Molecular simulations have been playing 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 rational 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**

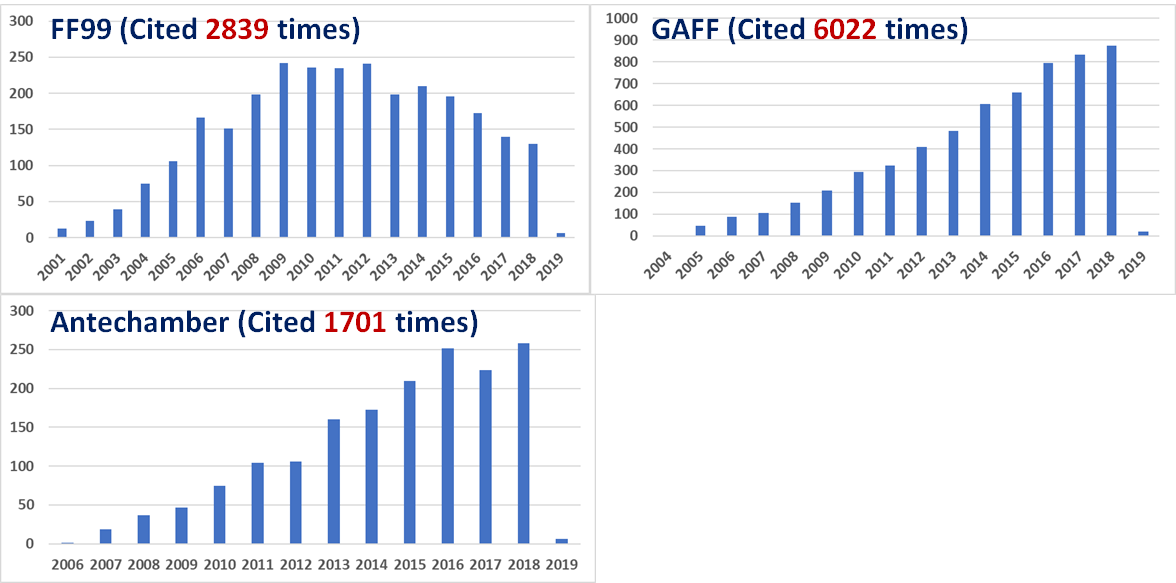
**A.1. Molecular Simulations and Force Fields.** Atomistic simulations have become an important tool for studying the structure, dynamics, interactions, and functions of biomolecular systems.[1-8](#_ENREF_1) Critical to their success is the accurate representations of potential energy (referred to as 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, bond angles, and dihedral angles, *req* and *θeq* are the equilibrium values of bond lengths and bond 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, which are for simulations of biological systems including proteins, DNAs & RNAs, etc., have been widely used in modeling biomolecules with more than 20,000 citations as of this writing.

**A.3. Qualification of the Research Team.** The principal investigator (**PI**), Dr. Junmei Wang, is the inventor of the widely-used 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) He is also one of the main contributors of the popular AMBER force field family (especially AMBER FF99[10](#_ENREF_10)). AMBER FF99, GAFF and Antechamber have been applied in numerous research studies with more than 10,500 citations in total (**Fig. 1**). Collaborator Dr. Tai-Sung Lee has a broad and strong background in free energy methodology, molecular simulations, and software development.[19](#_ENREF_19), [30-32](#_ENREF_30) Recently he developed and implemented pmemdGTI, a GPU-accelerated thermodynamic integration module into AMBER software package.[30](#_ENREF_30), [33](#_ENREF_33) Collaborator Dr. Xibing He is both an AMBER and CHARMM developer and he has extensive expertise and experience in force field development.[34-39](#_ENREF_34) In short, we have assembled a highly competitive team and have sufficient expertise and experience to pursue the proposed research.



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

**A.4 Additive vs Polarizable Force Fields.** Recently polarizable force fields which can model changing dielectric environment, are expected to extend our ability to more adequately study complex biomolecular systems. Atomic polarization effects play a critical role in the interactions of ions with nucleic acids, the environmental changes during protein folding, enzymatic mechanism, and low or heterogeneous dielectric environments.[40](#_ENREF_40) Towards the goal of properly including polarization, a great deal of effort has been directed to developing polarizable versions of force fields. A variety of methods have been explored, including induced dipole models,[22-25](#_ENREF_22), [40-42](#_ENREF_40) the fluctuating charge models,[43-47](#_ENREF_43) the Drude oscillator models,[48-52](#_ENREF_48) and the detailed multipole expansion model.[41](#_ENREF_41), [53](#_ENREF_53) In collaboration with Dr. Yong Duan and other AMBER FF consortium PIs, we have developed a set of polarizable FFs based on the framework of Thole’s atomic dipole interaction models. Four papers that document the solid progress on developing the AMBER polarizable FFs have been published.[22-25](#_ENREF_22)  Recently, we have developed a polarizable Gaussian Multipole (**pGM**) model.[54-59](#_ENREF_54) This model represents a novel framework, which applies Gaussian functions to seamlessly treat the multipole polarization and electron penetration effects. We have finished parameterization of atomic polarizability and screening factor parameters for pGM, which outperforms the previous Thole polarizable models in reproducing high-level *ab initio* polarizability anisotropy. As a member of AMBER consortium team, the PI will develop **pGM-GAFF**, a polarizable GAFF force field based on the framework of pGM.

While polarizable FFs are under active developments, serious issues of polarizable FFs, such as low computation efficiency and parameterization difficulty, make them unable to replace or beat the classical additive FFs in a wide range of applications, particularly in large-scale studies of protein-ligand and nucleic acid-ligand interactions and rational drug design. Therefore, we still put considerable effort on the development of additive FFs. Moreover, many FF developers including the PI believe that there is still a great room for the improvement of the accuracy of additive FFs, as demonstrated in the preliminary data of Section C.3.

**A.5. Development of a Set of 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 MMFF development. MMFF development has long been more of an “art” than a “science”.[60-62](#_ENREF_60) Atomic charge models, the cornerstones of MMFFs, are developed both empirically and physically (such as the RESP[63-64](#_ENREF_63) 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 molecules 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[65-76](#_ENREF_65) plays an essential role for us to achieve the above research goals. 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 binding site of the receptor for ΔGbind). The simplest, most ubiquitous and most important solvent is water. The calculation of ΔGsolv 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 primary target during FF parameterization, but 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,[75](#_ENREF_75) limiting the accuracy of predicting ΔGbind even with the state-of-the-art alchemical free energy calculation methods, such as thermodynamic integration (**TI**)[77](#_ENREF_77) and free energy perturbation (**FEP**),[78](#_ENREF_78) 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[79-80](#_ENREF_79) and access to the local and national supercomputing resources, particularly the GPU-TI[30](#_ENREF_30) code recently released in AMBER 18,[31](#_ENREF_31) it is timely for us to introduce TI reproduction of experimental solvation free energies into our FF parameterization scheme as a primary target. 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.6. 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,[81-82](#_ENREF_81) and the accuracy and reliability of the underlying force field (**FF**) that models the molecular interactions. The force fields contain large sets of parameters that were derived from quantum-chemical calculations and selected 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.[83-84](#_ENREF_83) 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, Dr. Lee et al. have developed a set of tools based on GPU-accelerated free energy methods, 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 AMBER 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.

