The Open Force Field Consortium: Project Overview

David L. Mobley[1,2], **Christopher I. Bayly**[3], **John D. Chodera**[4], **Michael R. Shirts**[5], **Michael K. Gilson**[6], **Lee-Ping Wang**[7], **Daniel G. A. Smith**[8]

[1]Department of Pharmaceutical Science, University of California, Irvine CA 92697; [2]Department of Chemistry, University of California, Irvine CA 92697; [3]OpenEye Scientific Software, Santa Fe NM 87507; [4]Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York NY 10065; [5]Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, CO 80309; [6]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego; [7]Department of Chemistry, University of California, Davis; [8]The Molecular Sciences Software Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060

***For correspondence:** dmobley@mobleylab.org (DLM)

Abstract Molecular mechanics force fields have diverse applications in pharmaceutical design and discovery. A key factor in determining their utility is the accuracy with which they model molecular interactions and the difficulty in predicting how reliably they can predict observables of interest, such as binding affinities. While a number of biomolecular force fields are currently in widespread use, their accuracies are still unsatisfactory in many cases. The pace of their development has been hindered by legacy constraints or proprietary restrictions, and they currently lack the ability to prospectively assess confidence in their predictions. The Open Force Field Consortium aims to solve these and other significant problems by developing a new generation of open force fields, along with the open software infrastructure and open datasets to advance the field and accelerate progress.

Specifically, the Consortium aims to (1) engineer a modern, open, sustainable, extensible, and well-supported framework for automated force field improvement and application; (2) use this to release rapid iteratively improved versions of an AMBER-compatible small molecule force field we have developed to take advantage of modern cheminformatics; (3) produce entirely new comprehensive force fields that break free of legacy accuracy limitations while maintaining compatibility with existing simulation software, providing dramatically improved accuracy for modeling predictions in diverse applications ranging from predictions of binding affinity, selectivity, and drug resistance, to partitioning, solubility, kinetics, and other properties; and (4) work closely with industry partners to ensure the development path follows that most relevant to R&D needs.

Here we present a project overview, covering key aspects of planned development efforts.

1 Introduction

Molecular simulations and modeling based on all-atom molecular mechanics force fields are a key component of modern pharmaceutical drug discovery workflows. Their use has grown considerably given the widespread success of free energy methods in predictive design settings in recent years [1, 6, 6–9, 27, 38]. These force fields, which describe the energy of a system as a function of the positions, provide simple approximations to the underlying physics in a way which can provide a reasonable balance of speed versus accuracy and achieve considerable predictive power that has been difficult for simpler models to achieve.

Current force fields are both wildly successful and very limited [27]—they have ushered in a new era where we attempt quantitative biomolecular design, using simulations and modeling to guide lead optimization [1, 8, 27] and predictions of other properties such as solvation [13, 23], partition and distribution coefficients [3], solubility [11, 24, 26], selectivity [2] and drug resistance. Failure cases are well known however and accuracy

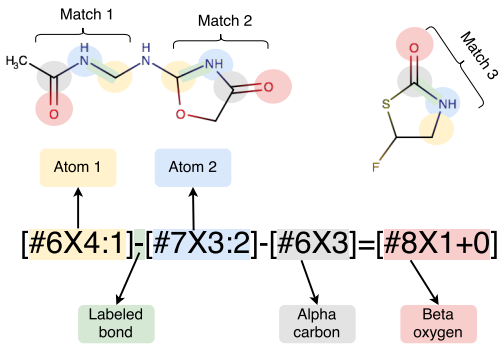


Figure 1. Using the SMIRKS chemical query language for direct chemical perception. SMIRKS allows for direct chemical perception via chemical substructure searches; here, a single SMIRKS pattern recognizes three different substructures in two different molecules (Match 1, 2, and 3), with matches shown by color coding of the atoms/SMIRKS patterns involved. Here, the relevant pattern is a carbon with four connections (`[#6X4:1]` (yellow) single bonded (–) to a trivalent nitrogen (`[#7X3:2]`, blue) which is single-bonded to a trivalent carbon (`[#6X3]`, gray) which itself is double bonded (=) to a neutral oxygen with a single connected atom (`#8X1+0]`, red). The first carbon and nitrogen are singled out for special treatment by having numerical atom labels (`:1` and `:2`) assigned to them, in this case because the SMIRKS pattern would be used to assign a bond parameter to the bond connecting the two labeled atoms. The bond connecting these labeled atoms (green) is singled out for special treatment because it connects two labeled atoms. Specifically, the force field format we introduce here allows us to use this SMIRKS pattern to refer specifically to parameters for the bond connecting these two labeled atoms.

is not yet what is needed [27], in part because the core of present day force fields (the nonbonded terms) still primarily dates from work done in the 1980s [17] and 1990s [16]. Human expertise, worked out over years and decades, has played a key role in developing present-day force fields and it is difficult to impossible for academics to obtain traditional grant funding to improve force fields; improvements are typically limited adjustments rather than entirely improved force fields [14, 15, 20, 25, 30]. Thus there are relatively few new force field efforts, and even those typically take a different approach (such as a different functional form, like AMOEBA or the Drude efforts) because of the difficulty of obtaining funding to improve existing force fields – which do still have clear room for improvement [12, 22]. Even COMPASS II and OPLS3 – which both involved significant commercial development efforts – both are an extension of earlier work rather than an overall refitting [14, 29].

Our vision is to remove much of the human effort and expertise involved in force field development, allowing *automated* development of new, improved force fields from the ground up so we can take advantage of the wealth of new data and computer power we now have rather than being limited by a foundation built 30 years ago.

2 Objectives

2.1 Overall goals

The Open Force Field Consortium aims to:

1. **Develop an open, scalable, extensible toolkit for automatically parameterizing and using force fields.** We aim to build a modern toolkit for improving parameters based on QM and physical property data (including refinement based on in-house datasets), applying parameters to biomolecular systems, and converting these for use in popular simulation packages.
2. **Generate/curate open datasets necessary for producing high-accuracy biomolecular force fields.** We aim to build an open quantum chemical fragment dataset for large regions of drug like chemical space, compile high-quality physical property datasets necessary for moving past current accuracy limitations (working with pharma partners to collect and compile such data as needed) and, as necessary, conduct novel targeted experimental measurements to provide missing information such as binding thermodynamics, liquid mixture densities/enthalpies of mixing, and logD/logP measurements.

3. **Release systematically-improved force fields on a rapid timecycle.** We aim to parameterize and release new force fields on a regular basis along with software/datasets used to build them; incorporate industry feedback each cycle including feedback on targeted chemistries, critical failures, and industry benchmarks; and prioritize easy accuracy improvements that maintain AMBER compatibility ahead of harder ones.

Modernizing the parameter application toolchain

We seek to simplify and modernize the toolchain used to parameterize biomolecular systems; a flexible, open tool should be able to parameterize systems for use with many biomolecular simulation codes (such as Amber, CHARMM, OpenMM, gromacs, and NAMD), rather than the insular package-specific tools often available for small molecule parameterization today.

Keeping force fields current

We seek to generate living, open, versioned force fields that consistently improve over rapid release cycles. While we need more accurate and thus more useful force fields *now* we want them to improve and extend over time as our physical property and quantum chemical datasets expand to encompass new chemistries and measurements.

Making force fields extensible

While the OFFC aims to release consistently improved force fields, we want to democratize access to force field parameterization and force field science through the release of open toolkits. Anyone should be able to use these toolkits to improve force field accuracy by fitting to new physical property or quantum chemical data, as well as experiment with new functional forms, or build a family of consistent force fields from the same datasets with minimal effort. Ultimately, pharmaceutical companies or others should be able to build force fields using internal proprietary data, if they so desire, or individual researchers should be able to build force fields tailored to specific problems. We want to enable all of these goals. A key part of this is data—good force fields require good data sets to build and test them—so we want to curate high-quality data sets for force field parameterization and benchmarking.

Enabling design of biopolymers and biologics

We also want to ultimately allow consistent parameterization of small molecules and biopolymers, including allowing graceful handling of covalent inhibitors, post-translational modifications, non-natural residues, and other use cases not well treated by today's force fields.

