JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Perspective

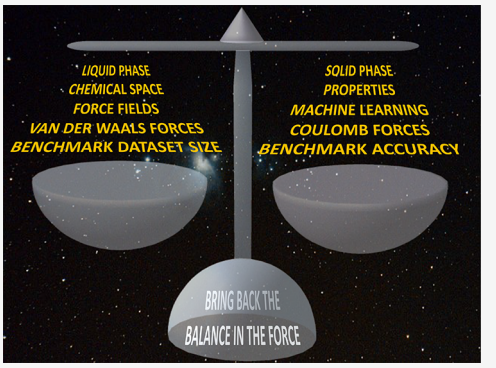
An Imbalance in the Force: The Need for Standardized Benchmarks for Molecular Simulation

Kristian Kříž, Lisa Schmidt, Alfred T. Andersson, Marie-Madeleine Walz, and David van der Spoel*

Cite This: J. Chem. Inf. Model. 2023, 63, 412−431 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Force fields (FFs) for molecular simulation have been under development for more than half a century. As with any predictive model, rigorous testing and comparisons of models critically depends on the availability of standardized data sets and benchmarks. While such benchmarks are rather common in the fields of quantum chemistry, this is not the case for empirical FFs. That is, few benchmarks are reused to evaluate FFs, and development teams rather use their own training and test sets. Here we present an overview of currently available tests and benchmarks for computational chemistry, focusing on organic compounds, including halogens and common ions, as FFs for these are the most common ones. We argue that many of the benchmark data sets from quantum chemistry can in fact be reused for evaluating FFs, but new gas phase data is still needed for compounds containing phosphorus and sulfur in different valence states. In addition, more nonequilibrium interaction energies and forces, as well as molecular properties such as electrostatic potentials around compounds, would be beneficial. For the condensed phases there is a large body of experimental data available, and tools to utilize these data in an automated fashion are under development. If FF developers, as well as researchers in artificial intelligence, would adopt a number of these data sets, it would become easier to compare the relative strengths and weaknesses of different models and to, eventually, restore the balance in the force.



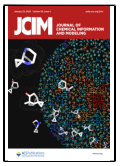
Downloaded via UPPSALA UNIV on February 26, 2026 at 15:28:34 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Over 50 years ago, Levitt and Lifson published one of the first energy calculations of a protein using a force field (FF).[1] These authors presented their work in a modest manner, mentioning that the FF used was a "gross approximation". Interestingly, the functional form of FFs used in most (bio)molecular simulations today is very similar to that used in the old paper. At the same time, the "real" functional form that a biomolecular FF would need is still considered unknown by some authors.[2] Here we will not dwell on the history of FF calculations but rather refer the reader to some excellent reviews on the topic by Dauber-Osguthorpe and Hagler.[3,4] In the second of these two reviews, Hagler describes how old lore has been forgotten and wheels reinvented. More in particular, he describes how different iterations of FF development start from different premises, adding new test systems and forgetting about older ones, which effectively leads to mending something while breaking something else.[4] The "loyalty" to the functional form of the FF potential from the 1960s[1] is remarkable, and Hagler is keen to point out that fixing torsion potentials cannot compensate for lacking physics in other parts of the model. Rather, he argues that more physics needs to be introduced in FFs.[4] For example, it has been known for a long time that the description of the repulsive part of the Lennard-Jones potential[5,6] is not very accurate and variations of the Buckingham potential[7] are to be preferred.[8−11]

Another important issue in this respect is the addition of explicit polarization, reviewed by Jing et al.,[12] as well as other approaches to improve the description of electrostatics[4,13] or dispersion interactions.[14−17] It has been shown that classical polarization models are able to capture the many-body interactions reasonably well.[18] However, the explicit inclusion of many-body interactions, that is, going beyond the pair potential by addition of three or three and four center potentials, improves the accuracy of a FF for predicting physicochemical properties for different phases to unprecedented levels. Much effort has been gone into water, where each molecule is considered a "body" in many-body potentials.[19−22] A somewhat more general approach was presented by Ströker et al. in a potential for argon including three-body dispersion[23] where a single argon atom is one "body". It is, therefore, fair to state that efforts to incorporate many-body effects are still limited to specific compounds or systems. Considerations on modeling solvents[24] as well as routes to systematic FF design have recently been reviewed elsewhere[25] and will not be discussed here.

Received: September 6, 2022
Published: January 11, 2023



© 2023 The Authors. Published by American Chemical Society

Two recent reviews compare FFs for small molecules. He and co-workers focus on the evaluation of free energies of solvation,[26] while Lewis-Atwell et al. describe how well the conformational energy landscape of small molecules in the gas phase is described by FFs.[27] Arguably, the main aim of a general FF is to predict the properties of molecules in any environment. It can be concluded from these two reviews, however, that different FFs are needed to predict different properties. This notion is implicitly confirmed by multiple force field development teams updating their routines for generating charges to take into account multiple dielectric environments. For instance, Schauperl et al. compute atomic charges using the restrained electrostatic potential (RESP) method[28,29] employing a polarizable continuum model (PCM) with two different dielectric constants, ε_r, one representing water ($\varepsilon_r \approx 80$) and one representing vacuum ($\varepsilon_r = 1$).[30] In their method, they then use a weighted sum of the charges computed in both dielectric environments. Similarly, Bleiziffer et al. used machine learning (ML) to predict charges for small molecules based on a data set computed using PCM with a dielectric constant $\varepsilon_r = 4$, a value that is somewhere between that of vacuum and that of water.[31] Rather than trying to generate charges that are a compromise between different environments, Hosseini and co-workers used one set of charges for a compound in aqueous solution and another one in a phospholipid membrane, when computing the potential of mean force for crossing the membrane.[32] These examples highlight that there is an imbalance in classical force fields that makes it difficult to accurately model compounds in different environments with one parameter set, unless polarizability is treated explicitly.[12,18] The same conclusion was drawn from benchmarks of liquids, where the enthalpy of vaporization of organic compounds (the calculation of which involves the gas-phase energy of a compound) was found to have a larger error in popular nonpolarizable force fields than pure condensed phase properties.[33,34]

Rather than using preconceived functional forms based on physically meaningful terms, ML can be used to derive a potential directly from data, using the relationship between input structure and reference values in the framework of the corresponding ML architecture.[35] ML in chemistry has advanced rapidly, e.g., see refs 36−56, and it has been questioned whether the profession of computational chemist has a future at all in the age of machine learning.[57] However, despite the rapid advance (or, rather, return[25]) of ML in computational modeling,[35] the laws of physics remain valid, and these can be used to improve the accuracy of predictive models.[58] Therefore, pursuing the holy grail of a FF that can be used in any environment remains a useful and even necessary proposition, even though the functional form for such a FF remains undecided.[2] To speed up progress in this pursuit, the whole field of computational chemistry would do well to overcome method-compartmentalization issues by introducing common standards of reference, i.e., benchmarks with a large variety of compounds in different phases. The challenges that are suggested to be solved by using standardized benchmarks are (1) avoiding cherry-picking of properties used for validation, (2) increased applicability range of force fields, (3) objective comparison of force fields, and (4) generation of well-balanced force fields. Obviously, standardized benchmarks need to be updated and extended regularly in order to prevent tuning models for the test set rather than a training set.

We are aware that there are many force fields specialized for different kinds of systems, for instance, for clays[59] or nanoparticles.[60] In this perspective, we focus our discussion on evaluations and benchmarks of general FFs that treat biomolecules and/or (small) organic compounds.[61,62] Halogen atoms need to be included since both drug compounds and pollutants[63] interacting with biomolecules often contain these and since there is lack of relevant data for some compound classes containing halogen atoms (such as perfluorinated compounds).[64] For reference, some of the most well-known FFs in this area are listed in Table 1. Obviously, the benchmark data sets can equally well be used for ML potentials, and indeed quite a few of the recently published data sets were developed with ML in mind as will be discussed in detail below.

Table 1. Some Recent Versions of Popular Force Fields

force field	Pol[a]	targets	refs
Amber ff14ipq	no	protein, RNA	68
Amber ff15ipq	no	protein, RNA	69
Amber ff14sb	no	protein, RNA	70
Amber ff19sb	no	protein, RNA	71
Amoeba	yes	protein, RNA, DNA	72, 73
GAFF	no	small molecules	74
GAFF2	no	small molecules	75
Charmm36	no	proteins, nucleic acids, lipids	76
CGenFF	no	small molecules	77
Charmm Drude	yes	protein, RNA	78
GROMOS 54A8	no	protein	79
GROMOS 2016H66	no	small molecules	80
TraPPE	no	liquids	81
OPLS-AA	no	liquids	82
OPLS3	no	drug-like small molecules, proteins	83
OPLS3e	no	drug-like small molecules	84
MM3,MM4	no	small molecules	85
MMFF94	no	small molecules, proteins	86−90
MMFF94S	no	small molecules, proteins	91, 92
Smirnoff99Frosst	no	small molecules	93
OpenFF-1.0, 1.1, 1.2, 2.0	no	small molecules	94
ReaxFF	yes	reactive systems	95

[a]Pol indicates whether polarizability is included in the FF explicitly.

Both FF development and ML within computational chemistry lean heavily on benchmarks from the field of electronic structure theory, because of the large amounts of data available. It should be noted, however, that the benchmarks used ultimately need to be scrutinized as well using experimental reference material.[65] Since FF methods are computationally cheap, they allow direct comparison with condensed phase properties based on experimental data,[33,34,66,67] circumventing potential uncertainties in quantum mechanical (QM) references.[65] In what follows we describe both the observables relevant for evaluating FFs and existing data sets or benchmarks. We start with physicochemical properties of monomers, then move to interaction energies of dimers and complexes, before addressing the condensed phases. The perspective ends with a discussion section containing recommendations for further work on development and application of data sets.

MONOMERIC COMPOUNDS

Thermochemistry. The study of chemical reactions using QM has led to the development of multiple methods to estimate thermochemistry values, such as the Gaussian-*n* methods,[96−100]

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

Table 2. Some of the Most Important Properties for Development and Validation of Force Fields and Training of Neural Networks

property	symbol	source[a]	importance
		Energetics	
energy relative to atoms	E	QM	for enthalpy of formation
excitation energy	E^*	QM	for reaction kinetics in reactive models
conformational energy relative to minimum energy	E_{conf}	QM	for intramolecular potentials
interaction energy in dimers and complexes	E_{int}	QM	for intermolecular interactions
force vector in nonequilibrium conformations or complexes	F	QM	for both intra- and intermolecular interactions
vibrational frequencies	ν	B	force constants and thermochemistry, dynamics
second virial coefficient	B	X	gas phase intermolecular interactions
		Thermochemistry	
enthalpy of formation	$\Delta_f H^\ominus$	B	validation of intramolecular interactions using experimental data
gas phase entropy	$S^\ominus$	B	for dynamics and free energy calculations
gas phase heat capacity at constant volume	C_V	B	for temperature dependence of (mainly) bonded potentials
		Electrostatics	
partial charges	Q	QM	to model electrostatic interactions
polarizability tensors	α	QM	to model polarization response
electrostatic potentials	V_{ESP}	QM	for training electrostatic models
dipole (or higher multipoles)	μ, θ, ...	B	validation of electrostatics
		Condensed Phase Properties	
lattice enthalpy	ΔH_{latt}	X	intermolecular interactions
enthalpy of vaporization, sublimation	ΔH_{vap}, ΔH_{sub}	X	intermolecular interactions and phase change
density	ρ	X	intermolecular interactions
solvation free energy	ΔG_{solv}	X	intermolecular interactions
heat capacities of liquid	C_V, C_P	X	temperature dependence of enthalpy
enthalpy of mixing	ΔH_{mix}	X	intermolecular interactions
excess molar volume of mixing	ΔV_{mix}	X	intermolecular interactions
melting point, boiling point	T_{melt}, T_{boil}	X	temperature dependence of intermolecular interactions and phase change
surface tension	γ	X	interface polarization and stiffness
dielectric constant	ε	X	balance between dynamics and interaction strength
viscosity and diffusion coefficient	η, D	X	temperature dependence of interaction strength
crystal structure coordinates and lattice parameters	r, a, b, c, α, β, γ	X	atomic radii as well as intermolecular interactions in the solid state

[a]The predominant source of data is indicated by either X (experiment), QM (quantum chemistry), or B (both).

the Weizmann methods,[101−105] and the complete basis set model chemistry[106,107] (not to be confused with extrapolation to the complete basis set limit discussed below). Of the properties these methods can predict, the standard enthalpy of formation $\Delta_f H^\ominus$, standard entropy $S^\ominus$, and heat capacity at constant volume C_V are particularly important for FF development. As per usual, the accuracy and computational cost of these methods vary significantly. The Gaussian-4 theory,[99] for example, was found to be a good compromise between cost and accuracy, reaching a root-mean-square deviation (RMSD) of 11−12 kJ/mol from experimental data for $\Delta_f H^\ominus$ for 600 compounds of up to 47 atoms.[108] Along with these method developments, benchmarks have been introduced that have been widely adopted, such as G3/05,[109,110] W4-17,[111] and GMTKN55.[112,113]

Reproduction of gas-phase properties such as the standard entropy $S^\ominus$ and heat capacity at constant volume C_V has been attempted using FFs as well with reasonable results[114] based on compounds from the Alexandria library.[108,115] FFs are known not to be very good at predicting vibrational frequencies.[114,116,117] In the specific case of united-atom force fields, a degree of error compensation can occur when computing the standard entropy $S^\ominus$ or heat capacity C_V, leading to results being quite good for the wrong reason.[80] Vibrational frequencies are indeed difficult to predict accurately by any computational chemistry methods, which has led to the introduction of scaled frequencies in quantum chemistry. Since this is beyond the scope of this perspective, we refer to Laury et al.[118] for an overview of frequency scaling factors. Beyond the paper mentioned,[114] the only attempts to predict thermochemical properties from classical FFs depend on *ad hoc* empirical corrections, such as in ref 119. Hence, experimental thermochemistry data presents an opportunity for improvements of FF as well as ML potentials, although the number of compounds for which this data is available is limited to about 7000.[120]

Molecular Structure. The lowest energy conformation of molecules, as determined experimentally or through high-level QM, provides yet another data point, since by definition the net forces on the atoms should be zero. Accurate structural data on bond lengths and covalent bond angles has traditionally been obtained from small molecule crystal structures,[121] and for instance, Allinger and co-workers have done much work on analyzing the conformational properties of small molecules.[85,122,123] The MM*x* FF family[85] developed by his group remains competitive for predicting these kinds of properties.[27] In another study, the OPLS3e[84] and OpenFF 1.2[94] FFs were found to be somewhat better at reproducing DFT optimized structures and conformational energies E_{conf}.[124] It should be noted, however, that the MM*x* FFs were optimized to reproduce experimental structures.

Aside from energy minimized structures, off-equilibrium structures are a very useful source of reference data, as they allow sampling the potential energy surface of molecules. For such

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

structures, the forces, that are now nonzero, can be computed in addition to energies and used in model development. Structures can be prepared with MD simulations or, as is often done in case of molecular complexes (section Gas Phase Dimers and Complexes), by scaling the intermolecular distance or a relevant angle by a constant. Examples of data sets containing off-equilibrium structures will be discussed below.

Data Sets on Monomeric Compounds. In this section, we present an overview of data sets including structures of (predominantly) monomers and corresponding energies, as well as other physicochemical properties useful for FF development (Table 3).

