thus, the pK_a of agonist/antagonist. The modeled receptor-ligand binding structures and calculated binding free energies consistently reveal that all of the antagonists bind more favorably with the closed-channel state of the receptor, whereas all of the agonists bind more favorably with the open-channel state. Depicted in Fig. 2 are the modeled structures of the closed-channel state binding with a representative antagonist and the open-channel state binding with a representative agonist. These results help to better understand why an agonist can open the channel, whereas an antagonist cannot. The binding free energies calculated for the favorable binding of antagonists with the closed-channel state and for the favorable binding of agonists with the open-channel state are all close to the corresponding experimentally-derived binding free energies [125]. The good agreement between the computational and experimental data suggests that the determined binding structures and calculated binding free energies are reasonable.

The computational results [125] led to propose a novel computational strategy and protocol that can be used to theoretically predict whether a nAChR ligand should be an agonist or antagonist. According to the computational protocol [125], one only needs to calculate the relative binding free energies for a ligand binding with both the open- and closed-channel states of the receptor and, thus, determine the most favorable channel state of the receptor binding the ligand, as the agonist and antagonist bind more favorably with the open- and closed-channel states, respectively. This protocol and the general computational strategy are expected to be valuable in structure-based rational design of novel agonists and antagonists of nAChRs as therapeutic agents. For example, a possibly more potent agonist of $\alpha 4\beta 2$ receptor may be designed to have a more favorable binding with the open-channel structure, whereas a possibly more potent antagonist of $\alpha 4\beta 2$ receptor may be designed to have a more favorable binding with the closed-channel receptor.

4.4 Determine the interaction between a drug and its catalytic antibody

A unique, efficient computational approach [126] has been developed to study competing reaction pathways and the corresponding free energy barriers for the chemical reaction of a substrate catalyzed by a catalytic antibody without performing the time-consuming QM/MM calculations. The computational approach has been used to study cocaine hydrolysis catalyzed by an anti-cocaine catalytic antibody (mAb 15A10) [126]. The efficient computational approach capable of studying the antibody catalysis is based on the re-

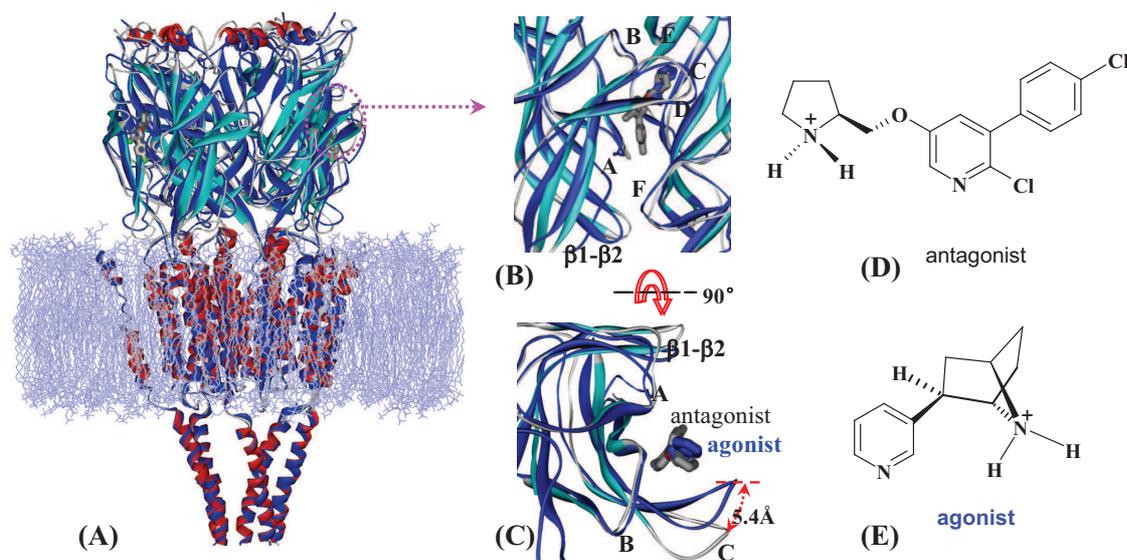


Figure 2: (A) A side view of the closed-channel structure of $\alpha 4 \beta 2$ nAChR binding with a representative antagonist. LBD refers to ligand binding domain, TMD to transmembrane domain, and ID to intracellular domain. For comparison, the corresponding open-channel structure is superimposed and in blue color. The phospholipid bilayer is in light blue color. (B) and (C) local view of one of the two equivalent ligand-binding sites at the extracellular ligand-binding domain of the receptor, in which the antagonist is shown in stick and the agonist in stick and blue color. Labeled are the functional loops in the right panels (A-F, and $\beta 1$ - $\beta 2$ loop). In comparison between the closed-channel and the open-channel structures, one of the most obvious differences exists in the motion of the C loop induced by the ligand binding. Such a difference is represented by the red-curved arrow and labeled with the distance. For clarity, the $\beta 2$ subunit and the hydrogen atoms of the agonist/antagonist are not shown. (D) Molecular structure of the antagonist in the binding structure. (E) Molecular structure of the agonist in the binding structure.

action coordinate calculations on the non-enzymatic hydrolysis of cocaine in solution by using an FPCM-based first-principles QM approach [106] and the MD simulations on the antibody binding with each of the possible molecular species of cocaine existing in the reaction process (including both the reactant and transition-state structures) [126].

The computational studies led to predict the free energy barriers for the competing reaction pathways of the cocaine hydrolysis catalyzed by mAb 15A10. On the basis of the calculated binding free energies, one can evaluate the free energy barrier shift from the cocaine hydrolysis in water to the antibody-catalyzed cocaine hydrolysis for each reaction pathway. The free energy barriers for the antibody-catalyzed cocaine hydrolysis were predicted to be the corresponding free energy barriers for the cocaine hydrolysis in water plus the calculated free energy barrier shifts. Based on the predicted free energy barriers, the dominant reaction pathway for the antibody-catalyzed cocaine hydrolysis was determined [126]. The calculated free energy barrier shift of -6.33 kcal/mol from the dominant reaction pathway of the cocaine benzoyl ester hydrolysis in water to the dominant reaction pathway of the antibody-catalyzed hydrolysis of cocaine benzoyl ester is in good agreement with the experimentally-derived free energy barrier shift of -5.93

kcal/mol (corresponding to the experimental rate acceleration $k_{cat}/k_0=23,000$), while the calculated binding free energy of -4.88 kcal/mol for the cocaine-antibody binding agrees with the experimentally-derived binding free energy of -4.97 kcal/mol (estimated from the experimental K_M value of $220 \mu\text{M}$) [126].