Making predictions with confidence

Another key goal of our Consortium is to provide an estimate of the confidence of predictions made by a force field, beyond just the statistical error. These estimates should be built in to the force field, allowing quantitative systematic error predictions for binding free energies and physical properties.

2.2 How is the effort open?

The effort is open in several critical aspects that are well-suited for an academia-industry partnership and are non-traditional in the field. First, it is **open source**, with free access to the tools and force fields developed provided to everyone under permissive licenses (MIT for software, CC-BY 4.0 for data and results). Second, it utilizes **open data**, with collected and curated datasets made available to everyone. OpenFF tools may be utilized by others to build force fields with proprietary data, but the Consortium itself focuses on truly open software and force fields driven by data we make or others make publicly available. Finally, the Consortium focuses on **open science**, with development done in the open, publicly, on GitHub, and all software and force fields available and open, along with the methodology used to develop them along with all source data, etc. The goal is to allow reproducible force field development. All code, data sets and force fields are made available publicly on https://github.com/openforcefield.

3 Breaking free of legacy constraints using direct chemical perception

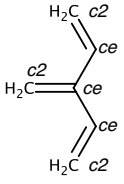
One key aspect of our effort is to reduce the dependence of force field development on human experts, and the human expertise required to build force fields. Typically, atom types are used to assign force field parameters. One problem with this approach is that atom type definitions have, to date, been crafted based on some combination of chemical intuition and analysis, without any rigorous basis for determining how many atom types are necessary and sufficient. Thus, although substantial effort has gone into parameter optimization, the science of atom typing is less developed, and does not offer confidence that current atom-types are near optimal. The replacement of human-annotated featurizations with fully automated ones has recently led to advances in both chemistry (e.g., DeepChem [39]) and image recognition (e.g., ImageNet [18]), and we see great potential in the use of such methods for force field definition and development.

Another set of problems has to do with the way atom types are used to assign parameters to molecules. In particular, most current approaches use a method we call *indirect chemical perception* to assign the parameters to most or all of their energy terms. Indirect chemical perception involves the assignment of a single identifier—an atom type—to each atom in a molecule, based on its local chemical environment, and the use of these identifiers to look up most or all of the required force field parameters (e.g., Lennard-Jones, bond-stretch, angle-bend, and torsions) in parameter tables. Thus, in indirect chemical perception, the set of atom types contains all the information required to parameterize a system. This approach leads to a number of difficulties.

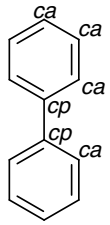
First, the creation of a new atom type can lead to undesired proliferation of other force field terms. For example, if new hydration free energy data leads one to add a new atom type to handle Lennard-Jones interactions better, this addition immediately requires one to create and assign parameters to all the bond, angle, and torsion types involving this new atom type. Often this leads to guessing or copying from existing "parent" parameters, without any solid basis for these choices. Furthermore, this addition of possibly needless parameters dramatically exacerbates the curse of dimensionality when one goes to optimize the parameters. In principle, this problem might be addressed by constraining parent and child parameters to be equal, but this is not standard practice, and, since the need for such constraints is not encoded in the force field files, such linkages are easily lost over time, leading back to the curse of dimensionality.

Furthermore, when indirect chemical perception is used, the need to differentiate bond-stretch or bond-torsion parameters can force the creation of new, otherwise needless, atom types, which must either be assigned Lennard-Jones parameters based on parent atom types or left as free parameters to be optimized. For example, in 3-methylenepenta-1,4-diene (Figure 2), the carbon atoms are all assigned the same Lennard-Jones parameters in GAFF, consistent with the fact that they are in similar chemical environments, with one double bond and two single bonds to another carbon or hydrogen atom. However, new atom types (see figure) had to be introduced, merely to encode the fact that some bonds are single and others double, since the bond-stretch parameters are inferred from the atom types. Figure 2b,c, and d [32, 33] shows additional cases in which chemically similar atoms with identical Lennard-Jones parameters had to be assigned different atom types to allow assignment of different bond-stretch parameters to single and double bonds. Similarly, in biphenyls and related molecules, atom typing makes it difficult to recognize which bonds should be rotatable without introducing new atom types. These additional atom types result, in turn, in a further proliferation of additional bond, angle and torsion types. To apply automated parameterization machinery when many similar parameters exist, such as the 16 sets of Lennard-Jones parameters for carbon in GAFF/GAFF2 which only have three distinct values [31]), a human expert would have to designate which parameters should be constrained to be identical versus which should be allowed to be distinct.

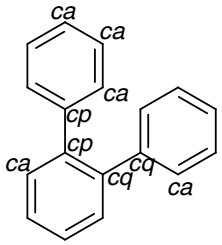
In addition to forcing the proliferation of parameters, indirect chemical perception can drive the occurrence of errors. For example, while the biphenyl cases just discussed have received careful attention to avoid incorrectly treating the bond between bridgehead carbons, this approach required a human expert to identify and solve problem cases and can fail for other systems which have not received such careful attention. For example, a similar scenario occurs for bonds between the GAFF/GAFF2 types `cc`–`cc`, `cc`–`cd`, and `cd`–`cd`, which have identical Lennard-Jones parameters and are used for carbons in non-pure aromatic systems (Figure 2e). Here, GAFF/GAFF2 give the torsions involving these single bonds a barrier height of



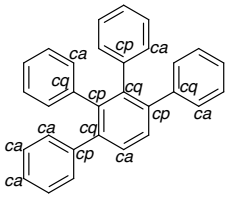
(a) 3-methylenepenta-1,4-diene



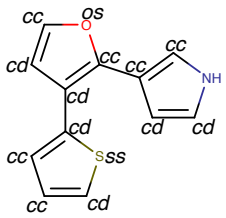
(b) biphenyl



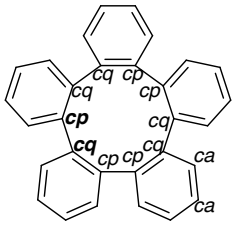
(c) 1,2-diphenylbenzene



(d) 1,2,3,4-tetra–phenylbenzene



(e) Bridgehead problem case



(f) Ring problem case

Figure 2. Even simple molecules can present challenging cases for indirect chemical perception via atom typing. Atom typing these molecules in a way which can lead to proper assignment of parameters can be challenging because of bonding patterns. Indicated atom types are GAFF/GAFF2 types. In **(a)**, all carbons are sp^2 yet are connected by alternating single and double bonds, forcing introduction of the `ce` atom type for the inner sp^2 carbons to allow single and double carbon-carbon bonds to be distinguished. In **(b)**, the bridgehead aromatic carbons in biphenyl must have a single bond joining them, forcing introduction of the `cp` atom type which is identical to the normal aromatic carbon (`ca`) except that `cp`–`cp` bonds are single and thus rotatable. Introduction of an additional phenyl ring, in 1,2-diphenylbenzene **(c)** further complicates matters since, with only the `ca` and `cp` types there would be a `cp`–`cp` (single) bond within the lower aromatic ring. GAFF/GAFF2 address this by introduction of a new type, `cq`, which is identical to `cp` and `ca` so that `cp`–`cp` bonds are single, `cq`–`cq` bonds are single, but `cq`–`cp` bonds (and bonds involving `ca` of any sort) are aromatic. This scheme can in principle handle even larger molecules like 1,2,3,4-tetraphenylbenzene **(d)** though, as discussed in the text, the massive proliferation of torsional parameters resulting from the numerous atom types employed leads to considerable potential for human error. However, this approach of introducing additional duplicate atom types requires a human expert to notice when such types will be needed. **(e)** shows a case where bridgehead bonds between five membered rings are typed as `cd`–`cd` or `cc`–`cc`, the same as bonds within aromatic rings. Thus, GAFF/GAFF2 make these bonds non-rotatable, even though they are single bonds. This could presumably be fixed in the same manner as the issues of (b) and (c) via introduction of additional atom types. However, some molecules, like **(f)**, are impossible to atom type properly in this framework [33]; incorrect atom types are shown in boldface and will lead to misassignment of parameters.

