**­­­­Development and Test of Molecular Mechanics Force Fields for Biomedical Research**

Molecular simulations have played essential roles in biochemical and biophysical sciences. A chief application of molecular simulations is to elucidate the molecular interactions between small molecule inhibitors and protein or nucleic acid targets, and then enhance or eradicate the functions of their targets via rationally design of high-potency agonists or antagonists. **However, nowadays it is still very challenging to accurately predict the free energies of such interactions and processes.** A key to the success of molecular simulation studies and structure-based rational drug designs is the quality of molecular mechanics force fields (**MMFFs**), and a major impediment to developing more accurate MMFF is to test them objectively in experiments. **The major goal of this project is to further develop the existing and develop a new generation of the general AMBER force fields (GAFF) for studying biomolecule-ligand interactions.** The new general-purpose MMFF (referred to here and thereafter as **GAFF3**) is based on a new charge model which is efficient, largely conformation-independent, and has high accuracy on solvation free energy calculations. In a long run, **a spectrum of high-quality GAFF3 compatible MMFFs will be developed for all types of biological macromolecules for biomedical research.** The new set of MMFFS will be highly transferable and self-consistent for studying protein-ligand and nucleic acid-ligand interactions, and will enable more accurate calculations of the protein-ligand and nucleic acid-ligand binding free energies and facilitate computer-aided drug discovery. The developed GAFF force fields will be rigorously scrutinized and critically assessed through direct comparisons with experiments in an extensive set of model systems. The large-scale FF assessment will be accelerated by using GPUs in both molecular dynamics simulations and alchemical free energy calculations.

**A. BACKGROUND AND SIGNIFICANCE**

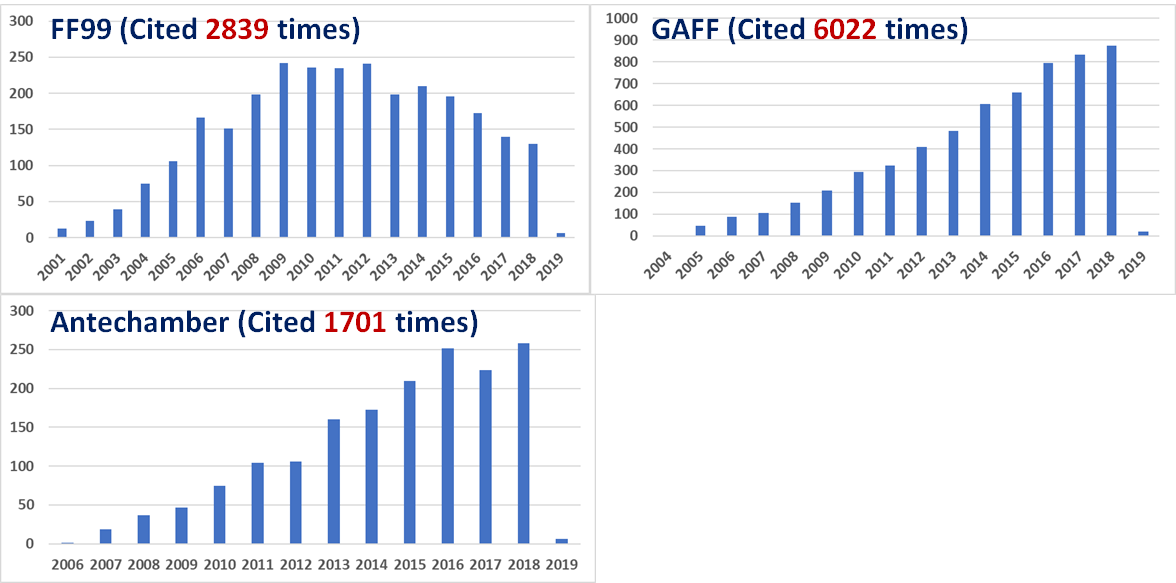
**A.1. Molecular Simulations and Force fields.** Atomistic simulations have become an important tool for studying the structure, dynamics, interactions, and functions of biomolecular systems.[1-8](#_ENREF_1) Critical to their success is the accurate representations of potential energy (referred to as the “force field”). Advancements in computer technology and simulation methodology have enabled molecular dynamics (**MD**) simulations of increasingly complex biomolecular systems over longer times and with larger molecules. With these advances, the need for accurate and efficient force fields that can represent diverse biomolecular systems has become much more critical.

A force field (**FF**) is composed of an energy function and its associated parameters. Because of its simplicity and robustness, the effective functional form (**Eq. 1**) has been extensively used in many biomolecular FFs, which include AMBER FF94,[9](#_ENREF_9) FF99,[10](#_ENREF_10) FF99SB,[11](#_ENREF_11) FF12SB,[12](#_ENREF_12) FF14SB,[13](#_ENREF_13) CHARMM,[3](#_ENREF_3) NAMD,[14](#_ENREF_14) OPLS,[8](#_ENREF_8), [15](#_ENREF_15) and GROMACS,[16](#_ENREF_16) etc.

Where *r*, *θ*, and *Φ* are the instant values of bond lengths, angles, and dihedral angles, *req* and *θeq* are the equilibrium values of bond lengths and angles, *n* and *γ* are the dihedral multiplicities and phases, *Kr*, *Kθ* and *Vn* are the respective force constants, *Rij* is the distance between atom *i* and atom *j*, *qi* and *qj* are their respective atomic charges, and *Aij* and *Bij* are van der Waals parameters.

**A.2. AMBER Force Fields.** “AMBER force fields” have been the primary force fields in the AMBER simulation packages and adapted to run directly in CHARMM and NAMD. Since the first description by Weiner and Kollman in 1981,[17](#_ENREF_17) AMBER-associated force fields have evolved from united-atom[18](#_ENREF_18) to all-atom models,[9-10](#_ENREF_9), [17](#_ENREF_17), [19-20](#_ENREF_19) from additive to polarizable,[21-25](#_ENREF_21) and from special (such as proteins[9](#_ENREF_9) and nucleic acids,[10](#_ENREF_10) lipids[26](#_ENREF_26) and carbohydrates[27](#_ENREF_27)) to general organic molecules.[28](#_ENREF_28) AMBER force fields have been widely used in modeling biomolecules with more than 20,000 citations as of this writing. The principal investigator (**PI**) is the main contributor of the popular AMBER force field family (including AMBER FF99[10](#_ENREF_10)), which is for simulations of biological systems including proteins, DNAs & RNAs, etc.