In light of the good agreement between the calculated energetic results and available experimental kinetic data, the computational protocol for calculating the free energy barrier shift from the cocaine hydrolysis in water to the antibody-catalyzed cocaine hydrolysis may be useful in future rational design of possible high-activity mutants of the catalytic antibody as anti-cocaine therapeutics. The general computational strategy for calculating the free energy barrier shift may also be valuable for studying a variety of chemical reactions catalyzed by other antibodies or proteins through non-covalent bonding interactions with the substrates [126].

4.5 Design new drugs based on computational modeling of protein-ligand interaction

Based on computational modeling of the detailed protein-ligand interactions, one can rationally design potentially valuable new drugs. A new drug to be designed may be a small-molecule ligand, such as inhibitor of an enzyme or agonist/antagonist of a receptor protein which aims to control/regulate the physiologic process involving the protein. The small-molecule drugs can be designed based on structural and energetic understanding of the protein-ligand interactions. The structure-based drug design includes virtual screening of known compounds collected in a library (for drug lead identification) [127] and *de novo* design which aims to design novel compounds (for drug lead identification and/or lead optimization).

A new drug to be designed may also be a protein mutant for a protein drug design effort. One type of protein drugs is the enzyme therapy which aims to detoxify the toxic compounds (such as abused drugs) in the body. For example, the FPCM-based first-principles QM calculations [97, 106] on the reaction mechanism for the non-enzymatic hydrolysis of cocaine in solution were followed by further MD simulations and QM/MM calculations [31, 128–132] on the detailed mechanism for the enzymatic hydrolysis of cocaine, leading to discovery of highly efficient cocaine hydrolases [133–138] as promising candidates for anti-cocaine therapeutics [139–145]. The first one [133] of these cocaine hydrolases discovered by Zhan's group has already been developed into an investigational new drug (known as TV-1380) by Teva Pharmaceutical Industries Ltd; the outcomes of the human clinical trials have revealed that this new drug is safe and efficacious for human [146].

The general concept of the enzyme therapy development for treatment of cocaine overdose and addiction may also be used to explore possible enzymes suitable for treatment of other drugs of abuse or detoxification of other toxic compounds (*e.g.*, chemical warfare nerve agents). In order to design a therapeutically useful enzyme which can metabolize a given ligand, one will first need to examine all possible metabolic pathways

of the ligand and identify a favorable metabolic pathway producing biologically inactive metabolites. If a favorable metabolic pathway and the corresponding native enzyme can be identified, then the general computational design approaches that have been used to design cocaine hydrolases may be employed to design high-activity mutants of the chosen ligand-metabolizing enzyme against the ligand. When necessary, further computational design will be performed to extend the *in vivo* half-life of the discovered enzyme so that the enzyme can be long-acting. In fact, thermostable mutants of bacterial cocaine esterase (CocE) have been designed and discovered successfully through computational modeling [147–150]. One of the designed and discovered CocE mutants (*i.e.* T172R/G173Q) [147] has been licensed to Reckitt Benckiser Pharmaceuticals Inc for cocaine overdose treatment. Investigational new drug (IND) application for human clinical use of the T172R/G173Q mutant product (RBP-8000) has been filed to the US Food and Drug Administration (FDA) in July 2011.

The encouraging outcomes of the drug discovery and development efforts based on computational modeling have demonstrated that computational drug design is valuable not only for small-molecule drug discovery, but also for protein drug discovery and development [139–144, 151].

5 Summary and concluding remarks

Computational modeling has been recognized as a powerful tool in understanding detailed protein-ligand interactions at molecular level and in rational drug design. To theoretically account for a protein binding with multiple molecular species of a ligand, one must accurately predict the relative free energies of all of the molecular species of the free ligand in solution and the corresponding microscopic binding free energies for all of the molecular species binding with the protein. A reliable first-principles QM method is required to predict the relative free energies of various molecular species of a ligand because the structural differences between different molecular species are usually associated with the differences in covalent bonds. As well known, it is extremely challenging to accurately determine the solvent effects in the first-principles QM calculations on molecules in solution. It has been a dream of the theoretical and computational chemists that QM-based electronic structure calculations on molecules in solution can achieve the chemical accuracy. However, the encouraging results obtained from the recently reported FPCM-based QM calculations suggest that the chemical accuracy of solvation-included first-principles QM calculation is now possible. The combined use of the FPCM-based QM calculations and other computational modeling and simulations enables us to accurately account for a protein binding with multiple molecular species of the ligand in solution.

FPCM-based QM calculations in combination with various other types of computational modeling and simulations have been carried out to study a variety of protein-ligand interactions, such as predicting the most favorable molecular species of ligand in-

teracting with a protein, the phenomenological binding affinity and its pH dependence, whether a ligand of a receptor protein should be an agonist or antagonist, the interaction between a drug and its catalytic antibody and associated catalytic reaction mechanism. Based on the computational modeling of the detailed protein-ligand interactions, possible new drugs may be designed rationally as either the small-molecule ligands of the protein (if the protein is a drug target) or engineered proteins (if the ligand is a toxic compound which must be removed from the body). The computational drug design has led to practical discovery and development of promising drugs. One may expect to see more and more practical applications of the computational modeling in understanding the detailed protein-ligand interactions and in rational drug design, discovery, and development.

Acknowledgments

This work was supported by the National Science Foundation (grant CHE-1111761) and the National Institutes of Health (grants R01 DA032910, R01 DA013930, R01 DA025100, R01 DA021416, and RC1 MH088480), Alzheimer's Drug Discovery Foundation (ADDA) and Institute for the Study of Aging (ISOA). The authors also acknowledge the Center for Computational Sciences (CCS) at University of Kentucky for supercomputing time on an IBM X-series supercomputer cluster with 340 nodes or 1,360 processors and a Dell supercomputer cluster with 388 nodes or 4,816 processors.