16.00 kcal/mol, which is identical to the aromatic bonds within the five membered rings and even higher than the 14.5 kcal/mol barrier height used for aromatic bonds in biphenyl. Thus, unexpected bridgehead atoms result in a rotatable single bond being treated as aromatic. Finally, for some molecules, it appears that consistent typing is impossible to achieve with indirect chemical perception [33]; while introducing new atom types (`cp` and `cq`) can resolve most problems in the biphenyl series of Figure 2b-d, with sufficiently complex molecules, it becomes impossible to avoid assigning a single bond as aromatic (Figure 2f).

4 Preliminary progress toward next-generation force fields

4.1 SMIRNOFF: A new force field format using direct chemical perception

Recently, we introduced a new force field specification format called SMIRKS Native Open Force Field (SMIRNOFF) which uses direct chemical perception, instead processing the molecular graph, with its varied bond types, and uses this to independently assign Lennard-Jones, bond, angle, and torsion parameters based on the local chemical environments of the respective atoms, bonds, angles, and torsions in the molecule. This process works via substructure queries and easily accommodates the addition of a new atom with distinct Lennard-Jones parameters without driving the creation of additional bonded types. The format is hierarchical

and last-one wins, simplifying extensibility (Figure 3). This format led to development of a complete general force field, called SMIRNOFF99Frosst, derived from AMBER parm99 and Merck's parm@Frosst [4], without any parameter refitting. We are able to show that SMIRNOFF99Frosst covers comparable chemical space to GAFF/GAFF2, with similar accuracy [21], in initial tests, meaning that it provides an excellent starting point for further optimization and parameter refitting as we discuss in our plans here.

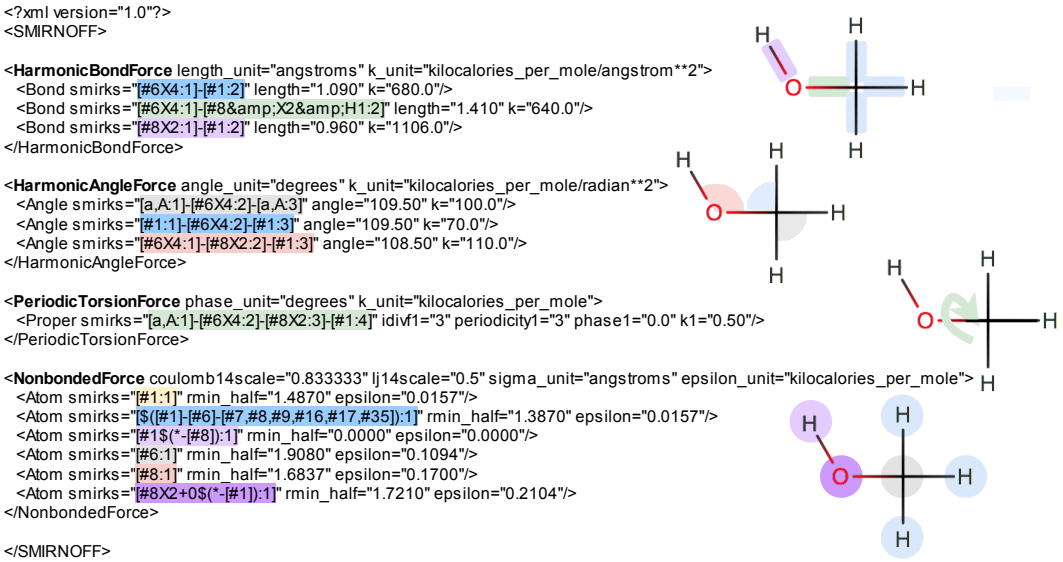
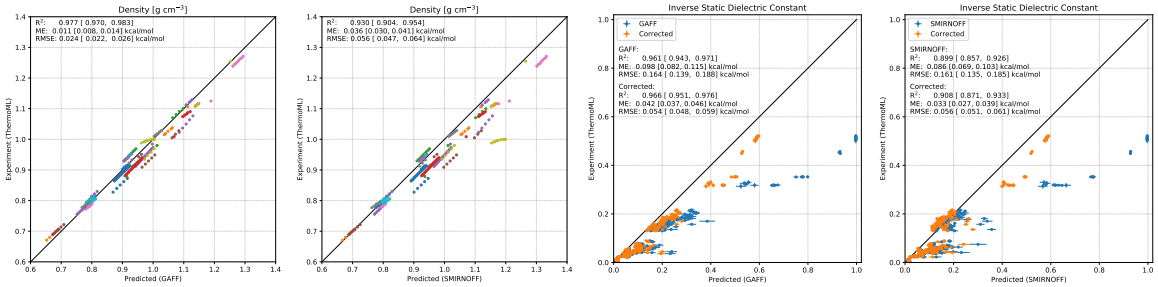


Figure 3. Application of a SMIRNOFF to methanol. Shown at left is an excerpt of the SMIRNOFF force field format XML for a test force field for our AlkEthOH test set, and the representations of methanol at right illustrate how SMIRNOFF provides the necessary force field parameters for methanol. For each force type or section in the XML (boldface), SMIRNOFF loops over the needed force terms for the molecule, and finds the last (most specialized) SMIRKS match in the XML, applying the parameters indicated there to the relevant force term. SMIRKS patterns are color coded, as are instances of the applied parameters in methanol. Force types run in sections from top to bottom, and for each of the relevant blocks (bond, angle, torsion, and nonbonded forces) here, a separate diagram is shown. For example, in the **HarmonicBondForce** section at top, for the O-H bond in methanol, the SMIRKS pattern `[#8X2:1-[#1:2]` (pink) matches and thus is selected and applied (pink highlighted bond, top right). Parameter assignments are color-coded by the element(s) involved – by the primary or central atom in the case of NonbondedForce and HarmonicAngleForce parameters (except when there is redundancy, in which case the second occurrence gets a color associated with non-central atoms), and by the central two atoms in the case of HarmonicBondForce and PeriodicTorsionForce parameters. Gray is used for carbon, red for oxygen, light green for oxygen-carbon, yellow for hydrogen, pink for hydrogen-oxygen, and light blue for hydrogen-carbon. Parameterization of some symmetry-equivalent angles is omitted in this diagram for simplicity. The hierarchical nature of parameterization is also shown here; for example, the **NonbondedForce** section contains a generic hydrogen SMIRKS pattern (`[#1:1]`, yellow) but this is overridden in the case of the hydroxyl hydrogen by a more specialized pattern (soft pink). Likewise, the generic oxygen (`[#8:1]`, red) is overridden by the more specialized neutral hydroxyl oxygen SMIRKS (magenta). Not shown here are sections for bond charge corrections (BCCs) and constraints, though specifications for these are provided at https://github.com/open-forcefield-group/openforcefield/blob/master/The-SMIRNOFF-force-field-format.md. It is also worth noting that the XML format is unit-bearing, allowing handling of units utilized by various different force fields.

4.2 SMIRNOFF99Frosst: A new force field with high coverage of chemical space

We examined the accuracy of our prototype SMIRNOFF force field, and its coverage of chemical space, as discussed in our preprint [21], and we find that it has better coverage of chemical space than other public force fields and comparable accuracy to the widely-used GAFF force field – even without any refitting of parameters (Figure 4 and 5). At the same time, it has far fewer parameters which would require fitting—only 335 parameter lines, where each line specifies a single force field term (e.g. one type of bond-stretch or torsion); whereas the files specifying other open force fields we compared to have 3600-6800 lines. Thus, this provides a much superior starting point for further optimization, a key part of our efforts. Additionally, it fixes problems with these other force fields [21].

Portability is a key consideration; any new force field needs to be easily accessible for a wide variety of simulation packages and software. SMIRNOFF-based force fields already have portability; it is designed to set up systems for simulation in the open-source OpenMM simulation package. However, via ParmEd (http://parmed.github.io/ParmEd) and InterMol (http://intermol.readthedocs.io/en/latest/), prepared `System` objects can easily be exported for use in essentially all common molecular simulation packages [28].