Accurate Models Require Big and Diverse Data. There are different philosophies on what kind of reference data to use, that is, experimental versus quantum chemistry and the availability of data from different sources is therefore indicated in Table 2. For instance, some force fields directly target the properties of interest, such as Gibbs energy of solvation, in their parameter development strategy.[75,125−127] A large body of experimental data is available from handbooks, e.g., refs 128−131. The majority of data sets presented below are based on QM calculations, however. Many of these data sets were designed to train neural networks for predicting energies and forces. Therefore, these models require a substantial number of data points to capture the chemical diversity of their target system, and to make this feasible the properties are usually calculated at moderate levels of theory (mostly DFT). While FF methods contain many fewer parameters (about 1500)[25] than neural network potentials (e.g., ≈325 000 parameters for the ANI-1ccx potential),[52] the ML-training data sets can still be used for FF development if they cover various physicochemical properties. That being said, we limit this perspective to properties that are useful in development of halo-organic FFs, see Table 2.

Long range interactions need special consideration in the development of a ML model.[132−135] For this purpose, data sets of interaction energies ΔE_{int} are particularly useful, and those are discussed in detail in the section Gas Phase Dimers and Complexes dedicated to noncovalent interactions.

It is worth noting that data sets may change over time, e.g., by adding more structures to the data set or recalculating properties at a higher level of theory. This has happened, for instance, for the GDML data set that had both an increase in number of structures and additional calculations with different levels of theory, essentially superseding its predecessor, the MD17 data set.[43,49,51,136]

QMx and Related Data Sets. There are several data sets called QM*x* (quantum mechanics *x*), where the *x* stands for the highest number of heavy atoms for that particular data set. The sources of these compounds are the generated database (GDB)-*x* data sets, that are enumerations of all possible compounds following some simple rules for chemical feasibility.[137] QM7 is one of the earliest such large data sets.[36] It is a subset of the organic molecules from the GDB-13 database,[137] and it consists of seven thousand molecules containing H, C, N, O, and S. An extended version of the data set called QM7b was released later, with 13 additional properties and including chlorine atoms.[37] Subsequently, QM8,[38,138] consisting of ≈21 000 structures, and QM9,[138,139] consisting of ≈134 000 structures, were introduced. QM9 includes several properties useful for FF development. Both the QM8 and QM9 data sets consist of small organic structures, in this case H, C, N, O, and F (but not Cl). The QM*x* data sets are well-known benchmarks in the field, and additions have been proposed. For instance, the QM9 data set contains additional molecules generated by the deep learning model G-SchNet[140] and the Alchemy[141] data set was provided to increase the number of heavy atoms from 9 to 14.

A drawback of the data sets above is that they have a limited sampling of the potential energy surfaces of the included molecules. The geometries in the QM7 data set were derived using the universal FF,[142] and the other QM*x* data sets were minimized using DFT. This results in geometries that are at or near equilibrium only. Both ML models and FFs are limited by their training data, so in order to truly predict the energies and forces during a simulation, where the system is not at equilibrium, off-equilibrium structures are required to explore a larger conformational space. The QM7b-T[143,144] is an example of such a data set, and it contains the same molecules as in QM7b but sampled using MD and calculated at coupled clusters level of theory. C7H10O2-17[145] and ISO17[139,145,146] are two additional examples that are based on the most common isomer from QM9, $C_7H_{10}O_2$, and evaluated at DFT level of theory.

ANI-1 Data Sets. Like the QM*x* data sets, the ANI-1 data sets are also based on a GDB database, in this case GDB-11.[147] Despite not being calculated at a very high level of theory, the original ANI-1[148] data set consists of 22 million off-equilibrium structures of various molecules. In 2019, the same authors released the updated ANI-1x[47,149] and ANI-1ccx[52,149] data sets. Both of these were improved using active learning, increasing the diversity of the included conformations while bringing the total number of conformations down.[150] ANI-1ccx was also calculated at a higher level of theory. With each of the published data sets, the authors also introduced a neural network potential trained on the respective data set presented.[42,47,52] To benchmark the performance of their potentials, they created a new benchmarking data set, COMP6,[47] consisting of six subsets of data. ANI-1 was later complemented by ANI-1E,[151] to supply the data set with equilibrium structures that were not provided in the original work.

Other Data Sets. Various other data sets that add different chemical environments and properties to the arsenal of the FF developer are available. OE62[152] and tmQM[153] cover a broad range of elements. The QMugs data set contains several useful properties of biologically and pharmacologically relevant molecules extracted from the ChEMBL database.[154] Large structural databases of molecules have also been used as a source of molecular structures for QM data sets, e.g., the TensorMol-0.1 network using molecules from ChemSpider[155,156] and the PC9 data set based on compounds from PubChem.[44,157] GMTKN55[113] is a database containing 55 data sets. In addition to data sets dedicated to noncovalent interactions, in part covered in Table 4, GMTKN55 encompasses other data sets with a variety of thermochemical properties, reaction energies, and reaction barrier heights, some of which are potentially useful for ML potentials or FF development. The Alexandria library[108,115,158] provides DFT-optimized structures of about 5000 compounds along with a range of properties such as molecular multipoles, polarizabilities, and electrostatic potentials. Yet another collection of data sets for benchmarking of QM methods is the Minnesota database by Truhlar and co-workers, which features various properties, computed or experimental where available.[159] One such example is a data set of solvation free energies, MNSOL. The latter two databases mentioned, along with several others, have been included in ACCDB meta-database.[160] The number of published data sets in this field is very large and describing every one in detail is out of scope of this review. However, an extensive list of data sets can be found

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

in Table 3, and we refer the reader to the references provided there.

GAS PHASE DIMERS AND COMPLEXES

The accurate description of noncovalent interactions is crucial for the modeling of phenomena in complex systems, such as those of interest in biology-related fields. This section presents an overview on the computational chemistry benchmark data sets dedicated to noncovalent interactions studied using dimers or multimeric complexes. Experimental data on noncovalent interactions is limited, and their reproduction by the computational methods is not straightforward.[65] Therefore, high-level *ab initio* QM calculations are often resorted to.

Interaction Energy. The concept of interaction energy (ΔE_{int}) is used for the unambiguous comparison of computational methods in their capability to predict the strength of interactions. The quantity is defined as the difference between the (gas phase) energy of infinitely separated monomers (E_A, E_B) and that of their complex (E_{AB}) at the distance at which monomers A and B interact. Neither zero-point vibrational energy (ZPVE) nor conformational changes upon binding are typically considered in studies mentioned below. For example, the geometries of monomers are taken directly from the minimum of the complex, without reoptimization.

$$\Delta E_{int} = E_{AB} - E_A - E_B \quad (1)$$

One of the approximations used by *ab initio* QM methods to make computations feasible is the use of truncated basis sets for the orbital description. However, calculation of the interaction energy in this way introduces a basis set superposition error, as the complex AB gains an extra stabilization compared to individual monomers A and B. There are approaches available to mitigate this artifact such as the symmetry adapted perturbation theory (SAPT) method which, however, fall outside of the scope of this perspective.[175,176]

The Accuracy Level. For data sets featuring smaller systems, a very high level of accuracy can be reached using more demanding methods. The coupled clusters singles, doubles, and perturbative triples, CCSD(T), with extrapolation to a complete basis set (CBS) are habitually used as reference data for this purpose. This level of theory is referred to as the "gold standard" of computational chemistry.[177] A common practice to reach this accuracy level is through a composite scheme, where the final benchmark energy is the sum of an MP2 energy in the CBS limit and a ΔCCSD(T) correlation correction to this energy. The correction is the difference between CCSD(T) and MP2 in the respective basis set. The MP2 component is extrapolated to a CBS from computations in (at least) two basis sets and the ΔCCSD(T) correlation component is then typically calculated in a smaller basis set. An example is an MP2 extrapolation from aTZ and aQZ basis sets with a ΔCCSD(T) in the aTZ basis set (here we use basis set designation where aXZ corresponds to aug-cc-pVXZ, with X being D, standing for double, T for triple, Q for quadruple, or 5 for pentuple zeta, respectively[178]). This setup, or approaches yielding comparable accuracy (see below), was deemed adequate[177] to reach the desired "gold standard", referred to as "gold" level from here on. Using a smaller aDZ basis set for the ΔCCSD(T) correction, which is an important factor for the accuracy, to obtain the CCSD(T)/CBS was defined as a "silver" level. Finally, going beyond the "gold" accuracy was termed the "platinum" level.[177,179]

Is Anyone Benchmarking the Benchmarks? The interaction energies calculated by CCSD(T)/CBS extrapolated methods were successfully tested as a part of the protocol to reproduce experimental values with errors of 0.15 to 0.3 kcal/mol for benzene−alkane clusters or 0.1 kcal/mol for small complexes of noble gases.[180,181] Authors of these publications concluded that the experimental determination of true minima alkanes or correct computation the ZPVE were larger sources of error than the CCSD(T) calculations themselves.

Certain approximations are generally employed to render these high-level QM computations tractable. Because the use of experimental data for noncovalent interactions is often impractical, as illustrated by the two studies above, to examine the validity of these approximations, the A24 data set was created. This data set provides ΔE_{int} on 24 small complexes of H, C, N, O atoms with the highest accuracy available at the time. For the benchmark values, a three-point extrapolation of CCSD(T) to the CBS using large basis sets was used, with the ΔCCSDT(Q) correlation correction up to the perturbative quadruple excitations, dropping the frozen-core approximation commonly employed and additionally accounting for relativistic effects for all the elements. The approach was suggested as a "platinum" level of accuracy.[177] The combined error due to neglecting of the following contributions (in this order of importance): coupled cluster treatment up to CCSDT(Q), correlation of core electrons, and relativistic effects was estimated to lie below 2% of benchmark interaction energy for the data set.[182] "Gold", using the aTZ to aQZ extrapolation with the aTZ ΔCCSD(T) term, and "silver", using the aTZ to aQZ extrapolation and the aDZ ΔCCSD(T) term, levels yielded errors of about 1% and 2% compared to the A24 benchmark, respectively, while using the nonaugmented DZ basis set to compute the ΔCCSD(T) term results in ∼6% error.[182,183] Regardless, both gold and silver levels of accuracy are indeed very high and definitely sufficient for applications in FF development.

Interaction Energy Data Sets. A selection of benchmark data sets for noncovalent interactions is presented in Table 4. One of the first benchmark data set publications for noncovalent interactions featured the JSCH2005 and S22 data sets.[184] JSCH2005 focused on nucleobases and amino acid complexes, while S22 was more general. The S66 data set, intended as an extended and more universal replacement of S22, became one of the most popular data sets for noncovalent interactions. These data sets received several updates over time, extending the coverage of potential surfaces and accuracy of the benchmark.[183,185−189] The S66 data set of equilibrium geometries uses aTZ to aQZ extrapolation for MP2/CBS. In the latest version,[183] the ΔCCSD(T) term is computed in haTZ, where the "h" letter for "heavy" signifies the use of the augmented basis set (with additional diffuse functions) for non-hydrogen elements only, with a minimal drop of accuracy compared to fully augmented basis set aTZ. The S101×7 data set was introduced for the development of a charge-penetration correction for the AMOEBA FF.[72] It uses the SAPT+/CBS method for the benchmark values, which in turn were tested on S66, where the FF delivered an error of 0.16 kcal/mol.[190] Recently, the S66 and S101 complexes have been included in the NENCI2021 data set,[191] which broadens the coverage of elements, includes charged complexes and also an extended mapping of the potential energy surface. The potential energy surface mapping of NENCI2021 focuses, in the fashion of S101, on repulsive regions, with intermolecular distances ranging from 70 to 110% of the equilibrium distance.

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

Table 3. Overview of Large Data Sets with Various Molecular Properties[a]

data set	level[b]	coverage	geometries[c]	properties	size	year	refs
QM7; QM7b	PBE0†; also ZINDO, SCS, GW	H, C, N, O, S. Up to 7 heavy atoms; incl. Cl	N/A; Opt	$\Delta_f H$; also E^*, α	7165; 7211	2012; 2013	36, 37, 137
MNSOL	experimental	H, C, N, O, F, Si, P, S, Cl, Br, I. Neutral and charged compounds in 92 solvents	Opt	ΔG_{solv}	790	2014	159
QM9	B3LYP/6-31G(2df,p), G4MP2	H, C, N, O, F. Up to 9 heavy atoms	Opt	E, E^*, $\Delta_f H$, S, C_V, ν, α, μ	133 885	2014	138, 139
QM8	CC2/def2-TZVP, TD-PBE0, and TD-CAM-B3LYP in def2-TZVP	H, C, N, O, F. Up to 8 heavy atoms	Opt	E, E^*	21 786	2015	38, 138
GDML	CCSD or CCSD(T)/cc-pV(D/T)Z or PBE (vdW-TS)[43]	H, C, N, O. Small compounds.	NE	E, F	3 875 468	2017; 2019	43, 49, 51, 136
ISO17	PBE[145,146] (vdW-TS)	$C_7H_{10}O_2$ isomers	NE	E, F	640 982	2017	139, 145, 146
ANI-1	ωB97x/6-31G(d)	H, C, N, O. Up to 8 heavy atoms	NE	E	22 057 374	2017	42, 148
Yao et al.	ωB97X-D/6-311G**	H, C, N, O. ChemSpider molecules with up to 35 atoms, water clusters	NE	E, F, μ	2 979 162, 370 844	2018	155, 156
MPCONF196	CCSD(T)/CBS (aTZ to aQZ, extrap., ΔCCSD(T)/haDZ) or MP2-F12/aDZ, ΔDLPNO-CCSD(T)/aDZ)	H, C, N, O. Macrocycles with up to 120 atoms	Opt	E_{conf}	13 196	2018	162
Alexandria Library	B3LYP/aug-cc-pVTZ, G4	Up to 4th row elements except K, Ca, incl. I. Mostly organic compounds	Opt	E, $\Delta_f H$, S, C_V, ν, Q, α, V_{ESP}, μ	5100	2018	115
COMP6	ωB97x/6-31G*	H, C, N, O. Six data sets, incl. drug-like compounds, peptides, and S66. Up to 312 atoms	NE	E, ΔE_{int}, F, Q, μ	56 182	2018	47
SN2 reactions	DSD-BLYP-D3(BJ)/def2-TZVP	H, C, F, Cl, Br, I. Halide anions, methyl halides	NE	E, F, μ	452 709	2019	163
solvated protein fragments	revPBE-D3(BJ)/def2-TZVP	H, C, N, O, S. Fragments up to 8 heavy atoms, systems up to 120 atoms, incl. ions	NE	E, F, μ	2 731 180	2019	163
QM7b-T	CCSD(T0)/cc-pVDZ, other methods	H, C, N, O, S, Cl. Up to 7 heavy atoms	NE	E	7211	2019	143144,
GDB13-T	MP2/cc-pVTZ	H, C, N, O, S, Cl. Up to 13 heavy atoms	NE	E	6000	2019	143, 144
ANI-1x	ωB97x/def2-TZVPP	H, C, N, O. Up to 63 atoms	NE	E, F, Q, μ	5 496 771	2019	47, 149
ANI-1ccx	DLPNO-CCSD(T)/CBS[52]	H, C, N, O. Up to 55 atoms	NE	E	489 571	2019	52, 149
G-SchNet data set	B3LYP/6-31G(2d,p)	H, C, N, O, F. Up to 9 heavy atoms	Opt	E, μ	9074	2019	140
Alchemy	B3LYP/6-31G(2df,p)	H, C, N, O, S, F, Cl. 9−14 heavy atoms	Opt	E, E^*, S, C_V, α, μ	119 487	2019	141
PC9	B3LYP/6-31G*	H, C, N, O, F. Up to 9 heavy atoms (singlets, doublets, triplets)	Opt	E, Q	99 234	2019	44, 157
Schütt et al.	PBE/def2-SVP, HF/def2-SVP	H, C, N, O. Hamiltonians and overlap matrices for conformations of several small compounds	NE	E, F	121 977	2019	53
OE62	PBE0† (vdW-TS), PBE0† (vdW-TS MPE−water), GW@PBE0/def2-(T/Q)ZVP	H, Li, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, Te, I. Up to 92 heavy atoms	Opt	E, Q	61 489, 30 876, 5239	2020	152
QMspin	CASSCF(2e,2o)/cc-pVDZ-F12, MRCISD+Q-F12/cc-pVDZ-F12	H, C, N, O, F. Carbenes, triplets, singlets	Opt	E, ν, μ	8062, 5021	2020	164
tmQM	TPSSh-D3BJ/def2-SVP, GFN2-xTB[165]	H, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, I. Transition metal complexes	Opt	E, E^*, Q, α, μ	86 665	2020	153
QM7-X	PBE0† (vdW-MBD)	H, C, N, O, S, Cl. Up to 7 heavy atoms	Opt/NE	E, E^* $\Delta_f H$, F, Q, α, μ	4 195 237	2021	166
ANI-1E	ωB97x/6-31G(d)	H, C, N, O. Up to 8 heavy atoms	Opt	E^*, $\Delta_f H$, S, C_V, α, μ	57 455	2021	167
Gastegger et al.	PBE0/def2-TZVP	H, C, O. Response properties in solvent, explicit (sampling) and implicit (PCM)	NE	E, F, α, μ	214 183	2021	56
BSE49	(RO)CBS-QB3[107]	H, B, C, N, O, F, Si, P, S, and Cl. Homolytic cleavage of 49 different bond types	Opt	ΔE	4502	2021	168