Besides the pathway method, a set of endpoint free energy methods, molecular mechanics Poisson-Boltzmann and Generalized Born Surface Area (MM-PB/GBSA),[32](#_ENREF_32), [85-87](#_ENREF_85) which achieve a good balance between computational efficiency and accuracy, have been widely used in structure-based drug designs. The performance of MM/PBSA and MM/GBSA depends on the quality of both the MMFFs and the implicit solvation models. We have conducted critical evaluations of both methods.[88-91](#_ENREF_88) To improve the performance of the two endpoint methods, a key is to develop force field-consistent solvation models. For the newly developed charge model for GAFF3, we plan to optimize the PBSA and GBSA radius parameters to reproduce measured solvation free energies. We will conduct large scale evaluation for a set of diverse drug targets that have multiple inhibitors and decoys through collaboration.

Moreover, our force field models have been and will continue to be assessed in numerous drug discovery projects where our predictions can be confirmed by experimental biologists. For most of the projects, we intend to develop novel compounds to activate (as agonists) or inhibit (as antagonists) protein targets. We have established a broad collaboration network with experimental scientists. Most collaborations resulted in grant application and publications.[92-97](#_ENREF_92)

**B. INNOVATION**

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

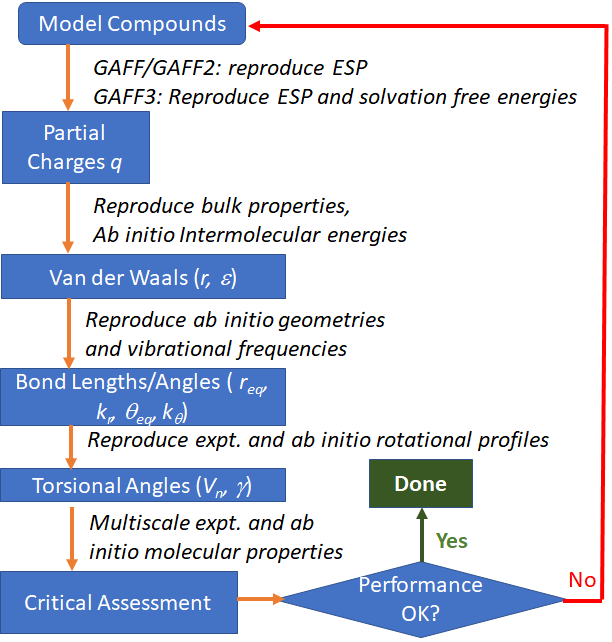
* In contrast to the traditional “bottom-up” strategy in MMFF development, the GAFF2/GAFF3 development is mainly guided by a novel top-down strategy to maximize the transferability and self-consistency.
* A novel charge method will be developed to largely overcome the limitations of the charge models 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 thus, they can boost the successful rate of structure-based rational drug design.
* Solvation free energy will be applied as a **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 compared 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/d𝞴* 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 scales of free energy simulations are needed.

**C. APPROACH**

**C.1 General Force Field Development Philosophy and Parameterization Strategy.**

The research goal of this application is to develop and evaluate a set of high-quality force fields that are highly transferable and self-consistent for studying protein-ligand and nucleic acid-ligand interactions. What are the traits of a good force field? We believe **transferability, self-consistency, accuracy, and applicability are the most important attributes of a good force field**. Even though most biomolecular force fields were developed using the effective function form (**Eq. 1**), the parameterization strategies are quite different. AMBER force fields distinguished themselves from others by using physical charges derived to fit the *ab initio* electrostatic potentials. This special feature, which makes AMBER more expandable as proven by the success of GAFF, will be kept in our newly developed FFs. Traditionally, most FFs were developed following a bottom-up strategy: FF parameterizations were performed one compound class after another, with an assumption that the developed FF parameters could be seamlessly transferred to new compound classes. However, this assumption may not always hold. To maximize the transferability and self-consistency, a novel top-down strategy was applied to develop the GAFF2/GAFF3 force fields: (1) the FF parameters were optimized for multiple compound classes simultaneously; (2) new atom types and new FF parameters were introduced only when they could significantly improve the fitting performance. Both the bottom-up and top-down strategies will be applied in our force field development in the future.

**Figure 2. Strategies of force field development.**



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, heats of vaporization, etc. 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.

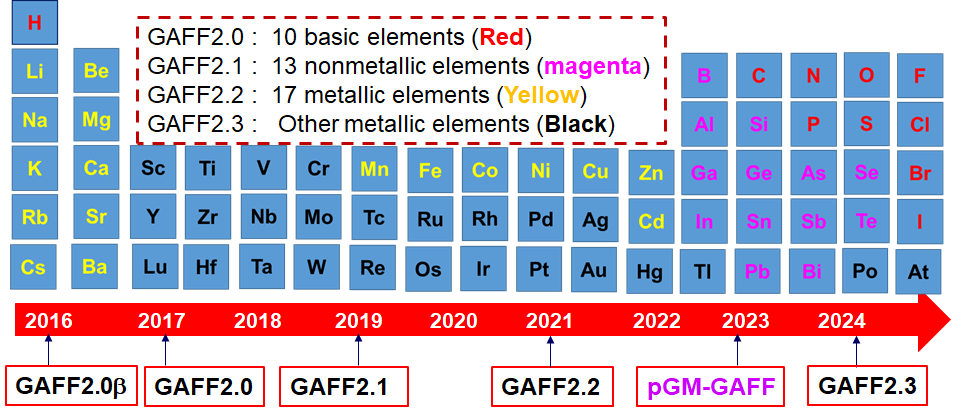
**C.2 Aim 1. Continue Development of the Second Generation of GAFF Force Field (GAFF2).**

**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. As most AMBER biomolecular FFs apply RESP charges,[63-64](#_ENREF_63) GAFF2 has been and will continue be developed using the RESP charge method.

**C.2.1 Current GAFF2 Development.**After its release, we received numerous suggestions and comments on GAFF, mainly coming from the AMBER mailing list. We have maintained GAFF and Antechamber for years and released several subversions (1.5, 1.6, 1.7 and 1.8) where recognized problems are remedied. Evidently, GAFF could only become a successful force field with continuous input from its users. Recently, we realized that the GAFF van der Waals parameters, which were adopted from the AMBER FF94/99 force fields, needed to be optimized to bring the performance of GAFF to a new higher level. As such, we decided to move forward to develop the second generation of GAFF – GAFF2. A beta version of GAFF2 – GAFF2β, which covers the chemical space of the common organic molecules, has been released with AMBER14 and 15. Later GAFF2.0 has been released with AMBER 16, 17 and 18. The major improvement achieved by GAFF2 includes: (1) reparameterization of the van der Waals parameters to better reproduce the high-level *ab initio* and high quality experimental data of a variety of molecular properties including intermolecular energy, liquid density, heat of vaporization, and hydration free energy (the RMS errors reduced 20 to 40% for the condensed molecular properties); (2) development of empirical functions and a set of associated parameters to directly calculate the bond stretching and bond bending force constants (Compared to GAFF, the RMS errors of vibration frequencies reduced ~50% in GAFF2); and (3) reparameterization of torsional angle parameters using a much larger set of model compounds for which the rotational profiles were calculated at the MP2/aug-cc-pVTZ level. GAFF2 is expected to be an even more successful general-purpose force field and GAFF2-based docking scoring functions could significantly improve the success rates of virtual screenings in target-based drug design. A manuscript describing GAFF2.0 development is in process.

