PROJECT SUMMARY/ABSTRACT

The study of biomolecular interactions and design of new therapeutics requires accurate physical models of the atomistic interactions between small molecules, biological macromolecules, and their complex biomolecular environments. Over the last few decades, molecular mechanics force fields have demonstrated the potential for physical models in quantitative biophysical modeling and predictive molecular design in drug discovery, vaccine development, and other discovery areas. However, despite substantial work by many researchers, significant technology gaps still exist in our ability to build systematically improvable force fields that achieve high accuracy and broad coverage of NIH-relevant chemistry, that self-consistently model heterogeneous biomolecular systems, and that can be broadly applied across a range of high-performance software packages.

This project builds on the success of the Open Force Field Initiative in its first project period, which developed open software infrastructure and open datasets, and built two successively improving generations of small molecule force fields (Parsley and Sage) whose quality and breadth have already led to significant adoption in academia and industry.

In this new project period, we aim to fully mature this technology, building new generations of accurate quantitative biomolecular force fields for molecular design in chemical biology and drug discovery, self-consistently covering nearly all NIH-relevant bioorganic chemistry. In **Aim 1**, we will extend our infrastructure to model essentially all NIH-relevant bioorganic chemistry using technologies that significantly expand generalizability and transferability. In **Aim 2**, we will enable new force field science by developing open-source technology capable of using new classes of experimental reference data to train and assess force fields and alternative functional forms in a rigorous and data-driven manner. In **Aim 3**, we will develop next-generation software infrastructure that accelerates force field parameterization and enables Bayesian inference to assess systematic error in predictions made with OpenFF force fields. In **Aim 4**, we will maximize the impact these developments have on the field by building communities around the establishment and deployment of high quality force fields, the construction and curation of open reference datasets, and the use of these datasets to train and assess force fields.

This research is significant in that the technology developed in this project has the potential to radically transform the study of biomolecular phenomena by providing highly accurate force fields with exceptionally broad chemical coverage via fully consistent parameterization of bioorganic chemistry. This will greatly improve our ability to study diverse biophysical processes at the molecular level, and to rationally design new small-molecule, protein, and nucleic acid therapeutics. This approach will bring statistical rigor to the field of force field construction and parameterization by providing greatly enhanced capabilities to make data-driven decisions, and enabling it to become a rigorous and reproducible process using a fully open infrastructure and datasets.

SPECIFIC AIMS

Computer simulations have become a critical tool to study molecular mechanisms in disease and health and to aid in the discovery of new therapeutics. However, the accuracy of molecular simulations has remained limited by approximations and inaccuracies in the molecular mechanics force fields they are built on. While recent years have seen enormous progress in computer hardware, data availability, and machine learning (ML) methods, force field science and infrastructure has yet to take full advantage of these opportunities.

We established the Open Force Field Initiative (OpenFF) to exploit these opportunities, aiming to produce modern, open, extensible infrastructure to create increasingly accurate force fields for small organic molecules and biomolecules and to curate the open quantum chemical and experimental datasets necessary to build and evaluate them. OpenFF force fields, downloaded over 200K times, have systematically improved modeling accuracy, and our work to date has generated 28 publications describing our open approach to modular, generalizable, self-consistent parameterization of biomolecular systems.

In this renewal, we will dramatically expand the ability of the OpenFF ecosystem to model with increasing accuracy the full range of chemistries relevant to the study of biomolecular mechanisms and drug discovery, incorporating novel data science and ML approaches into the physics-based modeling toolchain, via the following Aims:

Aim 1: Produce a transferable FF infrastructure to model virtually all NIH-relevant bioorganic chemistry. We will extend our self-consistent parameterization approaches to dramatically expand the domain of applicability of OpenFF force fields by enabling them to assign high-quality transferable parameters to virtually all NIH-relevant bioorganic chemistry. Our new force fields cover virtually of all bioorganic molecules (biopolymers *and* small molecules) including nucleic acids, carbohydrates, lipids, organic polymers, and complex mixtures—in addition to druglike small molecules and proteins currently supported by OpenFF. Our aim is to enable rapid, facile parameterization of complex heterogeneous systems (such as the ribosome or lipid encapsulated nucleic acids) and those including covalent modifications or conjugation (such as post-translational modifications, covalent ligands, and conjugated dyes), free of bespoke steps. We will achieve improved transferability by integrating both direct chemical perception (**Subaim 1.1**) and new machine-learning techniques that describe chemical environments using graph neural networks (**Subaim 1.2**)—approaches de-risked by prior work.

Aim 2: Enable and execute new force field science, systematically addressing alternate functional forms and the utility of experimental reference data. We will build a plugin infrastructure to rigorously evaluate the benefit provided by alternative functional forms for modeling bioorganic systems, accelerating progress within our own effort and in the community. Among other models, we will evaluate virtual sites, techniques for modifying torsions based on local conjugation, coupling of bonded potential terms, and neural net valence potentials (**Subaim 2.1**). We will build infrastructure to compare force fields using new data sources we have not yet tapped, including small molecule crystal diffraction data, osmotic and activity coefficients, and additional biomolecular experiments (**Subaim 2.2**).

Aim 3: Create next-generation tools for novel approaches to fitting force fields. We will migrate our force field fitting infrastructure to use well-supported, parallelizable machine learning frameworks that support automatic differentiation, hardware-accelerated computation, and the integration of machine learning valence potentials (**Subaim 3.1**). We will integrate surrogate models capable of dramatically reducing the number of molecular simulations required for force field parameterization using condensed phase data, which will in turn allow us to develop Bayesian model selection tools to make force field selection more rigorous (**Subaim 3.2**). We will build this infrastructure so it is capable of incorporating aspects of machine-learned potentials that are fast enough for large-scale biomolecular simulation (**Subaim 3.3**).

Aim 4: Build force field communities to maximize impact in biophysics and drug discovery. To expand the integration of OpenFF tools into biomolecular workflows, we will build **communities of software infrastructure** (**Subaim 4.1**) focused on extending interoperability with molecular simulation packages, integrating existing force fields into OpenFF, and enabling facile integration of OpenFF tools into existing workflows. To provide the practitioners with the datasets they need, we will build **communities of data** (**Subaim 4.2**), assembling larger and more diverse data sets, including quantum mechanical, physical property, and organic crystal datasets, as well as the tools for accessing them. To equip users with best practices for assessment, we will create **communities of validation** (**Subaim 4.3**), building on our success in developing community best practices for protein force field validation by establishing best practices for nucleic acids, lipids, sugars, and other important biomolecules.

Our new infrastructure and force fields will continue to transform the rapidly growing field of biomolecular modeling through significantly enhanced transferability, accuracy, and ease of use in a robust, data-enabled, automatable way, enabling research and discovery communities to meet modern challenges in biophysics and drug discovery.

PROGRESS REPORT (3/1/2020–3/5/2023) [Project start date: 3/1/2020 / Project end date: 2/29/2024]

Key results of this period included two major OpenFF force field generations, Parsley (1.x) and Sage (2.x). These force fields followed an extensive campaign of generating and curating training data and constructing benchmark assessments. Parsley[1] marked the first complete refit of all valence parameters of a force field based on direct perception of valence parameter types. This dramatically improved accuracy for small molecule geometries and energetics[1] relative to our starting point, SMIRNOFF99Frosst.[2] Sage, the second generation, incorporated an expanded quantum chemical training dataset along with co-optimization of Lennard-Jones (LJ) terms fit to condensed-phase properties, and provides a further improvement in accuracy.[3] The successful development of these force field releases was due to specific scientific and infrastructure advancements in the four aims:

Progress on Previous Aim 1: Create a modern, open software infrastructure for automatically generating and validating force fields and utilizing them broadly in modeling packages.

We developed an extensive force field optimization framework by adapting and extending ForceBalance[4] to re-fit force fields to numerous quantum chemical objectives using data we generated and disseminated via the public QCArchive repository[5] (**Figure 1**). This framework was used to fit Parsley[1] and all subsequent force field versions. To incorporate condensed phase physical property data, we built OpenFF Evaluator, a framework for efficient, automated calculation of these properties and their parameter gradients.[6] In parallel, we curated condensed phase property datasets appropriate for refitting nonbonded parameters. We demonstrated that mixture properties—specifically, heats of mixing and mixture densities from the NIST ThermoML Archive[7–9]—can produce higher quality force fields than pure condensed phase data typically used in force field parameterization.[10] Mixture property data are much more abundant than pure property data and free of issues with phase-dependent molecular repolarization. We used Evaluator and ForceBalance[11] to fully refit Lennard-Jones (LJ) and valence parameters for Sage.[3]

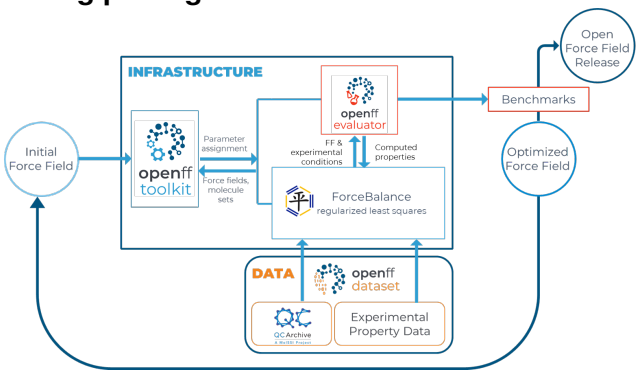

Figure 1. **OpenFF has built an open source software and data ecosystem for force field optimization.**

In pursuit of expanded generalization, we explored an alternative to our SMIRKS-based SMIRNOFF parameter assignment strategy, and developed a prototype end-to-end differentiable machine learning infrastructure (Espaloma) for force field parameterization (**Figure 2**).[12] Espaloma uses a graph convolutional neural network (GCNN) to assign continuous multidimensional vectors (or embeddings) representing atomic environments, rather than discrete atom or parameter types. These vectors can be fed to a modular neural framework for assigning parameters to atoms, bonds, angles, and torsions. This enables the entire force field typing and parameterization framework to become a fully continuous optimization problem, with smooth parameter interpolation between environments. This has the benefit of eliminating the challenges to Bayesian optimization encountered in our initial efforts using mixed discrete/continuous Bayesian optimization.[12;13] We used this strategy to produce a prototype fast, conformation-independent, drop-in replacement for AM1-BCC partial charges using GCNNs (EspalomaCharge) that self-consistently assigns high-quality, AM1-BCC-equivalent charges to all components of a biomolecular system. These approaches scale to thousands of atoms in fractions of a second, enabling fully consistent force fields to parameterize post-translational modifications, covalent ligands, covalently modified biomolecules, and other complex heterogeneous biomolecular systems.[14]

Progress on Aim 2: Construct open datasets and databases for next-generation force field development

Quantum chemical benchmark sets: We compiled QC benchmark datasets and used them to assess the quality of our force fields on broad datasets of pharmaceutical interest relative to quantum chemical data. Working with pharmaceutical industry collaborators, we compiled industry datasets for public release, deposited them in QCArchive, and built infrastructure for automated benchmarking of force field quality. **On public data outside our training dataset, Sage emerges as the best public/open small molecule force field in terms of molecular geometry and energies[3] (Figure 3); a pharmaceutical industry benchmark reached similar conclusions on proprietary internal datasets.[15]** To address the lack of additional open quantum chemical datasets with extensive coverage of relevant biomolecular space and equilibrium geometries for force field construct and assessment, we generated a large (∼1.1 million conformations) dataset of equilibrium geometries and quantum chemical energies with extensive coverage of peptides, druglike small molecules, and fragments of biomolecules at both high (ωB97M-D3BJ/def2-TZVPPD) and standard (B3LYP-D3BJ/DZVP) levels of theory in collaboration with OpenMM.[16]

Force Field	Geometries		Sublim. Heat	
	Mean ϵ	Mean RMSD	R^2	RMSD
OpenFF Sage (2.0)	1.52	0.31	0.65	4.9
OpenFF Parsley (1.3)	1.82	0.28		
GAFF 1.8	1.68	0.33		
GAFF 2.1	1.71	0.43	0.58	4.85
CGenFF	2.13	0.47	0.40	4.49
OPLS-AA	2.36	0.83	0.53	5.52

Table 1. **OpenFF Sage yielded highest accuracy across small molecule force fields tested, in terms of crystal geometries and energetics.** Geometries are assessed for crystal simulations of 13 amide or peptidic organic compounds. Heats of sublimation are computed for crystal structures of 280 compounds. ϵ is a metric of accuracy of the lattice, RMSD (Å) reflects time-averaged deviations of all non-H interatomic distances more than 3 bonds apart and 4 Å or less in the reference crystal structure. Standard errors of the mean are 0.42–0.48 for mean ϵ and 0.08–0.17 for mean RMSDs (0.40 for OPLS-AA).