References

- [1] C.-G. Zhan, O. Norberto de Souza, R. Rittenhouse, and R. L. Ornstein, Determination of two structural forms of catalytic bridging ligand in zinc-phosphotriesterase by molecular dynamics simulation and quantum chemical calculation, *J. Am. Chem. Soc.*, 121 (1999), 7279-7282.
- [2] C.-G. Zhan and F. Zheng, First computational evidence for a critical bridging hydroxide ion in phosphodiesterase active site, *J. Am. Chem. Soc.*, 123 (2001), 2835-2838.
- [3] J. Koca, C.-G. Zhan, R. Rittenhouse, and R. L. Ornstein, Mobility of the active site bound paraoxon and sarin in zinc-phosphotriesterase by molecular dynamics simulation and quantum chemical calculation, *J. Am. Chem. Soc.*, 123 (2001), 817-826.
- [4] F. Zheng, C.-G. Zhan, and R. L. Ornstein, Theoretical studies of reaction pathways and energy barriers for base-catalyzed hydrolysis of phosphotriesterase substrates paraoxon and related toxic nerve agents, *J. Chem. Soc. Perkin Trans. 2* (2001), 2355-2363.
- [5] Y. Xiong, H. Lu, Y. Li, G. Yang, and C.-G. Zhan, Characterization of a catalytic ligand bridging metal ions in phosphodiesterases 4 and 5 by molecular dynamics simulations and hybrid quantum mechanical/molecular mechanical calculations, *Biophys. J.*, 91 (2006), 1858-1867.
- [6] Y. Xiong, H.-T. Lu, and C.-G. Zhan, Dynamic structures of phosphodiesterase-5 active site by combined molecular dynamics simulations and hybrid quantum mechanical/molecular mechanical calculations, *J. Comput. Chem.*, 29 (2008), 1259-1267.

- [7] A. Hamza and C.-G. Zhan, Determination of the structure of human phosphodiesterase-2 in a bound state and its binding with inhibitors by molecular modeling, docking, and dynamics simulation, *J. Phys. Chem. B*, 113 (2009), 2896-2908.
- [8] S. Wu, D. Xu, and H. Guo, QM/MM studies of monozinc β -lactamase CphA suggest that the crystal structure of an enzyme-intermediate complex represents a minor pathway, *J. Am. Chem. Soc.*, 132 (2010), 17986-17988.
- [9] P. Bandyopadhyay and M. S. Gordon, A combined discrete/continuum solvation model: application to glycine, *J. Chem. Phys.*, 113 (2000), 1104-1109.
- [10] A. Warshel and M. Levitt, Theoretical studies of enzymic reactions - dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme, *J. Mol. Biol.*, 103 (1976), 227-249.
- [11] M. J. Field, P. A. Bash, and M. Karplus, M. A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations, *J. Comput. Chem.*, 11 (1990), 700-733.
- [12] J. Gao and X. Xia, A prior evaluation of aqueous polarization effects through Monte Carlo QM-MM simulation, *Science*, 258 (1992), 631-635.
- [13] T. Vreven and K. Morokuma, The ONIOM (our own N-layered integrated molecular orbital + molecular mechanics) method for the first singlet excited (S1) state photoisomerization path of a retinal protonated Schiff base, *J. Chem. Phys.*, 113 (2000), 2969-2975.
- [14] U. Singh and P. Kollman, A combined ab initio quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular-systems - applications to the $\text{CH}_3\text{Cl} + \text{Cl}^-$ exchange-reaction and gas-phase protonation of polyethers, *J. Comput. Chem.* 7 (1986), 718-730.
- [15] N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus, Frontier bonds in QM/MM methods: A comparison of different approaches, *J. Phys. Chem. A*, 104 (2000), 1720-1735.
- [16] Q. Cui and M. Karplus, Molecular properties from combined QM/MM methods. I. Analytical second derivative and vibrational calculations, *J. Chem. Phys.*, 112 (2000), 1133-1149.
- [17] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, and M. Karplus, A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method, *J. Phys. Chem. B*, 105 (2001), 569-585.
- [18] M. Trajbl, G. Y. Hong, and A. Warshel, Ab initio QM/MM simulation with proper sampling: "First principle" calculations of the free energy of the autodissociation of water in aqueous solution, *J. Phys. Chem. B*, 106 (2002), 13333-13343.
- [19] E. Rosta, M. Klahn, and A. Warshel, Towards accurate ab initio QM/MM calculations of free-energy profiles of enzymatic reactions, *J. Phys. Chem. B*, 110 (2006), 2934-2941.
- [20] A. Warshel, Computer simulations of enzyme catalysis: Methods, progress, and insights, *Annual Rev. Biophys. Biomol. Struct.* 32 (2003), 425-443.
- [21] Y. Zhang, T. Lee, and Yang, A pseudobond approach to combining quantum mechanical and molecular mechanical methods, *J. Chem. Phys.*, 110 (1999), 46-54.
- [22] Y. Zhang, Improved pseudobonds for combined ab initio quantum mechanical/molecular mechanical methods, *J. Chem. Phys.*, 122 (2005), 024114.
- [23] H. Liu, Y. Zhang, and W. Yang, How is the active site of enolase organized to catalyze two different reaction steps?, *J. Am. Chem. Soc.*, 122 (2000), 6560-6570.
- [24] Y. Zhang, J. Kua, and J. Mccammon, Role of the catalytic triad and oxyanion hole in acetylcholinesterase catalysis: An ab initio QM/MM study, *J. Am. Chem. Soc.*, 124 (2002), 10572-10577.
- [25] Y. Zhang, J. Kua, and J. Mccammon, Influence of structural fluctuation on enzyme reaction