**C.2.2 Further Development of GAFF2.** We propose to significantly improve the quality of GAFF2 by reparameterization and introduction new FF parameters, especially for the torsional terms. We plan to develop FF parameters for broader coverage of the chemical space of small molecules, going beyond the ten basic elements covered by GAFF and GAFF2.0. The road map of GAFF2 development is shown in **Fig. 3**. Eventually, GAFF2 will become a universal force field covering most elements in the periodic table except for lanthanoids and actinoids. All the developed FFs will be released with AMBER. The significant expansion of the chemical space coverage enables GAFF2 to study very complicated biomolecular systems consisting of different types of entities. It could also find applications in other research fields, such as material sciences.

**Figure 3. Roadmap of GAFF development**. The chemical space coverages are summarized in the dashed rectangle, and the tentative release dates of GAFF2 force fields are shown at the bottom of this figure.



**GAFF2.0** We plan to significantly improve the quality of GAFF2 by reparametrizing the existing and introducing new parameters. We will focus on the problematic molecules reported by our users and/or occurring to other general-purpose force fields.[98-99](#_ENREF_98) We plan to systematically evaluate GAFF2 in reproducing the *ab initio* rotational profiles for a great number of model compounds. In our effort of conducting drug likeness analysis, a brutal-force algorithm was applied to iteratively dissect drug molecules into fragments. We will collect all the fragments that have one rotatable bond to generate model compounds. We will then generate rotational profile at the MP2/aug-cc-pVTZ//MP2/6-31G\* level. Next, we will examine how well the GAFF2 rotational profiles reproduce the *ab initio* ones. New torsional angle parameters will be introduced if necessary.

**GAFF2.1.** Besides to improve the FF quality of GAFF2.0, we plan to expand the chemical space coverage to other elements occurring in biological, pharmaceutical and organic molecules. Very Frequently, our users raise questions on the missing force field parameters involved by the uncovered nonmetallic elements in GAFF2.0, such as selenium, silicon, boron, etc. FF parameterization for 13 nonmetallic elements in GAFF2.1 is ongoing. The same FF parameterization strategy of developing GAFF2.0 is utilized, as the liquid property data, such as density and heat of vaporization, are also available for those nonmetallic elements. We plan to release GAFF2.1 at the end of this year according to the roadmap.

**GAFF2.2/2.3**. FF development for metallic compounds is different from that for organic molecules. First, the functional forms are different: (1) a cosine Fourier expansion should be used for metallic bond angles as the harmonic functional form cannot address the multiple equilibrium bond angle problem;[100-101](#_ENREF_100) (2) the Morse function may describe the bond stretching better for metallic bonds;[100](#_ENREF_100) and (3) the 12-6 Lennard-Jones potential describing the van der Waals term may be replaced with the other potential forms.[102-104](#_ENREF_102) Second, metallic FF parameterization heavily depends on high-quality *ab initio* calculations. We will collaborate closely with Dr. Kenneth Merz at Michigan State University on the general metallic force field development.

The basic FF parameterization strategy is summarized in **Fig. 2**. We plan to perform FF parameterizations using a set of hierarchical data, which include high-level *ab initio* data (esp. interaction energy), and experimental gas and bulk properties including the thermodynamic data.[61](#_ENREF_61) This multiscale optimization, enabled by the rapidly increasing computer power, can facilitate the development of robust and highly transferrable force fields.[105](#_ENREF_105)

**C.3 Aim 2. Development of a New, Accurate, Highly Transferable, and Self-Consistent Force Field (GAFF3) 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 on ΔGsolv prediction will enable us to improve the accuracy of ΔGbind prediction, and the latter is extremely important for computer-aided drug design.

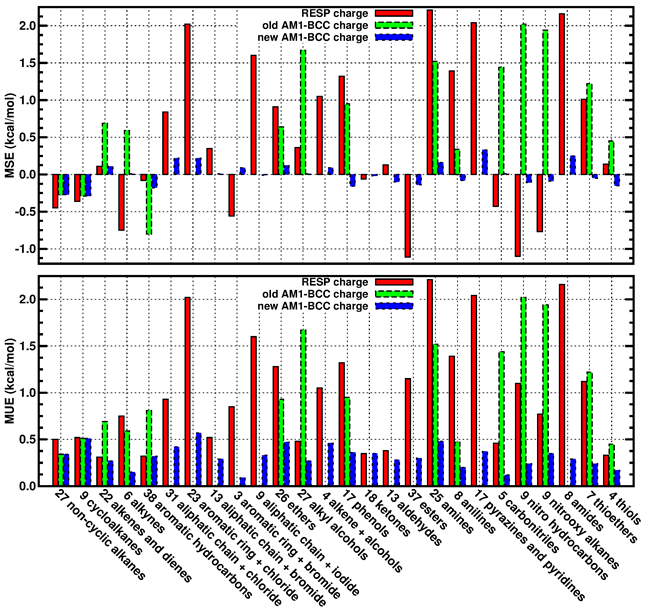
**C.3.1 Background.** Even though all FF parameters in an MMFF (**Eq. 1**) are developed in an iterative manner, the model of assigning atomic partial charges is always the first step due to its major role in determining the calculated intra- and inter-molecular energies. Traditionally, the AMBER force 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.[63-64](#_ENREF_63) 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[106-107](#_ENREF_106) 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 types in the molecule so that the resulting charges can emulate the RESP charges more closely than AM1 Mulliken.

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 (**Fig. 4**).

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 term parameters will be validated and if necessary, re-adjusted in accordance with newly developed BCC parameters (and corresponding new atomic charges).

**C.3.2 Preliminary Results.** 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[108](#_ENREF_108) which is bounded to AMBER force fields and GAFF. The calculation method is the TI as described in previous literature.[30](#_ENREF_30) As shown in **Fig. 4**, a lot of calculations using the RESP charges and old AM1-BCC charges 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 **Fig. 4**, 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 (**Fig. 4**). In a short summary, the accuracy of calculated ΔGhyd can be significantly improved by adjusting the BCC parameters.

**Figure 4. The mean signed errors (MSE, upper panel) and mean unsigned errors (MUE, lower panel) of calculated hydration free energies** (in kcal/mol) of various molecules using atomic charges from RESP, old AM1-BCC and new AM1-BCC methods. The BCC parameters for non-cyclic alkanes and cycloalkanes are not changed.



**C.3.3 Experimental Design.** The similar FF development philosophy and FF parameterization strategy for developing GAFF2 will be seamlessly introduced to GAFF3 force field development. GAFF3 will first be a general-purpose FF covering the 10 basic elements as GAFF2.0. In a long run, GAFF3 may be expanded to cover other elements given high-quality experimental solvation free energies are available for deriving BCC parameters.

**New Charge Model Development for GAFF3.** 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.[109-110](#_ENREF_109) 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.[109-110](#_ENREF_109) The newly developed charge model will be tested and adjusted against ΔGsolv in or partition coefficients between various solvents.

**van der Waals Parameterization for GAFF3.** 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 a polarizable GAFF[23](#_ENREF_23) will be utilized to tune VDW parameters to reproduce both the *ab initio* interaction energies and the liquid properties.

**Bonded Term Parameterization for GAFF3.** After finishing the development of the new charge model and the VDW parameters, 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 GAFF2, 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.

**Long-Term Goals on GAFF3-Compatible Force Field Development for Biomolecules.** The current approach of studying protein-ligand and nucleic acid-ligand interactions using a combination of the general-purpose and biomolecular force fields may have a consistency problem. 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 GAFF3 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.