Quantum chemical methodologies: Systematic investigation determined the level of quantum chemical theory (B3LYP-D3BJ/DZVP) providing an optimal balance of speed and accuracy for dataset generation.[18] In fitting fully consistent force fields to this data, we demonstrated that reasonable torsion profiles could be extracted from optimization trajectories initiated from enumerated conformers, avoiding computationally expensive systematic torsion drives.[12] We demonstrated inclusion of vibrational frequency data resulted in a degradation of force field quality, likely due to difficulties in matching vibrational frequencies between internal coordinate systems.[3] We also found deriving initial parameters directly from QM using the modified Seminario method, rather than optimizing from a prior force field, results in higher quality force fields.[19]

Condensed phase datasets: To incorporate physical properties into our fitting infrastructure, we curated data from the NIST ThermoML repository[7–9] and the Minnesota Solvation Database,[20] making these datasets programmatically accessible for automated benchmarking efforts through the OpenFF Evaluator API.[6]

Organic crystal data: We have also automated the simulation of periodic small molecule crystal structures from the Crystallography Open Database (COD)[21;22] and used these tools to compare four FFs, including two from OpenFF.[23] In all these measures, Sage demonstrates high accuracy, surpassing other public force fields (Table 1).

Progress on Aim 3: Develop Bayesian inference techniques to address key questions in force field physical modeling and predict systematic error

Automating SMIRNOFF type refinement: We have shown our approach to automated fitting of SMIRNOFF force fields can provide remarkable accuracy, but the choice of SMARTS classes defining these environments requires human experts. We developed tools to automatically suggest useful classes via a parsimonious search of bit-mapped molecular graphs.[24] Preliminary results suggest higher accuracy force fields with lower complexity than manually-defined types can be built, as suggested by other collaborations[25] We produced new general, transferable force fields from scratch in addition to extending and improving preexisting typing schemes.[24]

Fast surrogate models for optimization: One challenge of automated force field training is the expense of performing large numbers of molecular simulations. We used Gaussian process surrogate models fit to match physical properties predicted by simulation-only[26] force field optimization. We found that optimizations using these surrogate models allows rapid escape from local minima while maintaining equivalent levels of efficiency to simulation-only approaches.

Bayesian model selection: Another challenge is the difficulty of deciding whether a given change in force field model is truly an improvement. We showed how Bayesian inference can be used in a quantitative way to decide the utility of two different levels of theory in a study of the phase equilibria of diatomic molecules.[27] Although the approach is too expensive to apply to molecular simulations of physical properties directly, high-accuracy surrogate models such as those described above can make Bayesian analysis possible, improving such decisions as the number of parameter types necessary or the utility of polarizability in a more quantitative manner. In addition, the GCNN framework described in the Progress Report above will greatly simplify the development of Bayesian inference technologies in future work by changing a complex discrete inference problem to a continuous one.

A

	(a) dataset	Espaloma RMSE Train	Espaloma RMSE Test	Legacy FF RMSE (kcal/mol) (Test molecules) OpenFF 1.2.0	GAFF-1.81	GAFF-2.11	Amber ff14SB
	PhAlkEthOH (simple CHO)	$0.8656^{0.9131}_{0.8225}$	$1.1398^{1.2332}_{1.0715}$	$1.6071^{1.6915}_{1.5197}$	$1.7267^{1.7935}_{1.6543}$	$1.7406^{1.8148}_{1.6679}$	
	OpenFF Gen2 Optimization (druglike)	$0.7413^{0.7920}_{0.6914}$	$0.7600^{0.8805}_{0.6644}$	$2.1768^{2.3388}_{2.0380}$	$2.4274^{2.5207}_{2.3300}$	$2.5386^{2.6640}_{2.4370}$	
	VEHICLe (heterocyclic)	$0.4476^{0.4690}_{0.4273}$	$0.4233^{0.4414}_{0.4053}$	$8.0247^{8.2456}_{7.8271}$	$8.0077^{8.2313}_{7.7647}$	$9.4014^{9.6434}_{9.2135}$	
	PepConf (peptides)	$1.2714^{1.3616}_{1.1899}$	$1.8727^{1.9749}_{1.7309}$	$3.6143^{3.7288}_{3.4870}$	$4.4446^{4.5738}_{4.3386}$	$4.3356^{4.4641}_{4.1965}$	$3.1502^{3.1859,*}_{3.1117}$
joint	OpenFF Gen2 Optimization	$0.8264^{0.9007}_{0.7692}$	$1.8764^{1.9947}_{1.7827}$	$2.1768^{2.3388}_{2.0380}$	$2.4274^{2.5207}_{2.3300}$	$2.5386^{2.6640}_{2.4370}$	
	PepConf	$1.2038^{1.3056}_{1.1178}$	$1.7307^{1.8439}_{1.6053}$	$3.6143^{3.7288}_{3.4870}$	$4.4446^{4.5738}_{4.3386}$	$4.3356^{4.4641}_{4.1965}$	$3.1502^{3.1859,*}_{3.1117}$

B

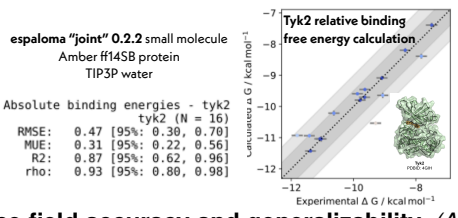

Figure 2. **Graph convolutional neural networks (GCNNs) can significantly improve force field accuracy and generalizability.** *(A)* Our GCNN prototype strategy (Espaloma, schematically shown in **Figure 5**) can produce MM force fields capable of reproducing quantum chemical energies significantly better than legacy force fields.[12] *(B)* Relative binding free energy calculations with Espaloma small molecule parameters produce significantly lower error with experiment (RMSE 0.47 [95% CI: 0.32, 0.68] kcal/mol)[12] than matched calculations with Schr odinger's FEP+ (RMSE 0.93±0.12 kcal/mol), which use a proprietary force field.[17]

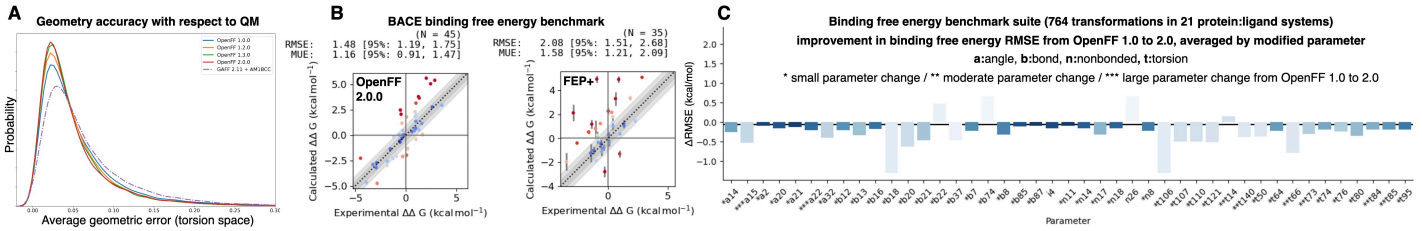

Figure 3. **Benchmarking of OpenFF shows continuous improvement and competitive performance.** *(A)* Small molecule geometries show systematic improvement with each OpenFF release. *(B)* Binding free energy benchmarks are competitive with commercial force fields. *(C)* Each generation of OpenFF shows significant improvement in binding free energy benchmarks.

Progress on Aim 4: Integrate the results of Aims 1–3 to construct and validate open source, transferable, and self-consistent force fields for small molecule interactions with complex biomolecular systems

New force fields integrating multiple improvements: We used the developments in Previous Aims to produce iteratively refined, versioned force field generations Parsley (1.x) and Sage (2.x). These significantly improved accuracy and coverage of chemical space for modeling drug-like small molecules[3;15] (**Figure 3**). In an industry-led benchmark, these OpenFF force fields, based on our SMIRNOFF chemical perception scheme,[28] successfully parameterized 99.7% of molecules in internal datasets and delivered clear improvements in accuracy of reproducing quantum chemical energies and geometries over public force fields.[15]

Extensive, consensus force field assessments: We evaluated the quality and transferability of our force fields in several ways. As Sage was a dramatic refit of nonbonded terms, it carried the risk that optimizing LJ parameters away from AMBER-family LJ parameters might result in reduced accuracy in binding free energies when used with AMBER protein FFs. To assess this, we carried out protein-ligand binding free energy benchmarks[3;19] and found we either maintained or modestly improved accuracy[3] (**Figure 3**). Sage also incorporated dramatic improvements to our methodology for selecting quantum chemical datasets for training, including automated methods for selecting diverse chemistries exercising the same parameters in order to ensure robustness and generality.[3]

Large user base and industry impact: Our tools and force fields are now widely used; our core toolkit has been donloaded >174,000 times,[29] and our force fields >204,000 times.[30] In addition to direct applications within the pharmaceutical industry, they are heavily used by software vendors, including OpenEye (Orion), Cresset (Flare), and other tool builders (e.g. OpenMM, OpenFE, AstraZeneca Icolos, MolSSI QCEngine, qubekit, OpenBioSim/BioSimSpace, and MolSysMT; see Letters).

Self-consistent protein:small molecule force fields with SMIRNOFF parameter assignment: We have built a preliminary version of OpenFF v3.0.0 (Rosemary) that self-consistently parameterizes small organic molecules and proteins, currently supporting the canonical 20 amino acids and protonation/tautomer variants. Rosemary supplements the Sage training set with quantum chemical (QC) data for capped amino acids and tripeptides, and uses a library of self-consistent AM1-BCC[32]ELF10[33] charges for amino acids. Preliminary evaluation of Rosemary against a test set of QC torsional scans for small peptides yielded an RMSE of 2.0 kcal/mol, compared with 3.6 kcal/mol for Amber ff14SB,[34] with no loss of accuracy (relative to Sage) on QC small molecule benchmarks. It shows similar performance as ff99SB for experimental NMR J-couplings over 500 ns aqeuous simulations of 13 representative peptides. Initial tests also show that the protein parameters provide stable μs simulations (**Figure 4**). We organized a community-driven overview of best practices for protein force field assessment[35] and are using these to drive further assessment.

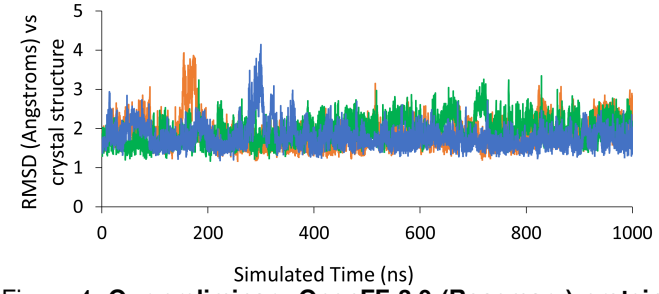

Figure 4. **Our preliminary OpenFF 3.0 (Rosemary) protein + small molecule force field is stable over μs timescales.** Color traces show RMSD of 3 replicates relative to NMR solution structure of GB3 (PDBID: 1P7E[31]) in explicit-solvent MD.

Self-consistent protein:small molecule force fields with GCNN parameter assignment: We also produced a prototype force field for proteins and small molecules using our GCNN framework (Espaloma), using similar QC datasets as Rosemary. This experimental force field demonstrates significant improvements (**Figure 2B**) over existing public force fields and significant reductions in error in protein:ligand alchemical free energy calculations.[12]

Bespoke tailoring for individual molecules or molecule sets: Our infrastructure work also allows easy development of tailor-made force fields for particular applications. For example, we partnered with Danny Cole to develop BespokeFit,[36] which allows custom refits of torsion parameters to provide especially high accuracy on specific chemistries of interest. This tool is already being used by several pharmaceutical companies and Cresset.

SIGNIFICANCE

Force fields now can drive discovery, yet still require better accuracy. Atomistic molecular simulations of biomolecules are growing ever more powerful and their impact in the biomedical sciences continues to expand.[37–39] In drug discovery, they are used to design orthosteric inhibitors,[40] discover transient allosteric binding pockets in proteins,[41] and improve the physical properties of biopharmaceuticals.[42] With increasing accuracy and chemical coverage, they could be used even more broadly, e.g., to predict off-target binding and resulting side-effects, or to design a ligand that specifically inhibits a targeted set of kinases or kinase mutants.[43;44] Simulations are frequently used to augment structural biology methods such as cryo-EM, NMR, and X-ray crystallography to elucidate the structures of macromolecules and their assemblages.[38;45] Given biomolecular structures, simulations are often used to help elucidate function and mechanisms of disease,[46] providing insights that feed into drug discovery.