**A.3. Qualification of the Research Team.** The PI is also the inventor of the popular general AMBER force field (GAFF)[28](#_ENREF_28) for general organic drug-like molecules, and the main developer of the corresponding tools for FF parameter assignment including Antechamber.[29](#_ENREF_29) AMBER FF99, GAFF and Antechamber have been widely used by researchers all over the world in numerous research studies with more than 10,500 citations in total (**Fig. 1**). Based on these solid foundations, we are going to develop the new set of MMFFs with similar but updated strategies and procedures which were used in the development of AMBER FF99 and GAFF.



**Figure 1. Annual citations of FF99, GAFF and Antechamber.** (webofknowledge.com)

**A.4. Development of a New, Accurate, Highly Transferable, and Self-Consistent Force fields for Studying Protein-Ligand and Nucleic Acid-Ligand Interactions.** It is a challenging and ambitious goal to develop such kinds of force fields given the extremely difficult nature of molecular mechanics force field development. MMFF development has long been more of an “art” than a “science”.[30-32](#_ENREF_30) Atomic charge model, the cornerstone of a MMFF, are developed both empirically and physically (such as the RESP[33-34](#_ENREF_33) charge of AMBER FFs). Ideally, a charge model for additive force fields would possess all the following features: efficient, conformation-independent, highly accurate in solvation free energy calculations, and can be applied to both small organic and macromolecules. Unfortunately, no charge method in the current FFs has all of these features. In this proposal, we propose to develop a new, efficient, largely conformation-independent and high-quality atomic charge model for our new AMBER force field development.

Accurate solvation free energy calculation plays an essential role for us to achieve the above research goal. Solvation free energy (**ΔGsolv**) calculation is essentially similar to that of receptor-ligand binding free energy (**ΔGbind**) calculation, as both deal with the free energy of transferring a molecule from one environment to another environment (from gas phase to solvent for ΔGsolv, and from solvent (usually water) to the inside of the receptor for ΔGbind). The simplest, most ubiquitous and most important solvent is water. The calculation of solvation free energy in water (hydration free energy, **ΔGhyd**) is still much more computationally demanding than that of other physicochemical properties such as density and heat of vaporization. Therefore, usually ΔGsolv (including ΔGhyd) is not used as a target during FF parameterization, but just as an auxiliary validation test. In such situation, the standard of justifying ΔGhyd is also not as strict as that of justifying density and heat of vaporization. Usually an unsigned error of 1.0 kcal/mol is considered as good, and 2.0 kcal/mol is still acceptable. The general difference between the calculated ΔGhyd with current popular force fields is > 1.2 kcal/mol or even larger,[35](#_ENREF_35) limiting the accuracy of predicting ΔGbind even with the state-of-the-art alchemical free energy calculation methods, such as thermodynamic integration (**TI**)[36](#_ENREF_36) and free energy perturbation (FEP),[37](#_ENREF_37) which are rigorous but very expensive. Loose criteria and poor quality of ΔGsolv prediction will ultimately impair the accurate prediction of ΔGbind no matter how much efforts are put on sampling protein and ligand conformations.

With the growing computer power, facilitated by both inexpensive GPU-based computer clusters[38-39](#_ENREF_38) and access to the PITT CRC and other national supercomputing resources, particularly the GPU-TI[40](#_ENREF_40) code recently released in AMBER 18,[41](#_ENREF_41) it is timely for us to introduce TI solvation free energies into our FF parameterization scheme. We set off to develop a new and high-quality atomic charge model, and based on that a spectrum of coherent FFs will be developed for modeling both general organic and biological molecules.

**A.5. Critical Assessment of GAFFs in Modeling Protein-Ligand and Nucleic Acid-Ligand Interactions.** Assessing a newly developed force field in realistic biological applications is a challenging task. Improving the accuracy of computational predictions is of critical importance for the drug design process. Binding constants depend exponentially on the binding free energy, thus relatively small errors (0.6 kcal/mol) can have a significant impact on the efficacy of a potential drug. The main factors that determine the accuracy of a free energy simulation are the extent to which important configurational degrees of freedom are sampled, and the accuracy and reliability of the underlying force field (FF) that models the molecular interactions. The force fields include large sets of parameters that were derived from quantum-chemical calculations and select experimental data such as density, heat of vaporization of neat liquids, or hydration free energies. However, to the best of our knowledge, no currently available force field has been tuned to consider quantitative binding free energy data in its parametrization scheme, due to the associated computational burden. Given that binding free energies represent one of the most important applications in computational chemistry, it is worthwhile to develop new methods that would enable it to enter into the force field parameterization/refinement process.

Very recently a set of tools based on GPU-accelerated free energy methods have been developed, i.e., the 𝞴 -SAMS approaches, which speedup the convergence of alchemical transformations by roughly three to five folds. Given that the speedup of the AMBE GPU TI code over the CPU code is over two orders of magnitudes, now it is possible to assess the performance of new force-field in free energy calculations on GPU roughly three orders of magnitudes faster than on CPU.

We will utilize the new GPU-accelerated free energy sampling (𝞴 -SAMS) to critically evaluate the proposed force fields: 1. Establishing extremely fast free energy simulation pipeline infrastructure; 2. Performing simulations on a current golden-standard of protein-ligand free energy simulation data set; 3. Producing meaningful error bars through multiple independent simulations.

**B. INNOVATION**

The novelty of this R01 project is highlighted from the following three aspects:

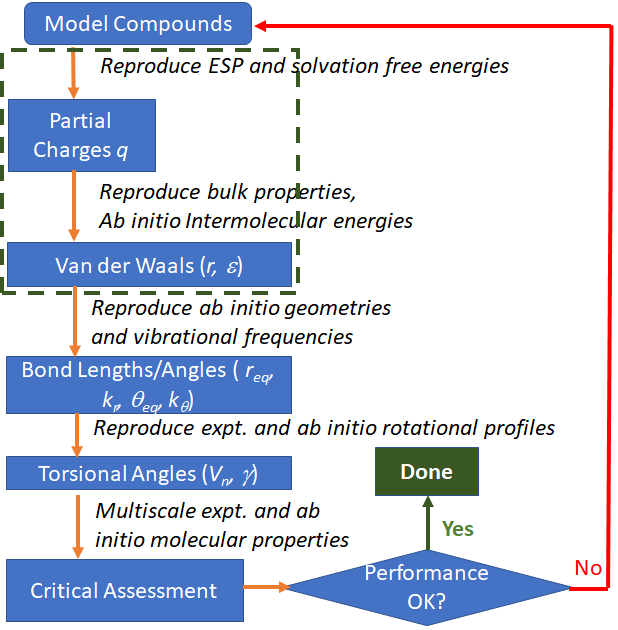
* FF development philosophe
* A novel charge method will be developed to largely overcome the limitations of the charge models in current mainstream FFs. This charge model is efficient, much less conformation-dependent, and can be applied to arbitrary small molecules and biomolecules. MMFFs based on this charge method will be coherent, self-consistent and have much better transferability. Those MMFFs will be able to much better model the protein-ligand and nucleic acid-ligand interactions, and as such, they can boost the successful rate of structure-based rational drug design.
* Solvation free energy will be applied as the primary target for the development of a charge model. As aforementioned, solvation free energy (including hydration free energy) has never been used as the primary target for force field development. This is partly due to the much higher computational cost of calculating solvation free energies comparing to that of calculating other physicochemical properties. This limitation can be alleviated due to the increasing computing power and computer resources nowadays. Furthermore, the high cost at the beginning stages like force field development will be rewarded with more accurate results at the final goal stages of applications on the biological systems, such as the calculation of protein-ligand binding free energies for the purpose of drug design.
* Advances on TI calculations and large-scale FF evaluations accelerated by GPU-Tis. The GPU-accelerated 𝞴-SAMS approach is novel in that it uses a statistically optimal estimation of the free energy difference between alchemical states through Rao-Blackwellization of the joint distribution of conformational and alchemical spaces. Here, we leverage the 𝞴 -SAMS approach to run the Gibbs sampler along a fine-grid in 𝞴 space to produce a statistically optimal distribution and collect dU/d 𝞴 statistics for thermodynamic integration.

Applications of the newly developed 𝞴 -SAMS approaches represent a major innovative advance in the field that would reach unprecedent milestones for various biological simulations and are perfectly suitable for force field development where large scale of free energy simulations are needed.

**C. APPROACH**

**C.1 General Force Field Parameterization Strategy.**

The parameterization for an FF is an iterative procedure as shown in **Fig. 2**. The targets of FF parameterization include the quantum mechanics (**QM**) calculations of geometries, dipole moments, vibrational frequencies, conformational and intermolecular energies, etc., and the experimental data of physicochemical properties, such as crystal geometries, crystal volumes, liquid densities, liquid heats of vaporization. The atomic partial charge assignment is usually the first step of FF parameterization, because the atomic charges are the main contributor to the intra- and inter-molecular energies, and other bulk properties including the solvation free energies.



**Figure 2. Strategies of force field development.** The FF parameterization of non-bonded terms, which is the focus of this R01 grant, are shown in dashed-rectangle.

**C.2 Aim 1. Continue Development of the Second Generation of GAFF Force Field (GAFF2)**

**Rational.** It is necessary to develop a general force field that is compatible to the mainstream AMBER biomolecular force fields in biomedical research for studying biomolecule-ligand interactions.

**Preliminary Result.**

To be added

**Experimental Design.**

To be added

**C.3 Aim 2. Development of a New, Accurate, Highly Transferable, and Self-Consistent Force fields for Studying Protein-Ligand and Nucleic Acid-Ligand Interactions**

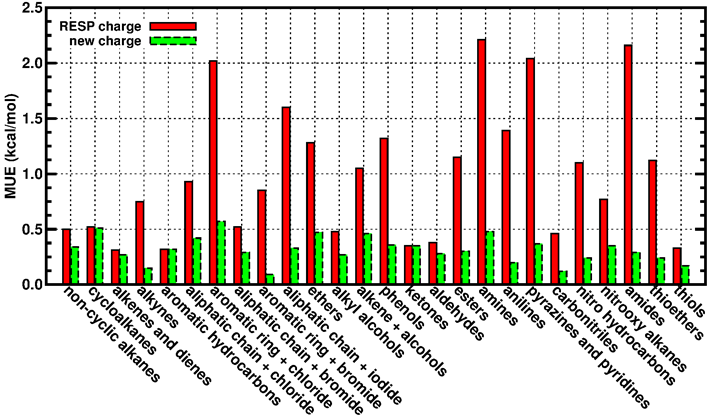
**Rational.** Considering the essential relationship between the solvation free energy ΔGsolv (including hydration free energy ΔGhyd) and the receptor-ligand binding free energy ΔGbind, the general accuracy of ΔGsolv calculation determines the best results that can be expected for ΔGbind calculation. Hence, a new set of force fields (**FFs**) with high accuracy of ΔGsolv prediction will enable us to improve the accuracy of ΔGbind prediction, and the latter is extremely important for computer-aided drug design.

**Background.** Even though all FF parameters in an MMFF (**Eq. 1**) are developed in an iterative manner, the model of assigning atomic partial charges is always the first step due to its major role in determining the calculated intra- and inter-molecular energies. Traditionally, the AMBER force fields and GAFF assign atomic charges utilizing the RESP charge model, which fits the electrostatic potential (ESP) of a molecule calculated by QM at the HF/6-31G\* level of theory.[33-34](#_ENREF_33) While RESP charge model produces suitable charges for condensed simulations, there are inherent problems, including but not limited to the high cost of QM calculations for every new molecule encountered in real projects. As an alternative solution, GAFF also adopts a faster and more efficient method named AM1-BCC[42-43](#_ENREF_42) to emulate the HF/6-31G\* ESP. The basic idea of AM1-BCC is as follows: the AM1 Mulliken charges, although incapable of satisfactorily reproducing the QM ESP, can capture underlying electronic features such as formal charge and electron delocalization; simple additive bond charge corrections (**BCCs**) are then applied based on the general atom and bond type in the molecule so that the resulting charges can emulate the RESP charges more closely than AM1.

While the AM1-BCC method has been widely and successfully applied along with AMBER force fields and GAFF, we’ve found that the calculated ΔGsolv for various organic solutes have a mean unsigned error (**MUE**) > 1 kcal/mol. Such a limitation is actually inherited from RESP charges, which cannot accurately reproduce ΔGsolv for some functional groups of organic molecules (**Table 1**).

Hence, we propose to re-adjust the BCC values for various atom types and bond types of organic molecules to directly target accurate reproduction of experimental data, which will lead to more accurate ΔGsolv and ΔGbind based on aforementioned discussion. The van der Waals and bonded terms parameters will be validated and if necessary, re-adjusted in accordance with newly developed BCC parameters (and corresponding new atomic charges).

**Preliminary Results**



**Figure 3. The mean unsigned errors (MUE) of calculated hydration free energies** for various types of organic solutes using atomic charges from the RESP model and our new AM1-BCC model.