(a) GAFF densities **(b)** SMIRNOFF densities **(c)** GAFF dielectrics **(d)** SMIRNOFF dielectrics

Figure 4. Densities and dielectric constants for pure solvents from GAFF and SMIRNOFF99Frosst. Shown are densities (top) and dielectric constants (bottom) for 45 compounds in 246 different conditions (near atmospheric pressure, at different temperatures), calculated as described in Beauchamp et al. [5] (data ours). GAFF data is shown in the left column ((a) and (c)) and SMIRNOFF99Frosst data is in the right column ((b) and (d)). In general accuracy is roughly comparable overall between the two force fields, with SMIRNOFF99Frosst densities having a slight additional systematic error relative to experiment but dielectric constants actually showing modestly improved errors (though slightly degraded correlation). All panels show error bars, but in the density plots error bars are typically smaller than the size of the data points. Statistics are shown in the inset on each panel, with brackets denoting 95% confidence intervals. In (a) and (b), each compound is shown in a different color, with multiple conditions represented in the same color. Of particular note is the density for flexible force field water, which is not typically used as a water model; in (b) this provides the most extreme set of outliers in SMIRNOFF99Frosst, around a predicted density of 1.2 g/mL and an actual density of 1.0 g/ml. A full list of compounds is available in the SI and at https://github.com/MobleyLab/SMIRNOFF_paper_code/blob/master/ThermoML_benchmark/results_GAFF/tables/data_with_metadata.csv

4.3 SMIRKY: Automated machine learning for chemical perception trees

A longer-term goal of our effort is to remove the dependence on human chemical intuition that underlies present-day force fields, improving automation. Currently, one place chemical intuition comes in is in the form of atom typing, deciding which chemical environments deserve separate treatment. Our SMIRNOFF99Frosst prototype force field uses chemical perception determined by a human expert, but we want to automate sampling over chemical perception so it can be determined as part of the force field fitting process. To this end, we recently developed SMARTY and SMIRKY, tools which can vary the chemical perception used in parameter assignment, so that we will ultimately be able to use these as part of the fitting process.

5 Planned development prioritizes high utility, low difficulty tasks

The Open Force Field Consortium will use these starting points to accomplish our goals of building steadily improved force fields while building infrastructure for to dramatically improve the accuracy of molecular design. Thus progress will always focus on two aspects to facilitate iterative improvements: Improving force fields, but also building infrastructure to make future cycles of improvement easier.

5.1 Refitting begins with incremental improvement of SMIRNOFF99Frosst

We already have an improved small molecule force field, SMIRNOFF99Frosst, which covers a vast chemical space with fewer than 350 lines of parameters, as opposed to the thousands of lines needed by GAFF and other modern force fields, providing an excellent starting point for further development, especially in view of its relative simplicity and reasonable accuracy. The initial phase of our effort will begin with re-parameterizing SMIRNOFF-family force fields from this starting point, beginning with the easiest tasks which will also bring the greatest short-term benefits (Figure 6). Short-term tasks will result in clear benefits

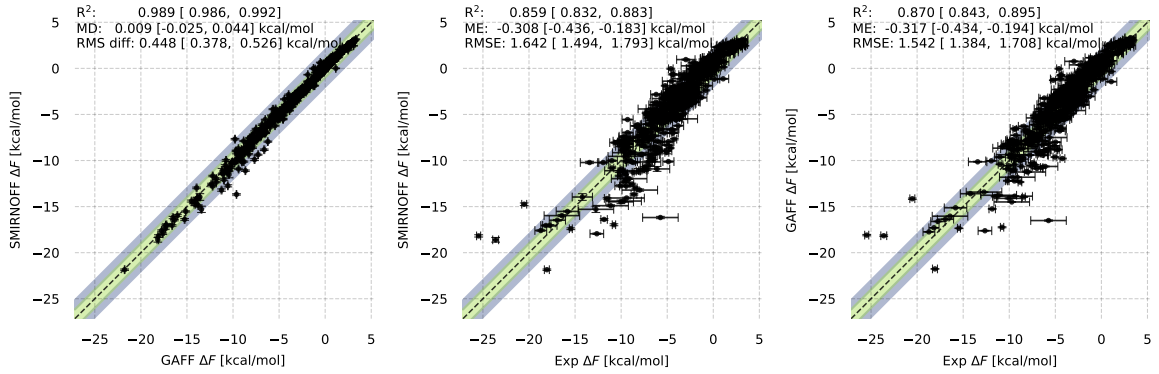


Figure 5. Hydration free energies for FreeSolv from GAFF and SMIRNOFF99Frosst. Shown are computed hydration free energies on the FreeSolv set for SMIRNOFF99Frosst from this work and previous work [10] with GAFF. The **left panel** shows SMIRNOFF99Frosst versus GAFF (left), the **middle panel** shows SMIRNOFF99Frosst versus experiment, and the **right panel** shows GAFF versus experiment. Statistics, with bootstrapped uncertainties representing 95% confidence intervals, are shown at the top of each panel. Here, the mean difference between SMIRNOFF99Frosst and GAFF is statistically indistinguishable from zero (left panel) though there is a significant discrepancy based on the RMS difference. However, compared to experimental values, the coefficient of determination R^2, mean error, and RMS error for GAFF and SMIRNOFF are within confidence intervals of one another (middle and right panels) indicating that the performance of SMIRNOFF99Frosst is essentially comparable on this dataset.

for force field accuracy; our starting point already has accuracy comparable to leading existing open force fields (e.g. GAFF and GAFF2) without any refitting, so this refitting should result in significant improvements given the improved data and underlying infrastructure being used.

5.2 We plan two force field generations

Initial efforts (for Generation 1) will focus on refitting small molecule parameters while maintaining compatibility with existing AMBER-family biopolymer force fields in order to deliver improved accuracy for small molecule interactions as rapidly as possible. To proceed further will require a full fitting of all Lennard-Jones parameters (start of Generation 2), which will necessarily break compatibility with the AMBER-family biopolymer force fields. The consortium, under the guidance and direction of the advisory board, will work with the established biopolymer force field groups, for example the group of Simmerling et al (FF14SB), to ensure compatibility with these force fields and planned replacements.

6 The project roadmap involves several interlocking lines of development

Overall, our project roadmap involves multiple lines of development which are designed to provide regular progress and iterative improvement, but also lay the groundwork for future innovations. Figure 8 shows the tentative overall project plan, along with key actions involved in each stage. Stages are labeled for reference later in Section 9. The are key opportunities for industry involvement and feedback at many points throughout the process, most notably in assisting with benchmarking of force fields, in identifying and prioritizing areas of chemistry for improvement, and even supplying molecules/chemistry (at least fragments) for inclusion in our QM database.

We plan to develop improved versions of SMIRNOFF99Frosst using experimental data primarily to tune nonbonded interactions, and quantum chemical data to tune bond-stretch, bond-angle, and torsional terms. We envision two aspects of this process: The first is a straightforward adjustment of parameters associated with the existing set of SMIRKS strings, with some human-curated changes to chemical perception and number of parameters; this will form the primary basis for Generation 1 force fields. Note that, because SMIRNOFF99Frosst has far fewer adjustable parameters than, e.g., GAFF and parm99@Frosst, it should be more amenable to automated parameter fitting methods, such as ForceBalance [19, 34–37]. Lee-Ping