However, we cannot yet design a ligand for a protein of known structure completely *in silico*, have a synthetic chemist make it, and find it binds with nanomolar affinity.[47] As computers grow faster, the remaining error increasingly stems from inaccuracy in the potential energy function, or force field (FF), used to model interactions within the biomolecular system:[48] These errors persist when the same methods used for computing free energies of protein-ligand systems are applied to much simpler host-guest systems—miniature receptors for which thorough sampling can be convincingly achieved.[49–52]

Many biomedically relevant systems are frustratingly difficult to model; we aim to change this. Heterogeneous systems remain challenging to simulate because of obstacles and uncertainties associated with assigning high-quality, self-consistent force field parameters. For example, parameterizing a lipid-encased mRNA nanoparticle currently requires major effort. Similar challenges arise for proteins with postranslational modifications, nonstandard residues, or covalent ligands. Generating force field parameters for heterogeneous systems is technically difficult, as is ensuring quality and consistency. Many chemistries also lack sufficiently high quality FFs—despite significant progress,[53–56] nucleic acid and carbohydrate FFs have received less attention than protein force FFs, and greater accuracy is needed to accurately model combinations, e.g., protein-DNA (e.g., transcription factors[57]) and protein-RNA-DNA (e.g., CRISPR[58;59]) interactions, and glycoproteins. Even protein force fields still require improvement[60;61] for simulation technologies to fulfill their ultimate promise of reliable simulation and design.

We will address these challenges by building accurate force fields that are, by design, self-consistent across molecule classes (e.g. protein, small molecule, nucleic acid), coupled with well-engineered software toolkits for force field optimization, assessment, and straightforward application to virtually the full array of biomedically relevant molecules. By *self-consistent*, we mean that parameters of the Thr sidechain should be very similar to those of ethanol, while lipid alkyl chains should be similar to those of small molecule alkanes, with any differences in parameters due to entirely to differences in chemical context. Likewise, peptidic moieties encountered in any molecule should the same parameters as those moieties in proteins, up to any changes in the chemical context.

This work will directly enable simulations to have far greater impact in biomedical science and drug design. Perhaps even more importantly, the infrastructure and tools generated by this fully open-source project will empower other researchers to do their own force field research and development, and we will build research communities that will help others make best use of our results and technologies.

Our infrastructure provides a rigorous way to evaluate and integrate advanced functional forms. We leverage automated parameterization, consistent datasets, and community-driven best assessment practices to rapidly and reproducibly test new approaches and determine the best path forward towards dramatically improved accuracy and transferability. This rigorous, reproducible approach provides the best way to weigh the benefits of proposed new approaches and functional forms, whether from inside our project or from the broader community, and thus to drive well-supported improvements in successive force field releases.

Our work impacts the entire molecular modeling community by improving the accuracy and domain of applicability of FFs and empowering researchers to independently explore and advance FF science. The community already recognizes this significance and is incorporating our FFs into a wide range of modeling tools and used across a broad swath of the pharmaceutical industry (see Letters).

INNOVATION

This project will yield the first open, scalable, data and software infrastructure designed to reproducibly build, apply, and assess, statistically robust force fields. It introduces significant innovations at multiple levels.

New ways to describe chemical environments for parameterization. Our novel force field approaches use more descriptive and robust ways to perceive atomic environments for parameter assignment. Traditionally, atom types, crafted by human experts, are used to assign force field parameters. This approach hampers extension into new

areas of chemical space, due to proliferation of parameters and consequent opportunities for human error.[28] For example, in the atom-typing paradigm, creating a new torsion type requires defining a new atom type, which then generates a need to define its LJ parameters, as well as the parameters for all bonds and angles that involve the new atom type.[28] Our innovative direct chemical perception approach, the SMIRKS Native Open Force Field (SMIRNOFF) specification,[28] abolishes this problem by separately typing each FF term directly from the chemical graph. Here, we will further innovate by building on data-driven methods for parsimoniously determining the most informative SMIRNOFF types, as well as novel parameter assignment technologies that use continuous, rather than discrete, representations of atomic environments using graph convolutional neural networks (GCNNs) that preserve chemical equivalences and symmetries (see Progress Report). These new methods promise to drastically improve our ability to match chemical space to force field parameters while also enabling Bayesian inference (Subaim 3.2) by removing the challenges of optimizing over discrete types.[12;13]

Force fields designed from the ground up for all bioorganic molecular systems. A second innovation is self-consistently derived force field parameterization across the full range of biomolecular classes, including small molecules, proteins, lipids, carbohydrates, and nucleic acids, and critical solvents like water fit to solvent-solute properties. This will enhance accuracy for data-limited chemical domains by leveraging data for related chemistries, and extend generalizability and transferability to new chemistries.

Use new types of chemical data to optimize and assess force fields. We will integrate hitherto unused or underused classes of chemical data into our fitting and assessment methods to enhance force field accuracy and generalizability. These include osmotic and activity coefficients (particularly useful for charged species in solution), speed of sound measurements, octanol/water partition coefficients, and small molecule X-ray crystal data. We will generate and curate new quantum chemical data sets, broadening coverage of chemistries for drug-like molecules and providing focused datasets needed to accurately describe biomolecules such as nucleic acids and sugars.

Rigorous validation of new physics in FFs. Our new infrastructure will provide built-in support for a broad range of physical interaction potentials well-supported by simulation packages but not extensively exercised by modern biomolecular force fields, such as virtual (off-atom) interaction sites, class II valence force field terms (e.g., term-term coupling),[62;63] and electronic polarizability. We will also enable integration of machine learning potentials for short-range interactions to replace valence terms, alongside physical interactions such as long-range electrostatics and dispersion interactions. Importantly, we will build a modular framework for automated, statistically rigorous, testing and validation of these potentials, using curated, open datasets.

Bayesian methods for data-driven force field construction and quantifying prediction errors. Although force fields are rooted in physics, they are fit to a limited quantity of experimental and quantum chemical data. We will leverage Bayesian inference approaches (Subaim 3.2) to address multiple issues that arise as a consequence. Bayesian inference enables us to generate multiple parameter sets that fit the data with different probabilities, and can be used to assess the systematic error in predictions arising from force field parameter uncertainty.[64–67] Bayesian model selection approaches can also bring statistical rigor to assessments of physical modeling choices, functional forms, mixing rules, and other choices arising in force field construction (Subaim 2.1).[27] Our proposed innovations in continuous typing[12] are particularly suited to Bayesian approaches, and promise to enable a new, data-driven approach to type definitions that avoids traditional by-hand type assignments, which, though chemically rational, may not be sufficiently descriptive.

Enabling community innovation: A fundamental goal is to create data and software infrastructure that enables researchers to innovate within force field research and development, exploring novel functional forms and new data sources. To maximize impact, we will help build topical research communities (Aim 4).

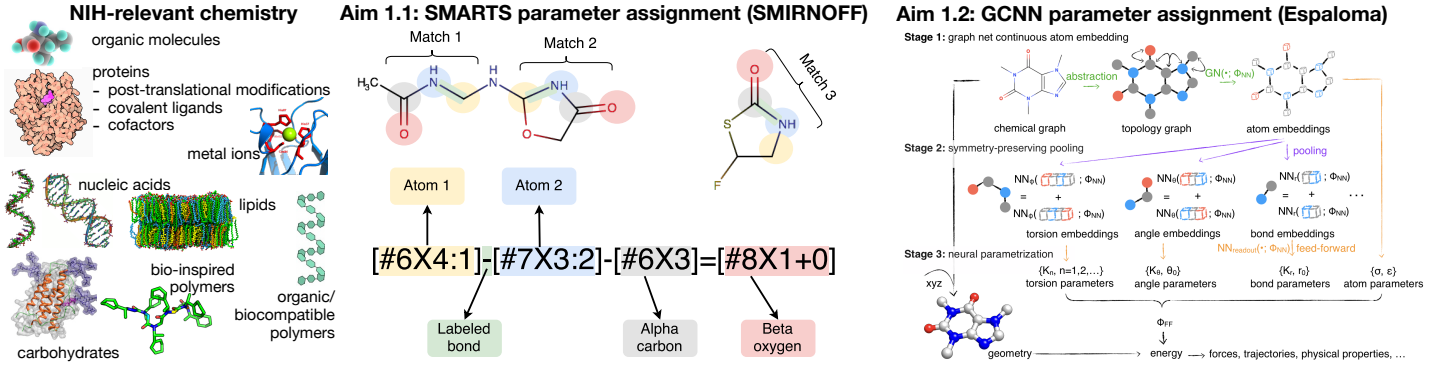

Figure 5. **Aim 1 will pursue two parameter assignment strategies to enable force field generalizion to nearly all NIH-relevant chemistry.** Subaim 1.1 will extend the SMIRNOFF SMARTS-based parameter assignment strategy, while Subaim 1.2 will explore graph convolutional neural network (GCNN)-based parameter assignment strategies to achieve superior generalization and interpolation.

APPROACH

We will build a modern, open infrastructure for the development of *general* force fields (FFs) that model organic small molecules, biomolecules, and chemical adducts self-consistently, recasting FF parameterization as a primarily automated machine learning problem that fully exploits known physics. By "modern" and "open", we mean modular, extensible, open source, and using and documenting open standards and best practices, as detailed in the Data Sharing Plan. This infrastructure is needed for transferable, general force fields that easily model new molecular systems.

Aim 1: Produce a transferable FF infrastructure to model virtually all NIH-relevant bioorganic chemistry.

With new approaches to chemical perception and transferability, our dramatically increased domain of applicability will mean users can assign high-quality parameters to essentially all NIH-relevant bioorganic chemistry. This includes parameterizing, self-consistently, all bioorganic molecules, including nucleic acids, sugars, lipids, organic polymers, and their derivatives, in addition to the drug-like small molecules and proteins currently supported. This will make simulations of systems including covalent dyes, glycosylations, lipid-encapsulated polymer nanoparticles drugs simple, rather than major research projects.

Subaim 1.1: Improving force field scope and transferability using direct chemical perception approaches to better describe chemical environments

We plan to provide accurate force fields describing all chemical space relevant to human health. Currently, our force fields cover small/drug-like molecules and, in our forthcoming Rosemary (OpenFF 3.0) force field release, they are being extended to proteins (see Progress Report). However, further work is needed to ensure we can accurately and self-consistently handle nucleic acids, post-translational modifications, non-natural amino acids, lipids, carbohydrates, cofactors, and common organic biological ions, as well as coordinated metals and ions. Our force fields already handle most of these in the sense of providing parameters, but prior work indicates that additional data and testing is required for specific application domains (e.g., lipid force fields[68–72] and nucleic acids[73–77]) to ensure particularly common molecules and building blocks are treated with sufficiently high accuracy, so we will generate this data in collaboration with the relevant communities (see Subaim 4.2).

Using this data, we will build self-consistent accurate parameters for these key molecules and building blocks. Thus, we are not tuning parameters separately for each class of bioorganic molecules, but rather adding more specific chemical perception (and associated data) to ensure that these critical building blocks are treated accurately and consistently across molecular types. Our basic approach thus involves extending our small molecule infrastructure to handle biopolymers composed of organic building blocks. First, we will engineer routes to preserve chemical information on import, extract it from databases, or infer it; this is crucial because common formats (e.g., PDB, PDBx/mmCIF) files often do not include a full description of chemical information, missing crucial information such as formal charge or bond order. Second, we will generalize our prototype graph neural net changing architecture, EspalomaCharge,[14] into a fast and accurate method for assigning charges to polymers of arbitrary size.

Another critical aspect of this work is to identify additional parameter types required beyond those provided by the small molecule force field. We will refine SMIRNOFF types as appropriate for nucleic acids and lipids drawing from extensive pioneering prior findings.[69;70;76;78–81] In other areas (e.g., sugars[82–84]) typing decisions are less certain, and we will utilize our `besmarts` (binary encoded SMARTS) framework to automatically determine when/where additional types are needed.[24]

Subaim 1.2: Improving transferability across chemical environments using graph convolutional neural networks

We and others have recently proposed a new approach to parameter assignment that uses a continuous latent embedding of atomic chemical environments[12;85] instead of traditional atom typing or the direct chemical perception[28] approach used in current OpenFF FFs. Using graph convolutional networks (GCNNs) that respect chemical equivalences (**Figure 5**), each atom can be assigned a continuous vector representation that is then combined in a symmetry-preserving fashion, enabling parameters to be generated directly by a neural network module.[12]

This approach provides several key advantages, the most significant of which is transforming a previously difficult mixed discrete-continuous optimization of discrete chemical environments and continuous parameters into a problem that can be solved by continuous optimization strategies, enabling both types and parameters to be fit simultaneously. In this way, new atom or parameter types can be learned simply by including representative molecules in the fitting process (**Figure 2**). New modules for different interaction types (e.g., point polarizability, alternative van der Waals representations, class II valence terms[62;63]) can easily be added and the force field

re-fit without the need to craft new discrete type classes, and the resulting parameters are smoothly interpolated based on changes in the automatically perceived chemical environment.