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

**Rational.** Force field validation is an on-going process. We choose different model systems to calibrate different properties for different force fields. We utilize protein-ligand and nucleic acid-ligand systems to evaluate both GAFF2/GAFF3 and the biomolecular force fields. The force field models will also be extensively assessed with real drug discovery projects. We plan to integrate the force field development and validation into drug discovery projects through collaborations.

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 thermodynamic integration (**TI**) and free energy perturbation (**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 (**C.4.1**), and we also plan to develop a user-friendly GUI toolkit to overcome the 3rd problem (**C.4.2**) to help non-expert users to set up the simulations and do analysis for various free energy calculations.

**C.4.1 New Sampling Algorithm for Thermodynamic Integration.** We propose to apply a recently developed innovative sampling algorithms to boost the convergence of TI calculations. The proposed *λ-*SAMS approach was applied to the thrombin ligand 1d->7a mutation[111](#_ENREF_111) in water as proof of principle. As shown in **Fig. 5**, the *λ-*SAMS, either with 100 *λ*-windows or 200 *λ-*windows reaches 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.

**Experimental Design.** The detailed plan on sampling algorithm development and evaluation is described as follows.

* Testing infrastructure: we will first establish an 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, the evaluation of free energy transformations will be performed using the current AMBER/GAFF, the AMBER/GAFF2 and AMBER/GAFF3 force fields. Results will be compared with benchmark results for all ligand force fields.
* Baseline: The second step will involve applying the established 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.[78](#_ENREF_78) 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.
* 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[112-113](#_ENREF_112) (which uses the maximum common substructure) and has been reported in the literature for the entire dataset.[78](#_ENREF_78) 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 GAFF3 force-field: We will perform binding free energy simulations on the data set mentioned above using GAFF3. 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 GAFF3. 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.

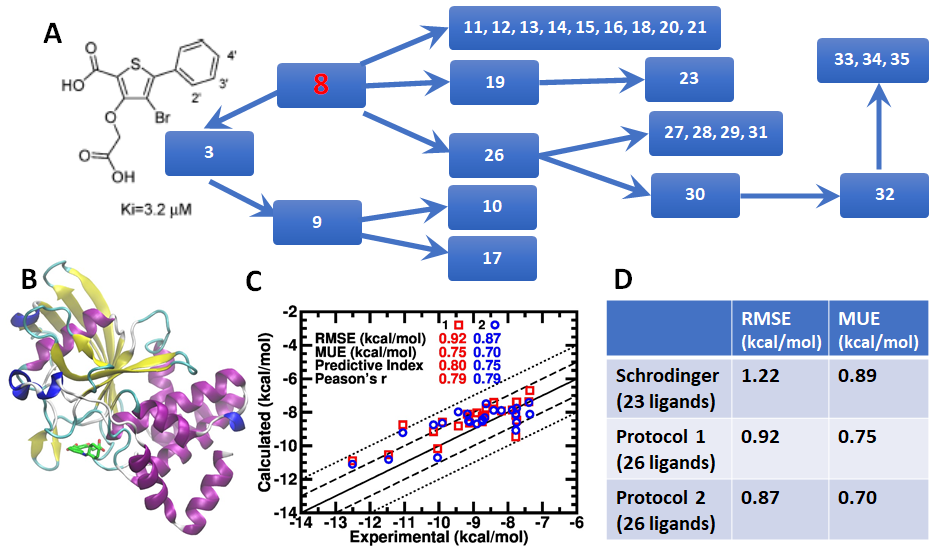


**Figure 5. Preliminary results of *λ*-SAMS approach.** The standard errors averaged 𝞓G from 16 independent simulations are shown for the thrombin ligand 1d->7a mutation in water. “No RepEx”: TI simulations without replica exchange; “RepEx at 100 fs”: TI simulations with the replica exchange (between *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.

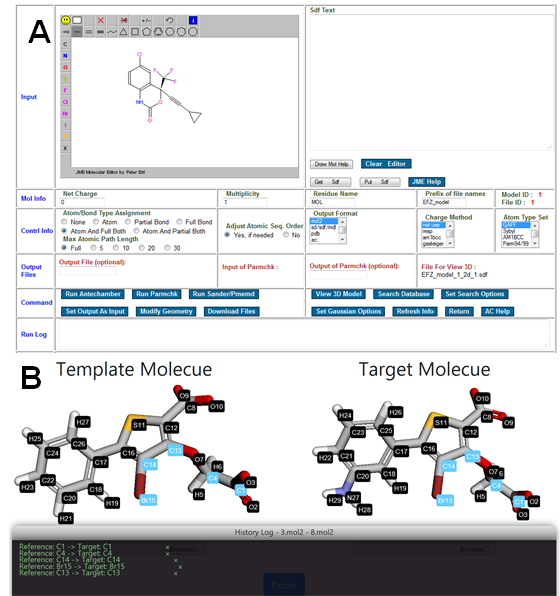
**C.4.2 Web-Based Toolkit for TI Preparation and Analysis.** According to our experience of talking to both academic researchers and industrial R&D personnel, such an automatic or semi-automatic GUI toolkit is in great demand.

**TI Setup.** 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/GAFF3 parameters to get accurate results. We finished testing a set of systems with the Protein Tyrosine Phosphatase 1B (PTP1B) as the receptor (**Fig. 6B**) and a series of 26 compounds (**Fig. 6A**) as the ligands[114](#_ENREF_114). 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 MUE than the reported results from the commercial FEP module in Schrodinger Suite[78](#_ENREF_78) (**Fig. 6C** and **6D**). Next, we need to integrate the in-house scripts and programs into the GUI toolkit.

**Figure 6**. **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 versus experimental data. (D) RMSEs and MUEs from our calculations compared to Schrodinger FEP.



**A GUI-Toolkit for TI Setup and Analysis**. 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 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 modified subroutines in the Antechamber programs which can precisely describe an atom’s chemical environment (including the atomic paths, atomic connectivity and atomic property), discriminate subtle local chemical environments, and automatically define complicated force field atom types, bond types. Sometimes multiple mapping is possible, such as the 2,6 positions or the 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.

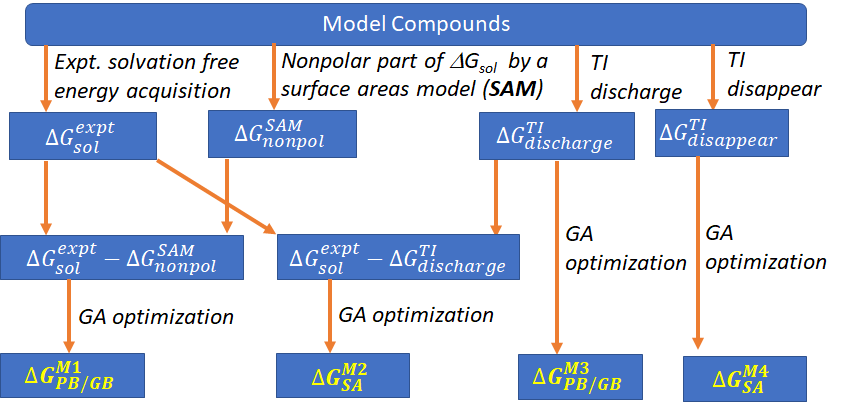


**Figure 7**. **GUIs for Graphic Antechamber (Panel A) and atom pair definition (Panel B)**

We have developed a set of GUIs for Antechamber programs using PHP (https://mulan.pharmacy.pitt.edu) to automatically get atomic charges and GAFF parameters for arbitrary organic molecules. The graphic antechamber program itself takes the input from JME Molecular Editor (**Fig. 7A**). Occasionally, users may want to define the matching atom pairs themselves. We have developed a graphic interface to facilitate users to define atom pairs (**Fig. 7B**). To use this tool, a molecule as the template and another molecule as the target need to be loaded first, then one can define matched atom pairs by clicking atom names. The defined atom pairs are shown in the bottom window and one can delete a pair by click the appended “×” symbol.