New Charge Model Development. We have calculated the solvation free energies of series of organic molecules in the solvent water (ΔGhyd), using the RESP charges, current (old) AM1-BCC charges and our newly developed AM1-BCC charges. The explicit water molecules are treated with the TIP3P water model which is bounded to AMBER force fields and GAFF. The calculation method is the Thermodynamic Integration (**TI**) as described in previous literature.[40](#_ENREF_40) As shown in **Table 1**, a lot of calculations have a mean signed error (**MSE**) with the absolute value very close or equal to MUE, which demonstrates a systematic bias, i.e., the calculated ΔGhyd results are systematically more positive or more negative than the experimental data. For example, the calculated ΔGhyd of 17 phenols are systematically more positive than experiment by 1.32 kcal/mol from RESP charges, and systematically more positive by 0.95 kcal/mol from old AM1-BCC charges. With our newly adjusted AM1-BCC model, the systematic error is greatly alleviated, and the MSE is only -0.16 kcal/mol, and MUE is reduced to only 0.36 kcal/mol. Among the 26 types of molecules in **Table 1**, 17 types show systematic errors for calculated ΔGhyd using RESP charges, and 11 of them have a systematic error greater than 1.0 kcal/mol in magnitude. With our new AM1-BCC model, all MSEs are close to 0 kcal/mol and most MUEs are less than 0.4 kcal/mol (**Table 1 and Fig. 3**). In a short summary, the accuracy of calculated ΔGhyd can be significantly improved by adjusting the BCC parameters.

**Experimental Design.**

New Charge Model Development. We plan to adjust the BCC parameters of the new charge model against ΔGsolv and partition coefficients in various solvents. We are going to calculate the ΔGsolv of various molecules in organic solvents, such as octanol, cyclohexane, benzene, etc.[44-45](#_ENREF_44) These solvents are important medium for studying chemical and biological processes, and mimicking the hydrophobic or amphiphilic environments inside of or on the surfaces of biological systems, such as membranes, proteins, DNAs and RNAs. The difference of ΔGsolv of a molecule in two immiscible solvents can be used to calculate an important property – partition coefficient, which is a measure of a solute’s hydrophobicity and a proxy for its membrane permeability in drug design. Especially, partition coefficients are relatively easy to measure experimentally compared to ΔGsolv, and more experimental data of partition coefficients are available. [44-45](#_ENREF_44) The newly developed charge model will be tested and adjusted against ΔGsolv in or partition coefficients between various solvents. After the new charge model is developed and evaluated, the van der Waals parameters of different atom types will be adjusted and validated in Aim 2.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1.** **The mean signed errors (MSE) and mean unsigned errors (MUE) of calculated hydration free energies** (in kcal/mol) of various molecules using atomic charges from RESP, old AM1-BCC and new AM1-BCC methods. ‘\*’: related BCC values are not changed; ‘/’: MSE & MUE are not calculated yet for this type of molecules. | | | | | | | |
| **Molecule Types** | **Mol #** | **RESP charges** | | **Old AM1-BCC charges** | | **New AM1-BCC charges** | |
| **MSE** | **MUE** | **MSE** | **MUE** | **MSE** | **MUE** |
| Non-cyclic alkanes \* | 27 | -0.45 | 0.50 | -0.27 | 0.34 | -0.27 | 0.34 |
| Cycloalkanes \* | 9 | -0.36 | 0.52 | -0.29 | 0.51 | -0.29 | 0.51 |
| Alkenes & dienes | 22 | 0.11 | 0.31 | 0.69 | 0.66 | 0.11 | 0.27 |
| Alkynes | 6 | -0.75 | 0.75 | 0.59 | 0.59 | 0.01 | 0.15 |
| Aromatic hydrocarbons | 38 | -0.08 | 0.32 | -0.80 | 0.81 | -0.18 | 0.32 |
| Aliphatic chain + chloride | 31 | 0.84 | 0.93 | / | / | 0.22 | 0.42 |
| Aromatic ring + chloride | 23 | 2.02 | 2.02 | / | / | 0.22 | 0.57 |
| Aliphatic chain + bromide | 13 | 0.35 | 0.52 | / | / | 0.01 | 0.29 |
| Aromatic ring + bromide | 3 | -0.56 | 0.85 | / | / | -0.09 | 0.09 |
| Aliphatic chain + iodide | 9 | 1.60 | 1.60 | / | / | -0.01 | 0.33 |
| ethers | 26 | 0.91 | 1.28 | 0.64 | 0.93 | 0.12 | 0.47 |
| Alkyl alcohols | 27 | 0.36 | 0.48 | 1.67 | 1.67 | 0.01 | 0.27 |
| Alkene + alcohols | 4 | 1.05 | 1.05 | / | / | 0.09 | 0.46 |
| Phenols | 17 | 1.32 | 1.32 | 0.95 | 0.95 | -0.16 | 0.36 |
| Ketones | 18 | -0.06 | 0.35 | / | / | -0.02 | 0.35 |
| Aldehydes | 13 | 0.13 | 0.38 | / | / | -0.10 | 0.28 |
| Esters | 37 | -1.11 | 1.15 | / | / | -0.14 | 0.30 |
| Amines | 25 | 2.21 | 2.21 | / | / | 0.16 | 0.48 |
| Anilines | 8 | 1.39 | 1.39 | / | / | -0.08 | 0.20 |
| Pyrazines & pyridines | 17 | 2.07 | 2.04 | / | / | 0.33 | 0.37 |
| Carbonitriles | 5 | -0.43 | 0.46 | / | / | 0.01 | 0.12 |
| Nitro hydrocarbons | 9 | -1.10 | 1.10 | / | / | -0.11 | 0.24 |
| Nitrooxy alkanes | 9 | -0.77 | 0.77 | / | / | -0.09 | 0.35 |
| Amides | 8 | 2.16 | 2.16 | / | / | 0.25 | 0.29 |
| Thioethers | 7 | 1.01 | 1.12 | / | / | -0.05 | 0.24 |
| Thiols | 4 | 0.14 | 0.33 | / | / | -0.15 | 0.17 |

van der Waals Parameterization.In a MMFF,van der Waals (**VDW**) interactions are strongly coupled to the employed charge method. We plan to validate, and adjust if necessary, the VDW parameters in the current GAFF. The target data include the interaction energies of dimers and the experimentally measured densities of liquids and heats of vaporization, and solvation free energies. We have finished calculating interaction energies at the MP2/aug-cc-pVTZ//MP2/6-311++G(d,p) level with counterpose corrections for 2954 dimers. We have collected a set of 241 diverse molecules that have measured densities and 308 molecules that have measured heats of vaporization. An efficient VDW parameterization scheme developed by us and has been successfully applied in the development of polarizable GAFF[23](#_ENREF_23) will be utilized to tune VDW parameters to reproduce both the *ab initio* interaction energies and the two pure liquid properties.