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

Table 3. continued

data set	level[b]	coverage	geometries[c]	properties	size	year	refs
Guan et al.	ωB97X-V/cc-pVTZ	Potential energy surface for 19 reaction channels for hydrogen combustion	NE	E, F	361 803	2022	169
QMugs	ωB97X-D/def2-SVP	H, C, N, O, P, S, F, Cl, Br, I. Up to 100 heavy atoms	Opt/NE	E, E^*, H, S, $\Delta_f H$, S, C_v, ν, Q, α, V_{ESP}, μ	2 004 003	2022	154
Thürlemann et al.	PBE0-D3BJ/def2-TZVP, MBIS[170]	H, C, N, O, F, S, Cl. Up to 20 heavy atoms	NE	F, V_{ESP}, μ	1 013 949	2022	171
VIBFREQ1295	Experiment, CCSD(T)(F12*)/cc-pVDZ-F12	H, C, N, O, P, S, F, Cl, B, Si, Al	Opt	ν	141, 1295 ν	2022	172
Chan	CCSD(T)/CBS (W1X-2)[173]	H, C, N, O. Up to 100 atoms from NIST database	Opt	$\Delta_f H$	1500	2022	174

[a]The properties of interest are explained in Table 2. [b]The † in PBE0† stands for PBE0 in FHI-AIMS,[161] with "tight"settings/"tier 2" basis set. For the definition of CCSD(T)/CBS, see next section. [c]Indicates if the geometries were optimized to reach the energy minimum (Opt) or if non-equilibrium (NE) structures were generated.

In these data sets, the "×" signifies different conformations generated from the equilibrium one by "scans", e.g. varying the bonding angle of an energy minimum structure or scaling the intermolecular distance by defined percentages, producing points along the dissociation curve. In this manner, the coverage of the potential energy surface is often extended to include nonequilibrium geometries (see Table 4).

The Noncovalent Interaction Atlas (NCIA)[192] is a collection of benchmarks containing specialized data sets for hydrogen bonds, e.g. neutral,[193] charged,[193] and neutral featuring sulfur, phosphorus, and halogens;[194] dispersion bound complexes featuring a wider coverage of elements;[195] nonequilibrium repulsive interactions at short distance;[196] and σ-hole interactions.[197] The series consistently applies the same "gold" level benchmark, and the data sets are complementary. The DES370K with about 3700 unique dimers and 370 000 points from QM optimizations, scans, and MD simulations is currently the largest data set for noncovalent interactions.[198] There, the ΔCCSD(T) correction varied from aDZ to aQZ, depending on the systems size. The authors provide a smaller, representative subset of this data set, DES15K, and another database of about 5 million geometries with benchmark energies computed by a ML approach trained on the DES370K data set. The HEAVY28 data set, designed to serve as the benchmark data set for DFT development, is unique in that it contains more exotic heavier elements.[199] Most of the dimer data sets neglect the deformation energies. An example of an exception is WATER27, which provides stabilization energies of water clusters and includes the energy corresponding to the conformational change upon complex formation.[200]

A few data sets are devoted to larger systems.[201−205] Due to the size of the systems, different high-level benchmarking strategies were employed in order to deviate as little as possible from the "gold" standard. For example in PLA15, where the ligand interacts with its protein environment, the energy of individual protein fragments constituting the environment calculated at higher level is simply summed up to yield the basis for the benchmark whereas the nonadditivity is addressed at a lower level of theory (DFT).[204] The ExL8 data set contains data on 8 different kinds of large complexes, including zeolite or boron nitride nanotubes.[205] The computational details such as the basis set usage for this data set varied with the system size. A different take on the reference values is presented by the data set of host−guest complexes, S30. The data set provides experimental association energy ΔG_a in various solvents complemented with its computational counterpart determined with a combination of different methods.[203]

Aside from the GMTKN55 database mentioned in the section Data Sets on Monomeric Compounds and the NCIA project, other useful repositories include BEGDB,[206] QCArchive,[207] and the Computational Chemistry Comparison and Benchmark DataBase.[208]

Force Field Approaches to Complexes. In order to illustrate how existing QM data sets can be used to validate FFs, we present a comparison between the FF interaction energies and benchmark energies of S66[228] (Table 4). The FF dimer interaction energies were computed according to eq 1, see Methods for details. The coordinates were taken directly from the data set (Figure 1, "without EM") and, in addition, from the result of a FF energy minimization of the input geometries (Figure 1, "with EM"). This was done since the FF energy minimum usually does not correspond to the QM energy minimum. Table S1 lists the energies of the complexes from QM

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

Table 4. Overview of High-Level Benchmark Data Sets for Noncovalent Interactions[a]

data set	level	coverage	size	year	refs
S22	Gold (aTZ to aQZ, ΔCCSD(T)/aTZ)	H, C, N, O	22	2006	184−186
Berka et al.	Silver (aDZ to aTZ, ΔCCSD(T)/6-31G*(0.25, 0.15))	Amino acid side chains	24	2009	209
WATER27	Silver (aXZ; x = 2, 3, 4 and 5, ΔCCSD(T)/aDZ)	Water clusters, neutral and charged	27	2009	200
HEAVY28	(CCSD(T)/CBS[199])	H, O, N, S, Pb, Sb, Bi, Te, Cl, Br, I	28	2010	199
S66; S66×8, S66a8	Gold; silver (aTZ to aQZ, ΔCCSD(T)/haTZ; aDZ[189])	H, C, N, O	66; 66×8	2011	183, 189, 210
HSG	Gold (aTZ to aQZ, ΔCCSD(T)/haTZ)	Amino acid side chains	21	2011	187, 211
HBC6	Gold (aTZ to aQZ, ΔCCSD(T)/aTZ)	H, C, N, O. H-bonds	6 (Dis. curves)	2011	187, 212
Karthikeyan et al.	Gold (aTZ to aQZ, ΔCCSD(T)/haTZ)	H, B, C, N, O, F, Cl. Charge transfer	11	2011	213, 214
NBC10	Gold or silver CCSD(T)/CBS[187]	Benzene, pyridine, H_2S, CH_4	10 (Dis. curves)	2011	187
Mintz and Parks	Gold (aTZ to aQZ, ΔCCSD(T)/aTZ)	H, C, N, O, S. S-interactions	14×8	2012	215
X40; X40×10	Gold; silver (aTZ to aQZ, ΔCCSD(T)/haTZ; aDZ)	H, C, N, O, F, Cl, Br, I. Halogen bonds	40×10	2012	216
Granatier et al.	Gold or silver (aTZ to aQZ, ΔCCSD(T)/aTZ or aDZ)	H, C. Dispersion int.	12	2012	217
A24	Platinum (aTZ, aQZ, a5Z, ΔCCSDT(Q)/aDZ, core corr., relat. eff.)	H, B, C, N, O, F, Ar. Small compounds	24	2013	182, 218
Bauzá et al.	CCSD(T)/aTZ	H, C, N, P, S, As, Se, F, Cl, Br. Incl. ions, σ-hole interactions	30	2013	219
XB18; XB51	Gold (aQZ to a5Z, ΔCCSD(T)/aTZ)	H, C, N, O, P, F, Cl, Br, I, Li, Pd. Halogen bonds	18; 51	2013	220
S101×7	SAPT2+/CBS-scaled[190]	H, C, N, O, P, S, F, Cl, Br. Incl. ions, charge penetration	101×7	2015	190
Parker and Sherrill	Silver (DW-CCSD(T**)-F12/aDZ) or "Pewter" SCS(MI)-MP2/TZ	Nucleobase dimers, tetramers	∼100, ∼30 000	2015	221
Hostaš et al.	Silver (aTZ to aQZ, ΔCCSD(T)/aDZ)	Nucleobase−amino acids	272	2015	222
Temelso et al.	Silver (aDZ to aTZ, ΔCCSD(T)/aDZ), CCSD(T)-F12/DZ variants	Binding energies of water clusters	62	2015	223
BBI; SSI	Silver (DW-CCSD(T**)-F12/aug-cc-pV(D+d)Z)	Protein backbone−backbone; side chain−side chain	100; 3380	2017	224
HB375×10; IHB100×10	Gold (aQZ to a5Z, ΔCCSD(T)/haTZ) or silver (aTZ to aQZ, ΔCCSD(T)/aDZ) scaled to gold	H, C, N, O. H-bonds; incl. ions	375×10, 100×10	2020	193
HB300SPX×10	Gold (aQZ to a5Z, ΔCCSD(T)/haTZ) or silver (aTZ to aQZ, ΔCCSD(T)/aDZ) scaled to gold	H, C, N, O, P, S, F, Cl, Br, I. H-bonds	300×10	2020	194
López et al.	Gold (aQZ to a5Z, ΔCCSD(T)/aTZ)	H, C, N, O, P, S compounds complexed with Li, Na, K, Be, Mg, Ca ions	26×6	2020	225
R739×5	Gold (aQZ to a5Z, ΔCCSD(T)/haTZ)	H, C, N, O, P, S, F, Cl, Br, I, He, Ne, Ar, Kr, Xe. Repulsive contacts	739×5	2021	196
NENCI2021	Gold (aTZ to aQZ, ΔCCSD(T)/haTZ)	H, C, N, O, F, P, S, Cl, Br, Li, Na. Incl. ions	141, 7763	2021	191
DES370K	Gold or silver (aTZ to aQZ, ΔCCSD(T)/aQZ or aTZ or aDZ)	H, C, N, O, P, S, F, Cl, Br, I, He, Ne, Ar, Kr, Xe, Li, Na, K, Mg, Ca. Incl. ions	3691 dimers, 370 959 conf.	2021	198
SH250×10	Gold (aQZ to a5Z, ΔCCSD(T)/haTZ) or silver (aTZ to aQZ, ΔCCSD(T)/aDZ) scaled to gold	H, C, N, O, P, S, As, Se, F, Cl, Br, I. σ-hole int.	250×10	2022	197
D1200; D442×10	Gold (aQZ to a5Z, ΔCCSD(T)/haTZ); or silver (aTZ to aQZ, ΔCCSD(T)/aDZ) scaled to gold	H, B, C, N, O, P, S, F, Cl, Br, I, He, Ne, Ar, Kr, Xe. Dispersion int.	1200; 442×10	2022	195
		Larger Systems			
L7	QCISD(T)/CBS (aDZ to aTZ, ΔQSISD(T)/631G* (0.25))	H, C, N, O. Up to ∼100 atoms	7	2013	201
S30	Exper. ΔG_a, PW6B95-D3/QZ′ or ωB97X-D3/QZ′,[203] HF-3c,[226] COSMO-RS	Supramolecular host−guest complexes. Up to ∼200 atoms, ions, solvation energies	30	2015	203
PLF547; PLA15	MP2-F12/DZ + ΔDLPNO−CCSD(T)/aDZ; Respective PLF547 frag. benchmarks summed with B3LYP-D3/DZVP-DFT nonadditivity[204]	Amino acid−ligand. Up to ∼100 atoms. Protein−ligand. Up to ∼500 atoms	547; 15	2020	204
ExL8	CIM-DLPNO-CCSD(T)/CBS[227]	Large complexes. Up to ∼1000 atoms, inc. Si, Al, B	8	2021	205

[a]For the definition of levels of accuracy see the text. If multiple levels of accuracy are used by the authors, only the most accurate/recent is stated. In CBS schemes basis sets used for extrapolation and for ΔCCSD(T) correlation energy are separated by comma. (a)XZ notation stands for (aug)-cc-pVXZ basis set, where X is D for double, T for triple, Q for quadruple, and 5 for pentuple zeta, respectively. In some cases, variants of these basis sets are used, such as pseudopotential versions for heavier elements (here, we include them under the XZ notation). The haXZ signifies a use of augmented basis set for all atoms except hydrogen. For the details of computational setup, we refer the reader to the corresponding citation.

as well as from the FF before and after minimization. It also lists the RMSD of the coordinates. Figure 2 shows the distribution of RMSD after minimization, to evaluate whether the structures remain well-conserved in the FF. Such comparisons give quantitative information about the performance of a FF and by analysis of its accuracy for different chemical compound classes detailed clues for improvement of FF models can be derived. The structures in the S66 data set have been used earlier to investigate electrostatic interactions by FF methods,[229] as well as for the development of models for charge penetration.[190,230,231] In addition, the S66×8 data set (Table 4) has been used for FF development by Vandenbrande and coworkers.[232]

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

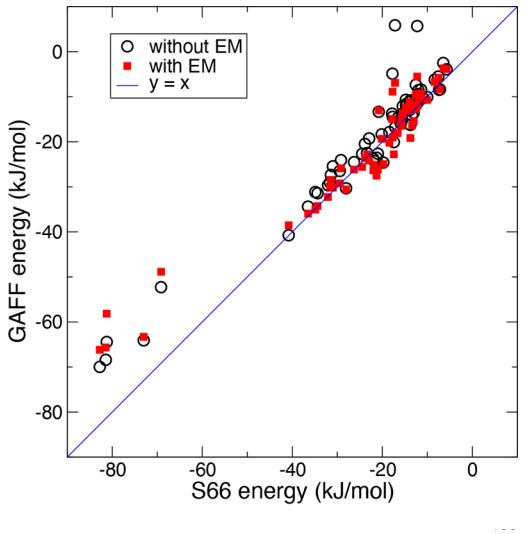


Figure 1. Correlation for interaction energies with S66 as reference[189] and GAFF as target. For raw data, see Table S1. EM stands for energy minimization. Statistic without EM, r^2 = 95.8%, slope a = 0.85; with EM, r^2 = 96.2%, slope a = 0.91. Slope computed from a fit to $y = ax$.