- energy barriers in combined quantum mechanical/molecular mechanical studies, *J. Phys. Chem. B*, 107 (2003), 4459-4463.
- [26] G. Cisneros, H. Liu, Y. Zhang, and W. Yang, Ab initio QM/MM study shows there is no general acid in the reaction catalyzed by 4-oxalocrotonate tautomerase, *J. Am. Chem. Soc.*, 125 (2003), 10384-10393.
- [27] G. Cisneros, M. Wang, P. Silinski, M. Fitzgerald, and W. Yang, The protein backbone makes important contributions to 4-oxalocrotonate tautomerase enzyme catalysis: Understanding from theory and experiment, *Biochemistry*, 43 (2004), 6885-6892.
- [28] C. Corminboeuf, P. Hu, M. E. Tuckerman, and Y. Zhang, Unexpected deacetylation mechanism suggested by a density functional theory QM/MM study of histone-deacetylase-like protein, *J. Am. Chem. Soc.*, 128 (2006), 4530-4531.
- [29] Y. Zhang, H. Liu, and W. Yang, Free energy calculations on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential, *J. Chem. Phys.*, 112 (2000), 3483-3492.
- [30] P. Hu and Y. Zhang, Catalytic mechanism and product specificity of the histone lysine methyltransferase SET7/9: An ab initio QM/MM-FE study with multiple initial structures, *J. Am. Chem. Soc.*, 128 (2006), 1272-1278.
- [31] J. Liu, A. Hamza, and C.-G. Zhan, Fundamental reaction mechanism and free energy profile for (-)-cocaine hydrolysis catalyzed by cocaine esterase, *J. Am. Chem. Soc.*, 131 (2009), 11964-11975.
- [32] J. Liu, Y. Zhang, C.-G. Zhan, Fundamental dephosphorylation mechanism of dimethylphosphonyl-inhibited human acetylcholinesterase, *J. Phys. Chem. B*, 113 (2009), 16226-16236.
- [33] Y. Xiong, J. Liu, G.-F. Yang, and C.-G. Zhan, Computational determination of fundamental pathway and free energy barriers for acetoxyacid synthase-catalyzed condensation reactions of α -keto acids, *J. Comput. Chem.*, 31 (2010), 1592-1602.
- [34] X. Chen, L. Fang, J. Liu, and C.-G. Zhan, Reaction pathway and free energy profile for butyrylcholinesterase-catalyzed hydrolysis of acetylcholine, *J. Phys. Chem. B*, 115 (2011), 1315-1322.
- [35] X. Chen, X. Zhao, Y. Xiong, J. Liu, and C.-G. Zhan, Fundamental Reaction Pathway and Free Energy Profile for Hydrolysis of Intracellular Second Messenger Adenosine 3',5'-Cyclic Monophosphate (cAMP) Catalyzed by Phosphodiesterase-4, *J. Phys. Chem. B*, 115 (2011), Oct. 5 (Epub ahead of print).
- [36] Z. Ke, G. K. Smith, Y. Zhang, H. Guo, Molecular mechanism for eliminylation, a newly discovered post-translational modification, *J. Am. Chem. Soc.*, 133 (2011), 11103-11105.
- [37] A. Nicholls and B. Honig, A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation, *J. Comput. Chem.* 12 (1991), 435-445.
- [38] D. Bashford and D. A. Case, Generalized Born models of Macromolecular Solvation Effects, *Annu. Rev. Phys. Chem.*, 51 (2000), 129-152.
- [39] A. Klamt, Conductor-like screening model for real solvents - a new approach to the quantitative calculation of solvation phenomena, *J. Phys. Chem.*, 99 (1995), 2224-2235.
- [40] X. Huang, Y. Pan, F. Zheng, and C.-G. Zhan, Reaction pathway and free energy profile for pre-chemical reaction step of human butyrylcholinesterase-catalyzed hydrolysis of (-)-cocaine by combined targeted molecular dynamics and potential of mean force simulations, *J. Phys. Chem. B*, 114 (2010), 13545-13554.
- [41] X. Huang, F. Zheng, and C.-G. Zhan, Human butyrylcholinesterase-cocaine binding pathway and free energy profiles by molecular dynamics and potential of mean force simula-

- tions, *J. Phys. Chem. B.*, 115 (2011), 11254-11260.
- [42] W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, and B. Honig, Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects, *J. Comput. Chem.* 23 (2002), 128-137.
- [43] C. Bertonati, B. Honig, and E. Alexov, Poisson-Boltzmann calculations of nonspecific salt effects on protein-protein binding free energies. *Biophys. J.*, 92 (2007), 1891-1899.
- [44] J. Zhu, E. Alexov, and B. Honig, Comparative study of generalized born models: Born radii and peptide folding, *J. Phys. Chem. B.*, 109 (2005), 3008-3022.
- [45] A. Hamza and C.-G. Zhan, How can (-)-epigallocatechin gallate from green tea prevent HIV-1 virus infection? Mechanistic insights from computational modeling and the implication for rational design of anti-HIV-1 entry inhibitors, *J. Phys. Chem. B.*, 110 (2006), 2910-2917.
- [46] M. D. M. AbdulHameed, A. Hamza, and C.-G. Zhan, Microscopic modes and free energies of 3-phosphoinositide-dependent kinase-1 (PDK1) binding with celecoxib and other inhibitors, *J. Phys. Chem. B.*, 110 (2006), 26365-26374.
- [47] P. Bargagna-Mohan, A. Hamza, Y.-E. Kim, Y. K. Ho, N. Mor-Vaknin, N. Wendschlag, J. Liu, R. M. Evans, D. M. Markovitz, C.-G. Zhan, K. B. Kim, and R. Mohan, The tumor inhibitor and anti-angiogenic agent withaferin A targets the intermediate filament protein vimentin, *Chemistry & Biology*, 14 (2007), 623-634.
- [48] X. Huang and C.-G. Zhan, How dopamine transporter interacts with dopamine: Insights from molecular modeling and simulation, *Biophys. J.*, 93 (2007), 3627-3639.
- [49] G.-F. Hao, X.-L. Zhu, F.-Q. Ji, G.-F. Yang, and C.-G. Zhan, Understanding mechanism of drug resistance due to a codon deletion in protoporphyrinogen oxidase through computational modeling, *J. Phys. Chem. B.*, 113 (2009), 4865-4875.
- [50] X. L. Zhu, G.-F. Hao, C.-G. Zhan, and G.-F. Yang, Computational simulations of the interactions between acetyl-coenzyme-A carboxylase and clodinafop: Resistance mechanism due to active and nonactive site mutations, *J. Chem. Inf. Modeling*, 49 (2009), 1936-1943.
- [51] X. Huang, H. Gu, and C.-G. Zhan, Mechanism for cocaine blocking the transport of dopamine: Insights from molecular modeling and dynamics simulations, *J. Phys. Chem. B.*, 113 (2009), 15057-15066.
- [52] P.-L. Zhao, L. Wang, X.-L. Zhu, X. Huang, C.-G. Zhan, J.-W. Wu, and G.-F. Yang, Subnanomolar inhibitor of cytochrome bc1 complex designed via optimizing interaction with conformationally flexible residues, *J. Am. Chem. Soc.*, 132 (2010), 185-194.
- [53] A. Hamza, M. Tong, M. D. M. AbdulHameed, H. Li, A. C. Goren, H.-H. Tai, and C.-G. Zhan, Understanding microscopic binding of human microsomal prostaglandin E synthase-1 (mPGES-1) trimer with substrate PGH2 and cofactor GSH: Insights from computational alanine scanning and site-directed mutagenesis, *J. Phys. Chem. B.*, 114 (2010), 5605-5616.
- [54] G.-F. Hao, G.-F. Yang, and C.-G. Zhan, Computational mutation scanning and drug resistance mechanisms of HIV-1 protease inhibitors, *J. Phys. Chem. B.*, 114 (2010), 9663-9676.
- [55] B. Lei, M. D. M. AbdulHameed, A. Hamza, M. Wehenkel, J. L. Muzyka, X.-J. Yao, K.-B. Kim, and C.-G. Zhan, Molecular basis of the selectivity of the immunoproteasome catalytic subunit LMP2-specific inhibitor revealed by molecular modeling and dynamics simulations, *J. Phys. Chem. B.*, 114 (2010), 12333-12339.
- [56] B. Yang, A. Hamza, Y. Wang, G. Chen, and C.-G. Zhan, Computational determination of binding structures and free energies of phosphodiesterase-2 with benzo[1,4]diazepin-2-one derivatives, *J. Phys. Chem. B.*, 114 (2010), 16020-16028.
- [57] P. Bargagna-Mohan, R. R. Paranthan, A. Hamza, N. Dimova, C. Srinivasan, G. I. Elliott, C.-G.