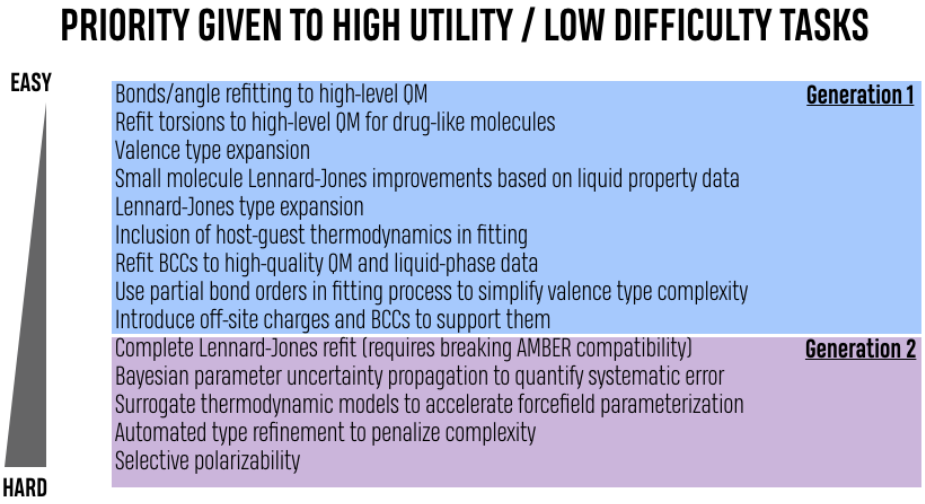


Figure 6. The Consortium prioritizes straightforward, high-impact tasks. The initial phase of work (**blue**) will focus on refitting small molecule force fields which maintain compatibility with AMBER-family protein force fields and focus on small molecules, beginning with the most straightforward tasks that will result in the highest impact such as refitting of bonded parameters to high-level quantum calculations. Limited Lennard-Jones refitting to condensed-phase properties will be done in the middle of this phase. Each aspect of fitting will be automated as completed, allowing it to easily be revisited when needed; some adjustments (such as the LJ type expansion) will require redoing earlier parts of the process (such as torsion fitting) but this will become trivial via automation, given our database infrastructure. A second generation of force fields (**purple**) will involve a complete refit of non-bonded parameters, breaking compatibility with AMBER-family protein force fields and meaning that we will begin at that point to also provide biomolecular force fields. This refit will leverage earlier advances in automation, allowing a corresponding refit of torsional and bonded parameters.

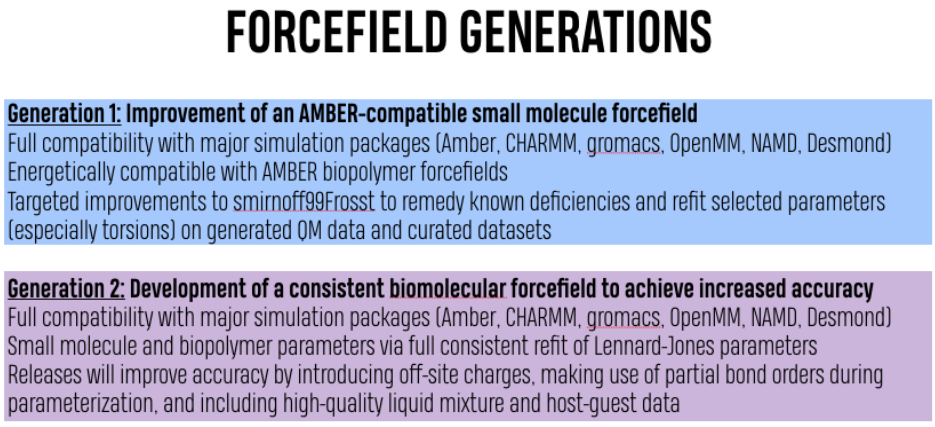


Figure 7. Two force field generations are planned. Initial work (**blue**) on Generation 1 small molecule force fields will maintain compatibility with AMBER-family protein force fields, whereas Generation 2 (**purple**) force fields will involve a full refit of non-bonded parameters, meaning that force fields will need to be used with a consistent set of biopolymer parameters which will be developed in collaboration with existing biopolymer force field experts. Both generations of force fields will maintain compatibility with typical molecular simulation packages.

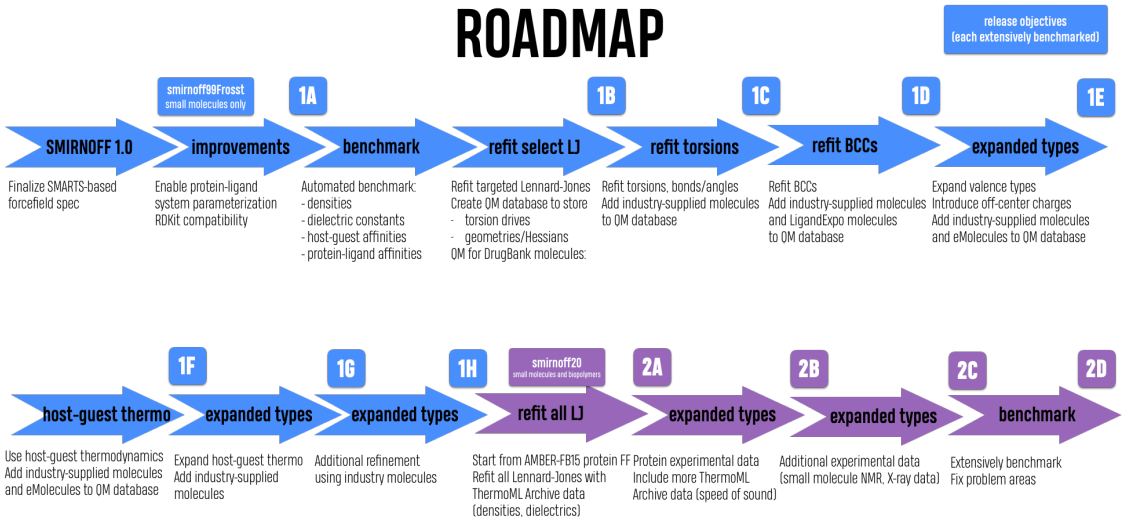


Figure 8. The Consortium aims to produce iteratively refined force fields on a rapid release cycle. Plans for Generation 1 (blue) and Generation 2 (purple) force fields, accompanied by major steps taken, benchmarking, types of data used, and major actions involved. Number/letter codes (top) mark major force field releases which will be rapid and regular. The exact pace of releases depends on the scope of funding as further detailed below.

Wang, the create of ForceBalance, is one of the Consortium investigators on this proposal and is an expert at automating this type of fitting.

The second aspect of our effort is to include automated optimization of the SMIRKS chemical perception trees for assigning atom, bond, angle and torsional terms, based on the SMARTY/SMIRKY framework. Our goal is to escape the traditional reliance on idiosyncratic chemical insight in assigning atom types. This process, which has not hitherto been amenable to automation, will be facilitated by our parallel development of the SMIRKY tool, which automatically carries out stochastic sampling and optimization over the SMIRKS strings used in the SMIRNOFF format [**?**]. Generation 2 force fields will draw heavily on this infrastructure.

7 Building open experimental and quantum chemical datasets

Force field development will draw extensively on both experimental and quantum chemical data (Figure 10). Existing and newly generated data will be organized into high-quality curated, versioned, DOI-identifiable datasets that will be released alongside the force field versions they were used to parameterize to guarantee reproducibility and extensibility.

7.1 Experimental data

The https://trc.nist.gov/ThermoML.html [**?**] provides a vast trove of machine-readable physical property condensed phase data vital for fitting nonbonded parameters. The OFFC has already demonstrated the utility of this data for automated benchmarking and validation of force fields [**? ?**]. We will build infrastructure to fit parameters using a wide range of physical properties stored in this database, including (but not limited to)