**C.4.3 Parameterization of GAFF3 Charge Model-Consistent Implicit Solvation Models.**

**Figure 8**. **Implicit solvation model development based on GAFF3 charges.**



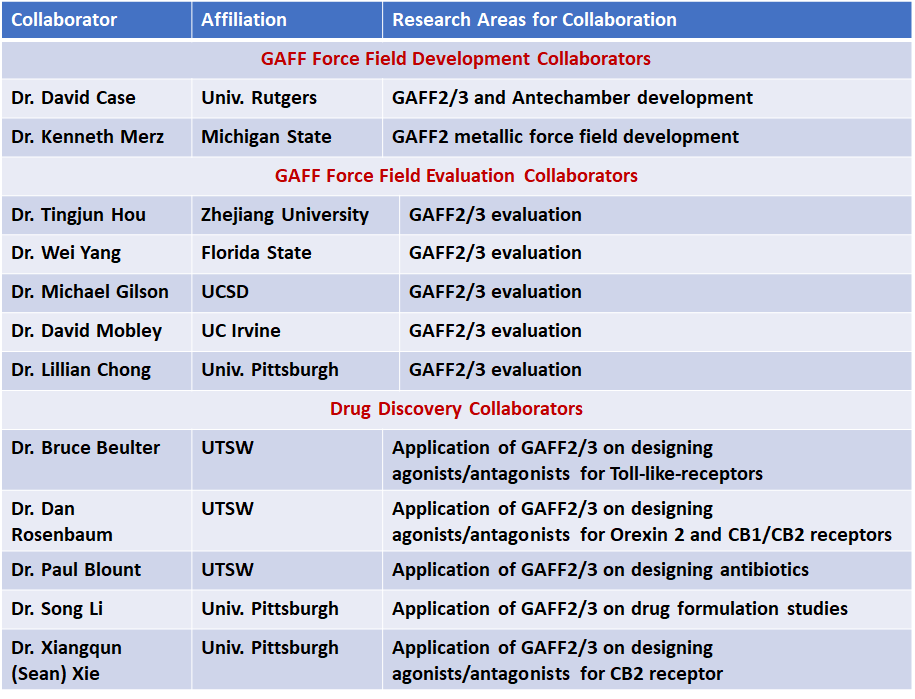
MM-PB/GBSA (Molecular Mechanics Poisson-Boltzmann/Generalized Born Surface Area) are two popular endpoint free energy methods in structure-based drug design.[85](#_ENREF_85), [87](#_ENREF_87), [115-118](#_ENREF_115) The performance of MM-PB/GBSA strongly depends on how well calculation errors can cancel out.[87](#_ENREF_87) Therefore, it is necessary to develop a set of solvation models to be consistent with the GAFF3 charge method. The strategy of PBSA and GBSA model development is summarized in **Fig. 8**. In detail, the polar part of a solvation free energy may come from either experimental solvation free energy () or TI free energy of the discharge procedure (). The nonpolar part has also two sources: the TI free energy of atom disappearing procedure () and the difference between the experimental and the TI free energy of the discharge procedure (). A genetic algorithm (GA) will be applied to optimize the radius parameters to reproduce the above reference data. Four model sets will be developed: Model Sets M1 and M3 are for the polar part of the solvation free energy, while Model Sets M2 and M4 are for the nonpolar part of solvation free energy. Model Sets M1 should be used with the employed nonpolar surface area model for preparing ; the combination of Model Sets M2 and M3 targets to reproduce the experimental solvation free energies while the combination of Model Sets M3 and M4 targets to reproduce the TI solvation free energies.

**C.4.4 Critical and Systemic Assessment of GAFF2/GAFF3 with TI.**

A comprehensive data set of relative binding affinities assembled was introduced by Schrödinger Inc. in 2015.[78](#_ENREF_78) 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 mitogen-activated protein kinase). This data set is the current “gold standard” for computational drug design. Another set of 10 to 20 well-studied protein targets[88-89](#_ENREF_88), [119](#_ENREF_119) and nucleic acid targets (such as DAPI,[120](#_ENREF_120)), which have measured *IC50* or *Ki* for several to tens of diverse small molecule ligands, will be selected to test our GAFF2/GAFF3 force fields.

**C.4.5 Systemic Assessment of GAFF2/GAFF3 with MM-PB/GBSA.**

In structure-based drug design, virtual screening becomes an indispensable technique. Endpoint free energy methods, like MM-PB/GBSA, are routinely applied in the last stage to re-rank docking scores. In many cases, the combination of molecular docking and MM/PB(GB)SA rescoring has been proven to be a more promising strategy in both the identification of the correct binding poses and the right ranking of the binding affinities of a series of ligands.[89](#_ENREF_89), [121-129](#_ENREF_121) To evaluate GAFF2/GAFF3 and the newly developed MM-PB/GBSA models, besides the aforementioned systems in **C.4.4**, we plan to study a set of drug targets which have tens, hundreds, or even thousands of actives and decoys. We would like to assess both the “Screening Power” and the “Docking Power” of the MM-PB/GBSA models in virtual screenings. We will collaborate closely with Dr. Tingjun Hou at Zhejiang University, who has systematically evaluated the MM/PB(GB)SA methods in virtual screenings for 98 drug targets,[89](#_ENREF_89) to fulfill the task.



**Table 1**. **List of GAFF Development Collaborators**

**C.4.6 Objectively Evaluation of GAFFs Using Drug Discovery Projects.**

Our force field models have been further critically assessed in numerous drug discovery projects where our predictions can be confirmed by experimental biologists. For most of the projects, we intend to develop novel compounds to activate (as agonists) or inhibit (as antagonists) protein targets. As most drug targets are GPCRs and only a few crystal structures are available, protein homology modeling and protein-ligand complex modeling are common tasks. Other common modeling tasks include high-throughput screenings,[130](#_ENREF_130) binding free energy calculations, MD simulations of bioprocesses. The best outcomes for us from those collaboration projects are experimental validations of our molecular modeling prediction (protein-ligand structure models, designed drug candidates, etc.). My collaborators working in wet labs are listed in bottom **Table 1**.

**C.4.7 Evaluation of Alternative General Force Fields and Comparison to GAFFs.**

We plan to do similar evaluation for alternative general force fields, such CGenFF[98](#_ENREF_98" \o "Vanommeslaeghe, 2010 #281) (CHARMM-compatible), 2016H66[131](#_ENREF_131) (GROMOS-compatible), SMIRNOFF Force Field class, etc., and compare them to GAFF2/GAFF3. The SMIRNOFF Force Field class uses the SMIRKS extension of the SMILES chemical perception language[132](#_ENREF_132) to label specific atoms within a chemical pattern. Its current released version SMIRNOFF99Frosst[133](#_ENREF_133) ports and adapts parameters drawn from AMBER FF99, GAFF2, and Merck’s parm@Frosst force field[134](#_ENREF_134). According to the road map of the developers,[135](#_ENREF_135) the future fixed charge model of SMIRNOFF will be either modified RESP targeting QM methods with a level higher than traditional HF/6-31G\* or a modified AM1-BCC targeting the modified RESP potentials, while our new AM1-BCC method directly targets both the QM ESP and the experimental data of solvation free energies.