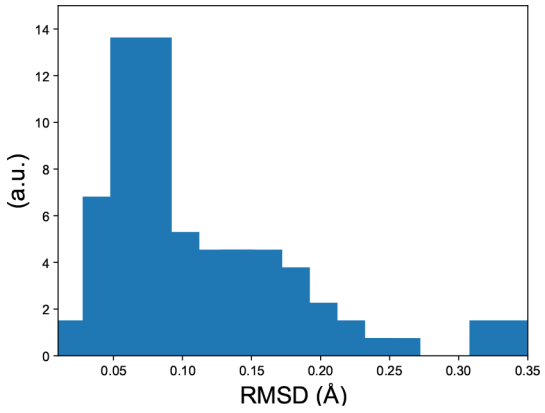


Figure 2. Root mean square deviation from S66 dimer coordinates after force field energy minimization. For raw data, see Table S1.

CONDENSED PHASES

Organic Crystals. A large amount of experimental data is available for the condensed phases that can be used for both developing and evaluating FFs. Hagler was one of the first to take advantage of molecular crystals for this purpose,[233,234] structures for which can be found in the Cambridge Structural Database.[235] Here, we ignore the related but specialized field of crystal structure prediction (reviewed recently here[236]) and focus on MD simulation studies on molecular crystals. The obvious advantage of crystals over liquids is that the position of the atoms is known (in most cases). This means that the properties of the models can be evaluated with more certainty than in the liquid phase,[4] that is, if the crystal structure is preserved in the FF simulation. Price and co-workers note that the weak intermolecular forces in some crystals put high demands on FFs.[237,238] Although it has been shown that special-purpose models can reproduce specific crystal polymorphs,[239] this is not generally possible. Indeed, Nemkevich et al.[240] conclude that molecular simulation of crystals can give qualitative results at best. The often small intermolecular forces in organic crystals may lead to changes in the relative orientation of compounds within the crystal, or even (partial) melting, if the forces are not well-balanced.[241] A further issue with simulations of crystals is that it may be cumbersome to evaluate many different crystals since it is nontrivial to assess whether a system is in the crystalline phase and, if so, what polymorph it is. In addition, some crystals are in a plastic crystal phase at certain conditions, that is the molecules occupy fixed lattice positions, with some more or less restrained degrees of freedom.[241,242] Some computational benchmarks based on organic crystals have been proposed (Table 5). Reilly and Tkatchenko extended the

Table 5. Computational Chemistry Benchmarks Using Organic Crystals[a]

data set	properties	MAE ($\Delta_{sub}H$)	method	size	year	ref
C21	$\Delta_{sub}H$	4.8	PBE and better	21	2012	243
X23	$\Delta_{sub}H$	3.9	PBE and better	23	2013	244
X23	$\Delta_{sub}H$, lattice	9.2	FIT[247]	23	2016	245
X23	$\Delta_{sub}H$	11.7	QM/MM	23	2019	246
Bernardes and Joseph	$\Delta_{sub}H$, ρ, lattice	5.5	OPLS/AA[126]	18	2015	248
Schmidt et al.	$\Delta_{sub}H$, ρ, T_{melt}	6.4	GAFF[74]	27	2022	241

[a]Size of the data set is given as well as the mean absolute error (MAE) of the best method for predicting $\Delta_{sub}H$ (kJ/mol) in the corresponding study. Lattice parameters are indicated by "lattice", solid density by ρ.

C21 database[243] to form the X23 database,[244] both studying molecular crystals using DFT. The X23 database has been evaluated using FFs by Nyman et al.[245] and by Teuteberg et al. using combined quantum mechanics/molecular mechanics (QM/MM).[246] Bernardes and Joseph[248] studied 18 drug-like aromatic compounds by performing experiments to evaluate the enthalpy of sublimation $\Delta_{sub}H$ and then compared the results to FF simulations. Apart from the enthalpy of sublimation, they also study the relative stability $\Delta_{trs}H$ of crystal polymorphs. Finally, Schmidt and co-workers studied 30 crystals of small organic molecules using long MD simulations,[241] and apart from $\Delta_{sub}H$ they determined melting temperatures T_{melt} and solid densities ρ.

Although all these papers use different data sets and computational methods, it is tempting to compare the accuracy for predicting $\Delta_{sub}H$. The most accurate DFT used was PBE0[249] with many-body dispersion correction,[250] yielding a MAE of 3.9 kJ/mol.[244] Bernardes and Joseph published a FF MAE of 5.5 kJ/mol,[248] whereas Schmidt et al. find 6.4 kJ/mol;[241] both of these used different data sets than X23, however, which means the numbers are not directly comparable. Data and inputs for the work presented by Schmidt et al. are available on GitHub.[251]

Organic Liquids. Historically, it was the study of (organic) liquids that has laid the foundation for MD simulations.[24,252] Many properties of bulk liquids have been measured over the last hundred or so years, and the results are available in databases and handbooks.[128−131] Since classical simulations of neat liquids are not very demanding in terms of computer resources, the available experimental data on liquids can be used to both derive[82,253,254] and evaluate FFs. Indeed, a paper by Caleman et al. scrutinizing FFs by computing seven bulk properties for about 150 liquids[33] found several systematic flaws, in particular for surface tensions and dielectric constants. As a result, a series of other papers has been published with the aim of addressing these

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

shortcomings, using the data provided by Caleman et al.[33,255] as a reference, e.g. for the improvement of the OPLS-AA FF[256] and the GROMOS 2016HH FF[80] and evaluation of the TRaPPE FF.[257] One important conclusion from follow-up papers,[66,258,259] is the recommendation to apply explicit long-range dispersion interactions using the particle-mesh Ewald algorithm for Lennard-Jones potential (LJ-PME).[260−262] The surface tension (vacuum−liquid) is particularly affected by these interactions[66,258] (for a review on surface tensions, see ref 263), but also systems as simple as liquid carbon dioxide.[259] Based on these lessons, LJ-PME was used in melting simulations of organic crystals.[241] Whether the use of LJ-PME is advantageous for simulations of biomolecules with current force fields is less clear.[264] Simulation studies of Gibbs energy of solvation of organic compounds in organic liquids were performed by Zhang et al.[34,67] based on earlier empirical work by Katritzky and coworkers.[265−267] These results were used in FF development and evaluation[268−270] as well as for comparison in FF based predictions of water−octanol partition coefficients.[64,271] It is clear that organic compounds are excellent diagnostic tools for FF development.[114] Nevertheless, due to the biological relevance of proteins, nucleic acids, and phospholipids, the community has put much more effort in evaluating biomolecular FFs for these kinds of molecules, e.g., see refs 272−282. Experimental reference data on thermophysical and thermochemical properties of (organic) liquids and liquid mixtures can be found, for example, in the NIST ThermoML Archive repository.[131] It contains various experimental data of complex systems−up to ternary mixtures. It has been expanded to include metals and their compounds, too, and recently encompassed about 8 000 000 data points.[283,284] This subtopic is so large, that it would warrant a specialized review, and we will therefore not discuss it further here. Likewise, there is an enormous body of work on solvation in water, that has been reviewed elsewhere.[24,285,286]

DISCUSSION

Enrico Clementi and his group pioneered force field development based on artificial intelligence and databases of QM data as early as the 1990s.[25,287,288] In modern times, the Parsley (Open) FF is perhaps a good example of a FF to be largely derived from QM calculations and (in part) validated based on liquid properties.[94] Although its functional form remains true to classic force fields,[1] systematic improvements using the ForceBalance algorithm[8] suggest that further accuracy gains may be possible. Interestingly, the Parsley Open FF and other well-known FFs were benchmarked[289] using data from the QCArchive.[207] Given the modeling target defined in the Introduction, that is halo-organic compounds plus biologically relevant inorganic compounds, what does the perfect data set look like? Is there a difference between data sets for model development and for benchmarking? An ideal data set would consist of enough structures to capture all of the diversity of the modeling target. Furthermore, relevant physicochemical properties have to be present (Table 2) and, for quantum chemical data sets, calculations for each and every structure should be carried out using a sufficiently high (*ab initio*) level of theory. The data set has to be easy to access and to work with, for instance, with a simple to use application programming interface such as in ref 284. Importantly, the amount of independent data points should be much larger than the amount of parameters to be optimized, which is a few thousand for classical force fields[25] but much larger for neural networks.[52] From the monomer databases listed in Table 3, both the Alexandria Library[231] and QMugs[154] cover all the targeted chemical elements (and some more) and provide at least nine properties relevant for model development per compound (Table 2). In contrast, OE62[152] supports all the target elements but only two properties. One of the properties especially advantageous for the development of models with accurate electrostatics is the electrostatic potential of a molecule. Both the Alexandria Library[115] and QMugs[154] provide this, but the number of compounds is limited to about 5000 for the former while the basis set used for the DFT calculations is relatively small for the latter, which may lead to poor predictions of polarizability.[115,290−292] Although both these data sets contain frequencies, which hold information on the curvature of the potential energy surface, they do not contain energies for off-equilibrium structures, which are needed to go beyond the harmonic approximation. A number of data sets mentioned in Table 3 have nonequilibrium structures but always with a limited coverage of the chemical elements (see section Recommendations for Data Sets). Swann and coauthors proposed to use machine learning to create a data set as a way to enforce diversity. Rather than picking representative compounds manually, their automated selection is based on molecular descriptors to cover the chemical space in an unbiased manner.[293] Data sets containing noncovalent interaction energies are particularly useful for evaluating and developing long-range interactions in force fields as well as ML models. DES370K is the largest data set for this purpose, making it suitable for training of ML methods.[198] The data sets in NCIA are smaller and dedicated to specific interaction types (Table 4). However, their advantage is that their benchmarks are at the same, consistent level of theory, which means that energies can be compared directly between data sets, and the data sets are complementary. The consistent level of theory is especially important for ML methods.[35] Considered as a whole, the NCIA to date contains 2206 unique dimers with 5 or 10 points along dissociation curves (Table 4). Of the interaction databases, both DES370K[198] and NENCI2021[191] cover all the target elements. The NCIA[192] also has a very broad coverage but does not yet feature complexes with inorganic ions. These, for example, can be found in a data set by López.[225] Data sets covering less common elements include HEAVY28[199] and the Exl8 data set of large complexes.[205] If a more compact data set is desirable, the S66 benchmark data set is already well-established, offering comparison with older benchmarking studies.[183,189,210] The use of standardized benchmark data sets (or their combinations) has several major advantages. First, it allows for consistent and straightforward comparison between different force fields. Improvements of a newly developed force field over its predecessors will be more clearly apparent if both are compared to the same benchmark data. It would prevent adding of new test systems while forgetting about the original ones and later "reinventing the wheel" situations, as alluded to in the Introduction and discussed in ref 4. Similarly, with standardized benchmark testing, it would be easier to see how well does a parametrization targeted to reproduce one specific property fare for other properties. That is, what trade-offs do researchers accept when adopting new methods or parameters? Thus, the habitual testing against multiple standardized benchmarks might encourage the development of balanced FFs with more general applicability. In practice, it can be expensive, and sometimes plain impossible, to calculate physicochemical properties for large compounds or clusters at a high level of theory. As a result there usually is a trade-off between more accurate calculations

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

using a high level of theory (dimers and complexes in Table 4) and more data points (monomeric compounds in Table 3). The more computationally expensive a data set is to prepare, the more important it is that is made accessible following the FAIR (Findable, Accessible, Interoperable, Reusable) principles,[294] not least to reduce the carbon footprint of the field.[54] Work toward making experimental data more FAIR is underway as well.[284] We argue here, that a higher level of theory is needed for benchmarking than for model development. At the end of the day, models need to be validated against experimental data or the highest level quantum chemistry that is available. However, there still are large amounts of information in accurate DFT calculations. If one is aware of potential pitfalls,[115] some of the large data sets in Table 3 can be used for model development. It should be added, that experimental data is not flawless either. Besides direct experimental uncertainty and conflicting numbers, errors may be introduced when values are manually transferred from (old) papers into large databases. For instance, during the development of the Alexandria library, a list of close to 200 errors in databases of physicochemical properties was collected (Table S2 in ref 108). To find the inadequacies of the models, such as a FF lacking essential physics in its functional form,[4] it is recommended to test models outside their "comfort zone" as well. For instance, how is the model affected by a change in temperature (and/or pressure), and taking it a step further, how does the model perform in different phases (gas, liquid, solid)? Or, if a model has been tuned to predict bulk properties well, how reasonably are interface properties described? If the focus was on kinetic properties, such as diffusion coefficients, how well are energy-related properties captured?[295] If a model was derived for equilibrium structures, how well are off-equilibrium structures described? In any case, after extending the model to better fit the problematic cases, it should be retested again on the original training set, if the general improvement of the model is to be sought.[4] In order to develop and validate FFs, condensed phase simulations have to be performed. In the case of small molecule force fields that are intended as, e.g., protein ligands, it may be sufficient to compute free energies of solvation in liquids, including water.[24] For general force fields it is possible, albeit costly, to optimize FFs using liquid simulations by slowly varying parameters until experimental observables are reproduced.[9,253,254,296−299] In this manner, a few parameters at a time can be tuned, against one or more observables.[300] In addition, inherently parallel methods such as the Multistate Bennett Acceptance Ratio method can be used to speed up parametrization of FFs for single compounds in the condensed phase.[301,302] Oliveira et al. proposed a framework to optimize a force field for a limited class of compounds with shared parameters at once using, e.g., the liquid density and enthalpy of vaporization as the reference.[303,304] Extension of such methods to a complete halo-organic force field remains to be done. If all else fails, scientific contests may be used to separate the wheat from the chaff. The Critical Assessment of Structure Prediction was one of the first such contests, where researchers were given amino acid sequences with the task to predict a protein structure.[305,306] Although force fields have successfully been used to predict three-dimensional structures of small proteins,[307] the computational cost is way too large to use protein structure prediction for force field development and benchmarking, although many partial comparisons exist.[272,273,308] The Industrial Fluid Properties Simulation Collective aims for researchers to predict properties of liquids and solvated compounds.[309] Properties such as the viscosity of alcohols and surface activity of compounds, that are difficult to predict, have been addressed in these challenges. Unfortunately, this challenge seems to be dormant at the time of writing. The Statistical Assessment of Modeling of Proteins and Ligands (SAMPL) challenge is still very much active, however.[310,311] It is a competition revolving around drug design but also more fundamental molecular properties such as distribution constants have been targeted. Such blind tests, where as of yet unpublished experimental data is the target of predictions, help to evaluate and diagnose models. Properties of pure liquids are readily accessible and therefore useful for evaluating force fields.[33,66,131,258,312] The data sets used by Caleman et al. are available[255] for reuse by other groups. Possible topics for blind tests would be prediction of excess properties of mixing two liquids,[94,313,314] or the Gibbs energy of solvation in non-water solvents.[34,67] The relatively poor performance of force fields for the prediction of octanol−water partition coefficients shows there is room for improvement.[64,271] Another potential target would be prediction of free energy of association of compounds in different solvents.[315]

Recommendations for Data Sets. The coverage of data sets presented in this perspective is extensive, and great progress has been made in recent years to increase the amount, quality, and accessibility of the data sets. For force fields targeting (halo)organic compounds, there is still room for improvement of data and tools. For instance,

- Broadening of chemical diversity,[293] in particular compounds containing phosphorus and sulfur at different valence states.
- Keeping in mind the importance of nonequilibrium structures with corresponding energies and forces.
- Benchmarking of more properties per compound, such as the electrostatic potential.
- Increasing the coverage of interaction energies (and forces) for complexes of neutral compounds with inorganic or molecular ions.
- The simultaneous use of the points above in data set construction or the use of comparable level of theory/consistent methodology in such a way that data sets may be used together.
- More ready-to-use properties of liquids like those available on the Virtual Chemistry Web site.[255] These can be used for validating force fields by comparing to, e.g., the ThermoML database.[131] Addition of liquid mixtures would provide a useful resource as well.