Bonded Term Parameterization for GAFF3.After finishing the development of the new charge model (**Aim 1**) and the VDW parameters (**Aim 2**), we will focus on the development of the bonded term parameters. We are going to test, and adjust if necessary, the current parameters of bonds, angle, and dihedrals in GAFF, to make sure they are consistent with the new charge model and the new VDW parameter set. The target data include experimentally measured (such as crystal data) or QM calculated geometries, dipole moments, conformational energies and rotational profiles of model compounds.

**C.4 Aim 3. Critical Evaluation of GAFF2/GAFF3 in Studying Biomolecule-Ligand Interactions**

**Rational:**

**Preliminary Result.**

**The efficiency of the** *λ-*SAMS algorithm: The proposed *λ-*SAMS approach wasapplied to the thrombin ligand 1d->7a mutation in water as proof of principle. Shown in Figure 5, the *λ-*SAMS, both with 100 *λ*-windows or 200 *λ-*windows reach the standard error of 0.01 kcal/mol in about 5 ns total simulation time, compared to 20 ns by the regular TI simulations and Hamiltonian (*λ-*window) replica exchange/TI simulations.

Figure XX. Preliminary results of *λ* -SAMS approach. The standard errors averaged 𝞓G from 16 independent simulations are shown for the thrombin ligand 1d->7a mutation in water. “No RepEx”: TI simulations without replica exchange; “RepEx at 100 fs”: TI simulations with the replica exchange (between *λ*-windows) with 100 fs trial frequency; “SAMS 100”: *λ* -SAMS with 100 *λ* -windows; “SAMS 200”: *λ* -SAMS with 200 *λ* -windows. TI calculations were performed with 12 *λ* -windows.

**Experimental Design**

* Testing infrastructure: we will first develop and test infrastructure that enables tuning of force field parameters for enhanced predictive capability using fast GPU-accelerated free energy methods based on the *λ-*SAMS algorithm. As the first step, free energy transformations evaluation will be performed using the current AMBER/GAFF and the AMBER/GAFF2 force fields. Results will be compared with benchmark results for both ligand force fields.
* Baseline: The second step will involve applying the developed framework and validated procedures to obtain the very first benchmark quality data and predictions (including the first meaningful statistical error estimates) of ligand binding potency against the current “gold standard” drug discovery dataset[90]. This dataset encompasses a broad range of target classes and a diverse set of 200 ligands. Toward this end, we will use the most commonly used publicly available force fields for ligand-protein binding in explicit solvent, including CHARMM/CGenFF, AMBER/GAFF and GAFF2, and OPLS-AA. Not only will the results of this work provide a valuable assessment of the quality of current publicly available force fields for ligand-protein binding, but it will create an extremely powerful dataset from which next-generation versions of these force fields for prospective drug discovery can evolve efficiently.

This most comprehensive data set of relative binding affinities assembled was introduced by Schrödinger Inc. in 2015.[90] It consists of 200 ligands and 10 different protein targets (Beta-secretase in Alzheimer’s disease, Tyrosine-protein phosphatase non-receptor type 1 in diabetes, thrombin in blood clotting, nonreceptor tyrosine-protein kinase in immunodeficiency, plus the cancer targets Cyclin-dependent kinase 2, Mitogen-activated protein kinase 8, induced myeloid leukemia cell differentiation protein, and P38 mitogenactivated protein kinase). This data set is the current “gold standard” for computational drug design.

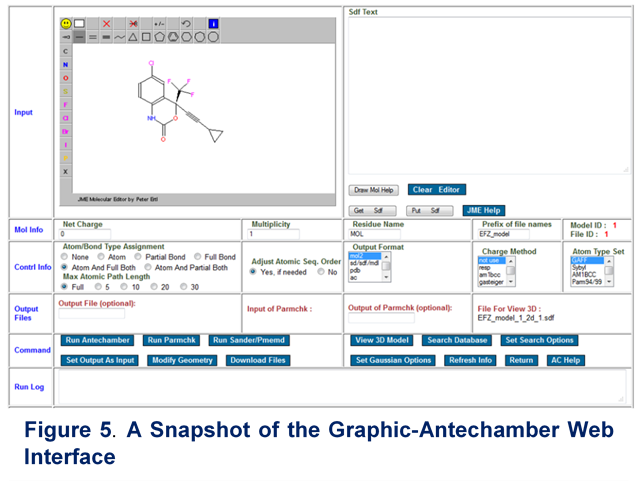
* Simulation Protocols:

Individual alchemical transformations for compound libraries that have a common protein target will be determined using a “thermodynamic map”. Following the same thermodynamic mapping procedure as Schrödinger Inc. which was obtained by the LOMAP algorithm[92] (which uses the maximum common substructure) and has been reported in the literature for the entire dataset[91].

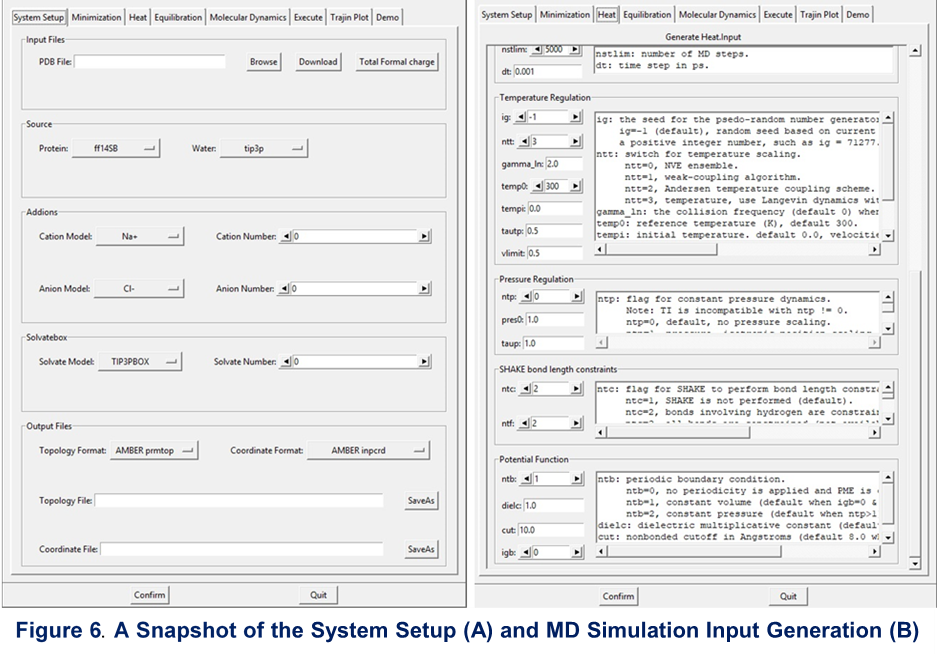
The GPU-accelerated 𝞴-SAMS approaches will be utilized to obtain statistical optimal estimation of relative free energies. Multiple independent sets (16 to 32) of simulations will be performed to establish meaningful error bars.