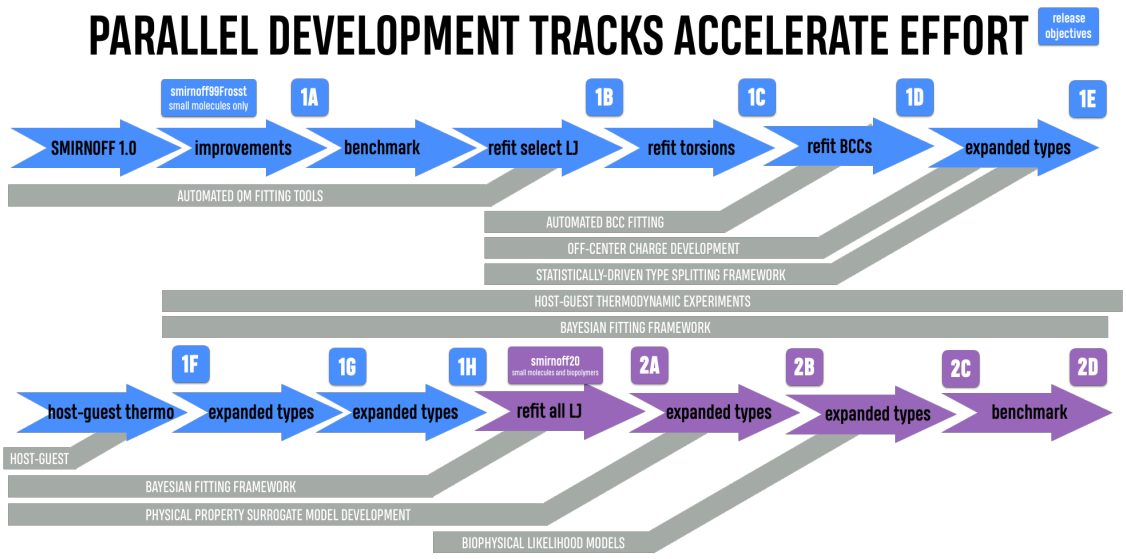


Figure 9. Parallel development tracks hide latency in force field improvement. Our effort involves several parallel development tracks to lay the groundwork for future efforts; for example, current work focuses on infrastructure for automated fitting of bonded parameters to quantum mechanical data, which will be first incorporated after our release 1B which refits select Lennard-Jones parameters. Likewise, longer-term efforts on Bayesian parameter fitting will proceed in parallel to the main track but only be incorporated for the 2A release, and similarly for several other parallel tracks.

temperature-dependent densities of pure and mixed solvents, enthalpies of mixing, transfer free energies, and other properties. Binding thermodynamics data (such as host-guest binding thermodynamics from the Gilson and Chodera laboratories, or stored in the http://bindingdb.org maintained by Gilson) will later be included, as will speed of sound data, and potentially crystallographic information such as from the CCSD.

7.2 Quantum chemical data

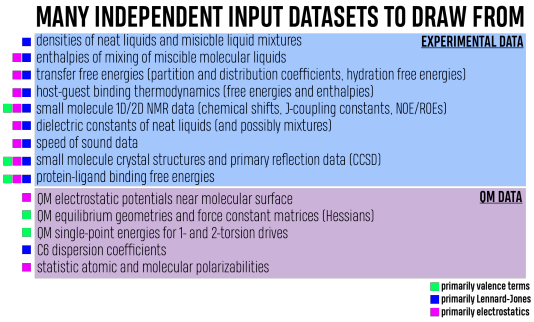
Quantum chemical data will be particularly important in parameterizing valence terms (bonds, angles, torsions) and developing improved partial charge models. Fundamentally, the effort will need to generate, store, search, and retrieve large quantities of quantum chemical data on small molecule fragments in multiple conformations.

We are partnering with the NSF-funded Molecular Sciences Software Sustainability Institute (MolSSI) to build a large-scale quantum chemical database and workflow engine which will serve for our effort, but which will also be useful to (and receive contributions from) the machine learning community. While a central open quantum chemical database will be operated for the community by MolSSI/OFFC, the software stack will be capable of being run internally by participants to generate and house quantum chemical data on private molecules that cannot be disclosed publicly.

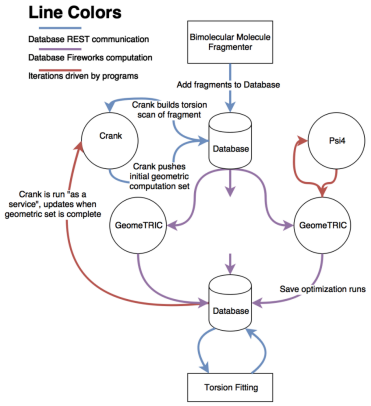
Initially, workflows will allow molecules to be provided for automated enumeration of protonation states, tautomers, and stereocenters prior to fragmentation and capping. Capped fragments will be subjected to advanced multidimensional torsion drives using generalizations of procedures used in the parameterization of AMBER15FB [**?**] to generate quantum chemical energy surfaces. Subsequent workflows will be developed to generate perturbations to other valence coordinates, electrostatic potentials, and fragment-fragment interaction energetics. Quantum chemical data will be stored in the MolSSI JSON QC schema standard.

8 More detailed development plans

As discussed in Section 6, we plan two generations of fore fields, with the first generation retaining compatability with existing AMBER-family biopolymer force fields, whereas generation 2 will mark a more substantial refit.



(a) Types of data and their impact



(b) QM infrastructure

Figure 10. We will heavily utilize both condensed-phase/experimental data and quantum mechanical data. (a) Shown are major types of data which will be utilized as part of the OpenFF Consortium, and what parameters they will primarily affect. (b) Shown is a QM database infrastructure and calculation pipeline currently under construction for torsion drives and parameter fitting; it utilizes Lee-Ping Wang's Crank torsional parameterization tool, and interfaces with a large-scale open QM database being developed together with the NSF-funded Molecular Sciences Software Sustainability Institute (MolSSI).

8.1 Detailed work planned in generation 1

Work in generation 1 will begin with bond/angle refitting to high-level QM and torsion refits to high-level QM (Figure 7), tasks which will be handled with existing infrastructure in the form of ForceBalance (for parameter optimization) and Crank (for torsion drives), both from the Wang lab, coupled with the quantum chemical dataset we are building with MolSSI (Section 7). This work will involve refitting existing parameters in our prototype SMIRNOFF99Frosst force field to improve accuracy.

Meanwhile, parallel work will laying the foundation for changing the number of valance parameters and refitting Lennard-Jones parameters. Work in the Shirts group is focusing on building hierarchical surrogate models in order to rapidly compute condensed phase properties, allowing large numbers of properties from NIST's ThermoML to be used in optimizing Lennard-Jones parameters. Work in the Mobley and Chodera groups is building infrastructure for inferring the number of necessary valence types based on sampling over chemical perception. Work in the Gilson lab is building infrastructure for including binding data in force field parameterization. At the same time we are also building in a framework for off-site charges.

These plans lead to a series of targeted SMIRNOFF force field and software versions shown in Figure 8. Specifically, our version 1A release will allow parameterization of protein-ligand systems and yield a fully open software framework, eliminating the current requirement for a license to the commercial OpenEye software to use the force fields. After extensive benchmarking of SMIRNOFF99Frosst, we will refit targeted Lennard-Jones parameters while preparing to do a large scale refit of torsions. A version 1B release will then include refits of selected Lennard-Jones parameters, then we will refit selected torsions and valance parameters yielding a 1C release. Benchmarking on industry-supplied molecules and/or fragments will yield additional material for further improvements to torsions; while those calculations are running, we will refit bond charge corrections (BCCs) based on condensed-phase properties for a 1D release. Version 1E will be the first which changes the *number* of valence types in our force field, adding new parameters only where most needed; it will also first introduce off-site charges in selected locations. 1F will include host-guest binding thermodynamics as one of the properties being fit, to improve binding free energy estimates. 1G will further expand binding thermodynamics and add additional industry supplied molecules/fragments to fitting criteria, and 1H will focus on refinement in problem cases identified in collaboration with industry. Benchmarking throughout this process (within the initiative and, to the extent desired, by industry partners in their own work) will provide continued opportunity to identify and fix problem cases.

At each stage of the process, we also focus not just on producing the next force field but building the

necessary software infrastructure to enable later versions and generations; for example, automated QM fitting tools are already being built, but will only be utilized for our 1B release 9. Similarly, infrastructure for off-center charges is already being developed but will only be used for our 1E release. Host-guest thermodyanmics are already being tested for parameterization inputs but will only be included in the 1F release, etc. This will ensure that we are able to make steady progress while also building towards larger, longer-term accuracy goals.