**D. TIMELINE AND MILESTONES**

We have proposed three aims this R01 project. The roadmap and milestones are summarized below:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Force Field Development Sub-Aims** | | **Y1** | | **Y2** | | **Y3** | | **Y4** | | **Y5** | |
| **1** | Aim 1a: Additional FF Parameterization for GAFF2.0 (10-basic elements) | |  |  |  | |  | |  | |  |  |
| ***2*** | Aim 1b: Extension to all nonmetallic elements (GAFF2.1) | |  | |  |  |  | |  | |  | |
| ***3*** | Aim 1c: Extension to key metallic elements (GAFF2.2, GAFF2.3) | |  | |  |  |  | |  | |  |  |
| ***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: TI toolkit and sampling technique development | |  | |  | |  | |  | |  |  |
| ***8*** | Aim 3b: GAFF3 charge compatible PB/GBSA parameterization | |  | |  |  |  |  |  | |  | |
| ***9*** | Aim 3c: Critical evaluation of GAFF2 in studying biomolecule-ligand interactions | |  | |  | |  | |  | |  |  |
| ***10*** | Aim 3c: Critical evaluation of GAFF3 in studying biomolecule-ligand interactions | |  | |  | |  | |  |  |  |  |
| ***11*** | Aim 3c: Critical evaluation of other general-purpose force fields | |  | |  | |  | |  | |  |  |
|  | **Planning Phase** | **Execution Phase** | **Closure Phase** | | | | | | | | | |

**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* (1-3), 1-41.

2. Salomon-Ferrer, R.; Case, D. A.; Walker, R. C., An overview of the Amber biomolecular simulation package. *Wires Comput Mol Sci* **2013,** *3* (2), 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* (10), 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* (1), 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* (16), 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* (4), 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* (2), 371-80.

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. *Journal of the American Chemical Society* **1996,** *118* (45), 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. *Journal of the American Chemical Society* **1995,** *117* (19), 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* (12), 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* (3), 853-6.

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* (10), 4770-9.

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* (8), 3696-713.

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* (16), 1781-802.

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* (7), 843-56.

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* (3), 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* (16), 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* (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* (28), 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* (24), 7088-101.

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* (12), 3091-9.

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* (12), 3100-11.

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* (2), 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* (4), 622-55.

28. Wang, J. M.; 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* (9), 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* (2), 247-60.

30. 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* (7), 3077-3084.

31. 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**.

32. 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* (12), 889-97.

33. Lee, T. S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M., GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J Chem Inf Model* **2018,** *58* (10), 2043-2050.

34. He, X. B.; Lopes, P. E. M.; MacKerell, A. D., Polarizable Empirical Force Field for Acyclic Polyalcohols Based on the Classical Drude Oscillator. *Biopolymers* **2013,** *99* (10), 724-738.

35. He, X. B.; Shinoda, W.; DeVane, R.; Anderson, K. L.; Klein, M. L., Paramaterization of a coarse-grained model for linear alkylbenzene sulfonate surfactants and molecular dynamics studies of their self-assembly in aqueous solution. *Chemical physics letters* **2010,** *487* (1-3), 71-76.

36. He, X. B.; Shinoda, W.; DeVane, R.; Klein, M. L., Exploring the utility of coarse-grained water models for computational studies of interfacial systems. *Mol Phys* **2010,** *108* (15), 2007-2020.

37. Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D., Jr., Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J Comput Chem* **2012,** *33* (31), 2451-68.

38. Patel, D. S.; He, X. B.; MacKerell, A. D., Polarizable Empirical Force Field for Hexopyranose Monosaccharides Based on the Classical Drude Oscillator. *J Phys Chem B* **2015,** *119* (3), 637-652.

39. Li, H.; Chowdhary, J.; Huang, L.; He, X. B.; MacKerell, A. D.; Roux, B., Drude Polarizable Force Field for Molecular Dynamics Simulations of Saturated and Unsaturated Zwitterionic Lipids. *J Chem Theory Comput* **2017,** *13* (9), 4535-4552.

40. Cieplak, P.; Dupradeau, F. Y.; Duan, Y.; Wang, J., Polarization effects in molecular mechanical force fields. *Journal of physics. Condensed matter : an Institute of Physics journal* **2009,** *21* (33), 333102.

41. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr.; Head-Gordon, M.; Clark, G. N.; Johnson, M. E.; Head-Gordon, T., Current status of the AMOEBA polarizable force field. *J Phys Chem B* **2010,** *114* (8), 2549-64.

42. Duan, Y.; Cai, Q.; Cieplak, P.; Wang, J.; Hsieh, M.; Luo, R., Development of polarizable force field for molecular mechanical calculations of proteins by quantitative comparison with experimental data. *Abstracts of Papers of the American Chemical Society* **2012,** *243*.

43. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A., Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. *J Comput Chem* **2002,** *23* (16), 1515-31.

44. Friesner, R. A., Modeling Polarization in Proteins and Protein-ligand Complexes: Methods and Preliminary Results. *Adv Protein Chem* **2005,** *72*, 79-104.

45. Patel, S.; Brooks, C. L., 3rd, CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J Comput Chem* **2004,** *25* (1), 1-15.

46. Patel, S.; Brooks, C. L., Development and application of a CHARMM fluctuating charge protein force field. *Abstracts of Papers of the American Chemical Society* **2004,** *227*, U1010-U1010.

47. Patel, S.; Mackerell Jr, A. D.; Brooks Iii, C. L., CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J Comput Chem* **2004,** *25* (12), 1504-1514.

48. Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D.; Schulten, K.; Roux, B., High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD. *Journal of Physical Chemistry Letters* **2011,** *2* (2), 87-92.

49. Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., A polarizable model of water for molecular dynamics simulations of biomolecules. *Chemical physics letters* **2006,** *418* (1-3), 245-249.

50. Lopes, P. E. M.; Lamoureux, G.; Roux, B.; MacKerell, A. D., Polarizable empirical force field for aromatic compounds based on the classical drude oscillator. *J Phys Chem B* **2007,** *111* (11), 2873-2885.

51. Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D., An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chemical Reviews* **2016,** *116* (9), 4983-5013.

52. Lamoureux, G.; Roux, B., Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J Chem Phys* **2003,** *119* (6), 3025-3039.

53. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *J Chem Theory Comput* **2013,** *9* (9), 4046-4063.

54. Chialvo, A. A.; Vlcek, L., Ewald Summation Approach to Potential Models of Aqueous Electrolytes Involving Gaussian Charges and Induced Dipoles: Formal and Simulation Results. *J Phys Chem B* **2014,** *118* (47), 13658-13670.

55. Cisneros, G. A., Application of Gaussian Electrostatic Model (GEM) Distributed Multipoles in the AMOEBA Force Field. *J Chem Theory Comput* **2012,** *8* (12), 5072-5080.

56. Cisneros, G. A.; Piquemal, J. P.; Darden, T. A., Generalization of the Gaussian electrostatic model: Extension to arbitrary angular momentum, distributed multipoles, and speedup with reciprocal space methods. *J Chem Phys* **2006,** *125* (18).