Preliminary work. We have shown that GCNNs can produce continuous descriptors of atom environments while preserving chemical equivalences and are expressive enough to learn existing human-assigned typing schemes.[12] These models are able to learn MM valence terms directly from quantum chemical potential energies.[12]

As with SMIRNOFF, this architecture is modular, enabling new parameter types (e.g., point polarizabilities) to be readily integrated. However, it offers the advantage of not having to craft parameter types, as the parameters can be learned from data as continuous functions of the atomic environments.[12] We have deployed a version of this for replacing AM1-BCC charges self-consistently to small molecules and biopolymers[14] and have used the approach to build a prototype self-consistent force field for proteins and small molecules that delivers significantly increased accuracy in a protein:ligand free energy benchmark.[12]

We will integrate this GCNN approach into the Open Force Field infrastructure and assess ways to improve the robustness, data efficiency, and ability to tailor these models to learn new chemical environments. This will include enabling our existing infrastructure for condensed-phase property calculations and quantum chemical targets to be used, bringing the GCNN approach on par with our existing SMIRNOFF-based infrastructure. We will then use this infrastructure to refit entire force fields using the same QC and experimental physical property data as our existing and new force field generations.

While our current prototypes already generalize well to new chemistries (**Figure 2**), further improvements in generalizability and data efficiency may be possible by replacing simple one-hot encoding of elements with continuous descriptors that capture the periodic table more completely.[86–88] We will also investigate semi-supervised training approaches, where unlabeled molecules are used to better learn chemical relationships.[89] Because complex, long-range conjugation can challenge GCNNs, we will also investigate alternative approaches to capturing most of this information.[90]

Other chemistries: Important classes of chemistries that don't fit as clearly into "organic" materials are water, main group ions, and metal ions. These will be addressed for both Subaim 1.1 and Subaim 1.2.

Currently, OpenFF force fields use the TIP3P water model[91] due to its historical success with AMBER-family force fields. OpenFF nonbonded refits to date have used aqueous mixture properties, fixing the choice of water model.[3] However, all current widely-used water models have been parameterized almost exclusively to replicate the properties of pure water. *We will co-optimize water along with all other molecular parameters*, capturing not only pure water properties but also interactions with cosolvents, solutes, and ions to better describe complex mixture thermodynamics. The data and infrastructure to perform this optimization already exists,[3] and there is good evidence[92] that there is enough parameter freedom in even rigid water models to still get pure water properties essentially correct while improving properties involving solutes.

We will incorporate new sources of condensed phase data into our fitting to better handle main group ions, group I and II metal cations, and halides, as described in Subaim 2.2. More challenging are coordinating transition metal ions, which are essential for modeling biomolecules.[93;94] At the present time, we will adopt a simple but widely-supported virtual site charge approach that provides a robust approach to model coordination geometry (the "cationic dummy model"[95]) and explore the possible use of well-supported LJ modifications for increased accuracy (e.g., LJ 12-6-4 potentials[96;97]). We will benchmark against stability and properties for relevant molecular systems in order to define the domain of applicability, as this approach may not extend to certain metal complexes (e.g., hydroxamic acid/Zn interactions[98]) without more advanced functional forms.

Assessment: Both Subaims 1.1 and 1.2 rely on strategies for assessing how well models effectively learn from limited data (data efficiency) and generalize to new chemistries. We will assemble assessment datasets (both quantum chemical and experimental) for relevant chemistry domains, and assess both accuracy and data efficiency on domain-specific benchmarks. When accuracy gaps are identified, we follow the same approach used in preparing our Rosemary release, which self-consistently handles proteins: identifying molecular sets where new quantum chemical calculations or experimental datasets need to become part of our training data and incorporating these into type inference and parameter fitting.

For protein FF assessment, we will curate published protein NMR data, including NOEs, through-bond ^{3}J couplings, Lipari-Szabo S^2 order parameters and residual dipolar couplings (RDCs) from the Biological Magnetic Resonance Data Bank (BMRB),[99] expanding on earlier work.[100] We will collaborate with existing force field efforts to incorporate extensive protein, nucleic acid, and lipid NMR datasets they have found useful in force field assessment (see Letters),[61;101;102] and implement best-practices for computing NMR observables. We will also

assess our protein FFs via simulations of selected room-temperature protein crystals.[34]

Aim 2: Enable and execute new force field science, systematically addressing alternate functional forms and the utility of experimental reference data.

To date, our force fields have used a well-accepted functional form with atom-centered, fixed partial charges to represent electrostatic interactions and the Lennard-Jones potential for van der Waals interactions—a form essentially unchanged for over four decades.[73;103;104] This widely-supported functional form, whose early selection was driven in part by limitations of 1980s-era computer architectures, neglects important physics, and may yield insufficient accuracy in some systems even when carefully parameterized against well-chosen data. Pioneering force field projects have explored more detailed physical interaction models, but it remains uncertain which models are most accurate for any given domain, and the trade-off between cost in computational speed and benefit in accuracy should be evaluated before committing to a proposed change. Little comparison data guides either users, in selecting optimal functional forms, or developers, in implementing new functional forms. This problem stems in part from the use of different training/testing datasets and methods by different researchers, and to the lack of robust statistical assessment frameworks. We argue that *rigorous assessment of functional forms requires comprehensive and consistent contruction and benchmarking of models using the same domain-relevant training and test data*. We will enable these rigorous comparisons by further expanding our efficient, automated, open-source infrastructure, comprising both software and data, to fit and assess force fields on consistent datasets and robustly evaluate alternate functional forms (Subaim 2.1).

We will also expand the dataset of experimental physical properties used to train FFs in order to improve FF accuracy and transferability. This work leverages modern computer architectures, automation, and the growing availability of large machine-readable collections of chemical property data. To take advantage of these opportunities, we will outfit OpenFF Evaluator with the ability to compute additional properties and make these advances available to the research community (Subaim 2.2). We will also use this data to explore the selection of optimal reference datasets to efficiently drive force field accuracy (Subaim 2.2).

Subaim 2.1: Build a plugin infrastructure enabling systematic parameterization and testing of force fields with varied functional forms, then use it to evaluate key functional forms

We have prototyped a plugin interface (`smirnoff-plugins`[105]) that allows OpenFF infrastructure (**Figure 1**) to accept external plugins that allow novel functional forms to be used (**Figure 6**). This enables force field scientists to parameterize complete force fields with different functional forms using the same dataset in just days to weeks.[105] This prototype has been used by collaborators[105;106] to explore the use of off-center partial charges[19] (below). We will extend this interface to handle functional forms supported by the widely used OpenMM simulation package,[107] which is uniquely flexible in its ability to handle a wide range of force field terms. This will enable research groups worldwide to rigorously study existing force field variants and explore new ones for proposed community adoption.

We will use this plugin infrastructure to test leading alternate functional forms, focusing on those which are of particular interest to the field, add little computational cost, and are easily implemented in or available in typical simulation codes (Figure 6). While these functional forms are not new, our focus is to determine how much impact they have when fairly trained/tested on the same data in comparison with more traditional FFs. Such systematic study will allow a next generation of more accurate and performant FFs. We plan to study variants such as the following, each of which will be trained consistently and benchmarked against our standard functional forms:

- **Off-atom (or virtual) interaction sites:** Partial charges situated not at the centers of atoms but at nearby sites defined in a local molecular frame of reference[108] may, for example, improve the treatment of lone pair electrons on carbonyl groups and the geometric preferences of hydrogen bonds. We will establish a compact set of candidate virtual sites motivated chiefly by lone-pair geometries and σ holes and assign partial charges analogously to the AM1-BCC method[32] in which a new set of bond-charge corrections (BCCs) are optimized

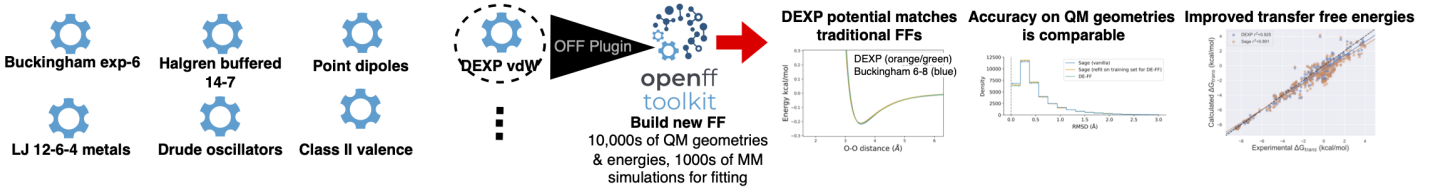

Figure 6. **A prototype plugin interface allows us to rapidly explore alternate functional forms.** Our `smirnoff-plugins`[105] interface allows alternate functional forms to be slotted into our toolkit and then trained using the same infrastructure and data as standard releases. This approach was used to assess the double exponential (DEXP) potential that avoids overly stiff repulsion and singularities as a replacement for Lennard-Jones; the resulting FF agrees better with QM and yields better transfer free energies than obtained with the baseline LJ model.[105]

to yield both atom and virtual site partial charges optimized to replicate QM electrostatic potentials (ESPs). We will then reoptimize all valence and LJ parameters against reference QM and experimental data, as in Sage,[3] and benchmark against baseline FFs trained against the same data using only atom-centered charges.

- **Alternative van der Waals models**: The conventional LJ potential yields an unrealistically steep representation of steric repulsion,[109] contains singularities, and may also provide an overly long-ranged representation of the dispersion interaction.[110] We will evaluate the benefits of moving to more realistic functional forms, such as the Buckingham exp-6[111] and double exponential (DEXP)[105] potentials. As softcore modifications of the LJ potential are normally required to eliminate singularities when running alchemical free energy calculations,[112] adopting a singularity-free van der Waals potential like DEXP could greatly simplify such calculations.
- **Angle/bond couplings**: Changes in bond angles affect the energetics of the associated bond stretches.[113] We will model this with the Urey-Bradley term,[113;114] typed according to our existing angle types.

If we find multiple variants yield clear improvements in accuracy, we will train and test FFs that combine them, assessing accuracy/performance tradeoffs. The same framework will also facilitate other explorations such as the use of conformation-dependent charges,[115–119] explicit treatments of electronic polarizability,[55;120–124] additional class II (i.e. angle-torsion coupling) force field valence terms,[62;63] use of neural net potentials,[125–127] etc. We anticipate that some of these will be studied by other research groups using our infrastructure and support.

Subaim 2.2: Build infrastructure to train and test force fields using untapped physical property data

Force field parameters are typically tuned against experimental reference data along with QC data. Using a large set of diverse data is expected to maximize accuracy and transferability. The experimental properties most commonly typically used to train FFs—densities, and heats of vaporization of pure organic liquids–do not report on the interactions of different compounds with each other, on interactions in aqueous solution, or on the properties of ionic species. We recently showed that enthalpies of mixing and densities of organic liquid mixtures provide considerable value for force field fitting.[10;26] Here, we propose to expand our OpenFF Evaluator[6] module to new properties so that additional, more diverse, data can be used for FF training and testing.

We will draw on diverse solution data from NIST's ThermoML Archive,[128] including osmotic coefficients ($\sim$6000 measurements over $\sim$300 data sets) and activity coefficients ($\sim$60K measurements over 10K data sets). Osmotic coefficients measure the change in free energy of a solvent in response to the presence of a solute, while activity coefficients measure solute-solute interactions that lead to nonideal solution behavior. Prior FF work based on Kirkwood-Buff solution theory has proven the power of solution data in FF development.[129] Data on the thermodynamic interactions of small, biologically relevant, molecules such as urea and glycols, with amino acids[130;131] and nucleic acid components[132] may also be of value. Such data can report on the interactions of ionized compounds, such as carboxylates and amines, in aqueous solution. We will also explore other hitherto untapped properties that offer a good balance of availability, relevance, and facile computability. These include octanol-water partition coefficients, which are available for many drug-like compounds,[133;134] and the speed of sound in liquid mixtures, which is closely related to compressibility and for which tens of thousands of data points are available. These can all be computed relatively quickly and are available for a wide range of compounds.

Once a force field has been created, it must be carefully assessed against domain-relevant data outside the training set. Unlike training-set properties, which are iteratively computed within an optimization loop, test-set properties need to be computed only once for a force field, so they can include important properties that are more complex and time-consuming to compute. We have already integrated the ability to compute aqueous host-guest binding free energies into Evaluator and used this computationally expensive but particularly relevant property, which is available for well over 100 systems in the literature[135] and databases like SupraBank and BindingDB,[136] for FF evaluation.[137] We will extend Evaluator to calculate NMR chemical shifts and scalar couplings, which report on conformational preferences, for small peptides in aqueous solution;[34] and to protein-small molecule binding free energies (of particular interest to drug discovery) computed on Folding@home.[17;138–141] We will also outfit Evaluator to carry out the small molecule crystal structure assessments described in the Progress Report.