- Zhan, D. Lau, F. Cambi, and R. Mohan, With aferin A targets intermediate filaments GFAP and vimentin in A model of retinal gliosis, *J. Biol. Chem.*, 285 (2010), 7657-7669.
- [58] S. Z. Fairchild, M. W. Peterson, A. Hamza, C.-G. Zhan, D. M. Cerasoli, and W. E. Chang, Computational characterization of how the VX nerve agent binds human serum paraoxonase 1, *J. Mol. Model.*, 17 (2011), 97-109.
- [59] X. Huang, F. Zheng, P. A. Crooks, L. P. Dvoskin, C.-G. Zhan, Modeling multiple species of nicotine and deschloroepibatidine interacting with $\alpha 4\beta 2$ nicotinic acetylcholine receptor: From microscopic binding to phenomenological binding affinity, *J. Am. Chem. Soc.*, 127 (2005), 14401-14414.
- [60] X. Huang, F. Zheng, X. Chen, P. A. Crooks, L. P. Dvoskin, and C.-G. Zhan, Modeling subtype-selective agonists binding with $\alpha 4\beta 2$ and $\alpha 7$ nicotinic acetylcholine receptors: Effects of local binding and long-range electrostatic interactions, *J. Med. Chem.*, 49 (2006), 7661-7674.
- [61] X. Huang, F. Zheng, C. Stokes, R. L. Papke, C.-G. Zhan, Modeling binding modes of $\alpha 7$ nicotinic acetylcholine receptor with ligands: The roles of Gln117 and other residues of the receptor in agonist binding, *J. Med. Chem.*, 51 (2008), 6293-6302.
- [62] J.-L. Rivail and D. Rinaldi, In computational chemistry: Reviews of current trends, Vol. 1; J. Leszczynski Ed., World Scientific, Singapore, 1996; pp. 139.
- [63] M. Orozco and F. J. Luque, Theoretical methods for the description of the solvent effect in biomolecular systems, *Chem. Rev.*, 100 (2000), 4187-4225.
- [64] J. Tomasi, B. Mennucci, and R. Cammi, Quantum mechanical continuum solvation models, *Chem. Rev.* 105 (2005), 2999-3093.
- [65] C. J. Cramer and D. G. Truhlar, A universal approaches to solvation modeling, *Acc. Chem. Res.*, 42 (2009), 493-497.
- [66] E. Clementi, *Computational Aspects of Large Chemical Systems*, Springer, Berlin, 1980.
- [67] J. Florian and A. Warshel, Langevin dipoles model for ab initio calculations of chemical processes in solution: Parametrization and application to hydration free energies of neutral and ionic solutes and conformational analysis in aqueous solution, *J. Phys. Chem. B*, 101 (1997), 5583-5595.
- [68] A. Broo, G. Pearl, and M. C. Zerner, Development of a hybrid quantum chemical and molecular mechanics method with application to solvent effects on the electronic spectra of uracil and uracil derivatives, *J. Phys. Chem. A*, 101 (1997), 2478-2488.
- [69] Y. Jung, C. H. Ho Choi, M. S. Gordon, Adsorption of water on the Si(100) surface: An ab initio and QM/MM cluster study, *J. Phys. Chem. B*, 105 (2001), 4039-4044.
- [70] T. Hori, H. Takahashi, M. Nakano, T. Nitta, and W. Yang, A QM/MM study combined with the theory of energy representation: Solvation free energies for anti/syn acetic acids in aqueous solution, *Chem. Phys. Lett.*, 419 (2006), 240-244.
- [71] O. Acevedo and W. L. Jorgensen, Advances in QM/MM simulations for organic and enzymatic reactions, *Acc. Chem. Res.*, 43 (2010), 142-151.
- [72] J. Tomasi and M. Persico, Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent, *Chem. Rev.*, 94 (1994), 2027-2094.
- [73] C. J. Cramer and D. G. Truhlar, Implicit solvation models: Equilibria, structure, spectra, and dynamics, *Chem. Rev.*, 99 (1999), 2161-2200.
- [74] M. Kinoshita and F. Hirata, Application of the reference interaction site model theory to analysis on surface-induced structure of water, *J. Chem. Phys.*, 104 (1996), 8807-8815.
- [75] D. S. Palmer, V. P. Sergiievskiy, F. Jensen, and M. V. Fedorov, Accurate calculations of the hydration free energies of druglike molecules using the reference interaction site model, *J.*