57. Elking, D. M.; Cisneros, G. A.; Piquemal, J. P.; Darden, T. A.; Pedersen, L. G., Gaussian Multipole Model (GMM). *J Chem Theory Comput* **2010,** *6* (1), 190-202.

58. Elking, D. M.; Perera, L.; Duke, R.; Darden, T.; Pedersen, L. G., Atomic Forces for Geometry-Dependent Point Multipole and Gaussian Multipole Models. *J Comput Chem* **2010,** *31* (15), 2702-2713.

59. Elking, D. M.; Perera, L.; Duke, R.; Pedersen, L. G., Modeling molecular charge density with Gaussian multipoles. *Abstracts of Papers of the American Chemical Society* **2010,** *240*.

60. 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.

61. Wang, J.; Kollman, P. A., Automatic parameterization of force field by systematic search and genetic algorithms. *J Comput Chem* **2001,** *22* (12), 1219-1228.

62. Aronica, P. G. A.; Fox, S. J.; Verma, C. S., Comparison of Charge Derivation Methods Applied to Amino Acid Parameterization. *Acs Omega* **2018,** *3* (4), 4664-4673.

63. 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-Us* **1993,** *97* (40), 10269-10280.

64. 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.

65. Yildirim, A.; Wassenaar, T. A.; van der Spoel, D., Statistical efficiency of methods for computing free energy of hydration. *J Chem Phys* **2018,** *149* (14).

66. Riquelme, M.; Lara, A.; Mobley, D. L.; Verstraelen, T.; Matamala, A. R.; Vohringer-Martinez, E., Hydration Free Energies in the FreeSolv Database Calculated with Polarized Iterative Hirshfeld Charges. *J Chem Inf Model* **2018,** *58* (9), 1779-1797.

67. Boulanger, E.; Huang, L.; Rupakheti, C.; MacKerell, A. D.; Roux, B., Optimized Lennard-Jones Parameters for Druglike Small Molecules. *J Chem Theory Comput* **2018,** *14* (6), 3121-3131.

68. Matos, G. D. R.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L., Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. *J Chem Eng Data* **2017,** *62* (5), 1559-1569.

69. Zhang, J.; Tuguldur, B.; van der Spoel, D., Force Field Benchmark of Organic Liquids. 2. Gibbs Energy of Solvation. *J Chem Inf Model* **2015,** *55* (6), 1192-1201.

70. Jambeck, J. P. M.; Lyubartsev, A. P., Update to the General Amber Force Field for Small Solutes with an Emphasis on Free Energies of Hydration. *J Phys Chem B* **2014,** *118* (14), 3793-3804.

71. Rustenburg, A. S.; Dancer, J.; Lin, B.; Feng, J. A.; Ortwine, D. F.; Mobley, D. L.; Chodera, J. D., Measuring experimental cyclohexane-water distribution coefficients for the SAMPL5 challenge. *J Comput Aided Mol Des* **2016,** *30* (11), 945-958.

72. Skyner, R. E.; McDonagh, J. L.; Groom, C. R.; van Mourik, T.; Mitchell, J. B., A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Physical chemistry chemical physics : PCCP* **2015,** *17* (9), 6174-91.

73. Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Dill, K. A., Predictions of hydration free energies from all-atom molecular dynamics simulations. *J Phys Chem B* **2009,** *113* (14), 4533-7.

74. Mobley, D. L.; Guthrie, J. P., FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J Comput Aided Mol Des* **2014,** *28* (7), 711-20.

75. 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* (2), 350-358.

76. Koziara, K. B.; Stroet, M.; Malde, A. K.; Mark, A. E., Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. *J Comput Aided Mol Des* **2014,** *28* (3), 221-33.

77. Beveridge, D. L.; DiCapua, F. M., Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu Rev Biophys Biophys Chem* **1989,** *18*, 431-92.

78. 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* (7), 2695-703.

79. 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* (5), 1542-1555.

80. 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* (9), 3878-88.

81. Miao, Y.; McCammon, J. A., Unconstrained Enhanced Sampling for Free Energy Calculations of Biomolecules: A Review. *Mol Simul* **2016,** *42* (13), 1046-1055.

82. Mobley, D. L.; Klimovich, P. V., Perspective: Alchemical free energy calculations for drug discovery. *J Chem Phys* **2012,** *137* (23), 230901.

83. Yang, W., Advanced Sampling for Molecular Simulation is Coming of Age. *J Comput Chem* **2016,** *37* (6), 549-549.

84. Zheng, L. Q.; Chen, M. G.; Yang, W., Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *P Natl Acad Sci USA* **2008,** *105* (51), 20227-20232.

85. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* **2015,** *10* (5), 449-61.

86. Onufriev, A.; Case, D. A.; Bashford, D., Effective Born radii in the generalized Born approximation: the importance of being perfect. *J Comput Chem* **2002,** *23* (14), 1297-304.

87. Wang, J. M.; Hou, T. J.; Xu, X. J., Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Curr Comput-Aid Drug* **2006,** *2* (3), 287-306.

88. 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* (1), 69-82.

89. 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* (5), 866-77.

90. Chen, F.; Sun, H.; Wang, J.; Zhu, F.; Liu, H.; Wang, Z.; Lei, T.; Li, Y.; Hou, T., Assessing the performance of MM/PBSA and MM/GBSA methods. 8. Predicting binding free energies and poses of protein-RNA complexes. *RNA* **2018,** *24* (9), 1183-1194.

91. 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* (28), 8408-21.

92. Lee, J. Y.; Kinch, L. N.; Borek, D. M.; Wang, J.; Wang, J.; Urbatsch, I. L.; Xie, X. S.; Grishin, N. V.; Cohen, J. C.; Otwinowski, Z.; Hobbs, H. H.; Rosenbaum, D. M., Crystal structure of the human sterol transporter ABCG5/ABCG8. *Nature* **2016,** *533* (7604), 561-4.

93. Shao, Z. H.; Yin, J.; Chapman, K.; Grzemska, M.; Clark, L.; Wang, J. M.; Rosenbaum, D. M., High-resolution crystal structure of the human CB1 cannabinoid receptor. *Nature* **2016,** *540* (7634), 602-+.

94. Wray, R.; Iscla, I.; Gao, Y.; Li, H.; Wang, J. M.; Blount, P., Dihydrostreptomycin Directly Binds to, Modulates, and Passes through the MscL Channel Pore. *PLoS biology* **2016,** *14* (6).

95. Suno, R.; Kimura, K. T.; Nakane, T.; Yamashita, K.; Wang, J. M.; Fujiwara, T.; Yamanaka, Y.; Im, D.; Horita, S.; Tsujimoto, H.; Tawaramoto, M. S.; Hirokawa, T.; Nango, E.; Tono, K.; Kameshima, T.; Hatsui, T.; Joti, Y.; Yabashi, M.; Shimamoto, K.; Yamamoto, M.; Rosenbaum, D. M.; Iwata, S.; Shimamura, T.; Kobayashi, T., Crystal Structures of Human Orexin 2 Receptor Bound to the Subtype-Selective Antagonist EMPA. *Structure* **2018,** *26* (1), 7-+.

96. Yin, J.; Chapman, K.; Clark, L. D.; Shao, Z.; Borek, D.; Xu, Q.; Wang, J.; Rosenbaum, D. M., Crystal structure of the human NK1 tachykinin receptor. *Proc Natl Acad Sci U S A* **2018,** *115* (52), 13264-13269.