We will also build out infrastructure for benchmarking on miniprotein thermostability. Collaborator Gabriel Rocklin (see Letters) has conducted high-throughput stability studies of thousands of designed mini-proteins,[142] so we will use this dataset for protein force field assessment. We will also include melting curves for several small proteins and peptides, such as the villin headpiece, the helical $(AAQAA)_3$ peptide, Trp-cage, and the CLN025 hairpin, which have been experimentally obtained using standard techniques .[143–145]

The choice of training dataset composition dramatically impacts FF quality, but this area of science is mostly

unexplored. We will assess how dataset composition impacts FF accuracy while seeking ideal minimally expensive collections of experimental data to train FFs, exploring diverse data types and their impact on predictive power.

Aim 3: Create next-generation tools for novel approaches to fitting force fields.

Subaim 3.1 Build a modern software infrastructure to improve FF training, assessment, and tailoring

OpenFF has now built multiple generations of production FFs from curated quantum chemical and experimental datasets using ForceBalance,[146;147] a Python FF fitting tool. However, a revolution in advanced Python infrastructure for ML—supported by massive investments in commercial software development—presents a new opportunity to rearchitect our fitting infrastructure to take advantage of investments in scalability, accelerated hardware, modularity, automatic differentiation, and just-in-time compilation, simplifying our tooling and greatly expanding capabilities. In addition, ML frameworks enable integration of accurate new ML-based potentials for describing local valence interactions, such as SchNet,[127] ANI,[148;149] and NequIP,[150] which can be fast enough for biomolecular simulations.

Preliminary data: We have developed prototype force field fitting infrastructures[12] (**Figure 2**) in both PyTorch (espaloma[151]) and JAX (espalomax[152]). PyTorch is a popular, well-supported Python ML framework (developed by Meta) with extensive community support, while JAX (developed by Google) leverages similarities with the popular numpy ecosystem. Both frameworks offer parallelization, hardware acceleration, and automatic differentiation. The last enables computed fitting targets to be differentiated with respect to either parameters, configurations, or both, which has the potential to drastically simplify the code and accelerate the implementation of new features and fitting targets.

We will migrate our fitting infrastructure to one of these ML frameworks, drastically simplifying the code in the process, as (1) numerous features we have been maintaining are built-in features of these frameworks, and (2) parameter and position gradients need not be implemented because they are automatically generated by automatic differentiation. We will focus on modularity to enable new fitting targets and functional forms to be rapidly prototyped, evaluated, and used. We will support both SMIRNOFF (Subaim 1.1) and GCNN (Subaim 1.2) type perception and parameter assignment modalities. With ready access to both built-in and community packages of optimizers, we will compare the performance and suitability of our current parameter optimization strategies with those popular in ML (such as Adam, stochastic gradient descent, and others[153]) and new optimizers based on Langevin dynamics.[154] To enable users to rapidly tailor our force fields with additional quantum chemical data for their chemistries of interest, we will assess various strategies for tailoring our released force fields using extensive parallelization and accelerator hardware. Our current workflow for condensed phase simulations uses OpenMM—which cannot take full advantage of modern GPUs for small condensed-phase systems and does not support analytical gradient computation in all parameters. We will therefore explore the potential to significantly enhance throughput by using the ML framework to pack many simulations onto each GPU and analytically compute any desired parameter gradients during optimization. Finally, we will ensure our tools are rapidly installable (via conda-forge), deployable, scalable, and well-documented with tutorials, so the community can rapidly experiment with force field parameterization and assessment.

Subaim 3.2: Investigate the use of surrogate models to reduce the number of molecular simulations required for force field parameterization and to enable Bayesian inference.

To fit FF parameters against ensemble-averaged equilibrium properties, such as densities of organic liquids, we need to run hundreds to thousands of molecular simulations. Some properties, such as density or speed of sound, require only a single simulation while others, such as transfer free energies or activity coefficients, require multiple simulations. However, performing Bayesian analysis requires stochastic sampling over FF parameters, repeating this process thousands of times. A standard machine learning approach to overcome this problem is to develop a surrogate model which approximates the simulation results at far lower computational cost, and to integrate this model with full simulations to create a rigorous, efficient, multifidelity approach to parameter optimization.

As described in the Progress Report, we have shown that, by using adaptively-expanded Gaussian process surrogates, we can perform much more complex and thorough optimizations, finding parameter combinations that give as good or better results than those afforded by the local gradient descent with simulations alone,[26] with similar computational cost. We will integrate surrogate models into the ML framework, so that we can perform both optimization and Bayesian sampling with such a multifidelity approach.

Traditional FF parameterization seeks a single model that best fits the training data. However, a given training set may not tightly constrain all parameters, particularly when parameters co-vary. For example, different combinations of ϵ and σ LJ parameters can yield equally accurate hydration free energies.[155;156] Importantly, even if two models

are equally consistent with training data, they may lead to very different predictions of observables not well-represented during training. Thus, Bayesian approaches will allow us to understand how uncertainties in the parameters can propagate to uncertainties in predictions.

Bayesian inference also allows comparison of statistical evidence for different families of force field models via Bayes factors,[157] with data-driven decisions for each modeling choice, even if model families possess different numbers of parameters. We have demonstrated this approach with model systems,[27] and will use this strategy to investigate the statistical support for alternative van der Waals functional forms, partial charge models, selective multipoles, selective off-center charges, site polarizability models, and deep learning models for valence terms.

Subaim 3.3: Incorporate machine-learned potentials fast enough for large-scale biomolecular simulation

While molecular mechanics force fields have traditionally used low-order Taylor and Fourier expansions to model valence interactions,[62;63] the significant limitations imposed by these approximations have inspired the community to look toward more flexible approximations to these important quantum chemical contributions to molecular potential energies. Recently, there has been much excitement around the rapid progress in using machine learning (ML) potentials that combine featurization of local atomic environments with flexible neural networks to model valence energies.[127;148–150;158;159] Inspired by pioneering work on hybrid physics/ML potentials,[159–162] we will explore the use of short-range ML potentials to model valence interactions, replacing Taylor and Fourier expansion terms, in combination with standard physical models of nonbonded (e.g., van der Waals and electrostatic) interactions.

Preliminary data: We have extensively experimented with the ANI2x[149] ML potential, illustrating how replacing standard ligand MM valence terms with ML valence terms can significantly improve accuracy in benchmark protein:ligand alchemical free energy calculations.[163] In addition, we have shown how condensed-phase properties can be used together with quantum chemical data to optimize ML potentials.[164;165] In collaboration with OpenMM, we have demonstrated how force evaluations for ML potentials as SchNet[127] and ANI[148;149] can be accelerated to within $\sim$2.5 the speed of GPU-accelerated MM for modeling small molecules in solvated protein:ligand systems.[166]

Using our ML fitting framework (Subaim 3.1), we will explore strategies that replace the valence terms based on Taylor (harmonic bonds, angles) and Fourier expansions (periodic torsions) with ML potentials, with a focus on fast ML potentials that offer competitive performance with MM,[166] such as SchNet[127] and ANI[148;149] and others that offer good compromises of simplicity, speed, and accuracy. We will build on our preliminary work which simultaneously fit to quantum chemical and ML valence with physical Lennard-Jones and electrostatics models.[163]

Aim 4: Build force field communities to maximize impact in biophysics and drug discovery

Community infrastructure accelerates force field science

Our high quality, well-documented tools accelerate community force field science. For example, Danny Cole's group built an entire self-consistent force field that replaced the LJ 12-6 functional form with a double exponential (DEXP) functional form[105] in a matter of weeks using our infrastructure, then proceeded to test it on hydration free energies, an effort which would have taken years of human time before OpenFF. Similarly, Brian Space's group uses OpenFF infrastructure for novel force fields for metal organic frameworks (see Letters)[106] Thus, our infrastructure already enables exploration and innovation far beyond what we can achieve ourselves.

Our infrastructure drives community innovation, which folds back into setting our future directions; for example, if the proof-of-concept work on DEXP force fields (**Figure 6**) continues to show value, this may drive mainline force fields to move in this direction because of benefits for free energy calculations. Similarly, our proof-of-concept work using off-site charges[19] is slated for inclusion into a future force field release after 3.0.

Overall, this approach allows the community to rapidly prototype, implement, and de-risk ideas before they are brought into full production. These early successes already demonstrate the value of nucleating community, and our community-building efforts will thus focus in three main areas detailed below.

Subaim 4.1: Build communities of software infrastructure

To continue leveraging quality, fully open-source, software infrastructure, we will build communities around this infrastructure. To some extent, this is already well underway given the successes described above. To further extend our software community, we will build a regular community-facing seminar where we invite people using our tools and infrastructure to talk about their successes, failures, results, and future directions, facilitating exchange of ideas and further collaboration around our infrastructure, while helping build community and accelerating progress.

We will also help the community use the OpenFF network to grow and expand their projects and build partnerships. Many innovators don't have access to a broad range of potentially interested parties and collaborators who want to put their innovations to use. OpenFF has a network of pharmaceutical industry users/collaborators who are

keen to take advantage of new innovations and we will use this to grow and expand the community around our projects and infrastructure. As a concrete example, Micaela Matta (King's College London, see Letters) was applying for funding for a Ph.D. student to use OpenFF infrastructure to do innovative new work on biomaterials, but needed an industry partner for the work; we used our network to circulate a white paper and connect her with collaborators at Janssen Pharmaceuticals so her project could be funded and move forward, further expanding our software/science community. We plan to formalize a process for circulating collaboration proposals to our network.

OpenFF also maintains a strong connection with the Open Molecular Software Foundation (OMSF), a 501(c)(3) nonprofit focusing on building high quality, open source software and sustainable communities for research software development. OMSF will further help build our community; for example, it also houses the Open Free Energy project (OpenFE) which takes advantage of our tools and infrastructure to deploy scalable binding free energy calculations. We plan to continue to partner with OMSF to help grow the relevant community.

An important way to grow our community is better interoperation with diverse, widely used, molecular simulation engines, such as OpenMM, AMBER, GROMACS, CHARMM, NAMD, and LAMMPS. Thus, we will enable export from OpenFF tools to a wider range of simulation engines and create tools to interoperate with automated workflows; our Interchange framework already is helping in this regard and we plan to continue its development. We will also develop tools to put other force fields into the OpenFF workflow.

Subaim 4.2: Build communities of data

We will continue to assemble larger and more diverse data sets, including quantum mechanical, physical property, and organic crystal data sets, as well as the tools for accessing them.

We will continue to construct an open, comprehensive QC database of high-accuracy, *ab initio* datasets for parameterizing small molecules and biomolecules. Force fields generally rely on QC calculations for determining torsional potentials, valence terms, and partial charges. Typically, these datasets have been produced at great cost, used once, and then discarded—despite their considerable value to the community for force field improvements, new force field efforts, and machine learning research.[148;167–169] However, in OpenFF we now store all such data and make it easily reusable via QCArchive so it is available for diverse applications, not just our own. Already, we have made a large number of QC geometries and energies and torsional potential energy scans available for a wide variety of pharmaceutically relevant small molecules and biomolecules.

In particular, drawing on community expertise, we will curate and extend our QC datasets to provide enhanced coverage of nucleic acid and lipid building blocks and structures, much as we've done for protein building blocks.[35] This will allow us to build, for example, high quality lipid and nucleic acid force fields, when used in combination with experimental data. Additionally, as we introduce and improve small molecule parameters, we will continue our current approach[3] of systematically extending our small molecule training data to ensure each individual parameter is trained across a broad range of chemical contexts that exercise that parameter. These data will be available to the molecular modeling community, and take advantage of input from that community. We will also curate extensive experimental datasets, including novel data types as discussed in Aim 2, including those which are too expensive to use in parameterization, and provide tools for assessing biomolecular force fields. As described in Subaim 2.1, we will integrate new property computation plug-ins into our physical property computation pipeline for each new property class, documenting best practices.

Subaim 4.3: Create communities of validation

We will build on our success in developing community best practices for protein force field validation[34] to establish best practices in validation for nucleic acids, lipids, sugars, and other biologically important molecules. We will convene experts from our scientific advisory board, and in the US and abroad to collate and annotate systems most useful for validation.