8.2 Detailed work planned in generation 2

Generation 2 will mark a more substantial departure from earlier force fields, as we are expecting to need a complete refit of Lennard-Jones parameters in a data-driven manner; Lennard-Jones refits in Generation 1 will only be those necessary for accuracy and will be kept relatively small without a full refit. When we reach the need for a full refit we will partner with existing biopolymer force field groups to ensure our small molecule force field is able to remain compatible with next-generation biopolymer force fields. A version 2A force field will involve a refit of *all* Lennard-Jones parameters in the force field based on condensed phase properties 8. Version 2B will introduce more parameters in a data-driven manner, and also begin including additional untapped data from ThermoML such as speed of sound data, resulting in a 2B release. We will then begin including small molecule NMR data and X-ray data in fitting data for a 2C release, and then after extensive benchmarking focus on targeted refitting to improve any problem areas for a 2D release.

8.2.1 A bit about software tools

Our work involves software infrastructure in several main areas. First is our SMIRNOFF force fields, which are applied to molecules via the general `openforcefield` python package which parses force fields in the SMIRNOFF format; this is one of our core deliverables. Parameter fitting involves several components; initially we are leaning heavily on Wang's ForceBalance software tool, which can already optimize force fields based on a fairly wide variety of target properties. Wang's Crank tool also handles torsion drives, which will be vital for torsion refitting, and these will remain part of our effort. We are also developing a general `PropertyCalculator` framework which can efficiently calculate all or almost all condensed-phase properties available in ThermoML, to allow these to ultimately be included in parameter fitting efforts. This will likely connect to ForceBalance. The Shirts group, experts on efficiency and reweighting techniques, is playing a vital role in making the property calculation framework efficient via the use of reweighting and surrogate models.

9 Personnel, project management, and financials

9.1 Management plans, governance and personnel

Overall, the Consortium will be led by a Governing Board consisting of the five initial primary investigators (PIs) (Chodera, Gilson, Mobley, Shirts, Wang) along with two industry representatives, advised by an Advisory Board (which elects the industry representatives to the Governing Board) consisting of industry representatives from pharma Partners, as separately detailed in the Consortium bylaws and research agreement. The Governing Board will manage science priorities, determine allocation of funds and specific research directions to accomplish this plan, and otherwise handle leadership of project personnel.

Project personnel are expected to be roughly equally distributed across the five sites, with some exceptions. Notably, OpenFF plans to hire a Software Scientist who to coordinate overall software architecture, serve as the main point of support contact for pharma partners, develop key infrastructure, and oversee the review of all code committed to the main source code repositories to ensure it conforms to best practices (established in coordination with MolSSI). The Software Scientist will ensure the software we produce conforms to a high level of quality and aid project personnel in coordinating their work to achieve shared goals. The Scientist will be headquartered at one of our institutions, partly dictated by personnel availability. Currently it seems likely this position will be based in Southern California, with relatively regular travel to the other participating institutions (three of these are in California) to facilitate coordination.

The remainder of project personnel will primarily be graduate students and postdocs in the different participating academic labs, with the balance between graduate students and postdocs determined by the Governing Board based on availability and expertise of suitable personnel, as well as funding-related issues. For example, postdoc hiring typically takes longer and postdoc positions may not last as long, and postdocs are substantially more expensive; graduate student training takes longer but they are typically involved longer and can be substantially cheaper. When suitable graduate students are already available in a research group and can be diverted from other projects, these may be an ideal choice for project personnel because of the quick startup time, low cost and potential long-term involvement, though when no such suitable students are available, hiring of postdocs may be preferable.

The exact number of personnel will depend on the scope of funding, which will dictate overall project velocity (Section 10). We tentatively plan that the overall project will involve one software scientist, and then differ in distribution of postdocs/graduate students depending on total initiative funding. At $750K/yr, we would expect to hire roughly two postdoctoral fellows and three graduate students (roughly one per group) in addition to the software scientist, whereas at $1.2M/yr we would expect to hire roughly five postdoctoral fellows and six graduate students (close to two per group). Personnel (except the software scientist) will be supported via fellowships from MolSSI, the initiative's coordinating intermediary.

9.2 Expertise and areas of groups involved

This project is a collaborative effort that blends synergistic capabilities of multiple laboratories. The Chodera lab (Sloan Kettering) brings extensive experience in the use of statistical inference in molecular simulations and automated laboratory measurements for biophysical data collection. The Shirts group (University of Colorado Boulder) brings experience in developing efficient reweighting techniques for rapid computation of physical properties given altered force field parameters, accurate thermodynamic models from chemical engineering, and experience with extensive benchmarks of force field performance. The Mobley lab (UC Irvine) has extensive experience in force field evaluation exercises and improvement, extensive experience in binding free energy calculations and in benchmarks and validation, and produced SMIRNOFF99Frosst. The Gilson lab (UC San Diego) brings expertise in the development and use of host-guest systems to test and improve force fields, both experimentally and computationally. The Wang group (UC Davis) brings valuable infrastructure for automated force field improvement, including the ForceBalance tool for automated parameterization. Success in this project requires close coordination among all of these researchers. Personnel will typically be headquartered with the group which best matches their research; e.g. torsion fitting work is primarily being done in the Chodera and Wang groups, valence type expansion between the Mobley and Chodera groups, and refitting of Lennard-Jones paramters in the Shirts group.

10 Timeline and project velocity

Our project timeline depends on the amount of funding/number of personnel involved, but we imagine modulating between two basic scenarios (as detailed in the preceding section) depending on amount of funding, which we describe as the Fast and Slow scenarios. The basic project plan in both scenarios is as detailed in Figure 8(a), but the timing of major milestones and releases (as designated in the figure) will differ. The following gives a tentative timeline based on a start of funding in Q3 2018, with dates marking the END of the designated quarter:

1. Fast velocity:
    - 1A: Q3 2018
    - 1B: Q1 2019
    - 1C: Q2 2019
    - 1D: Q3 2019
    - 1E: Q4 2019
    - 1F: Q1 2020
    - 1G: Q2 2020
    - 1H: Q3 2020

- 2A: Q4 2020
- 2B: Q1 2021
- 2C: Q2 2021

2. Slow velocity:

- 1A: Q3 2018
- 1B: Q3 2019
- 1C: Q1 2020
- 1D: Q2 2020
- 1E: Q4 2020
- 1F: Q1 2021
- 1G: Q2 2021
- 1H: Q4 2021
- 2A: Q2 2022
- 2B: Q4 2022
- 2C: Q2 2023

11 Conclusions

Overall, this initiative aims to democratize force field development and force field science, providing not just improved force fields but infrastructure to continue improving force fields. Already we have a prototype small molecule force field which is competitive with existing public force fields, and we have a clear plan to continue improving this via selective refitting (such as with Wang's ForceBalance) while laying the groundwork for larger-scale refitting and parameterization work which will dramatically improve accuracy by using new data and computing power to replace foundations of today's force fields which were laid in the 1980s. This promises steady progress towards dramatically improved force fields. We are excited to work with the community to help make this happen.

References

[1] **Abel R**, Wang L, Harder ED, Berne BJ, Friesner RA. Advancing Drug Discovery through Enhanced Free Energy Calculations. Acc Chem Res. 2017 Jul; 50(7):1625–1632. doi: 10.1021/acs.accounts.7b00083.

[2] **Aldeghi M**, Heifetz A, Bodkin MJ, Knapp S, Biggin PC. Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. J Am Chem Soc. 2017 Jan; 139(2):946–957. doi: 10.1021/jacs.6b11467.

[3] **Bannan CC**, Burley KH, Chiu M, Shirts MR, Gilson MK, Mobley DL. Blind Prediction of Cyclohexane–water Distribution Coefficients from the SAMPL5 Challenge. JComput Aided Mol Des. 2016 Sep; 30(11):1–18. doi: 10.1007/s10822-016-9954-8.

[4] **Bayly C**, McKay D, Truchon J, An Informal AMBER Small Molecule Force Field: Parm@ Frosst; 2010. http://www.ccl.net/cca/data/parm_at_Frosst/.

[5] **Beauchamp KA**, Behr JM, Rustenburg AS, Bayly CI, Kroenlein K, Chodera JD. Towards Automated Benchmarking of Atomistic Forcefields: Neat Liquid Densities and Static Dielectric Constants from the ThermoML Data Archive. Journal of Physical Chemistry B. 2015 Sep; 119(40):12912–12920.