97. Wray, R. I., I.; Kovacs, Z.; Wang, J.; Blount, P. , Novel compounds that specifically bind and modulate MscL: insights into channel gating mechanisms. *FASEB Journal* **2019**, Article ASAP

98. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **2010,** *31* (4), 671-90.

99. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J Chem Theory Comput* **2016,** *12* (1), 281-96.

100. Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M., Uff, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations. *Journal of the American Chemical Society* **1992,** *114* (25), 10024-10035.

101. Shi, S. H.; Yan, L.; Yang, Y.; Fisher-Shaulsky, J.; Thacher, T., An extensible and systematic force field, ESFF, for molecular Modeling of organic, inorganic, and organometallic systems. *J Comput Chem* **2003,** *24* (9), 1059-1076.

102. Li, P. F.; Song, L. F.; Merz, K. M., Parameterization of Highly Charged Metal Ions Using the 12-6-4 LJ-Type Nonbonded Model in Explicit Water. *J Phys Chem B* **2015,** *119* (3), 883-895.

103. Li, P. F.; Song, L. F.; Merz, K. M., Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J Chem Theory Comput* **2015,** *11* (4), 1645-1657.

104. Peters, M. B.; Yang, Y.; Wang, B.; Fusti-Molnar, L.; Weaver, M. N.; Merz, K. M., Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J Chem Theory Comput* **2010,** *6* (9), 2935-2947.

105. Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T., Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J Chem Theory Comput* **2016,** *12* (8), 3926-3947.

106. 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* (2), 132-146.

107. 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* (16), 1623-41.

108. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926.

109. 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>.

110. 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* (11), 927-944.

111. Silva, V. D.; Cargnelutti, M. T.; Giesel, G. M.; Palmieri, L. C.; Monteiro, R. Q.; Verli, H.; Lima, L. M. T. R., Structure and Behavior of Human alpha-Thrombin upon Ligand Recognition: Thermodynamic and Molecular Dynamics Studies. *PLoS One* **2011,** *6* (9).

112. Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.; Mobley, D. L., Lead optimization mapper: automating free energy calculations for lead optimization. *J Comput Aided Mol Des* **2013,** *27* (9), 755-70.

113. Klimovich, P. V.; Mobley, D. L., A Python tool to set up relative free energy calculations in GROMACS. *J Comput Aided Mol Des* **2015,** *29* (11), 1007-14.

114. Wilson, D. P.; Wan, Z. K.; Xu, W. X.; Kirincich, S. J.; Follows, B. C.; Joseph-McCarthy, D.; Foreman, K.; Moretto, A.; Wu, J.; Zhu, M.; Binnun, E.; Zhang, Y. L.; Tam, M.; Erbe, D. V.; Tobin, J.; Xu, X.; Leung, L.; Shilling, A.; Tam, S. Y.; Mansour, T. S.; Lee, J., Structure-based optimization of protein tyrosine phosphatase 1B inhibitors: from the active site to the second phosphotyrosine binding site. *J Med Chem* **2007,** *50* (19), 4681-98.

115. Homeyer, N.; Stoll, F.; Hillisch, A.; Gohlke, H., Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context. *J Chem Theory Comput* **2014,** *10* (8), 3331-3344.

116. Homeyer, N.; Gohlke, H., Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol Inform* **2012,** *31* (2), 114-22.

117. Hansen, N.; van Gunsteren, W. F., Practical Aspects of Free-Energy Calculations: A Review. *J Chem Theory Comput* **2014,** *10* (7), 2632-47.

118. Ashida, T.; Kikuchi, T., Overview of binding free energy calculation techniques for elucidation of biological processes and for drug discovery. *Med Chem* **2015,** *11* (3), 248-53.

119. Wang, J.; Hou, T., Develop and test a solvent accessible surface area-based model in conformational entropy calculations. *J Chem Inf Model* **2012,** *52* (5), 1199-212.

120. Spackova, N.; Cheatham, T. E., 3rd; 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* (7), 1759-69.

121. Sun, H.; Li, Y.; Shen, M.; Tian, S.; Xu, L.; Pan, P.; Guan, Y.; Hou, T., Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Physical chemistry chemical physics : PCCP* **2014,** *16* (40), 22035-45.

122. Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M., Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J Comput Chem* **2010,** *31* (4), 797-810.

123. Lindstrom, A.; Edvinsson, L.; Johansson, A.; Andersson, C. D.; Andersson, I. E.; Raubacher, F.; Linusson, A., Postprocessing of docked protein-ligand complexes using implicit solvation models. *J Chem Inf Model* **2011,** *51* (2), 267-82.

124. Steinbrecher, T.; Case, D. A.; Labahn, A., A multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase. *J Med Chem* **2006,** *49* (6), 1837-44.

125. Thompson, D. C.; Humblet, C.; Joseph-McCarthy, D., Investigation of MM-PBSA rescoring of docking poses. *J Chem Inf Model* **2008,** *48* (5), 1081-91.

126. Zhang, X.; Wong, S. E.; Lightstone, F. C., Toward fully automated high performance computing drug discovery: a massively parallel virtual screening pipeline for docking and molecular mechanics/generalized Born surface area rescoring to improve enrichment. *J Chem Inf Model* **2014,** *54* (1), 324-37.

127. Greenidge, P. A.; Kramer, C.; Mozziconacci, J. C.; Sherman, W., Improving docking results via reranking of ensembles of ligand poses in multiple X-ray protein conformations with MM-GBSA. *J Chem Inf Model* **2014,** *54* (10), 2697-717.

128. Zhu, T.; Lee, H.; Lei, H.; Jones, C.; Patel, K.; Johnson, M. E.; Hevener, K. E., Fragment-based drug discovery using a multidomain, parallel MD-MM/PBSA screening protocol. *J Chem Inf Model* **2013,** *53* (3), 560-72.

129. Karami, M.; Jalali, C.; Mirzaie, S., Combined virtual screening, MMPBSA, molecular docking and dynamics studies against deadly anthrax: An in silico effort to inhibit Bacillus anthracis nucleoside hydrolase. *J Theor Biol* **2017,** *420*, 180-189.

130. Wang, J.; Kang, X.; Kuntz, I. D.; Kollman, P. A., Hierarchical database screenings for HIV-1 reverse transcriptase using a pharmacophore model, rigid docking, solvation docking, and MM-PB/SA. *J Med Chem* **2005,** *48* (7), 2432-44.

131. Horta, B. A.; Merz, P. T.; Fuchs, P. F.; Dolenc, J.; Riniker, S.; Hunenberger, P. H., A GROMOS-Compatible Force Field for Small Organic Molecules in the Condensed Phase: The 2016H66 Parameter Set. *J Chem Theory Comput* **2016,** *12* (8), 3825-50.

132. OpenSMILES specification. <http://opensmiles.org/opensmiles.html> (accessed Feb 01, 2019).

133. Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K., Escaping Atom Types in Force Fields Using Direct Chemical Perception. *J Chem Theory Comput* **2018,** *14* (11), 6076-6092.

134. Bayly, C.; McKay, D.; Truchon, J. An Informal AMBER Small Molecule Force Field: Parm@Frosst. <http://www.ccl.net/cca/data/parm> (accessed Feb 01, 2019).

135. Schauperl, M.; Nerenberg, P.; Jang, H.; Wang, L.-P.; Gilson, M. K., Electrostatics: Fixed Point Charges and Polarization. In *Open Force Field Consortium Workshop*, 2019.