Final Words. MD simulations of crystals highlight that there indeed is an imbalance in the forces predicted by force field models.[237,238,240,241] This imbalance can manifest itself in, for instance, changes in the unit cell shape or incorrect energies or melting points.[241] Similar problems have been described in simulations of organic liquids[33,34,66,67] and gas phase studies by force fields.[114,124] To improve on this in a systematic manner, it would be good if the force field community would adopt a number of the data sets described in this perspective. Addition of new data sets is still needed to cover particular areas of chemical space but much can be done with the existing data sets already. For the development of classical force fields as well as ML models it would be advantageous to

1. Perform training and testing including single molecule properties such as conformational energies, vibrational frequencies, and electrostatic properties. Examples of data sets targeting these properties are listed in Table 3.

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

2. Compare FF predictions of interaction energies to values of high-level benchmark data sets (in a way exemplified in Figures 1 and 2). Examples of such data sets are listed in Table 4. This may contribute to a more consistent and transferable evaluation of FFs. The NCIA and DES370 are particularly relevant.
3. Include the experimental properties of condensed phases in the analysis of performance, such that the transferability to bigger and more complex systems of interest may be improved.

METHODS

Energy calculations were performed using the GROMACS software[316] (version 2021) using the generalized Amber FF (GAFF).[74] Molecular topology files were downloaded from the Virtual Chemistry Web site,[255] based on the paper on thermochemistry calculations using FFs.[114] No cut-offs were used for Coulomb and van der Waals interactions. The dimer coordinates were taken from the S66 database.[210] The energy of all dimers and their constituting monomer was calculated using GROMACS. For the potential energy of minimized systems, the molecules were energy minimized until the potential energy had converged. Energies as well as the root-mean-square deviation of coordinates are given in Table S1.

ASSOCIATED CONTENT

Data Availability Statement

Input topologies corresponding to Figure 1 are available from github.[251]

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c01127.

Table containing interaction energies corresponding to Figure 1 and coordinate RMSDs for dimers from the S66 data set (PDF)

AUTHOR INFORMATION

Corresponding Author

David van der Spoel − *Department of Cell and Molecular Biology, Uppsala University, SE-75124 Uppsala, Sweden;* orcid.org/0000-0002-7659-8526; Phone: +46 18 471 4205; Email: david.vanderspoel@icm.uu.se

Authors

Kristian Kříž − *Department of Cell and Molecular Biology, Uppsala University, SE-75124 Uppsala, Sweden;* orcid.org/0000-0001-9072-0949

Lisa Schmidt − *Faculty of Biosciences, University of Heidelberg, Heidelberg 69117, Germany;* orcid.org/0000-0001-7058-3107

Alfred T. Andersson − *Department of Cell and Molecular Biology, Uppsala University, SE-75124 Uppsala, Sweden;* orcid.org/0000-0002-6101-4823

Marie-Madeleine Walz − *Department of Cell and Molecular Biology, Uppsala University, SE-75124 Uppsala, Sweden;* orcid.org/0000-0002-3932-5780

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c01127

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Levitt, M.; Lifson, S. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.* **1969**, *46*, 269−279.

(2) Cailliez, F.; Pernot, P.; Rizzi, F.; Jones, R.; Knio, O.; Arampatzis, G.; Koumoutsakos, P.Bayesian calibration of force fields for molecular simulations. *Uncertainty Quantification in Multiscale Materials Modeling*; Elsevier, 2020; pp 169−227.

(3) Dauber-Osguthorpe, P.; Hagler, A. 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*, 133−203.

(4) Hagler, A. 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*, 205−264.

(5) Jones, J. E. On the Determination of Molecular Fields. -I. From the variation of the viscosity of a gas with temperature. *Proc. Royal Soc. London A* **1924**, *106*, 441−462.

(6) Jones, J. E. On the determination of molecular fields. -II. From the equation of state of a gas. *Proc. Royal Soc. London A* **1924**, *106*, 463−477.

(7) Buckingham, R. A. The Classical Equation of State of Gaseous Helium, Neon and Argon. *Proc. R. Soc. London A* **1938**, *168*, 264−283.

(8) Wang, L.-P.; Chen, J.; Van Voorhis, T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.* **2013**, *9*, 452−460.

(9) Walz, M. M.; Ghahremanpour, M. M.; van Maaren, P. J.; van der Spoel, D. Phase-transferable force field for alkali halides. *J. Chem. Theory Comput.* **2018**, *14*, 5933−5948.

(10) Burrows, S. A.; Korotkin, I.; Smoukov, S. K.; Boek, E.; Karabasov, S. Benchmarking of Molecular Dynamics Force Fields for Solid-Liquid and Solid-Solid Phase Transitions in Alkanes. *J. Phys. Chem. B* **2021**, *125*, 5145−5159.

(11) Bernhardt, M. P.; Nagata, Y.; van der Vegt, N. F. A. Where Lennard-Jones Potentials Fail: Iterative Optimization of Ion-Water Pair Potentials Based on Ab Initio Molecular Dynamics Data. *J. Phys. Chem. Lett.* **2022**, *13*, 3712−3717.

(12) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annu. Rev. Biophys.* **2019**, *48*, 371−394.

(13) Shaimardanov, A. R.; Shulga, D. A.; Palyulin, V. A. Is an Inductive Effect Explicit Account Required for Atomic Charges Aimed at Use within the Force Fields? *J. Phys. Chem. A* **2022**, *126*, 6278.

(14) Mohebifar, M.; Johnson, E. R.; Rowley, C. N. Evaluating Force-Field London Dispersion Coefficients Using the Exchange-Hole Dipole Moment Model. *J. Chem. Theory Comput.* **2017**, *13*, 6146−6157.

(15) Walters, E.; Mohebifar, M.; Johnson, E. R.; Rowley, C. N. Evaluating the London Dispersion Coefficients of Protein Force Fields Using the Exchange-Hole Dipole Moment Model. *J. Phys. Chem. B* **2018**, *122*, 6690−6701.

(16) Bashardanesh, Z.; van der Spoel, D. Impact of Dispersion Coefficient on Simulations of Proteins and Organic Liquids. *J. Phys. Chem. B* **2018**, *122*, 8018−8027.

(17) Qiu, Y.; Shan, W.; Zhang, H. Force Field Benchmark of Amino Acids. 3. Hydration with Scaled Lennard-Jones Interactions. *J. Chem. Inf. Model.* **2021**, *61*, 3571−3582.

(18) Liu, C.; Qi, R.; Wang, Q.; Piquemal, J.-P.; Ren, P. Capturing Many-Body Interactions with Classical Dipole Induction Models. *J. Chem. Theory Comput.* **2017**, *13*, 2751−2761.

(19) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Gotz, A. W.; Paesani, F. On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *J. Chem. Phys.* **2016**, *145*, 194504.

(20) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Yu, Q.; Bowman, J. M. A CCSD(T)-Based 4-Body Potential for Water. *J. Phys. Chem. Lett.* **2021**, *12*, 10318−10324.

(21) Yu, Q.; Qu, C.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. q-AQUA: A Many-Body CCSD(T) Water Potential, Including Four-Body Interactions, Demonstrates the Quantum Nature of Water from Clusters to the Liquid Phase. *J. Phys. Chem. Lett.* **2022**, *13*, 5068−5074.
(22) Cisneros, G. A.; Wikfeldt, K. T.; Ojamae, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartok, A. P.; Csanyi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. *Chem. Rev.* **2016**, *116*, 7501−7528.
(23) Ströker, P.; Hellmann, R.; Meier, K. Thermodynamic properties of argon from Monte Carlo simulations using ab initio potentials. *Phys. Rev. E* **2022**, *105*, 064129.
(24) van der Spoel, D.; Zhang, J.; Zhang, H. Quantitative predictions from molecular simulations using explicit or implicit interactions. *WIREs Comput. Mol. Sci.* **2022**, *12*, No. e1560.
(25) van der Spoel, D. Systematic design of biomolecular force fields. *Curr. Opin. Struct. Biol.* **2021**, *67*, 18−24.
(26) He, X.; Walker, B.; Man, V. H.; Ren, P.; Wang, J. Recent progress in general force fields of small molecules. *Curr. Opin. Struct. Biol.* **2022**, *72*, 187−193.
(27) Lewis-Atwell, T.; Townsend, P. A.; Grayson, M. N. Comparisons of different force fields in conformational analysis and searching of organic molecules: A review. *Tetrahedron* **2021**, *79*, 131865.
(28) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.
(29) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(30) Schauperl, M.; Nerenberg, P. S.; Jang, H.; Wang, L.-P.; Bayly, C. I.; Mobley, D. L.; Gilson, M. K. Non-bonded force field model with advanced restrained electrostatic potential charges (RESP2). *Commun. Chem.* **2020**, *3*, 44.
(31) Bleiziffer, P.; Schaller, K.; Riniker, S. Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations. *J. Chem. Inf. Model.* **2018**, *58*, 579−590.
(32) Hosseini, A.; Lund, M.; Ejtehadi, M. R. Electronic polarization effects on membrane translocation of anti-cancer drugs. *Phys. Chem. Chem. Phys.* **2022**, *24*, 12281−12292.
(33) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Compressibility, Expansion Coefficient and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61−74.
(34) Zhang, J.; Tuguldur, B.; van der Spoel, D. Force field benchmark II: Gibbs energy of solvation of organic molecules in organic liquids. *J. Chem. Inf. Model.* **2015**, *55*, 1192−1201.
(35) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142−10186.
(36) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
(37) Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* **2013**, *15*, 095003.
(38) Ramakrishnan, R.; Hartmann, M.; Tapavicza, E.; Von Lilienfeld, O. A. Electronic spectra from TDDFT and machine learning in chemical space. *J. Chem. Phys.* **2015**, *143*, 084111.
(39) Rupp, M.; Ramakrishnan, R.; von Lilienfeld, O. A. Machine Learning for Quantum Mechanical Properties of Atoms in Molecules. *J. Phys. Chem. Lett.* **2015**, *6*, 3309−3313.
(40) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, 170901.
(41) Li, Y.; Li, H.; Pickard, F. C.; Narayanan, B.; Sen, F. G.; Chan, M. K. Y.; Sankaranarayanan, S. K. R. S.; Brooks, B. R.; Roux, B. Machine Learning Force Field Parameters from Ab Initio Data. *J. Chem. Theory Comput.* **2017**, *13*, 4492−4503.
(42) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192−3203.
(43) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, e1603015.
(44) Nakata, M.; Shimazaki, T. PubChemQC Project: A Large-Scale First-Principles Electronic Structure Database for Data-Driven Chemistry. *J. Chem. Inf. Model.* **2017**, *57*, 1300−1308.
(45) Bereau, T.; DiStasio, R. A.; Tkatchenko, A.; von Lilienfeld, O. A. Non-covalent interactions across organic and biological subsets of chemical space: Physics-based potentials parametrized from machine learning. *J. Chem. Phys.* **2018**, *148*, 241706.
(46) Gastegger, M.; Schwiedrzik, L.; Bittermann, M.; Berzsenyi, F.; Marquetand, P. wACSF-Weighted atom-centered symmetry functions as descriptors in machine learning potentials. *J. Chem. Phys.* **2018**, *148*, 241709.
(47) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148*, 241733.
(48) Hughes, Z. E.; Ren, E.; Thacker, J. C. R.; Symons, B. C. B.; Silva, A. F.; Popelier, P. L. A. A FFLUX Water Model: Flexible, Polarizable and with a Multipolar Description of Electrostatics. *J. Comput. Chem.* **2019**, 619.
(49) Chmiela, S.; Sauceda, H. E.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. sGDML: Constructing accurate and data efficient molecular force fields using machine learning. *Comput. Phys. Commun.* **2019**, *240*, 38−45.
(50) Wilkins, D. M.; Grisafi, A.; Yang, Y.; Lao, K. U.; DiStasio, R. A.; Ceriotti, M. Accurate molecular polarizabilities with coupled cluster theory and machine learning. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 3401−3406.
(51) Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *J. Chem. Phys.* **2019**, *150*, 114102.
(52) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **2019**, *10*, 2903.
(53) Schütt, K. T.; Gastegger, M.; Tkatchenko, A.; Müller, K.-R.; Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **2019**, *10*, 5024.
(54) Heinen, S.; Schwilk, M.; von Rudorff, G. F.; von Lilienfeld, O. A. Machine learning the computational cost of quantum chemistry. *Mach. Learn.: Sci. Technol.* **2020**, *1*, 025002.
(55) Bogojeski, M.; Vogt-Maranto, L.; Tuckerman, M. E.; Müller, K.-R.; Burke, K. Quantum chemical accuracy from density functional approximations via machine learning. *Nat. Commun.* **2020**, *11*, 5223.
(56) Gastegger, M.; Schütt, K. T.; Müller, K.-R. Machine learning of solvent effects on molecular spectra and reactions. *Chem. Sci.* **2021**, *12*, 11473−11483.
(57) Kulik, H. J. What's Left for a Computational Chemist To Do in the Age of Machine Learning? *Isr. J. Chem.* **2022**, *62*, No. e202100016.
(58) Willcox, O.; Ghattas, K. E.; Heimbach, P. The imperative of physics-based modeling and inverse theory in computational science. *Nat. Comput. Sci.* **2021**, *1*, 166−168.
(59) Greathouse, J.; Cygan, R. In *Handbook of Clay Science*; Bergaya, F., Lagaly, G., Eds.; Developments in Clay Science; Elsevier, 2013; Vol. *5*, pp 405−423.
(60) Padilla Espinosa, I. M.; Jacobs, T. D. B.; Martini, A. Evaluation of Force Fields for Molecular Dynamics Simulations of Platinum in Bulk and Nanoparticle Forms. *J. Chem. Theory Comput.* **2021**, *17*, 4486−4498.