- Chem. Phys., 133 (2010), 044104.
- [76] T. Miyata, Y. Ikuta, and F. Hirata, Accurate calculations of the hydration free energies of druglike molecules using the reference interaction site model, *J. Chem. Phys.*, 133 (2010), 044114.
- [77] W. Chen and M. S. Gordon, The effective fragment model for solvation: Internal rotation in formamide, *J. Chem. Phys.*, 105 (1996), 11081-11090.
- [78] T. Kerdcharoen and K. Morokuma, ONIOM-XS: An extension of the ONIOM method for molecular simulation in condensed phase, *Chem. Phys. Lett.*, 355 (2002), 257-262.
- [79] T. Kerdcharoen and K. Morokuma, Combined quantum mechanics and molecular mechanics simulation of Ca^{2+} /ammonia solution based on the ONIOM-XS method: Octahedral coordination and implication to biology, *J. Chem. Phys.*, 118 (2003), 8856-8862.
- [80] C.-G. Zhan, D. W. Landry, and R. L. Ornstein, Reaction pathways and energy barriers for alkaline hydrolysis of carboxylic acid esters in water studied by a hybrid supermolecule polarizable continuum approach, *J. Am. Chem. Soc.*, 122 (2000), 2621-2627.
- [81] C.-G. Zhan and D. A. Dixon, Absolute hydration free energy of the proton from first-principles electronic structure calculations, *J. Phys. Chem. A*, 105 (2001), 11534-11540.
- [82] C.-G. Zhan and D. A. Dixon, First-principles determination of absolute hydration free energy of hydroxide ion, *J. Phys. Chem. A*, 106 (2002), 9737-9744.
- [83] C.-G. Zhan and D. A. Dixon, The nature and absolute hydration free energy of the solvated electron in water, *J. Phys. Chem. B*, 107 (2003), 4403-4417.
- [84] C.-G. Zhan and D. A. Dixon, Hydration of the fluoride anion: Structures and absolute hydration free energy from first-principles electronic structure calculations, *J. Phys. Chem. A*, 108 (2004), 2020-2029.
- [85] Y. Xiong and C.-G. Zhan, Theoretical studies of the transition state structures and free energy barriers for base-catalyzed hydrolysis of amides, *J. Phys. Chem. A*, 110 (2006), 12644-12652.
- [86] D. M. Chipman, Charge penetration in dielectric models of solvation, *J. Chem. Phys.*, 106 (1997), 10194-10206.
- [87] M. Cossi, V. Barone, R. Cammi, and J. Tomasi, Ab initio study of solvated molecules: A new implementation of the polarizable continuum model, *Chem. Phys. Lett.*, 255 (1996), 327-335.
- [88] J. B. Foresman, T. A. Keith, K. B. Wiberg, J. Snoonian, and M. J. Frisch, Solvent effects 5. Influence of cavity shape, truncation of electrostatics and electron correlation on ab initio reaction field calculation, *J. Phys. Chem.*, 100 (1996), 16098-16104.
- [89] E. Cancès, B. Mennucci, and J. Tomasi, A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics, *J. Chem. Phys.*, 107 (1997), 3032-3041.
- [90] J. Tomasi, B. Mennucci, and E. Cancès, The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM ab initio level, *J. Mol. Struct. (Theochem)*, 464 (1999), 211-226.
- [91] A. Klamt and V. Jonas, Treatment of the outlying charge in continuum solvation models, *J. Chem. Phys.*, 105 (1996), 9972-9981.
- [92] V. Barone and M. Cossi, Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model, *J. Phys. Chem. A*, 102 (1998), 1995-2001.
- [93] C.-G. Zhan, J. Bentley, and D. M. Chipman, Volume polarization in reaction field theory, *J. Chem. Phys.* 108 (1998), 177-192.
- [94] C.-G. Zhan and D. M. Chipman, Cavity size in reaction field theory, *J. Chem. Phys.*, 109 (1998), 10543.
- [95] C.-G. Zhan and D. M. Chipman, Reaction field effects on nitrogen shielding, *J. Chem. Phys.*,