OpenFF already benefits from extensive communities of validation, especially from the pharmaceutical industry. Our partnership with the (separately funded) OpenFF Consortium means that new innovations which may impact near-term work in the drug discovery area are presented to the Consortium's advisory board for discussion and input, and those innovations which are the most promising are potentially rolled into our mainline force fields and then see real-world validation and use.[15] This work is so important to industry that industry collaborators often test tools and infrastructure during alpha and beta testing stages; for example, Cresset was a key early tester of our BespokeFit technology.[36]

References

[1] Qiu Y, Smith DGA, Boothroyd S, Jang H, Hahn DF, Wagner J, Bannan CC, Gokey T, Lim VT, Stern CD, Rizzi A, Tjanaka B, Tresadern G, Lucas X, Shirts MR, Gilson MK, Chodera JD, Bayly CI, Mobley DL, Wang LP. Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. J Chem Theory Comput. 2021; 17(10):6262–6280. doi:10.1021/acs.jctc.1c00571.

[2] Mobley DL, Bannan CC, Rizzi A, Bayly CI, Chodera JD, Lim VT, Lim NM, Beauchamp KA, Slochower DR, Shirts MR, Gilson MK, Eastman PK. Escaping Atom Types in Force Fields Using Direct Chemical Perception. J Chem Theory Comput. 2018; 14(11):6076–6092. doi:10.1021/acs.jctc.8b00640.

[3] Boothroyd S, Behara PK, Madin O, Hahn D, Jang H, Gapsys V, Wagner J, Horton J, Dotson D, Thompson M, Maat J, Gokey T, Wang LP, Cole D, Gilson M, Chodera J, Bayly C, Shirts M, Mobley D. Development and Benchmarking of Open Force Field 2.0.0 — the Sage Small Molecule Force Field. . 2022; doi:10.26434/chemrxiv-2022-n2z1c.

[4] Wang B, He X, Merz KM. Quantum Mechanical Study of Vicinal J Spin–Spin Coupling Constants for the Protein Backbone. J Chem Theory Comput. 2013; 9(10):4653–4659. doi:10.1021/ct400631b.

[5] Smith DGA, Altarawy D, Burns LA, Welborn M, Naden LN, Ward L, Ellis S, Pritchard BP, Crawford TD. The MolSSI QCArchive Project: An Open-Source Platform to Compute, Organize, and Share Quantum Chemistry Data. WIREs Comput Mol Sci. 2021; 11(2):e1491. doi:10.1002/wcms.1491.

[6] Boothroyd S, Wang LP, Mobley DL, Chodera JD, Shirts MR. Open Force Field Evaluator: An Automated, Efficient, and Scalable Framework for the Estimation of Physical Properties from Molecular Simulation. J Chem Theory Comput. 2022; 18(6):3566–3576. doi:10.1021/acs.jctc.1c01111.

[7] Frenkel M, Chiroco RD, Diky V, Dong Q, Marsh KN, Dymond JH, Wakeham WA, Stein SE, Königsberger E, Goodwin ARH. XML-based IUPAC standard for experimental, predicted, and critically evaluated thermodynamic property data storage and capture (ThermoML) (IUPAC Recommendations 2006). Pure Appl Chem. 2006; 78(3):541–612. doi:10.1351/pac200678030541.

[8] Frenkel M, Chirico RD, Diky VV, Dong Q, Frenkel S, Franchois PR, Embry DL, Teague TL, Marsh KN, Wilholt RC. ThermoML—An XML-based approach for storage and exchange of experimental and critically valuated thermophysical and thermochemical property data. 1. Experimental data. J Chem Eng Data. 2003; 48:2–13.

[9] Chirico RD, Frenkel M, Diky VV. ThermoML—An XML-based approach for storage and exchange of experimental and critically valuated thermophysical and thermochemical property data. 2, Uncertainties. J Chem Eng Data. 2003; 48:1344–1359.

[10] Boothroyd S, Madin OC, Mobley DL, Wang LP, Chodera JD, Shirts MR. Improving Force Field Accuracy by Training against Condensed-Phase Mixture Properties. J Chem Theory Comput. 2022; 18(6):3577–3592. doi:10.1021/acs.jctc.1c01268.

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

[12] Wang Y, Fass J, Kaminow B, Herr JE, Rufa D, Zhang I, Pulido I, Henry M, Macdonald HEB, Takaba K, et al. End-to-end differentiable construction of molecular mechanics force fields. Chem Sci. 2022; 13(41):12016–12033.

[13] Zanette C, Bannan CC, Bayly CI, Fass J, Gilson MK, Shirts MR, Chodera JD, Mobley DL. Toward Learned Chemical Perception of Force Field Typing Rules. J Chem Theory Comput. 2019; 15(1):402–423. doi:10.1021/acs.jctc.8b00821.

[14] Wang Y, Pulido I, Takaba K, Kaminow B, Scheen J, Wang L, Chodera JD, EspalomaCharge: Machine Learning-Enabled Ultra-Fast Partial Charge Assignment. arXiv; 2023. doi:10.48550/arXiv.2302.06758.

[15] D'Amore L, Hahn DF, Dotson DL, Horton JT, Anwar J, Craig I, Fox T, Gobbi A, Lakkaraju SK, Lucas X, Meier K, Mobley DL, Narayanan A, Schindler CEM, Swope WC, in 't Veld PJ, Wagner J, Xue B, Tresadern G. Collaborative Assessment of Molecular Geometries and Energies from the Open Force Field. J Chem Inf Model. 2022; 62(23):6094–6104. doi:10.1021/acs.jcim.2c01185.

[16] Eastman P, Behara PK, Dotson DL, Galvelis R, Herr JE, Horton JT, Mao Y, Chodera JD, Pritchard BP, Wang Y, et al. SPICE, A Dataset of Drug-like Molecules and Peptides for Training Machine Learning Potentials. Sci Data. 2023; 10(1):11.

[17] 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; 137(7):2695–2703. doi:10.1021/ja512751q.

[18] Behara PK, Jang H, Horton J, Dotson D, Boothroyd S, Cavender C, Gapsys V, Gokey T, Hahn D, Maat J, Madin O, Pulido I, Thompson M, Wagner J, Wang L, Chodera J, Cole D, Gilson M, Shirts M, Bayly C, et al., Benchmarking QM Theory for Drug-like Molecules to Train Force Fields; 2022. doi:10.5281/zenodo.7548777.
[19] Mobley D, Wagner J, Wang L, Nolasco D, 5th Open Force Field Workshop; 2022. doi:10.5281/zenodo.6774452.
[20] Marenich AV, Kelly CP, Thompson JD, Hawkins GD, Chambers CC, Giesen DJ, Winget P, Cramer CJ, Truhlar DG, Minnesota Solvation Database Home Page; 2012. https://comp.chem.umn.edu/mnsol/.
[21] Gražulis S, Chateigner D, Downs RT, Yokochi AFT, Quirós M, Lutterotti L, Manakova E, Butkus J, Moeck P, Le Bail A. Crystallography Open Database – an Open-Access Collection of Crystal Structures. J Appl Cryst. 2009; 42(4):726–729. doi:10.1107/S0021889809016690.
[22] Vaitkus A, Merkys A, Gražulis S. Validation of the Crystallography Open Database Using the Crystallographic Information Framework. J Appl Cryst. 2021; 54(2):661–672. doi:10.1107/S1600576720016532.
[23] Hüfner T, Gilson M, Probing Forcefields Using Small Molecule Crystal Structures; 2023. doi:10.5281/zenodo.7693780.
[24] Gokey T, Mobley D, Optimizing General Molecular Mechanics Force Fields in Chemical Space Using Binary-Encoded SMARTS; 2023. doi:10.5281/zenodo.7700973.
[25] Schauperl M, Kantonen SM, Wang LP, Gilson MK. Data-Driven Analysis of the Number of Lennard–Jones Types Needed in a Force Field. Commun Chem. 2020; 3(1):1–12. doi:10.1038/s42004-020-00395-w.
[26] Madin O, Shirts M, Using Physical Property Surrogate Models to Perform Multi-Fidelity Global Optimization of Force Field Parameters. ChemRxiv; 2022. doi:10.26434/chemrxiv-2022-7bmzv-v2.
[27] Madin OC, Boothroyd S, Messerly RA, Fass J, Chodera JD, Shirts MR. Bayesian-Inference-Driven Model Parametrization and Model Selection for 2CLJQ Fluid Models. J Chem Inf Model. 2022; 62(4):874–889. doi:10.1021/acs.jcim.1c00829.
[28] Mobley DL, Bannan CC, Rizzi A, Bayly CI, Chodera JD, Lim VT, Lim NM, Beauchamp KA, Slochower DR, Shirts MR, et al. Escaping Atom Types in Force Fields Using Direct Chemical Perception. J Chem Theory Comput. 2018; 14(11):6076–6092.
[29] OpenFF Force Field Toolkit download count from conda-forge;. https://anaconda.org/conda-forge/openff-toolkit/badges.
[30] OpenFF Force Field force field releases download count from conda-forge;. https://anaconda.org/conda-forge/openff-forcefields/badges.
[31] Ulmer TS, Ramirez BE, Delaglio F, Bax A. Evaluation of Backbone Proton Positions and Dynamics in a Small Protein by Liquid Crystal NMR Spectroscopy. J Am Chem Soc. 2003; 125(30):9179–9191. doi:10.1021/ja0350684.
[32] Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. J Comput Chem. 2000; 21(2):132–146.
[33] OpenEye Toolkit theory: ELF Conformer Selection;. https://docs.eyesopen.com/toolkits/python/quacpactk/molchargetheory.html#elf-conformer-selection.
[34] Cavender CE, Behara PK, Boothroyd S, Dotson DL, Gokey T, Horton JT, Mitchell JA, Pulido IJ, Thompson MW, Wagner J, Wang L, Chodera JD, Cole DJ, Mobley DL, Shirts MR, Gilson MK, Development of a Self-Consistent Force Field for Proteins and Small Molecules from the Open Force Field Initiative; 2023. doi:10.5281/zenodo.7694828.
[35] Cavender CE, Case DA, Chen JCH, Chong LT, Keedy DA, Lindorff-Larsen K, Mobley DL, Ollila OHS, Oostenbrink C, Robustelli P, Voelz VA, Wall ME, Wych DC, Gilson MK. Structure-Based Experimental Datasets for Benchmarking of Protein Simulation Force Fields. Zenodo; 2023. doi:10.5281/zenodo.7697462.
[36] Horton JT, Boothroyd S, Wagner J, Mitchell JA, Gokey T, Dotson DL, Behara PK, Ramaswamy VK, Mackey M, Chodera JD, Anwar J, Mobley DL, Cole DJ. Open Force Field BespokeFit: Automating Bespoke Torsion Parametrization at Scale. J Chem Inf Model. 2022; 62(22):5622–5633. doi:10.1021/acs.jcim.2c01153.
[37] Dror RO, Dirks RM, Grossman J, Xu H, Shaw DE. Biomolecular simulation: a computational microscope for molecular biology. Annu Rev Biophys. 2012; 41:429–452.
[38] Bottaro S, Lindorff-Larsen K. Biophysical experiments and biomolecular simulations: A perfect match? Science. 2018; 361(6400):355–360.
[39] Huggins DJ, Biggin PC, Dämgen MA, Essex JW, Harris SA, Henchman RH, Khalid S, Kuzmanic A, Laughton CA, Michel J, et al. Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2019; 9(3):e1393.
[40] Schindler CE, Baumann H, Blum A, Böse D, Buchstaller HP, Burgdorf L, Cappel D, Chekler E, Czodrowski