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

(61) Riniker, S. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *J. Chem. Inf. Model.* **2018**, *58*, 565−578.
(62) Brooks, C. L.; Case, D. A.; Plimpton, S.; Roux, B.; van der Spoel, D.; Tajkhorshid, E. Classical molecular dynamics. *J. Chem. Phys.* **2021**, *154*, 100401.
(63) Manzetti, S.; van der Spoel, E. R.; van der Spoel, D. Chemical Properties, Environmental Fate, and Degradation of Seven Classes of Pollutants. *Chem. Res. Toxicol.* **2014**, *27*, 713−737.
(64) van der Spoel, D.; Zhang, H.; Manzetti, S.; Klamt, A. Prediction of partition coefficients of environmental toxins using computational chemistry methods. *ACS Omega* **2019**, *4*, 13772−13781.
(65) Mata, R. A.; Suhm, M. A. Benchmarking Quantum Chemical Methods: Are We Heading in the Right Direction? *Angew. Chem.-Int. Ed.* **2017**, *56*, 11011−11018.
(66) Fischer, N. M.; van Maaren, P. J.; Ditz, J. C.; Yildirim, A.; van der Spoel, D. Properties of liquids in Molecular Dynamics Simulations with explicit long-range Lennard-Jones interactions. *J. Chem. Theory Comput.* **2015**, *11*, 2938−2944.
(67) Zhang, J.; Tuguldur, B.; van der Spoel, D. Correction to Force field benchmark II: Gibbs energy of solvation of organic molecules in organic liquids. *J. Chem. Inf. Model.* **2016**, *56*, 819−820.
(68) Cerutti, D. S.; Swope, W. C.; Rice, J. E.; Case, D. A. ff14ipq: A Self-Consistent Force Field for Condensed-Phase Simulations of Proteins. *J. Chem. Theory Comput.* **2014**, *10*, 4515−4534.
(69) Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **2016**, *12*, 3926−3947.
(70) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(71) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16*, 528−552.
(72) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.
(73) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.* **2018**, *14*, 2084−2108.
(74) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General AMBER Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(75) He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **2020**, *153*, 114502.
(76) Huang, J.; Simmonett, A. C.; Pickard IV, F. C.; MacKerell, A. D., Jr.; Brooks, B. R. Mapping the Drude polarizable force field onto a multipole and induced dipole model. *J. Chem. Phys.* **2017**, *147*, 161702.
(77) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(78) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D., Jr. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theory Comput* **2013**, *9*, 5430−5449.
(79) Reif, M. M.; 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*, 1247−1264.
(80) Horta, B. A. C.; Merz, P. T.; Fuchs, P. F. J.; Dolenc, J.; Riniker, S.; Huenenberger, P. H. A GROMOS-Compatible Force Field for Small Organic Molecules in the Condensed Phase: The 2016H66 Parameter Set. *J. Chem. Theory Comput.* **2016**, *12*, 3825−3850.
(81) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569−2577.
(82) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(83) Harder, E.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281−296.
(84) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1863−1874.
(85) Allinger, Y.; Yuh, N. L.; Lii, J. Molecular mechanics - THE MM3 force-field for hydrocarbons 0.1. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.
(86) Halgren, T. A. Merck molecular force field. 1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(87) Halgren, T. A. Merck molecular force field. 2. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.
(88) Halgren, T. A. Merck molecular force field. 3. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553−586.
(89) Halgren, T. A.; Nachbar, R. B. Merck molecular force field. IV. Conformational energies and geometries for MMFF94. *J. Comput. Chem.* **1996**, *17*, 587−615.
(90) Halgren, T. A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* **1996**, *17*, 616−641.
(91) Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.* **1999**, *20*, 730−748.
(92) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720−729.
(93) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Escaping Atom Types in Force Fields Using Direct Chemical Perception. *J. Chem. Theory Comput.* **2018**, *14*, 6076−6092.
(94) Qiu, Y.; et al. Development and Benchmarking of Open Force Field v1.0.0 - the Parsley Small-Molecule Force Field. *J. Chem. Theory Comput.* **2021**, *17*, 6262−6280.
(95) Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; Verstraelen, T.; Grama, A.; van Duin, A. C. T. The ReaxFF reactive force-field: development, applications and future directions. *npj Comput. Mater.* **2016**, *2*, 15011.
(96) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. Gaussian-1 theory: A general procedure for prediction of molecular energies. *J. Chem. Phys.* **1989**, *90*, 5622−5629.
(97) Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. Gaussian-1 theory of molecular energies for second-row compounds. *J. Chem. Phys.* **1990**, *93*, 2537−2545.
(98) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *J. Chem. Phys.* **1998**, *109*, 7764−7776.
(99) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 theory. *J. Chem. Phys.* **2007**, *126*, 084108.
(100) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gn theory. *WIREs. Comput. Mol. Sci.* **2011**, *1*, 810−825.

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

(101) Martin, E. J.; Critchlow, R. E. Beyond Mere Diversity: Tailoring Combinatorial Libraries for Drug Discovery. *J. Comb. Chem.* **1999**, *1*, 32−45.
(102) Parthiban, S.; Martin, J. M. L. Fully ab initio atomization energy of benzene via Weizmann-2 theory. *J. Chem. Phys.* **2001**, *115*, 2051−2054.
(103) Barnes, E. C.; Petersson, G. A.; Montgomery, J. A.; Frisch, M. J.; Martin, J. M. L. Unrestricted Coupled Cluster and Brueckner Doubles Variations of W1 Theory. *J. Chem. Theory Comput.* **2009**, *5*, 2687−2693.
(104) Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions. *J. Chem. Phys.* **2006**, *125*, 144108.
(105) Karton, A.; Daon, S.; Martin, J. M. W4−11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem. Phys. Lett.* **2011**, *510*, 165−178.
(106) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J. Chem. Phys.* **1999**, *110*, 2822−2827.
(107) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VII. Use of the minimum population localization method. *J. Chem. Phys.* **2000**, *112*, 6532−6542.
(108) Ghahremanpour, M. M.; van Maaren, P. J.; Ditz, J.; Lindh, R.; van der Spoel, D. Large-scale calculations of gas phase thermochemistry: enthalpy of formation, standard entropy and heat capacity. *J. Chem. Phys.* **2016**, *145*, 114305.
(109) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-3 and density functional theories for a larger experimental test set. *J. Chem. Phys.* **2000**, *112*, 7374−7383.
(110) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Assessment of Gaussian-3 and density-functional theories on the G3/05 test set of experimental energies. *J. Chem. Phys.* **2005**, *123*, 124107.
(111) Karton, A.; Sylvetsky, N.; Martin, J. M. L. W4-17: A diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods. *J. Comput. Chem.* **2017**, *38*, 2063−2075.
(112) Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals-Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 291−309.
(113) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184−32215.
(114) van der Spoel, D.; Ghahremanpour, M. M.; Lemkul, J. Small Molecule Thermochemistry: A Tool For Empirical Force Field Development. *J. Phys. Chem. A* **2018**, *122*, 8982−8988.
(115) Ghahremanpour, M. M.; van Maaren, P. J.; van der Spoel, D. The Alexandria library: a quantum chemical database of molecular properties for force field development. *Sci. Data* **2018**, *5*, 180062.
(116) Henschel, H.; Andersson, A. T.; Jespers, W.; Mehdi Ghahremanpour, M.; van der Spoel, D. Theoretical Infrared Spectra: Quantitative Similarity Measures and Force Fields. *J. Chem. Theory Comput.* **2020**, *16*, 3307−3315.
(117) Henschel, H.; van der Spoel, D. An Intuitively Understandable Quality Measure for Theoretical Vibrational Spectra. *J. Phys. Chem. Lett.* **2020**, *11*, 5471−5475.
(118) Laury, M. L.; Boesch, S. E.; Haken, I.; Sinha, P.; Wheeler, R. A.; Wilson, A. K. Harmonic vibrational frequencies: Scale factors for pure, hybrid, hybrid meta, and double-hybrid functionals in conjunction with correlation consistent basis sets. *J. Comput. Chem.* **2011**, *32*, 2339−2347.
(119) Allinger, N. L.; Schmitz, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K. Heats of formation of organic molecules. 2. The basis for calculations using either ab initio or molecular mechanics methods. Alcohols and ethers. *J. Am. Chem. Soc.* **1992**, *114*, 2880−2883.
(120) Yaws, C. L.*Yaws' Handbook of Thermodynamic Properties for Hydrocarbons and Chemicals*; Knovel: 2009.
(121) Engh, R. A.; Huber, R. Accurate Bond and Angle Parameters for X-Ray Protein Structure Refinement. *Act. Crystallogr. A* **1991**, *47*, 392−400.
(122) Allinger, N. L.; Chen, K. S.; Katzenellenbogen, J. A.; Wilson, S. R.; Anstead, G. M. Hyperconjugative effects on carbon-carbon bond lengths in molecular mechanics (MM4). *J. Comput. Chem.* **1996**, *17*, 747−755.
(123) Allinger, N. L.; Chen, K. S.; Lii, J. H. An improved force field (MM4) for saturated hydrocarbons. *J. Comput. Chem.* **1996**, *17*, 642−668.
(124) Lim, V. T.; Hahn, D. F.; Tresadern, G.; Bayly, C. I.; Mobley, D. L. Benchmark assessment of molecular geometries and energies from small molecule force fields. *F1000Research* **2020**, *9*, 1390.
(125) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.
(126) Jorgensen, W. L.; Tirado-Rives, J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665−6670.
(127) Pechlaner, M.; Reif, M. M.; Oostenbrink, C. Reparametrisation of united-atom amine solvation in the GROMOS force field. *Mol. Phys.* **2017**, *115*, 1144−1154.
(128) Rumble, J.*CRC Handbook of Chemistry and Physics*, 103rd ed.; CRC Press: Gaitherburg, MD, 2022.
(129) Rowley, R. L.; Wilding, W. V.; Oscarson, J. L.; Yang, Y.; Giles, N. F.*Data Compilation of Pure Chemical Properties*; Design Institute for Physical Properties, American Institute for Chemical Engineering: New York, 2012.
(130) Yaws, C. L.*Yaws' Handbook of Physical Properties for Hydrocarbons and Chemicals*; Knovel, 2008.
(131) *NIST ThermoML Archive*. 2002; https://trc.nist.gov/ThermoML/.
(132) Grisafi, A.; Ceriotti, M. Incorporating long-range physics in atomic-scale machine learning. *J. Chem. Phys.* **2019**, *151*, 204105.
(133) Staacke, C. G.; Heenen, H. H.; Scheurer, C.; Csanyi, G.; Reuter, K.; Margraf, J. T. On the Role of Long-Range Electrostatics in Machine-Learned Interatomic Potentials for Complex Battery Materials. *ACS Appl. Energy Mater.* **2021**, *4*, 12562−12569.
(134) Gao, A.; Remsing, R. C. Self-consistent determination of long-range electrostatics in neural network potentials. *Nat. Commun.* **2022**, *13*, 1572.
(135) Zhang, L.; Wang, H.; Muniz, M. C.; Panagiotopoulos, A. Z.; Car, R.; E, W. A deep potential model with long-range electrostatic interactions. *J. Chem. Phys.* **2022**, *156*, 124107.
(136) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 3887.
(137) Blum, L. C.; Reymond, J.-L. 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *J. Am. Chem. Soc.* **2009**, *131*, 8732−8733.
(138) Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864−2875.
(139) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, 140022.
(140) Gebauer, N.; Gastegger, M.; Schütt, K. Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules. *Adv. Neural Inf. Proc. Syst.* **2019**, *32*, 4132.
(141) Chen, G.; Chen, P.; Hsieh, C.-Y.; Lee, C.-K.; Liao, B.; Liao, R.; Liu, W.; Qiu, J.; Sun, Q.; Tang, J.; et al.Alchemy: A quantum chemistry dataset for benchmarking ai models. *arXiv* preprint, arXiv:1906.09427, 2019, https://arxiv.org/abs/1906.09427.
(142) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular

Mechanics and Molecular Dynamics. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(143) Welborn, M.; Cheng, L.; Miller, T. F., III Transferability in machine learning for electronic structure via the molecular orbital basis. *J. Chem. Theory Comput.* **2018**, *14*, 4772−4779.
(144) Cheng, L.; Welborn, M.; Christensen, A. S.; Miller, T. F., III A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules. *J. Chem. Phys.* **2019**, *150*, 131103.
(145) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **2017**, *8*, 13890.
(146) Schütt, K.; Kindermans, P.-J.; Sauceda Felix, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *Adv. Neural Inf Process Syst* **2017**, *30*, 992.
(147) Fink, T.; Bruggesser, H.; Reymond, J.-L. Virtual Exploration of the Small-Molecule Chemical Universe below 160 Da. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 1504−1508.
(148) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. *Sci. Data* **2017**, *4*, 170193.
(149) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Sci. Adv.* **2019**, *5*, No. eaav6490.
(150) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Sci. Data* **2020**, *7*, 134.
(151) Vazquez-Salazar, L. I.; Boittier, E. D.; Unke, O. T.; Meuwly, M. Impact of the Characteristics of Quantum Chemical Databases on Machine Learning Prediction of Tautomerization Energies. *J. Chem. Theory Comput.* **2021**, *17*, 4769−4785.
(152) Stuke, A.; Kunkel, C.; Golze, D.; Todorović, M.; Margraf, J. T.; Reuter, K.; Rinke, P.; Oberhofer, H. Atomic structures and orbital energies of 61,489 crystal-forming organic molecules. *Sci. Data* **2020**, *7*, 58.
(153) Balcells, D.; Skjelstad, B. B. tmQM Dataset-quantum geometries and properties of 86k transition metal complexes. *J. Chem. Inf. Model.* **2020**, *60*, 6135−6146.
(154) Isert, C.; Atz, K.; Jiménez-Luna, J.; Schneider, G. QMugs, quantum mechanical properties of drug-like molecules. *Sci. Data* **2022**, *9*, 273.
(155) Herr, J. E.; Yao, K.; McIntyre, R.; Toth, D. W.; Parkhill, J. Metadynamics for training neural network model chemistries: A competitive assessment. *J. Chem. Phys.* **2018**, *148*, 241710.
(156) Yao, K.; Herr, J. E.; Toth, D. W.; Mckintyre, R.; Parkhill, J. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.* **2018**, *9*, 2261−2269.
(157) Glavatskikh, M.; Leguy, J.; Hunault, G.; Cauchy, T.; Da Mota, B. Dataset's chemical diversity limits the generalizability of machine learning predictions. *J. Cheminf.* **2019**, *11*, 69.
(158) Ghahremanpour, M. M.; van Maaren, P. J.; van der Spoel, D. Alexandria Library [Data set]. *Zenodo* **2017**, 1004711.
(159) Peverati, R.; Truhlar, D. G. Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics. *Philos. T. R. Soc. A* **2014**, *372*, 20120476.
(160) Morgante, P.; Peverati, R. ACCDB: A collection of chemistry databases for broad computational purposes. *J. Comput. Chem.* **2019**, *40*, 839−848.
(161) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **2009**, *180*, 2175−2196.
(162) Řezáč, J.; Bím, D.; Gutten, O.; Rulíšek, L. Toward accurate conformational energies of smaller peptides and medium-sized macrocycles: MPCONF196 benchmark energy data set. *J. Chem. Theory Comput.* **2018**, *14*, 1254−1266.
(163) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678−3693.
(164) Schwilk, M.; Tahchieva, D. N.; von Lilienfeld, O. A.Large yet bounded: Spin gap ranges in carbenes. *arXiv* preprint, arXiv:2004.10600, 2020, https://arxiv.org/abs/2004.10600.
(165) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1−86). *J. Chem. Theory Comput.* **2017**, *13*, 1989−2009.
(166) Hoja, J.; Medrano Sandonas, L.; Ernst, B. G.; Vazquez-Mayagoitia, A.; DiStasio, R. A., Jr; Tkatchenko, A. QM7-X, a comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules. *Sci. Data* **2021**, *8*, 43.
(167) Vazquez-Salazar, L. I.; Meuwly, M. ANI-1E: An equilibrium database from the ANI-1 database. *Zenodo* **2021**, 5549536.
(168) Prasad, V. K.; Khalilian, M. H.; Otero-de-la Roza, A.; DiLabio, G. A. BSE49, a diverse, high-quality benchmark dataset of separation energies of chemical bonds. *Sci. Data* **2021**, *8*, 300.
(169) Guan, X.; Das, A.; Stein, C. J.; Heidar-Zadeh, F.; Bertels, L.; Liu, M.; Haghighatlari, M.; Li, J.; Zhang, O.; Hao, H.; et al. A benchmark dataset for Hydrogen Combustion. *Sci. Data* **2022**, *9*, 215.
(170) Lillestolen, T. C.; Wheatley, R. J. Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *ChemComm* **2008**, 5909−5911.
(171) Thürlemann, M.; Böselt, L.; Riniker, S. Learning Atomic Multipoles: Prediction of the Electrostatic Potential with Equivariant Graph Neural Networks. *J. Chem. Theory Comput.* **2022**, *18*, 1701−1710.
(172) Zapata Trujillo, J. C.; McKemmish, L. K. VIBFREQ1295: A New Database for Vibrational Frequency Calculations. *J. Phys. Chem. A* **2022**, *126*, 4100−4122.
(173) Chan, B.; Radom, L. W1X-1 and W1X-2: W1-Quality Accuracy with an Order of Magnitude Reduction in Computational Cost. *J. Chem. Theory Comput.* **2012**, *8*, 4259−4269.
(174) Chan, B. High-Level Quantum Chemistry Reference Heats of Formation for a Large Set of C, H, N, and O Species in the NIST Chemistry Webbook and the Identification and Validation of Reliable Protocols for Their Rapid Computation. *J. Phys. Chem. A* **2022**, *126*, 4981−4990.
(175) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. Comparing Counterpoise-Corrected, Uncorrected, and Averaged Binding Energies for Benchmarking Noncovalent Interactions. *J. Chem. Theory Comput.* **2014**, *10*, 49−57.
(176) McDaniel, J. G.; Schmidt, J. Next-Generation Force Fields from Symmetry-Adapted Perturbation Theory. *Annu. Rev. Phys. Chem.* **2016**, *67*, 467−488.
(177) Kodrycka, M.; Patkowski, K. Platinum, gold, and silver standards of intermolecular interaction energy calculations. *J. Chem. Phys.* **2019**, *151*, 070901.
(178) Dunning, T. H., Jr. Gaussian-basis sets for use in correlated molecular calculations. 1. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(179) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (SCS), explicitly correlated (F12), and specialized wavefunction approaches. *J. Chem. Phys.* **2014**, *141*, 234111.
(180) Fujii, A.; Hayashi, H.; Park, J. W.; Kazama, T.; Mikami, N.; Tsuzuki, S. Experimental and theoretical determination of the accurate CH/π interaction energies in benzene − alkane clusters: correlation between interaction energy and polarizability. *Phys. Chem. Chem. Phys.* **2011**, *13*, 14131−14141.
(181) Černý, J.; Tong, X.; Hobza, P.; Müller-Dethlefs, K. State of the art theoretical study and comparison to experiment for the phenol ··· argon complex. *J. Chem. Phys.* **2008**, *128*, 114319.
(182) Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the "Gold