* Assessment of the to-be-developed force-field: We will perform binding free energy simulations on the data set mentioned above using the to-be-developed force-field. Since the baseline results developed have benchmark quality and have rigorous statistical error estimates based on multiple independent sets of simulations, we will be able to assess the quality of the to-be-developed force-field. Trajectory snapshots from alchemical free energy simulations for these transformations will be saved at 1 ps intervals creating an incredibly rich dataset that can be efficiently post-processed to obtain both TI and FEP/MBAR metadata for further free energy analysis.

Toolkit Development for TI. One common problem and task in real drug design projects is to pre-select which compounds to synthesize and further test with experimental assays which are expensive in terms of both time and cost. Physics-based alchemical free energy calculation methods such as TI and FEP are expected to be very useful in the pre-selecting (screening, lead identification and lead optimization) stages. However, several factors bring hurdles to the routine usages of free energy calculation methods in real drug design projects: (1) availability and accuracy of FF parameters of query compounds; (2) demanding computing resource; (3) difficulty of automatic system setup and posterior analysis. We are working on Aim 1 (continue development of GAFF2) and Aim 2 (development of GAFF3) in this proposal to solve the 1st problem, developing and improving GPU accelerated TI algorithms to alleviate the 2nd problem, and we also plan to develop a user-friendly GUI toolkit to overcome the 3rd problem to help non-expert users to set up the simulations and do analysis for various free energy calculations. According to our experience of talking to both academic researchers and industrial R&D personnel, such an automatic or semi-automatic GUI toolkit is in great demand.

We have developed a set of GUIs for some Antechamber programs using PHP (<http://www.php.net>) to automatically get atomic charges and GAFF parameters for arbitrary organic molecules. The antechamber program itself has two web interface pages, one takes input from a file like the standard alone program, the other takes the input from JME Molecular Editor as shown in **Figure 5** To enable the communication between those pages, PHP $\_SESSION will be used to store parameters, structures in SMILES strings, etc. Open Babel (www.openbabel.org) will be used to generate the canonical SMILES string for a molecule. It is emphasized that those web applications are still in the testing stage and need to improve dramatically both on their functionality and web security.

A set of GUIs have been developed with Python for generating inputs and system topologies for normal MD simulations using AMBER software package. The pages for molecular system setup and MD input generation for AMBER in shown **Figure 6.** We have also been working on a set of semi-automatic scripts and protocols to set up inputs and topologies for Ti simulations, which will be described in the Experimental Design below.



Endpoint Method. PBSA/GBSA methods to be added by Junmei

**Experimental Design.**

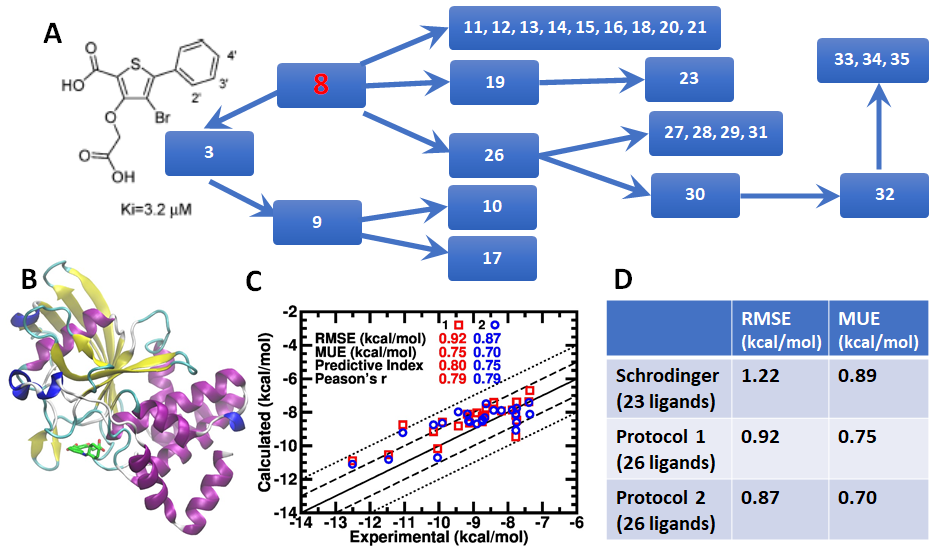
GPU-TI. To be added by Taisung

Toolkit Development for TI. TI calculations are often be used to calculate the relative binding free energy, where an atom or a group of atoms in ligand A is/are mutated to another atom or another group of atoms in ligand B. The designed GUI toolkit will have the following functions: (1) Open a graphic window and load the reference ligand A with 3D coordinates (sdf, mol, mol2, pdb, restart files from various packages). (2) Load the target ligand B in the format of SMILES/SMARTS string or 2D/3D coordinates, automatically generate a 3D structure if the input is not 3D, and show it on a second graphic window. (3) Automatically calculate the maximum common substructure (MCS) between ligands A and B, and map the MCS atoms. The algorithms behind the manipulations are based on the subroutines in the Antechamber program which can precisely describe an atom’s chemical environment (including the atomic paths, atomic connectivity and atomic property) and automatically define complicated force field atom types, bond types and discriminate subtle local chemical environments. Sometimes multiple mapping is possible, such the 2,6 positions or 3,5 positions on a benzene ring; in such cases, users are allowed to click several atoms on ligands A and B to reduce the mapping possibility to what they want. (4) Re-set the atom coordinates of ligand B automatically so that the positions of MCS atoms in ligand B are the same of corresponding ones in ligand A, and the non-MCS atoms in ligand B obtain reasonable new coordinates. (5) Save the new coordinates of ligand B, superpose ligand B to ligand A and show them in one graphic window to let the users have a quick visual check. (6) Write the topology file, parameter file, configuration file and control file which are needed for TI calculations for A→B mutation using the AMBER package. (7) Upon user’s request, convert all necessary files from AMBER style to the style of another simulation package.