P, Dorsch D, et al. Large-scale assessment of binding free energy calculations in active drug discovery projects. J Chem Info Model. 2020; 60(11):5457–5474.
[41] Kuzmanic A, Bowman GR, Juarez-Jimenez J, Michel J, Gervasio FL. Investigating cryptic binding sites by molecular dynamics simulations. Accounts Chem Res. 2020; 53(3):654–661.
[42] Salo-Ahen OM, Alanko I, Bhadane R, Bonvin AM, Honorato RV, Hossain S, Juffer AH, Kabedev A, Lahtela-Kakkonen M, Larsen AS, et al. Molecular dynamics simulations in drug discovery and pharmaceutical development. Processes. 2020; 9(1):71.
[43] Albanese SK, Chodera JD, Volkamer A, Keng S, Abel R, Wang L. Is structure-based drug design ready for selectivity optimization? J Chem Info Model. 2020; 60(12):6211–6227.
[44] Hauser K, Negron C, Albanese SK, Ray S, Steinbrecher T, Abel R, Chodera JD, Wang L. Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. Comm Bio. 2018; 1(1):70.
[45] Bottaro S, Bengtsen T, Lindorff-Larsen K. Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach. Structural bioinformatics: methods and protocols. 2020; p. 219–240.
[46] Karplus M, Kuriyan J. Molecular dynamics and protein function. Proc Nat Acad Sci USA. 2005; 102(19):6679–6685.
[47] Ackloo S, Al-Awar R, Amaro RE, Arrowsmith CH, Azevedo H, Batey RA, Bengio Y, Betz UA, Bologa CG, Chodera JD, et al. CACHE (Critical Assessment of Computational Hit-finding Experiments): A public–private partnership benchmarking initiative to enable the development of computational methods for hit-finding. Nat Rev Chem. 2022; 6(4):287–295.
[48] Hagler AT. Force Field Development Phase II: Relaxation of Physics-Based Criteria... or Inclusion of More Rigorous Physics into the Representation of Molecular Energetics. J Comput Aided Mol Des. 2019; 33(2):205–264. doi:10.1007/s10822-018-0134-x.
[49] Muddana HS, Fenley AT, Mobley DL, Gilson MK. The SAMPL4 host–guest blind prediction challenge: an overview. J Comput-Aid Mol Des. 2014; 28:305–317.
[50] Yin J, Henriksen NM, Slochower DR, Shirts MR, Chiu MW, Mobley DL, Gilson MK. Overview of the SAMPL5 host–guest challenge: Are we doing better? J Comput-Aid Mol Des. 2017; 31:1–19.
[51] Amezcua M, El Khoury L, Mobley DL. SAMPL7 Host–Guest Challenge Overview: Assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations. J Comput-Aid Mol Des. 2021; 35(1):1–35.
[52] Amezcua M, Setiadi J, Ge Y, Mobley DL. An overview of the SAMPL8 host–guest binding challenge. J Comput-Aid Mol Des. 2022; 36(10):707–734.
[53] Galindo-Murillo R, Robertson JC, Zgarbová M, Šponer J, Otyepka M, Jurečka P, Cheatham TE. Assessing the Current State of Amber Force Field Modifications for DNA. J Chem Theory Comput. 2016; 12(8):4114–4127. doi:10.1021/acs.jctc.6b00186.
[54] Tan D, Piana S, Dirks RM, Shaw DE. RNA Force Field with Accuracy Comparable to State-of-the-Art Protein Force Fields. Proc Natl Acad Sci USA. 2018; p. 201713027. doi:10.1073/pnas.1713027115.
[55] Lin FY, Lopes PEM, Harder E, Roux B, MacKerell AD. Polarizable Force Field for Molecular Ions Based on the Classical Drude Oscillator. J Chem Inf Model. 2018; doi:10.1021/acs.jcim.8b00132.
[56] Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. J Comput Chem. 2008; 29(4):622–655. doi:10.1002/jcc.20820.
[57] MacKerell Jr AD, Nilsson L. Molecular dynamics simulations of nucleic acid–protein complexes. Curr Opin Struc Biol. 2008; 18(2):194–199.
[58] Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. Proc Nat Acad Sci USA. 2017; 114(28):7260–7265.
[59] Palermo G, Ricci CG, Chen JS, Miao Y, Jinek M, Doudna JA, McCammon JA. Molecular Mechanism of Off-Target Effects in CRISPR-Cas9. Biophys J. 2019; 116(3):319a.
[60] Nerenberg PS, Head-Gordon T. New Developments in Force Fields for Biomolecular Simulations. Current Opinion in Structural Biology. 2018; 49:129–138. doi:10.1016/j.sbi.2018.02.002.
[61] Robustelli P, Piana S, Shaw DE. Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. Proc Natl Acad Sci. 2018; 115(21):E4758–E4766. doi:10.1073/pnas.1800690115.
[62] Dauber-Osguthorpe P, Hagler AT. Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there? J Comput-Aid Mol Des. 2019; 33(2):133–203.
[63] Hagler AT. Force field development phase II: Relaxation of physics-based criteria... or inclusion of more

rigorous physics into the representation of molecular energetics. J Comput-Aid Mol Des. 2019; 33(2):205–264.
[64] Rizzi F, Najm HN, Debusschere BJ, Sargsyan K, Salloum M, Adalsteinsson H, Knio OM. Uncertainty quantification in MD simulations. Part II: Bayesian inference of force-field parameters. Multiscale Model Sim. 2012; 10(4):1460–1492.
[65] Angelikopoulos P, Papadimitriou C, Koumoutsakos P. Bayesian uncertainty quantification and propagation in molecular dynamics simulations: a high performance computing framework. J Chem Phys. 2012; 137(14):144103.
[66] Wu S, Angelikopoulos P, Papadimitriou C, Moser R, Koumoutsakos P. A hierarchical Bayesian framework for force field selection in molecular dynamics simulations. Philos T R Soc A. 2016; 374(2060):20150032.
[67] Messerly RA, Shirts MR, Kazakov AF. Uncertainty quantification confirms unreliable extrapolation toward high pressures for united-atom Mie λ-6 force field. J Chem Phys. 2018; 149(11):114109.
[68] Klauda JB, Venable RM, MacKerell Jr AD, Pastor RW. Considerations for lipid force field development. Curr top membranes. 2008; 60:1–48.
[69] Pastor R, MacKerell Jr A. Development of the CHARMM force field for lipids. J Phys Chem Lett. 2011; 2(13):1526–1532.
[70] Dickson CJ, Madej BD, Skjevik ÅA, Betz RM, Teigen K, Gould IR, Walker RC. Lipid14: the amber lipid force field. J Chem Theory Comput. 2014; 10(2):865–879.
[71] Poger D, Caron B, Mark AE. Validating lipid force fields against experimental data: Progress, challenges and perspectives. BBA-Biomembranes. 2016; 1858(7):1556–1565.
[72] Leonard AN, Wang E, Monje-Galvan V, Klauda JB. Developing and testing of lipid force fields with applications to modeling cellular membranes. Chem rev. 2019; 119(9):6227–6269.
[73] Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc. 1984; 106(3):765–784.
[74] MacKerell Jr AD, Banavali N, Foloppe N. Development and current status of the CHARMM force field for nucleic acids. Biopolymers: original Research on biomolecules. 2000; 56(4):257–265.
[75] Šponer J, Cang X, Cheatham III TE. Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures. Methods. 2012; 57(1):25–39.
[76] Cheatham III TE, Case DA. Twenty-five years of nucleic acid simulations. Biopolymers. 2013; 99(12):969–977.
[77] Galindo-Murillo R, Robertson JC, Zgarbova M, Sponer J, Otyepka M, Jurecka P, Cheatham III TE. Assessing the current state of amber force field modifications for DNA. J Chem Theory Comput. 2016; 12(8):4114–4127.
[78] Ricci CG, de Andrade AS, Mottin M, Netz PA. Molecular dynamics of DNA: comparison of force fields and terminal nucleotide definitions. J Phys Chem B. 2010; 114(30):9882–9893.
[79] Reif MM, Winger M, Oostenbrink C. Testing of the GROMOS force-field parameter set 54A8: structural properties of electrolyte solutions, lipid bilayers, and proteins. J Chem Theory Comput. 2013; 9(2):1247–1264.
[80] Robertson MJ, Tirado-Rives J, Jorgensen WL. Improved treatment of nucleosides and nucleotides in the OPLS-AA force field. Chem Phys Lett. 2017; 683:276–280.
[81] Robertson MJ, Qian Y, Robinson MC, Tirado-Rives J, Jorgensen WL. Development and Testing of the OPLS-AA/M Force Field for RNA. J Chem Theory Comput. 2019; 15(4):2734–2742.
[82] Guvench O, Mallajosyula SS, Raman EP, Hatcher E, Vanommeslaeghe K, Foster TJ, Jamison FW, MacKerell Jr AD. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling. J Chem Theory Comput. 2011; 7(10):3162–3180.
[83] Kirschner KN, Lins RD, Maass A, Soares TA. A glycam-based force field for simulations of lipopolysaccharide membranes: parametrization and validation. J Chem Theory Comput. 2012; 8(11):4719–4731.
[84] Foley BL, Tessier MB, Woods RJ. Carbohydrate force fields. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2012; 2(4):652–697.
[85] Th urlemann M, oselt L, Riniker S. Regularized by Physics: Graph Neural Network Parametrized Potentials for the Description of Intermolecular Interactions. J Chem Theory Comput. 2023; .
[86] Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. Chem Mater. 2019; 31(9):3564–3572. doi:10.1021/acs.chemmater.9b01294.
[87] Chen C, Ong SP. A Universal Graph Deep Learning Interatomic Potential for the Periodic Table. Nat Comput Sci. 2022; 2(11):718–728. doi:10.1038/s43588-022-00349-3.
[88] Kusaba M, Liu C, Koyama Y, Terakura K, Yoshida R. Recreation of the Periodic Table with an Unsupervised

Machine Learning Algorithm. Sci Rep. 2021; 11(1):4780. doi:10.1038/s41598-021-81850-z.
[89] Hao Z, Lu C, Huang Z, Wang H, Hu Z, Liu Q, Chen E, Lee C. ASGN: An active semi-supervised graph neural network for molecular property prediction. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*; 2020. p. 731–752.
[90] Matlock MK, Hoffman M, Dang NL, Folmsbee DL, Langkamp LA, Hutchison GR, Kumar N, Sarullo K, Swamidass SJ. Deep Learning Coordinate-Free Quantum Chemistry. J Phys Chem A. 2021; 125(40):8978–8986.
[91] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79(2):926–935.
[92] Yin J, Henriksen NM, Muddana HS, Gilson MK. Bind3P: Optimization of a Water Model Based on Host–Guest Binding Data. J Chem Theory Comput. 2018; 14(7):3621–3632. doi:10.1021/acs.jctc.8b00318.
[93] Harding MM, Nowicki MW, Walkinshaw MD. Metals in protein structures: a review of their principal features. Crystallograp Rev. 2010; 16(4):247–302.
[94] Tran JB, Krezel A. InterMetalDB: a database and browser of intermolecular metal binding sites in macromolecules with structural information. J Proteome Res. 2021; 20(4):1889–1901.
[95] Peters MB, Yang Y, Wang B, Füsti-Molnár L, Weaver MN, Merz KMJ. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). J Chem Theory Comput. 2010; 6(9):2935–2947. doi:10.1021/ct1002626.
[96] Li P, Song LF, Merz Jr KM. 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.
[97] Kantakevičius P, Mathiah C, Johannissen LO, Hay S. Chelator-Based Parameterization of the 12-6-4 Lennard-Jones Molecular Mechanics Potential for More Realistic Metal Ion–Protein Interactions. J Chem Theory Comput. 2022; 18(4):2367–2374.
[98] Ducháčková L, Roithová J. The Interaction of Zinc (II) and Hydroxamic Acids and a Metal-Triggered Lossen Rearrangement. Chemistry–A European Journal. 2009; 15(48):13399–13405.
[99] Biological Magnetic Resonance Bank;. http://brmb.wisc.edu.
[100] Beauchamp KA, Lin YS, Das R, Pande VS. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. J Chem Theory Comp. 2012; 8(4):1409–1414.
[101] Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, Jr ADM. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. Nat Methods. 2017; 14(1):71–73. doi:10.1038/nmeth.4067.
[102] Best RB, Zheng W, Mittal J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. J Chem Theory Comput. 2014; 10(11):5113–5124. doi:10.1021/ct500569b.
[103] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan Sa, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem. 1983; 4(2):187–217.
[104] Mayo SL, Olafson BD, Goddard WA. DREIDING: a generic force field for molecular simulations. J Phys Chem. 1990; 94(26):8897–8909.
[105] Horton J, Boothroyd S, Behara PK, Mobley D, Cole D, Functional Form Exploration: The Double Exponential Force Field; 2023. doi:10.5281/zenodo.7669506.
[106] Hogan A, Space B, Next-Generation Accurate, Transferable, and Polarizable Potentials for Materials Simulations; 2022. doi:10.5281/zenodo.7641569.
[107] Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, Wang LP, Simmonett AC, Harrigan MP, Stern CD, Wiewiora RP, Brooks BR, Pande VS. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. PLOS Comput Bio. 2017; 13(7):e1005659. doi:10.1371/journal.pcbi.1005659.
[108] Cole DJ, Vilseck JZ, Tirado-Rives J, Payne MC, Jorgensen WL. Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning. J Chem Theory Comput. 2016; 12(5):2312–2323. doi:10.1021/acs.jctc.6b00027.
[109] Wilson AL, Popelier PLA. Exponential Relationships Capturing Atomistic Short-Range Repulsion from the Interacting Quantum Atoms (IQA) Method. J Phys Chem A. 2016; 120(48):9647–9659. doi:10.1021/acs.jpca.6b10295.
[110] Elstner M, Hobza P, Frauenheim T, Suhai S, Kaxiras E. Hydrogen Bonding and Stacking Interactions of Nucleic Acid Base Pairs: A Density-Functional-Theory Based Treatment. J Chem Phys. 2001; 114(12):5149–5155. doi:10.1063/1.1329889.
[111] Allinger N. Calculation of molecular structure and energy by force-field methods. In: *Advances in physical organic chemistry*, vol. 13 Elsevier; 1976.p. 1–82.