Standard," CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151−2155.
(183) Řezáč, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016**, *116*, 5038−5071.
(184) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.
(185) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. Basis set consistent revision of the S22 test set of noncovalent interaction energies. *J. Chem. Phys.* **2010**, *132*, 144104.
(186) Riley, K. E.; Hobza, P. Assessment of the MP2Method, along with Several Basis Sets, for the Computation of Interaction Energies of Biologically Relevant Hydrogen Bonded and Dispersion Bound Complexes. *J. Phys. Chem. A* **2007**, *111*, 8257−8263.
(187) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction, $\delta_{MP2}^{CCSD(T)}$: Best practices for benchmarking non-covalent interactions and the attendant revision of the S22, NBC10, HBC6, and HSG databases. *J. Chem. Phys.* **2011**, *135*, 194102.
(188) Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. Comparative Study of Selected Wave Function and Density Functional Methods for Noncovalent Interaction Energy Calculations Using the Extended S22 Data Set. *J. Chem. Theory Comput.* **2010**, *6*, 2365−2376.
(189) Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466−3470.
(190) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 2609−2618.
(191) Sparrow, Z. M.; Ernst, B. G.; Joo, P. T.; Lao, K. U.; DiStasio, R. A. NENCI-2021. I. A large benchmark database of non-equilibrium non-covalent interactions emphasizing close intermolecular contacts. *J. Chem. Phys.* **2021**, *155*, 184303.
(192) *Non-Covalent Interactions Atlas*. 2020; http://www.nciatlas.org/.
(193) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.* **2020**, *16*, 2355−2368.
(194) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 2: Hydrogen Bonding in an Extended Chemical Space. *J. Chem. Theory Comput.* **2020**, *16*, 6305−6316.
(195) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 5: London Dispersion in an Extended Chemical Space. *Phys. Chem. Chem. Phys.* **2022**, *24*, 14780−14793.
(196) Kříž, K.; Nováček, M.; Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 3: Repulsive Contacts. *J. Chem. Theory Comput.* **2021**, *17*, 1548−1561.
(197) Kříž, K.; Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets 4: σ-Hole Interactions. *Phys. Chem. Chem. Phys.* **2022**, *24*, 14794−14804.
(198) Donchev, A. G.; Taube, A. G.; Decolvenaere, E.; Hargus, C.; McGibbon, R. T.; Law, K.-H.; Gregersen, B. A.; Li, J.-L.; Palmo, K.; Siva, K.; Bergdorf, M.; Klepeis, J. L.; Shaw, D. E. Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Sci. Data* **2021**, *8*, 55.
(199) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(200) Bryantsev, V. S.; Diallo, M. S.; van Duin, A. C. T.; Goddard, W. A. Evaluation of B3LYP, X3LYP, and M06-Class Density Functionals for Predicting the Binding Energies of Neutral, Protonated, and Deprotonated Water Clusters. *J. Chem. Theory Comput.* **2009**, *5*, 1016−1026.
(201) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364−3374.
(202) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Eur. J. Chem.* **2012**, *18*, 9955−9964.
(203) Sure, R.; Grimme, S. Comprehensive Benchmark of Association (Free) Energies of Realistic Host−Guest Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 3785−3801.
(204) Kříž, K.; Řezáč, J. Benchmarking of Semiempirical Quantum-Mechanical Methods on Systems Relevant to Computer-Aided Drug Design. *J. Chem. Inf. Model.* **2020**, *60*, 1453−1460.
(205) Ni, Z.; Guo, Y.; Neese, F.; Li, W.; Li, S. Cluster-in-Molecule Local Correlation Method with an Accurate Distant Pair Correction for Large Systems. *J. Chem. Theory Comput.* **2021**, *17*, 756−766.
(206) *Benchmark Geometry and Energy Database*. 2008; http://www.begdb.org/.
(207) *QCArchive − Machine Learning Datasets Repository*. https://qcarchive.molssi.org/apps/ml_datasets/, (accessed on 06/29/2022).
(208) *Computational Chemistry Comparison and Benchmark DataBase*. https://cccbdb.nist.gov/introx.asp, (accessed on 06/29/2022).
(209) Berka, K.; Laskowski, R.; Riley, K. E.; Hobza, P.; Vondrášek, J. Representative Amino Acid Side Chain Interactions in Proteins. A Comparison of Highly Accurate Correlated ab Initio Quantum Chemical and Empirical Potential Procedures. *J. Chem. Theory Comput.* **2009**, *5*, 982−992.
(210) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.
(211) Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Kennedy, M. R.; Sherrill, C. D.; Merz, K. M. Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein-Ligand Complexes. *J. Chem. Theory Comput.* **2011**, *7*, 790−797.
(212) Thanthiriwatte, K. S.; Hohenstein, E. G.; Burns, L. A.; Sherrill, C. D. Assessment of the Performance of DFT and DFT-D Methods for Describing Distance Dependence of Hydrogen-Bonded Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 88−96.
(213) Karthikeyan, S.; Sedlak, R.; Hobza, P. on the Nature of Stabilization in Weak, Medium, and Strong Charge-Transfer Complexes: CCSD(T)/CBS and SAPT Calculations. *J. Phys. Chem. A* **2011**, *115*, 9422−9428.
(214) Řezáč, J.; de la Lande, A. Robust, Basis-Set Independent Method for the Evaluation of Charge-Transfer Energy in Noncovalent Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 528−537.
(215) Mintz, B. J.; Parks, J. M. Benchmark Interaction Energies for Biologically Relevant Noncovalent Complexes Containing Divalent Sulfur. *J. Phys. Chem. A* **2012**, *116*, 1086−1092.
(216) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285−4292.
(217) Granatier, J.; Pitoňák, M.; Hobza, P. Accuracy of Several Wave Function and Density Functional Theory Methods for Description of Noncovalent Interaction of Saturated and Unsaturated Hydrocarbon Dimers. *J. Chem. Theory Comput.* **2012**, *8*, 2282−2292.
(218) Řezáč, J.; Dubecký, M.; Jurečka, P.; Hobza, P. Extensions and applications of the A24 data set of accurate interaction energies. *Phys. Chem. Chem. Phys.* **2015**, *17*, 19268−19277.
(219) Bauzá, A.; Alkorta, I.; Frontera, A.; Elguero, J. On the Reliability of Pure and Hybrid DFT Methods for the Evaluation of Halogen, Chalcogen, and Pnicogen Bonds Involving Anionic and Neutral Electron Donors. *J. Chem. Theory Comput.* **2013**, *9*, 5201−5210.
(220) Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918−1931.
(221) Parker, T. M.; Sherrill, C. D. Assessment of Empirical Models versus High-Accuracy Ab Initio Methods for Nucleobase Stacking: Evaluating the Importance of Charge Penetration. *J. Chem. Theory Comput.* **2015**, *11*, 4197−4204.
(222) Hostaš, J.; Jakubec, D.; Laskowski, R. A.; Gnanasekaran, R.; Řezáč, J.; Vondrášek, J.; Hobza, P. Representative Amino Acid Side-Chain Interactions in Protein−DNA Complexes: A Comparison of

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

Highly Accurate Correlated Ab Initio Quantum Mechanical Calculations and Efficient Approaches for Applications to Large Systems. *J. Chem. Theory Comput.* **2015**, *11*, 4086−4092.
(223) Temelso, B.; Renner, C. R.; Shields, G. C. Importance and Reliability of Small Basis Set CCSD(T) Corrections to MP2 Binding and Relative Energies of Water Clusters. *J. Chem. Theory Comput.* **2015**, *11*, 1439−1448.
(224) Burns, L. A.; Faver, J. C.; Zheng, Z.; Marshall, M. S.; Smith, D. G. A.; Vanommeslaeghe, K.; MacKerell, A. D.; Merz, K. M.; Sherrill, C. D. The BioFragment Database (BFDb): An open-data platform for computational chemistry analysis of noncovalent interactions. *J. Chem. Phys.* **2017**, *147*, 161727.
(225) López, R.; Díaz, N.; Suárez, D. Alkali and Alkaline-Earth Cations in Complexes with Small Bioorganic Ligands: Ab Initio Benchmark Calculations and Bond Energy Decomposition. *ChemPhysChem* **2020**, *21*, 99−112.
(226) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013**, *34*, 1672−1685.
(227) Li, S.; Shen, J.; Li, W.; Jiang, Y. An efficient implementation of the "cluster-in-molecule" approach for local electron correlation calculations. *J. Chem. Phys.* **2006**, *125*, 074109.
(228) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.
(229) Kumar, P.; Bojarowski, S. A.; Jarzembska, K. N.; Domagala, S.; Vanommeslaeghe, K.; MacKerell, A. D., Jr.; Dominiak, P. M. A Comparative Study of Transferable Aspherical Pseudoatom Databank and Classical Force Fields for Predicting Electrostatic Interactions in Molecular Dimers. *J. Chem. Theory and Comput.* **2014**, *10*, 1652−1664.
(230) Bojarowski, S. A.; Kumar, P.; Dominiak, P. M. A Universal and Straightforward Approach to Include Penetration Effects in Electrostatic Interaction Energy Estimation. *Comput. Phys. Commun.* **2016**, *17*, 2455−2460.
(231) Ghahremanpour, M. M.; van Maaren, P. J.; Caleman, C.; Hutchison, G. R.; van der Spoel, D. Polarizable drude model with s-type gaussian or slater charge density for general molecular mechanics force fields. *J. Chem. Theory Comput.* **2018**, *14*, 5553−5566.
(232) Vandenbrande, S.; Waroquier, M.; Speybroeck, V. V.; Verstraelen, T. The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions. *J. Chem. Theory Comput.* **2017**, *13*, 161−179.
(233) Hagler, A. T.; Huler, E.; Lifson, S. Energy Functions for Peptides and Proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J. Am. Chem. Soc.* **1974**, *96*, 5319−5327.
(234) Hagler, A. T.; Lifson, S.; Dauber, P. Consistent force-field studies of inter-molecular forces in hydrogen-bonded crystals 0.2. Benchmark for the objective comparison of alternative force-fields. *J. Am. Chem. Soc.* **1979**, *101*, 5122−5130.
(235) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta. Crystallogr. B* **2016**, *72*, 171−179.
(236) Bowskill, D. H.; Sugden, I. J.; Konstantinopoulos, S.; Adjiman, C. S.; Pantelides, C. C. Crystal Structure Prediction Methods for Organic Molecules: State of the Art. *Annu. Rev. Chem. Biomol. Eng.* **2021**, *12*, 593−623.
(237) Price, S. L.; Hamad, S.; Torrisi, A.; Karamertzanis, P. G.; Leslie, M.; Catlow, C. R. A. Applications Of Dl_poly And Dl_multi To Organic Molecular Crystals. *Mol. Simul.* **2006**, *32*, 985−997.
(238) Price, S. L. Predicting crystal structures of organic compounds. *Chem. Soc. Rev.* **2014**, *43*, 2098−2111.
(239) Jordan, P. C.; van Maaren, P. J.; Mavri, J.; van der Spoel, D.; Berendsen, H. J. C. Towards Phase Transferable Potential Functions: Methodology and Application to Nitrogen. *J. Chem. Phys.* **1995**, *103*, 2272−2285.
(240) Nemkevich, A.; Bürgi, H.-B.; Spackman, M. A.; Corry, B. Molecular dynamics simulations of structure and dynamics of organic molecular crystals. *Phys. Chem. Chem. Phys.* **2010**, *12*, 14916−14929.
(241) Schmidt, L.; van der Spoel, D.; Walz, M.-M. Probing phase transitions in organic crystals using atomistic MD simulations. *ACS Phys. Chem. Au* **2022**, DOI: 10.1021/acsphyschemau.2c00045.
(242) Gavezzotti, A. Dynamic simulation of orientational disorder in organic crystals: methyl groups, trifluoromethyl groups and whole molecules. *Act. Crystallogr. B* **2022**, *78*, 333−343.
(243) Otero-de-la Roza, A.; Johnson, E. R. A benchmark for non-covalent interactions in solids. *J. Chem. Phys.* **2012**, *137*, 054103.
(244) Reilly, A. M.; Tkatchenko, A. Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals. *J. Chem. Phys.* **2013**, *139*, 024705.
(245) Nyman, J.; Pundyke, O. S.; Day, G. M. Accurate force fields and methods for modelling organic molecular crystals at finite temperatures. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15828−15837.
(246) Teuteberg, T. L.; Eckhoff, M.; Mata, R. A. A full additive QM/MM scheme for the computation of molecular crystals with extension to many-body expansions. *J. Chem. Phys.* **2019**, *150*, 154118.
(247) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *J. Phys. Chem.* **1996**, *100*, 7352−7360.
(248) Bernardes, C. E. S.; Joseph, A. Evaluation of the OPLS-AA Force Field for the Study of Structural and Energetic Aspects of Molecular Organic Crystals. *J. Phys. Chem. A* **2015**, *119*, 3023−3034.
(249) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.
(250) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 73005.
(251) Schmidt, L.; van der Spoel, D.; Walz, M.-M.*Molecular Dynamics Benchmark*. https://github.com/dspoel/MDBenchmark, 2022 (Date accessed 2022-11-02).
(252) Allen, M. P.; Tildesley, D. J.*Computer Simulation of Liquids*; Oxford Science Publications: Oxford, 1987.
(253) van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. C. A systematic study of water models for molecular simulation. *J. Chem. Phys.* **1998**, *108*, 10220−10230.
(254) van Maaren, P. J.; van der Spoel, D. Molecular dynamics simulations of water with a novel shell-model potential. *J. Phys. Chem. B* **2001**, *105*, 2618−2626.
(255) van der Spoel, D.; van Maaren, P. J.; Caleman, C. GROMACS molecule & liquid database. *Bioinformatics* **2012**, *28*, 752−753.
(256) Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14* CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121*, 3864−3870.
(257) Núñez-Rojas, E.; Aguilar-Pineda, J. A.; Pérez de la Luz, A.; de Jesús González, E. N.; Alejandre, J. Force Field Benchmark of the TraPPE_UA for Polar Liquids: Density, Heat of Vaporization, Dielectric Constant, Surface Tension, Volumetric Expansion Coefficient, and Isothermal Compressibility. *J. Phys. Chem. B* **2018**, *122*, 1669−1678.
(258) Zubillaga, R. A.; Labastida, A.; Cruz, B.; Martínez, J. C.; Sánchez, E.; Alejandre, J. Surface Tension of Organic Liquids Using the OPLS/AA Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 1611−1615.
(259) van der Spoel, D.; Henschel, H.; van Maaren, P. J.; Ghahremanpour, M. M.; Costa, L. T. A potential for molecular simulation of compounds with linear moieties. *J. Chem. Phys.* **2020**, *153*, 084503.
(260) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(261) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8592.
(262) Wennberg, C. L.; Murtola, T.; Hess, B.; Lindahl, E. Lennard-Jones Lattice Summation in Bilayer Simulations Has Critical Effects on