We have been developing a series of in-house scripts and programs for setting up TI calculations in AMBER and we have explored protocols of selecting TI 𝛌 windows, mutation paths, integration methods, and GAFF2 parameters to get accurate results. We finished testing a set of systems with the Protein Tyrosine Phosphatase 1B (PTP1B) as the receptor (Figure. 7B) and a series of 26 compounds (Figure. 7A) as the ligands [x1]. The experimental values of Ki measurements range over 4 orders of magnitude (from several µM to less than 1 nM). We achieved lower root mean square error (RMSE) and mean of unsigned errors (MUE) than the reported results from the commercial FEP module in Schrodinger Suite [x2] (Figures. 7C and 7D). Next, we need to integrate the in-house scripts and programs into the GUI toolkit.

[X1]. Wilson DP, Wan ZK, Xu WX, Kirincich SJ, Follows BC, Joseph-McCarthy D, Foreman K, Moretto A, Wu J, Zhu M, Binnun E, Zhang YL, Tam May, Erbe DV, Tobin J, Xu X, Leung L, Shilling A, Tam SY, Mansour TS, Lee J (2007). *J. Med. Chem.* 50: 4681-4698.

[x2]. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren M, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, Murcko M, Frye L, Farid R, Lin T, Mobley DL, Jorgensen WL, Berne BJ, Friesner RA, Abel R (2015) Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. J Am Chem Soc 137:2695-2703.



**Figure 7**. **TI setup and results for PTP1B receptor with a series of 26 ligands.** (A) mutation paths of 26 ligands. (B) 3D structure of PTP1B with a ligand. (C) Calculated binding free energies VS experimental data. (D) RMSEs and MUEs from our calculations compared to Schrodinger FEP.

Endpoint Method. PBSA/GBSA methods: to be added by Junmei

Biological Systems. To evaluate the GAFFs in modeling protein-ligand and nucleic acid-ligand interactions, we have conducted a set of systematic studies to critically assess the MM-PB/GBSA (Molecular Mechanics Poisson-Boltzmann/Generalized Born Surface Area) method in absolute and relative binding free energy calculations.[46-49](#_ENREF_46) We are among the pioneers of applying MM-PB/GBSA to model protein-ligand complexes[50](#_ENREF_50) and to identify drug leads through virtual screenings.[50](#_ENREF_50) The new generation of GAFF and the associated biomolecular force fields will be evaluated by studying proteins and nucleic acids interacting with drugs. A set of well-studied protein targets which have measured *IC50* or *Ki* for several to tens of diverse small molecule ligands will be selected to test how well new GAFF performs in calculating the absolute and relative binding free energies using both the MM-PB/GBSA[46-48](#_ENREF_46), [51](#_ENREF_51) and thermodynamic integration methods.[52-53](#_ENREF_52) DNA minor groove binding drugs, such as DAPI,[54](#_ENREF_54) will also be studied.

**D. Long-Term Goals on New AMBER Force Field Development**

**New GAFF-Compatible Force Fields for Biomolecules.** Our long-term plan is to develop a set of new GAFF-based biomolecular force fields for proteins, nucleic acids, lipids and carbohydrates. The spectrum of specific force fields will be developed in the spirit of the top-down FF parameterization strategy, i.e. force field parameters are mostly inherited from the new-GAFF and special force field parameters are introduced for biological building blocks only when it is necessary. This development will ultimately lead to a set of highly transferrable, self-consistent, and coherent force fields for studying biomolecule-ligand interactions.

**E. TIMELINE AND MILESTONES**

We have proposed two aims to develop the nonbonded terms of new GAFF for this R01 project (**Aims 1-2**), and five aims to develop the new GAFF and new GAFF-compatible biomolecular force fields (**Aims 3-7**). The roadmap and milestones are summarized below:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Objective** | **Force Field Development Aims** | | **Y1** | | **Y2** | | **Y3** | | **Y4** | | **Y5** | |
| **1** | Aim 1a: Additional Torsional Angle Parameterization | |  |  |  | |  | |  | |  | |
| ***2*** | Aim 1b: Extension to all nonmetallic elements | |  | |  |  |  | |  | |  | |
| ***3*** | Aim 1c: Extension to key metallic elements | |  | |  |  |  |  |  | |  | |
| ***4*** | Aim 2a: Charge model development for GAFF3 | |  | |  | |  |  |  |  |  |  |
| ***5*** | Aim 2b: van der Waals parameterization for GAFF3 | |  | |  | |  |  |  |  |  |  |
| ***6*** | Aim 2c: Bonded term parameterization for GAFF3 | |  | |  | |  |  |  |  |  |  |
| ***7*** | Aim 3a: New TI-GPU technique development | |  | |  | |  |  |  |  |  |  |
| ***8*** | Aim 3b: Parameterization for GBSA and PBSA | |  | |  | |  |  |  |  |  |  |
| ***9*** | Aim 3c: Critical evaluation of GAFF2 in studying biomolecule-ligand interactions | |  | |  | |  |  |  |  |  |  |
| ***10*** | Aim 3c: Critical evaluation of GAFF3 in studying biomolecule-ligand interactions | |  | |  | |  |  |  |  |  |  |
| **Preparation** | **First Round Parameterization** | **Refinement** | **Evaluation** | | | | | | | | | |

**REFERENCES**

1. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995,** *91*, 1-41.

2. Salomon-Ferrer, R.; Case, D. A.; Walker, R. C., An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2013,** *3*, 198-210.

3. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009,** *30*, 1545-614.

4. Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K., NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **1999,** *151*, 283-312.

5. Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; Van Gunsteren, W. F., The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **2005,** *26*, 1719-1751.

6. Schmid, N.; Christ, C. D.; Christen, M.; Eichenberger, A. P.; van Gunsteren, W. F., Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. *Comput. Phys. Commun.* **2012,** *183*, 890-903.

7. Hu, Z.; Jiang, J., Assessment of biomolecular force fields for molecular dynamics simulations in a protein crystal. *J. Comput. Chem.* **2010,** *31*, 371-380.

8. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996,** *118*, 11225-11236.

9. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995,** *117*, 5179-5197.

10. Wang, J., Cieplak, P., Kollman, P.A., How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000,** *21*, 1049-1074.

11. Wickstrom, L.; Okur, A.; Simmerling, C., Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophys. J.* **2009,** *97*, 853-856.

12. Perez, A.; MacCallum, J. L.; Brini, E.; Simmerling, C.; Dill, K. A., Grid-based backbone correction to the ff12SB protein force field for implicit-solvent simulations. *J. Chem. Theory Comput.* **2015,** *11*, 4770-4779.

13. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.

14. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26*, 1781-1802.

15. Kaminski, G., Friesner, R.A., Tirado-Rives, J., Jorgensen, W.L., Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001,** *105*, 6474-6487.

16. Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F., Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011,** *40*, 843-856.

17. Weiner, P. K.; Kollman, P. A., Amber - Assisted Model-Building with Energy Refinement - a General Program for Modeling Molecules and Their Interactions. *J. Comput. Chem.* **1981,** *2*, 287-303.

18. Luo, R.; Wang, J. M.; Kollman, P. A., Development of a more accurate amber united-atom force field for protein folding and large-scale biomolecular simulations. *Abstracts of Papers of the American Chemical Society* **2002,** *224*, U470-U471.

19. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003,** *24*, 1999-2012.

20. Cerutti, D. S.; Swope, W. C.; Rice, J. E.; Case, D. A., ff14ipq: A Self-Consistent Force Field for Condensed-Phase Simulations of Proteins. *J. Chem. Theory Comput.* **2014,** *10*, 4515-4534.

21. Cieplak, P., Caldwell, J., Kollman, P.A., Molecular Mechanical Models for Organic and Biological Systems. Going Beyond the Atom Centered Two Body Additive Approximation: Aqueous Solution Free Energies of Methanol and N-Methyl Acetamide, Nucleic Acid Bases and Amide Hydrogen Bonding and Chloroform/Water Partition Coefficients of the Nucleic Acid Bases. *J. Copmut. Chem.* **2001,** *22*, 1048-1057.

22. Wang, J.; Cieplak, P.; Cai, Q.; Hsieh, M. J.; Wang, J.; Duan, Y.; Luo, R., Development of polarizable models for molecular mechanical calculations. 3. Polarizable water models conforming to Thole polarization screening schemes. *J. Phys. Chem. B* **2012,** *116*, 7999-8008.

23. Wang, J.; Cieplak, P.; Li, J.; Cai, Q.; Hsieh, M. J.; Luo, R.; Duan, Y., Development of polarizable models for molecular mechanical calculations. 4. van der Waals parametrization. *J. Phys. Chem. B* **2012,** *116*, 7088-7101.

24. Wang, J.; Cieplak, P.; Li, J.; Hou, T.; Luo, R.; Duan, Y., Development of polarizable models for molecular mechanical calculations I: parameterization of atomic polarizability. *J. Phys. Chem. B* **2011,** *115*, 3091-3099.

25. Wang, J.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M.; Lei, H.; Luo, R.; Duan, Y., Development of polarizable models for molecular mechanical calculations II: induced dipole models significantly improve accuracy of intermolecular interaction energies. *J. Phys. Chem. B* **2011,** *115*, 3100-3111.

26. Dickson, C. J.; Madej, B. D.; Skjevik, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C., Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014,** *10*, 865-879.

27. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29*, 622-655.

28. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25*, 1157-1174.

29. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006,** *25*, 247-260.

30. Bowen, J. P.; Allinger, N. L., Molecular Mechanics: The Art and Science of Parameterization. In *Reviews in Computational Chemistry*, John Wiley & Sons, Inc.: 1991; pp 81-97.

31. Wang, J.; Kollman, P. A., Automatic parameterization of force field by systematic search and genetic algorithms. *J. Comput. Chem.* **2001,** *22*, 1219-1228.

32. Aronica, P. G. A.; Fox, S. J.; Verma, C. S., Comparison of Charge Derivation Methods Applied to Amino Acid Parameterization. *ACS Omega* **2018,** *3*, 4664-4673.

33. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges- the RESP model. *J. Phys. Chem.* **1993,** *97*, 10269-10280.

34. Cieplak, P., Cornell, W.D., Bayly, C.I., Kollman, P.A., Application of the Multimolecule and Multiconformation RESP Methodology to Biopolymers - Charge Derivation for DNA, RNA and Proteins. *J. Comput. Chem.* **1995,** *16*, 1357-1377.

35. Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A., Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009,** *5*, 350-358.

36. Beveridge, D. L.; DiCapua, F. M., Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989,** *18*, 431-492.

37. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015,** *137*, 2695-2703.

38. Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012,** *8*, 1542-1555.

39. Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013,** *9*, 3878-3888.

40. Lee, T. S.; Hu, Y.; Sherborne, B.; Guo, Z.; York, D. M., Toward Fast and Accurate Binding Affinity Prediction with pmemdGTI: An Efficient Implementation of GPU-Accelerated Thermodynamic Integration. *J. Chem. Theory Comput.* **2017,** *13*, 3077-3084.

41. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, I., T.E. ; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; D.M., Y.; Kollman, P. A., AMBER 2018. *University of California, San Francisco* **2018**.

42. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000,** *21*, 132-146.

43. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002,** *23*, 1623-1641.

44. Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G., Minnesota Solvation Database – version 2012. *University of Minnesota, Minneapolis* **2012**, <https://comp.chem.umn.edu/mnsol>.

45. Bannan, C. C.; Burley, K. H.; Chiu, M.; Shirts, M. R.; Gilson, M. K.; Mobley, D. L., Blind prediction of cyclohexane-water distribution coefficients from the SAMPL5 challenge. *J. Comput. Aided Mol. Des.* **2016,** *30*, 927-944.

46. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2011,** *51*, 69-82.

47. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* **2011,** *32*, 866-877.

48. Wang, J.; Hou, T.; Xu, X., Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Curr. Comput. Aided-Drug Des.* **2006,** *2*, 287-306.

49. Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T., Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models. *J. Phys. Chem. B* **2013,** *117*, 8408-8421.

50. Wang, J.; Morin, P.; Wang, W.; Kollman, P. A., Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. *J. Am. Chem. Soc.* **2001,** *123*, 5221-5230.

51. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000,** *33*, 889-897.

52. Meirovitch, H., Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation. *Curr. Opin. Struct. Biol.* **2007,** *17*, 181-186.

53. Peter, C.; Oostenbrink, C.; van Dorp, A.; van Gunsteren, W. F., Estimating entropies from molecular dynamics simulations. *J. Chem. Phys.* **2004,** *120*, 2652-2661.

54. Spackova, N.; Cheatham, T. E.; Ryjacek, F.; Lankas, F.; van Meervelt, L.; Hobza, P.; Sponer, J., Molecular dynamics simulations and thermodynamics analysis of DNA-drug complexes. Minor groove binding between 4 ',6-diamidino-2-phenylindole and DNA duplexes in solution. *J. Am. Chem. Soc.* **2003,** *125*, 1759-1769.