- 110 (1999), 1611-1622.
- [96] C.-G. Zhan, D. W. Landry, and R. L. Ornstein, Energy barriers for alkaline hydrolysis of carboxylic acid esters in aqueous solution by reaction field calculations, *J. Phys. Chem. A*, 104 (2000), 7672-7678.
- [97] C.-G. Zhan and D. W. Landry, Theoretical studies of competing reaction pathways and energy barriers for alkaline ester hydrolysis of cocaine, *J. Phys. Chem. A*, 105 (2001), 1296-1301.
- [98] C.-G. Zhan, S. Niu, and R. L. Ornstein, Theoretical studies of fundamental reaction pathways for the three stages of carboxylation of ribulose-1,5-bisphosphate, *J. Chem. Soc. Perkin Trans.*, 2 (2001), 23-29.
- [99] D. A. Dixon, D. Feller, C.-G. Zhan, and J. S. Francisco, Decomposition pathways of peroxynitrous acid: Gas-phase and solution energetics, *J. Phys. Chem. A*, 106 (2002), 3191-3196.
- [100] C.-G. Zhan, D. A. Dixon, M. I. Sabri, M.-S. Kim, and P. S. Spencer, Theoretical determination of chromophores in the chromogenic effects of neurotoxicants, *J. Am. Chem. Soc.*, 124 (2002), 2744-2752.
- [101] D. A. Dixon, D. Feller, C.-G. Zhan, and S. F. Francisco, The gas and solution phase acidities of HNO, HOONO, HONO, and HONO₂, *Int. J. Mass Spectrom*, 227 (2003), 421-438.
- [102] C.-G. Zhan, D. A. Dixon, and P. S. Spencer, Computational insights into the chemical structures and mechanisms of the chromogenic and neurotoxic effects of aromatic γ -diketones, *J. Phys. Chem. B*, 107 (2003), 2853-2861.
- [103] X. Chen and C.-G. Zhan, Fundamental reaction pathways and free energy barriers for ester hydrolysis of intracellular second messenger 3', 5'-cyclic nucleotide, *J. Phys. Chem. A*, 108 (2004), 3789-3797.
- [104] X. Chen and C.-G. Zhan, Theoretical determination of activation free energies for alkaline hydrolysis of cyclic and acyclic phosphodiester in aqueous solution, *J. Phys. Chem. A*, 108 (2004), 6407-6413.
- [105] C. G. Zhan, P. S. Spencer, and D. A. Dixon, Chromogenic and neurotoxic effects of aliphatic γ -diketone: Computational insights into the molecular structures and mechanism, *J. Phys. Chem. B*, 108 (2004), 6098-6104.
- [106] C.-G. Zhan, S.-X. Deng, J. G. Skiba, B. A. Hayes, S. M. Tschampel, and G. C. Shields, and D. W. Landry, First-principle studies of intermolecular and intramolecular catalysis of protonated cocaine, *J. Comput. Chem.*, 26 (2005), 980-986.
- [107] H.-T. Lu, X. Chen, and C.-G. Zhan, First-principles calculation of pKa for cocaine, nicotine, neurotransmitters, and anilines in aqueous solution, *J. Phys. Chem. B*, 111 (2007), 10599-10605.
- [108] X. Chen and C.-G. Zhan, First-principles determination of molecular conformations of cyclic adenosine 3',5'-monophosphate in gas phase and aqueous solution, *J. Phys. Chem. B*, 112 (2008), 16851-16859.
- [109] F. Zheng, L. P. Dwoskin, P. A. Crooks, and C.-G. Zhan, First-principles determination of molecular conformations of indolizidine (-)-235B' in solution, *Theo. Chem. Acc.*, 124 (2009), 269-278.
- [110] D. M. Chipman, Simulation of volume polarization in reaction field theory, *J. Chem. Phys.* 110 (1999), 8012-8018.
- [111] D. M. Chipman, Reaction field treatment of charge penetration, *J. Chem. Phys.*, 112 (2000), 5558-5565.
- [112] D. M. Chipman, Comparison of solvent reaction field representations, *Theo. Chem. Acta*, 107 (2002), 80-89.
- [113] D. M. Chipman, Simulation of volume polarization for the influence of solvation on chem-

- ical shielding, *Theo. Chem. Acta*, 111 (2004), 61-65.
- [114] Y. Xiong and C.-G. Zhan, Reaction pathways and free energy barriers for alkaline hydrolysis of insecticide 2-trimethylammonioethyl methylphosphonofluoridate and related organophosphorus compounds: Electrostatic and steric effects, *J. Org. Chem.*, 69 (2004), 8451-8458.
- [115] M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, General atomic and molecular electronic structure system, *J. Comput. Chem.*, 14 (1993), 1347-1363.
- [116] M. J. Vilkas and C.-G. Zhan, An efficient implementation for determining volume polarization in self-consistent reaction field theory, *J. Chem. Phys.*, 129 (2008), 194109.
- [117] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.
- [118] C. J. Cramer and D. G. Truhlar, A universal approach to solvation modeling, *Acc. Chem. Res.*, 41 (2008), 760-768.
- [119] J. Liu, C. P. Kelly, A. C. Goren, A. V. Marenich, C. J. Cramer, D. G. Truhlar, and C.-G. Zhan, Free energies of solvation with surface, volume, and local electrostatic effects and atomic surface tensions to represent the first solvation shell, *J. Chem. Theory Comput.*, 6 (2010), 1109-1117.
- [120] C. P. Kelly, C. J. Cramer, and D. G. Truhlar, SM6: A density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters, *J. Chem. Theory Comput.*, 1 (2005), 1133-1152.
- [121] D. Li, X. Huang, K. Han, and C.-G. Zhan, Catalytic mechanism of cytochrome P450 for 5'-hydroxylation of nicotine: Fundamental reaction pathways and stereoselectivity, *J. Am. Chem. Soc.* 133 (2011), 7416-7427.
- [122] J. Wu, Q. Liu, K. Yu, J. Hu, Y. P. Kuo, M. Sequerberg, P. A. St. John, and R. J. Lukas, Roles of nicotinic acetylcholine receptor beta subunits in function of human alpha4-containing nicotinic receptors, *J. Physiol.*, 576 (2006), 103-118.
- [123] A. Karlin and G. G. Wilson, Acetylcholine receptor channel structure in the resting, open, and desensitized states probed with the substituted-cysteine-accessibility method, *Proc. Natl. Acad. Sci. U.S.A.*, 98 (2001), 1241-1248.
- [124] X. Cheng, H. Wang, B. Grant, S. M. Sine, and J. A. McCammon, Targeted molecular dynamics study of C-loop closure and channel gating in nicotinic receptors, *PLoS Comput. Biol.*, 2 (2006), 1173-1184.
- [125] X. Huang, F. Zheng, and C.-G. Zhan, Modeling differential binding of $\alpha 4\beta 2$ nicotinic acetylcholine receptor with agonists and antagonists, *J. Am. Chem. Soc.* 130 (2008), 16691-16696.