[6] **Christ CD**, Binding Affinity Prediction from Molecular Simulations: A New Standard Method in Structure-Based Drug Design?; 2016. dx.doi.org/10.7490/f1000research.1112651.1.

[7] **Christ CD**, Fox T. Accuracy Assessment and Automation of Free Energy Calculations for Drug Design. J Chem Inf Model. 2014 Jan; 54(1):108–120. doi: 10.1021/ci4004199.

[8] **Cournia Z**, Allen B, Sherman W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. J Chem Inf Model. 2017 Dec; 57(12):2911–2937. doi: 10.1021/acs.jcim.7b00564.

[9] **Cui G**, Affinity Predictions with FEP+: A Different Perspective on Performance and Utility; 2016. dx.doi.org/10.7490/f1000research.1112773.1.

[10] **Duarte Ramos Matos G**, Kyu DY, Loeffler HH, Chodera JD, Shirts MR, Mobley DL. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. J Chem Eng Data. 2017 May; 62(5):1559–1569. doi: 10.1021/acs.jced.7b00104.

[11] **Duarte Ramos Matos G**, Mobley D. Challenges of the Use of Atomistic Simulations to Predict Solubilities of Drug-like Molecules. ChemRxiv. 2018 May; doi: 10.26434/chemrxiv.6241946.v1.

[12] **Fennell CJ**, Wymer KL, Mobley DL. A Fixed-Charge Model for Alcohol Polarization in the Condensed Phase, and Its Role in Small Molecule Hydration. J Phys Chem B. 2014 Jun; 118(24):6438–6446. doi: 10.1021/jp411529h.

[13] **Geballe MT**, Guthrie JP. The SAMPL3 Blind Prediction Challenge: Transfer Energy Overview. J Comput Aided Mol Des. 2012 Apr; 26(5):489–496. doi: 10.1007/s10822-012-9568-8.

[14] **Harder E**, Damm W, Maple J, Wu C, Reboul M, Xiang JY, Wang L, Lupyan D, Dahlgren MK, Knight JL, Kaus JW, Cerutti DS, Krilov G, Jorgensen WL, Abel R, Friesner RA. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput. 2016 Jan; 12(1):281–296. doi: 10.1021/acs.jctc.5b00864.

[15] **Hornak V**, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. Proteins. 2006 Nov; 65(3):712–725. doi: 10.1002/prot.21123.

[16] **Jorgensen WL**, Maxwell DS, 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 Jan; 118(45):11225–11236. doi: 10.1021/ja9621760.

[17] **Jorgensen WL**, Tirado-Rives J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. Journal of the American Chemical Society. 1988 Mar; 110(6):1657–1666. doi: 10.1021/ja00214a001.

[18] **Krizhevsky A**, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. *Advances in Neural Information Processing Systems 25* Curran Associates, Inc.; 2012.p. 1097–1105.

[19] **Laury ML**, Wang LP, Pande VS, Head-Gordon T, Ponder JW. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. J Phys Chem B. 2015 Jul; 119(29):9423–9437. doi: 10.1021/jp510896n.

[20] **Maier JA**, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015 Aug; 11(8):3696–3713. doi: 10.1021/acs.jctc.5b00255.

[21] **Mobley D**, Bannan CC, Rizzi A, Bayly CI, Chodera JD, Lim VT, Lim NM, Beauchamp KA, Shirts MR, Gilson MK, Eastman PK. Open Force Field Consortium: Escaping Atom Types Using Direct Chemical Perception with SMIRNOFF v0.1. bioRxiv. 2018 Mar; p. 286542. doi: 10.1101/286542.

[22] **Mobley DL**, Bayly CI, Cooper MD, Shirts MR, Dill KA. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. J Chem Theory Comput. 2009 Feb; 5(2):350–358. doi: 10.1021/ct800409d.

[23] **Mobley DL**, Liu S, Cerutti DS, Swope WC, Rice JE. Alchemical Prediction of Hydration Free Energies for SAMPL. J Comput Aided Mol Des. 2012 Jan; 26(5):551–562. doi: 10.1007/s10822-011-9528-8.

[24] **Park J**, Nessler I, McClain B, Macikenas D, Baltrusaitis J, Schnieders MJ. Absolute Organic Crystal Thermodynamics: Growth of the Asymmetric Unit into a Crystal via Alchemy. J Chem Theory Comput. 2014 Jul; 10(7):2781–2791. doi: 10.1021/ct500180m.

[25] **Piana S**, Klepeis JL, Shaw DE. Assessing the Accuracy of Physical Models Used in Protein-Folding Simulations: Quantitative Evidence from Long Molecular Dynamics Simulations. Current Opinion in Structural Biology. 2014 Feb; 24:98–105. doi: 10.1016/j.sbi.2013.12.006.

[26] **Schnieders MJ**, Baltrusaitis J, Shi Y, Chattree G, Zheng L, Yang W, Ren P. The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field. J Chem Theory Comput. 2012 May; 8(5):1721–1736. doi: 10.1021/ct300035u.

[27] **Sherborne B**, Shanmugasundaram V, Cheng AC, Christ CD, DesJarlais RL, Duca JS, Lewis RA, Loughney DA, Manas ES, McGaughey GB, Peishoff CE, van Vlijmen H. Collaborating to Improve the Use of Free-Energy and Other Quantitative Methods in Drug Discovery. J Comput Aided Mol Des. 2016 Dec; 30(12):1139–1141. doi: 10.1007/s10822-016-9996-y.

[28] **Shirts MR**, Klein C, Swails JM, Yin J, Gilson MK, Mobley DL, Case DA, Zhong ED. Lessons Learned from Comparing Molecular Dynamics Engines on the SAMPL5 Dataset. J Comput Aided Mol Des. 2017 Jan; 31(1):147–161. doi: 10.1007/s10822-016-9977-1.

[29] **Sun H**, Jin Z, Yang C, Akkermans RLC, Robertson SH, Spenley NA, Miller S, Todd SM. COMPASS II: Extended Coverage for Polymer and Drug-like Molecule Databases. J Mol Model. 2016 Feb; 22(2):47. doi: 10.1007/s00894-016-2909-0.

[30] **Tan D**, Piana S, Dirks RM, Shaw DE. RNA Force Field with Accuracy Comparable to State-of-the-Art Protein Force Fields. PNAS. 2018 Jan; p. 201713027. doi: 10.1073/pnas.1713027115.

[31] **Wang J**, A Snapshot of GAFF2 Development; 2017.

[32] **Wang J**, Cieplak P, Kollman PA. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? J Comput Chem. 2000 Sep; 21(12):1049–1074. doi: 10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F, parm99.

[33] **Wang J**, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and Testing of a General AMBER Force Field. J Comput Chem. 2004 Jul; 25(9):1157–1174. doi: 10.1002/jcc.20035.

[34] **Wang LP**, Chen J, Van Voorhis T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. J Chem Theory Comput. 2013 Jan; 9(1):452–460. doi: 10.1021/ct300826t.

[35] **Wang LP**, Head-Gordon T, Ponder JW, Ren P, Chodera JD, Eastman PK, Martinez TJ, Pande VS. Systematic Improvement of a Classical Molecular Model of Water. J Phys Chem B. 2013 Aug; 117(34):9956–9972. doi: 10.1021/jp403802c.

[36] **Wang LP**, Martinez TJ, Pande VS. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. J Phys Chem Lett. 2014 Jun; 5(11):1885–1891. doi: 10.1021/jz500737m.

[37] **Wang LP**, McKiernan KA, Gomes J, Beauchamp KA, Head-Gordon T, Rice JE, Swope WC, Martínez TJ, Pande VS. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. J Phys Chem B. 2017 Apr; 121(16):4023–4039. doi: 10.1021/acs.jpcb.7b02320.

[38] **Wang L**, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, et al. 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 Feb; 137(7):2695–2703. doi: 10.1021/ja512751q.

[39] **Wu Z**, Ramsundar B, N Feinberg E, Gomes J, Geniesse C, S Pappu A, Leswing K, Pande V. MoleculeNet: A Benchmark for Molecular Machine Learning. Chemical Science. 2018; 9(2):513–530. doi: 10.1039/C7SC02664A.