[112] Beutler TC, Mark AE, van Schaik RC, Gerber PR, Van Gunsteren WF. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. Chem Phys Lett. 1994; 222(6):529–539.
[113] Urey HC. The Vibrations of Pentatonic Tetrahedral Molecules. Phys Rev. 1931; 38(11):1969–1978. doi:10.1103/PhysRev.38.1969.
[114] MacKerell ADJ, Bashford D, Bellott M, Dunbrack RLJ, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B. 1998; 102(18):3586–3616. doi:10.1021/jp973084f.
[115] Dinur U, Hagler AT. Geometry-Dependent Atomic Charges: Methodology and Application to Alkanes, Aldehydes, Ketones, and Amides. J Comput Chem. 1995; 16(2):154–170. doi:10.1002/jcc.540160204.
[116] Decius JC. An Effective Atomic Charge Model for Infrared Intensities. J Mol Spectrosc. 1975; 57(3):348–362. doi:10.1016/0022-2852(75)90296-9.
[117] Stern PS, Chorev M, Goodman M, Hagler AT. Computer Simulation of the Conformational Properties of Retro–Inverso Peptides. II. Ab Initio Study, Spatial Electron Distribution, and Population Analysis of N-formylglycine Methylamide, N-formyl N'-Acetyldiaminomethane, and N-methylmalonamide. Biopolymers. 1983; 22(8):1901–1917. doi:10.1002/bip.360220807.
[118] Cieplak P, Kollman P. On the Use of Electrostatic Potential Derived Charges in Molecular Mechanics Force Fields. The Relative Solvation Free Energy of Cis- and Trans-N-methyl-acetamide. J Comput Chem. 1991; 12(10):1232–1236. doi:10.1002/jcc.540121010.
[119] Stouch Tr, Williams DE. Conformational Dependence of Electrostatic Potential Derived Charges of a Lipid Headgroup: Glycerylphosphorylcholine. J Comput Chem. 1992; 13(5):622–632. doi:10.1002/jcc.540130512.
[120] Applequist J, Carl JR, Fung KK. Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and Determination of Atom Polarizabilities. J Am Chem Soc. 1972; 94(9):2952–2960. doi:10.1021/ja00764a010.
[121] Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T. Current Status of the AMOEBA Polarizable Force Field. J Phys Chem B. 2010; 114(8):2549–2564. doi:10.1021/jp910674d.
[122] Stern HA, Kaminski GA, Banks JL, Zhou R, Berne BJ, Friesner RA. Fluctuating Charge, Polarizable Dipole, and Combined Models: Parameterization from Ab Initio Quantum Chemistry. J Phys Chem B. 1999; 103(22):4730–4737. doi:10.1021/jp984498r.
[123] Rick SW, Berne BJ. Dynamical Fluctuating Charge Force Fields: The Aqueous Solvation of Amides. J Am Chem Soc. 1996; 118(3):672–679. doi:10.1021/ja952535b.
[124] 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):3091–3099. doi:10.1021/jp112133g.
[125] Smith JS, Isayev O, Roitberg AE. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. Chem Sci. 2017; 8(4):3192–3203. doi:10.1039/C6SC05720A.
[126] Unke OT, Meuwly M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. J Chem Theory Comput. 2019; 15(6):3678–3693. doi:10.1021/acs.jctc.9b00181.
[127] Schütt KT, Sauceda HE, Kindermans PJ, Tkatchenko A, Müller KR. Schnet–a deep learning architecture for molecules and materials. J Chem Phys. 2018; 148(24):241722.
[128] The ThermoML Archive;. http://trc.nist.gov/ThermoML.html.
[129] Ploetz EA, Karunaweera S, Bentenitis N, Chen F, Dai S, Gee MB, Jiao Y, Kang M, Kariyawasam NL, Naleem N, Weerasinghe S, Smith PE. Kirkwood–Buff-Derived Force Field for Peptides and Proteins: Philosophy and Development of KBFF20. J Chem Theory Comput. 2021; 17(5):2964–2990. doi:10.1021/acs.jctc.1c00075.
[130] Knowles DB, Shkel IA, Phan NM, Sternke M, Lingeman E, Cheng X, Cheng L, O'Connor K, Record MT. Chemical Interactions of Polyethylene Glycols (PEGs) and Glycerol with Protein Functional Groups: Applications to Effects of PEG and Glycerol on Protein Processes. Biochemistry. 2015; 54(22):3528–3542. doi:10.1021/acs.biochem.5b00246.
[131] Diehl RC, Guinn EJ, Capp MW, Tsodikov OV, Record MT. Quantifying Additive Interactions of the Osmolyte Proline with Individual Functional Groups of Proteins: Comparisons with Urea and Glycine Betaine, Interpretation of m-Values. Biochemistry. 2013; 52(35):5997–6010. doi:10.1021/bi400683y.
[132] Guinn EJ, Schwinefus JJ, Cha HK, McDevitt JL, Merker WE, Ritzer R, Muth GW, Engelsgjerd SW, Mangold KE, Thompson PJ, Kerins MJ, Record MTJ. Quantifying Functional Group Interactions That De-

termine Urea Effects on Nucleic Acid Helix Formation. J Am Chem Soc. 2013; 135(15):5828–5838. doi:10.1021/ja400965n.
[133] Mansouri K, Grulke CM, Richard AM, Judson RS, Williams AJ. An Automated Curation Procedure for Addressing Chemical Errors and Inconsistencies in Public Datasets Used in QSAR Modelling. SAR QSAR Environ Res. 2016; 27(11):911–937. doi:10.1080/1062936X.2016.1253611.
[134] Sangster J. Octanol-Water Partition Coefficients of Simple Organic Compounds. J Phys Chem Ref Data. 1989; 18(3):1111–1229. doi:10.1063/1.555833.
[135] Rekharsky MV, Inoue Y. Complexation Thermodynamics of Cyclodextrins. Chem Rev. 1998; 98(5):1875–1918.
[136] Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: A Web-Accessible Database of Experimentally Determined Protein–Ligand Binding Affinities. Nucleic Acids Res. 2007; 35(Database issue):D198–D201. doi:10.1093/nar/gkl999.
[137] Slochower DR, Henriksen NM, Wang LP, Chodera JD, Mobley DL, Gilson MK. Binding Thermodynamics of Host–Guest Systems with SMIRNOFF99Frosst 1.0.5 from the Open Force Field Initiative. J Chem Theory Comput. 2019; 15(11):6225–6242. doi:10.1021/acs.jctc.9b00748.
[138] Tembe BL, McCammon JA. Ligand-Receptor Interactions. Comput Chem. 1984; 8:281–283.
[139] Cournia Z, Allen B, Sherman W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. J Chem Inf Model. 2017; 57(12):2911–2937. doi:10.1021/acs.jcim.7b00564.
[140] Miyamoto S, Kollman PA. Absolute and Relative Binding Free Energy Calculations of the Interaction of Biotin and Its Analogs with Streptavidin Using Molecular Dynamics/Free Energy Perturbation Approaches. Prot Struct Func Gen. 1993; 16:226–245.
[141] Gilson MK, Given JA, Bush BL, McCammon JA. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. Biophys J. 1997; 72:1047–1069.
[142] Rocklin GJ, Chidyausiku TM, Goreshnik I, Ford A, Houliston S, Lemak A, Carter L, Ravichandran R, Mulligan VK, Chevalier A, et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. Science. 2017; 357(6347):168–175.
[143] Shalongo W, Dugad L, Stellwagen E. Distribution of Helicity within the Model Peptide Acetyl(AAQAA)(3)amide. J Am Chem Soc. 1994; 116(18):8288–8293. doi:–10.1021/ja00097a039˝.
[144] Streicher WW, Makhatadze GI. Unfolding thermodynamics of Trp-cage, a 20 residue miniprotein, studied by differential scanning calorimetry and circular dichroism spectroscopy. Biochemistry. 2007; 46(10):2876–2880. doi:10.1021/bi602424x.
[145] Davis CM, Xiao S, Raeigh DP, Dyer RB. Raising the Speed Limit for beta-Hairpin Formation. J Am Chem Soc. 2012; 134(35):14476–14482. doi:10.1021/ja3046734.
[146] Wang LP, Chen J, Van Voorhis T. Systematic parametrization of polarizable force fields from quantum chemistry data. J Chem Theory Comput. 2013; 9(1):452–460.
[147] Wang LP, Martinez TJ, Pande VS. Building force fields: An automatic, systematic, and reproducible approach. J Phys Chem Lett. 2014; 5(11):1885–1891.
[148] Smith JS, Isayev O, Roitberg AE. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. Chem Sci. 2017; 8(4):3192–3203.
[149] Devereux C, Smith JS, Huddleston KK, Barros K, Zubatyuk R, Isayev O, Roitberg AE. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. J Chem Theory Comput. 2020; 16(7):4192–4202.
[150] Batzner S, Musaelian A, Sun L, Geiger M, Mailoa JP, Kornbluth M, Molinari N, Smidt TE, Kozinsky B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. Nat commun. 2022; 13(1):2453.
[151] GitHub repository: Espaloma: Extensible Surrogate Potential Optimized by Message-passing Algorithms in PyTorch;. https://github.com/choderalab/espaloma.
[152] GitHub repository: EspalomaX: Extensible Surrogate Potential Optimized by Message-passing Algorithms in JAX;. https://github.com/choderalab/espalomax.
[153] Desai C. Comparative analysis of optimizers in deep neural networks. International Journal of Innovative Science and Research Technology. 2020; 5(10):959–962.
[154] Leimkuhler B, Matthews C, Vlaar T. Partitioned integrators for thermodynamic parameterization of neural networks. arXiv preprint arXiv:190811843. 2019; .
[155] Naden LN, Shirts MR. Rapid Computation of Thermodynamic Properties over Multidimensional Nonbonded Parameter Spaces Using Adaptive Multistate Reweighting. J Chem Theory Comput. 2016; 12(4):1806–1823.

doi:10.1021/acs.jctc.5b00869.
[156] Joung IS, Cheatham TE. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. J Phys Chem B. 2008; 112(30):9020–9041. doi:10.1021/jp8001614.
[157] Wasserman L. Bayesian model selection and model averaging. J Math Psych. 2000; 44(1):92–107.
[158] Pinheiro M, Ge F, Ferré N, Dral PO, Barbatti M. Choosing the right molecular machine learning potential. Chem Sci. 2021; 12(43):14396–14413.
[159] Unke OT, Meuwly M. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. J Chem Theory Comput. 2019; 15(6):3678–3693.
[160] Zubatyuk R, Smith JS, Leszczynski J, Isayev O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. Sci Adv. 2019; 5(8):eaav6490.
[161] Ko TW, Finkler JA, Goedecker S, Behler J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. Nat commun. 2021; 12(1):398.
[162] Zubatiuk T, Isayev O. Development of multimodal machine learning potentials: toward a physics-aware artificial intelligence. Accounts Chem Res. 2021; 54(7):1575–1585.
[163] Rufa DA, Bruce Macdonald HE, Fass J, Wieder M, Grinaway PB, Roitberg AE, Isayev O, Chodera JD. Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning/molecular mechanics potentials. BioRxiv. 2020; p. 2020–07.
[164] Wieder M, Fass J, Chodera JD. Fitting quantum machine learning potentials to experimental free energy data: predicting tautomer ratios in solution. Chem Sci. 2021; 12(34):11364–11381.
[165] Wieder M, Fass J, Chodera JD. Teaching free energy calculations to learn from experimental data. bioRxiv. 2021; p. 2021–08.
[166] Galvelis R, Varela-Rial A, Doerr S, Fino R, Eastman P, Markland TE, Chodera JD, De Fabritiis G. NNP/MM: Fast molecular dynamics simulations with machine learning potentials and molecular mechanics. arXiv preprint arXiv:220108110. 2022; .
[167] Botu V, Batra R, Chapman J, Ramprasad R. Machine learning force fields: construction, validation, and outlook. J Phys Chem C. 2016; 121(1):511–522.
[168] Chmiela S, Tkatchenko A, Sauceda HE, Poltavsky I, Schütt KT, Müller KR. Machine learning of accurate energy-conserving molecular force fields. Science Adv. 2017; 3(5):e1603015.
[169] Podryabinkin EV, Shapeev AV. Active learning of linearly parametrized interatomic potentials. Comput Mat Sci. 2017; 140:171–180.
[170] QCArchive: An open community database for quantum chemistry;. https://qcarchive.molssi.org.