- [126] Y. Pan, D. Gao, and C.-G. Zhan, Modeling the catalysis of anti-cocaine catalytic antibody: Competing reaction pathways and free energy barriers, *J. Am. Chem. Soc.*, 130 (2008), 5140-5149.
- [127] A. Hamza, X. Zhao, M. Tong, H.-H. Tai, and C.-G. Zhan, Novel human mPGES-1 inhibitors identified through structure-based virtual screening, *Bioorg. Med. Chem.*, 19 (2011), 6077-6086.
- [128] C.-G. Zhan, F. Zheng, and D. W. Landry, Fundamental reaction mechanism for cocaine hydrolysis in human butyrylcholinesterase, *J. Am. Chem. Soc.*, 125 (2003), 2462-2474.
- [129] A. Hamza, H. Cho, H.-H. Tai, and C.-G. Zhan, Molecular dynamics simulation of cocaine binding with human butyrylcholinesterase and its mutants, *J. Phys. Chem. B*, 109(2005), 4776-4782.
- [130] D. Gao and C.-G. Zhan, Modeling effects of oxyanion hole on the ester hydrolysis catalyzed by human cholinesterases, *J. Phys. Chem. B*, 109 (2005), 23070-23076.
- [131] C.-G. Zhan and D. Gao, Catalytic mechanism and energy barriers for butyrylcholinesterase-catalyzed hydrolysis of cocaine, *Biophys. J.*, 89 (2005), 3863-3872.
- [132] D. Gao and C.-G. Zhan, Modeling evolution of hydrogen bonding and stabilization of transition states in the process of cocaine hydrolysis catalyzed by human butyrylcholinesterase, *Proteins*, 62 (2006), 99-110.
- [133] Y. Pan, D. Gao, W. Yang, H. Cho, G.-F. Yang, H.-H. Tai, and C.-G. Zhan, Computational redesign of human butyrylcholinesterase for anticocaine medication, *Proc. Natl. Acad. Sci. USA*, 102 (2005), 16656-16661.
- [134] D. Gao, H. Cho, W. Yang, Y. Pan, G.-F. Yang, H.-H. Tai, and C.-G. Zhan, Computational design of a human butyrylcholinesterase mutant for accelerating cocaine hydrolysis based on the transition-state simulation, *Angew. Chem. Int. Ed.*, 45 (2006), 653-657.
- [135] Y. Pan, D. Gao, W. Yang, H. Cho, and C. G. Zhan, Free energy perturbation (FEP) simulation on the transition states of cocaine hydrolysis catalyzed by human butyrylcholinesterase and its mutants, *J. Am. Chem. Soc.*, 129 (2007), 13537-13543.
- [136] F. Zheng, W. Yang, M.-C. Ko, J. Liu, H. Cho, D. Gao, M. Tong, H.-H. Tai, J. H. Woods, and C.-G. Zhan, Most efficient cocaine hydrolase designed by virtual screening of transition states, *J. Am. Chem. Soc.*, 130 (2008), 12148-12155.
- [137] W. Yang, Y. Pan, L. Fang, D. Gao, F. Zheng, and C.-G. Zhan, Free energy perturbation simulation on transition states and high-activity mutants of human butyrylcholinesterase for (-)-cocaine hydrolysis, *J. Phys. Chem. B*, 114 (2010), 10889-10896.
- [138] W. Yang, Y. Pan, F. Zheng, H. Cho, H. H. Tai, and C.-G. Zhan, Free-energy perturbation simulation on transition states and redesign of butyrylcholinesterase, *Biophys. J.*, 96 (2009), 1931-1938.
- [139] F. Zheng and C.-G. Zhan, Structure-and-mechanism-based design and discovery of therapeutics for cocaine overdose, *Org. Biomol. Chem.*, 6 (2008), 836-843.
- [140] F. Zheng and C.-G. Zhan, Rational design of an enzyme mutant for anti-cocaine therapeutics, *J. Computer-Aided Mol. Design*, 22 (2008), 661-671.
- [141] W. Yang, L. Xue, L. Fang, and C.-G. Zhan, Characterization of a high-activity mutant of human butyrylcholinesterase against (-)-cocaine, *Chemico-Biological Interactions*, 187 (2010), 148-152.
- [142] F. Zheng, W. Yang, L. Xue, S. Hou, J. Liu, and C.-G. Zhan, Design of high-activity mutants of human butyrylcholinesterase against (-)-cocaine: Structural and energetic factors affecting the catalytic efficiency, *Biochemistry*, 49 (2010), 9113-9119.
- [143] F. Zheng and C.-G. Zhan, Recent progress in protein drug design and discovery with a focus

- on novel approaches to the development of anti-cocaine medications, *Future Med. Chem.* 1 (2009), 515-528.
- [144] F. Zheng and C.-G. Zhan, Enzyme therapy approaches for treatment of drug overdose and addiction, *Future Med. Chem.*, 3 (2011), 9-13.
- [145] L. Xue, M.-C. Ko, M. Tong, W. Yang, S. Hou, F. Zheng, J. H. Woods, H. H. Tai, and C.-G. Zhan, Rational design, preparation and characterization of high-activity mutants of human butyrylcholinesterase against (-)-cocaine, *Mol. Pharmacol.*, 79 (2011), 290-297.
- [146] Teva Pharmaceutical Industries Ltd, BChE Albumin Fusions for the Treatment of Cocaine Abuse, PCT WO 2011/071926 A1 (International Publication Date: June 16, 2011).
- [147] D. Gao, D. L. Narasimhan, J. Macdonald, M.-C. Ko, D. W. Landry, J. H. Woods, R. K. Sunahara, and C.-G. Zhan, Thermostable variants of cocaine esterase for long-time protection against cocaine toxicity, *Mol. Pharmacol.*, 75 (2009), 318-323.
- [148] G. T. Collins, R. L. Brim, D. Narasimhan, M.-C. Ko, R. K. Sunahara, C.-G. Zhan, and J. H. Woods, Cocaine esterase prevents cocaine-induced toxicity and the ongoing intravenous self-administration of cocaine in rats, *J. Pharm. Exp. Ther.* 331 (2009), 445-455.
- [149] R. L. Brim, M. R. Nance, D. W. Youngstrom, D. Narasimhan, C.-G. Zhan, J. J. G. Tesmer, R. K. Sunahara, and J. H. Woods, A thermally stable form of bacterial cocaine esterase: A potential therapeutic agent for treatment of cocaine abuse, *Mol. Pharmacol.*, 77 (2010), 593-600.
- [150] D. Narasimhan, M. R. Nance, D. Gao, M.-C. Ko, J. Macdonald, P. Tamburi, D. Yoon, D. M. Landry, J. H. Woods, C.-G. Zhan, J. J. G. Tesmer, and R. K. Sunahara, Structural analysis of thermostabilizing mutations of cocaine esterase, *Protein Eng. Des. Sel.* 23 (2010), 537-547.
- [151] C.-G. Zhan, Novel pharmacological approaches to treatment of drug overdose and addiction, *Expert Rev. Clinical Pharmacology*, 2 (2009), 1-4.