Surface Tension and Lipid Properties. *J. Chem. Theory Comput.* **2013**, *9*, 3527−3537.
(263) Ghoufi, A.; Malfreyt, P.; Tildesley, D. J. Computer modelling of the surface tension of the gas-liquid and liquid-liquid interface. *Chem. Soc. Rev.* **2016**, *45*, 1387−1409.
(264) Bashardanesh, Z.; Lötstedt, P. Efficient Green's Function Reaction Dynamics (GFRD) simulations for diffusion-limited, reversible reactions. *J. Comput. Phys.* **2018**, *357*, 78−89.
(265) Katritzky, A. R.; Oliferenko, A. A.; Oliferenko, P. V.; Petrukhin, R.; Tatham, D. B.; Maran, U.; Lomaka, A.; Acree, W. E. A general Treatment of Solubility. 1. The QSPR Correlation of Solvation Free Energies of Single Solutes in Series of Solvents. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1794−1805.
(266) Katritzky, A. R.; Oliferenko, A. A.; Oliferenko, P. V.; Petrukhin, R.; Tatham, D. B.; Maran, U.; Lomaka, A.; Acree, W. E. A General Treatment of Solubility. 2. QSPR Prediction of Free Energies of Solvation of Specified Solutes in Ranges of Solvents. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1806−1814.
(267) Katritzky, A. R.; Tulp, I.; Fara, D. C.; Lauria, A.; Maran, U.; Acree, W. E. A General Treatment of Solubility. 3. Principal Component Analysis (PCA) of the Solubilities of Diverse Solutes in Diverse Solvents. *J. Chem. Inf. Model.* **2005**, *45*, 913−923.
(268) Sweere, A. J. M.; Fraaije, J. G. Accuracy Test of the OPLS-AA Force Field for Calculating Free Energies of Mixing and Comparison with PAC-MAC. *J. Chem. Theory Comput.* **2017**, *13*, 1911−1923.
(269) Kashefolgheta, S.; Oliveira, M. P.; Rieder, S. R.; Horta, B. A. C.; Acree, W. E.; Hünenberger, P. H. Evaluating Classical Force Fields against Experimental Cross-Solvation Free Energies. *J. Chem. Theory Comput.* **2020**, *16*, 7556−7580.
(270) Kashefolgheta, S.; Wang, S.; Acree, W. E.; Hünenberger, P. H. Evaluation of nine condensed-phase force fields of the GROMOS, CHARMM, OPLS, AMBER, and OpenFF families against experimental cross-solvation free energies. *Phys. Chem. Chem. Phys.* **2021**, *23*, 13055−13074.
(271) Bannan, C. C.; Calabró, G.; Kyu, D. Y.; Mobley, D. L. Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water. *J. Chem. Theory Comput.* **2016**, *12*, 4015−4024.
(272) van der Spoel, D.; Lindahl, E. Brute-Force Molecular Dynamics Simulations of Villin Headpiece: Comparison with NMR Parameters. *J. Phys. Chem. B* **2003**, *107*, 11178−11187.
(273) Lange, O. F.; van der Spoel, D.; de Groot, B. L. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophys. J.* **2010**, *99*, 647−655.
(274) Antila, H. S.; M. Ferreira, T.; Ollila, O. H. S.; Miettinen, M. S. Using Open Data to Rapidly Benchmark Biomolecular Simulations: Phospholipid Conformational Dynamics. *J. Chem. Inf. Model.* **2021**, *61*, 938−949.
(275) Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8*, 1409−1414.
(276) Li, N.; Gao, Y.; Qiu, F.; Zhu, T. Benchmark Force Fields for the Molecular Dynamic Simulation of G-Quadruplexes. *Molecules* **2021**, *26*, 5379.
(277) Minhas, V.; Sun, T.; Mirzoev, A.; Korolev, N.; Lyubartsev, A. P.; Nordenskiöld, L. Modeling DNA Flexibility: Comparison of Force Fields from Atomistic to Multiscale Levels. *J. Phys. Chem. B* **2020**, *124*, 38−49.
(278) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS One* **2012**, *7*, e32131.
(279) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, E4758−E4766.
(280) Wang, L.; O'Mara, M. L. Effect of the Force Field on Molecular Dynamics Simulations of the Multidrug Efflux Protein P-Glycoprotein. *J. Chem. Theory Comput.* **2021**, *17*, 6491−6508.
(281) Vangaveti, S.; Ranganathan, S. V.; Chen, A. A. Advances in RNA molecular dynamics: a simulator's guide to RNA force fields. *WIREs RNA* **2017**, *8*, No. e1396.
(282) Salsbury, A. M.; Lemkul, J. A. Recent developments in empirical atomistic force fields for nucleic acids and applications to studies of folding and dynamics. *Curr. Opin. Struct. Biol.* **2021**, *67*, 9−17.
(283) Frenkel, M.; Chirico, R. D.; Diky, V. V.; Dong, Q.; Frenkel, S.; Franchois, P. R.; Embry, D. L.; Teague, T. L.; Marsh, K. N.; Wilhoit, R. C. ThermoMLAn XML-Based Approach for Storage and Exchange of Experimental and Critically Evaluated Thermophysical and Thermochemical Property Data. 1. Experimental Data. *J. Chem. Eng. Data* **2003**, *48*, 2−13.
(284) Riccardi, D.; Trautt, Z.; Bazyleva, A.; Paulechka, E.; Diky, V.; Magee, J. W.; Kazakov, A. F.; Townsend, S. A.; Muzny, C. D. Towards improved FAIRness of the ThermoML Archive. *J. Comput. Chem.* **2022**, *43*, 879−887.
(285) Duarte Ramos Matos, G.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. *J. Chem. Eng. Data* **2017**, *62*, 1559−1569.
(286) Hossain, S.; Kabedev, A.; Parrow, A.; Bergstrom, C. A. S.; Larsson, P. Molecular simulation as a computational pharmaceutics tool to predict drug solubility, solubilization processes and partitioning. *Eur. J. Pharmac. Biopharmac.* **2019**, *137*, 46−55.
(287) Clementi, E.; Corongiu, G.; Bahattacharya, D.; Feuston, B.; Frye, D.; Preiskorn, A.; Rizzo, A.; Xue, W. Selected topics in ab initio computational chemistry in both very small and very large chemical systems. *Chem. Rev.* **1991**, *91*, 679−699.
(288) Aida, M.; Corongiu, G.; Clementi, E. Ab initio force field for simulations of proteins and nucleic acids. *Int. J. Quantum Chem.* **1992**, *42*, 1353−1381.
(289) Smith, D. G. A.; Altarawy, D.; Burns, L. A.; Welborn, M.; Naden, L. N.; Ward, L.; Ellis, S.; Pritchard, B. P.; Crawford, T. D. The MolSSI QCArchive project: An open-source platform to compute, organize, and share quantum chemistry data. *WIREs Comput. Mol. Sci.* **2021**, *11*, No. e1491.
(290) Champagne, B.; Perpête, E. A.; van Gisbergen, S. J. A.; Baerends, E.-J.; Snijders, J. G.; Soubra-Ghaoui, C.; Robins, K. A.; Kirtman, B. Assessment of conventional density functional schemes for computing the polarizabilities and hyperpolarizabilities of conjugated oligomers: An ab initio investigation of polyacetylene chains. *J. Chem. Phys.* **1998**, *109*, 10489−10498.
(291) Limacher, P. A.; Mikkelsen, K. V.; Lüthi, H. P. On the accurate calculation of polarizabilities and second hyperpolarizabilities of polyacetylene oligomer chains using the CAM-B3LYP density functional. *J. Chem. Phys.* **2009**, *130*, 194114.
(292) Huzak, M.; Deleuze, M. S. Benchmark theoretical study of the electric polarizabilities of naphthalene, anthracene, and tetracene. *J. Chem. Phys.* **2013**, *138*, 024319.
(293) Swann, E. T.; Fernandez, M.; Coote, M. L.; Barnard, A. S. Bias-Free Chemically Diverse Test Sets from Machine Learning. *ACS Comb. Sci.* **2017**, *19*, 544−554.
(294) Wilkinson, M.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018.
(295) Zhang, H.; Jiang, Y.; Yan, H.; Yin, C.; Tan, T.; van der Spoel, D. Free Energy Calculations of Ionic Hydration Consistent with the Experimental Hydration Free Energy of the Proton. *J. Phys. Chem. Lett.* **2017**, *8*, 2705−2712.
(296) Sun, H.; Ren, P.; Fried, J. The COMPASS force field: parameterization and validation for phosphazenes. *Comput. Theor. Polymer Sci.* **1998**, *8*, 229−246.
(297) Njo, S. L.; van Gunsteren, W. F.; Müller-Plathe, F. Determination of force field parameters for molecular simulation by molecular simulation: An application of the weak-coupling method. *J. Chem. Phys.* **1995**, *102*, 6199−6207.
(298) Di Pierro, M.; Elber, R. Automated Optimization of Potential Parameters. *J. Chem. Theory Comput.* **2013**, *9*, 3311−3320.

https://doi.org/10.1021/acs.jcim.2c01127
J. Chem. Inf. Model. 2023, 63, 412−431

(299) Di Pierro, M.; Mugnai, M. L.; Elber, R. Optimizing Potentials for a Liquid Mixture: A New Force Field for a tert-Butanol and Water Solution. *J. Phys. Chem. B* **2015**, *119*, 836−849.
(300) Stroet, M.; Koziara, K. B.; Malde, A. K.; Mark, A. E. Optimization of Empirical Force Fields by Parameter Space Mapping: A Single-Step Perturbation Approach. *J. Chem. Theory Comput.* **2017**, *13*, 6201−6212.
(301) Naden, L. N.; Shirts, M. R. Rapid Computation of Thermodynamic Properties over Multidimensional Nonbonded Parameter Spaces Using Adaptive Multistate Reweighting. *J. Chem. Theory Comput.* **2016**, *12*, 1806−1823.
(302) Messerly, R. A.; Razavi, S. M.; Shirts, M. R. Configuration-Sampling-Based Surrogate Models for Rapid Parameterization of Non-Bonded Interactions. *J. Chem. Theory Comput.* **2018**, *14*, 3144−3162.
(303) Oliveira, M. P.; Andrey, M.; Rieder, S. R.; Kern, L.; Hahn, D. F.; Riniker, S.; Horta, B. A. C.; Hünenberger, P. H. Systematic Optimization of a Fragment-Based Force Field against Experimental Pure-Liquid Properties Considering Large Compound Families: Application to Saturated Haloalkanes. *J. Chem. Theory Comput.* **2020**, *16*, 7525−7555.
(304) P. Oliveira, M.; Hünenberger, P. H. Systematic optimization of a fragment-based force field against experimental pure-liquid properties considering large compound families: application to oxygen and nitrogen compounds. *Phys. Chem. Chem. Phys.* **2021**, *23*, 17774−17793.
(305) Moult, J. A. decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* **2005**, *15*, 285−289.
(306) Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. Critical assessment of methods of protein structure prediction (CASP) - Round XIV. *Proteins: Struct. Funct. Bioinf.* **2021**, *89*, 1607−1617.
(307) Nassar, R.; Dignon, G. L.; Razban, R. M.; Dill, K. A. The Protein Folding Problem: The Role of Theory. *J. Mol. Biol.* **2021**, *433*, 167126.
(308) Jana, K.; Kepp, K. P. Force-Field Benchmarking by Alternatives: A Systematic Study of Ten Small α- and β-Proteins. *bioRxiv* **2020**, DOI: 10.1101/2020.03.03.974477.
(309) *Industrial Fluid Properties Challenge*. http://www.fluidproperties.org, 2006.
(310) Tielker, N.; Eberlein, L.; Hessler, G.; Schmidt, K. F.; Güssregen, S.; Kast, S. M. Quantum-mechanical property prediction of solvated drug molecules: what have we learned from a decade of SAMPL blind prediction challenges? *J. Comput. Aid. Mol. Des.* **2021**, *35*, 453−472.
(311) Grosjean, H.; Isik, M.; Aimon, A.; Mobley, D.; Chodera, J.; von Delft, F.; Biggin, P. C. SAMPL7 protein-ligand challenge: A community-wide evaluation of computational methods against fragment screening and pose-prediction. *J. Comput. Aid. Mol. Des.* **2022**, *36*, 291−311.
(312) Oliveira, M. P.; Hünenberger, P. H. Force fields optimized against experimental data for large compound families using CombiFF: Validation considering non-target properties and polyfunctional compounds. *J Mol Graph Model* . **2023**, *118*, 108312.
(313) Wensink, E. J. W.; Hoffmann, A. C.; van Maaren, P. J.; van der Spoel, D. Dynamic Properties of Water/Alcohol Mixtures Studied by Computer Simulation. *J. Chem. Phys.* **2003**, *119*, 7308−7317.
(314) van der Spoel, D.; van Maaren, P. J.; Larsson, P.; Tîmneanu, N. Thermodynamics of hydrogen bonding in hydrophilic and hydrophobic media. *J. Phys. Chem. B* **2006**, *110*, 4393−4398.
(315) Zhang, H.; Tan, T.; Feng, W.; van der Spoel, D. Molecular recognition in different environments: β-cyclodextrin dimer formation in organic solvents. *J. Phys. Chem. B* **2012**, *116*, 12684−93.
(316) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*−*2*, 19−25.

Recommended by ACS

Computing the Volume, Surface Area, Mean, and Gaussian Curvatures of Molecules and Their Derivatives
Patrice Koehl, Herbert Edelsbrunner, *et al.*
JANUARY 13, 2023
JOURNAL OF CHEMICAL INFORMATION AND MODELING READ

Insight into the Role of Active Site Protonation States and Water Molecules in the Catalytic Inhibition of DPP4 by Vildagliptin
Jeisson D. Corredor, Verónica A. Jiménez, *et al.*
FEBRUARY 09, 2023
JOURNAL OF CHEMICAL INFORMATION AND MODELING READ

Treating Polarization Effects in Charged and Polar Bio-Molecules Through Variable Electrostatic Parameters
Qiang Zhu, Jing Ma, *et al.*
JANUARY 02, 2023
JOURNAL OF CHEMICAL THEORY AND COMPUTATION READ

What is the Optimal Dipole Moment for Nonpolarizable Models of Liquids?
Miguel Jorge, Daniel J. Cole, *et al.*
FEBRUARY 24, 2023
JOURNAL OF CHEMICAL THEORY AND COMPUTATION READ

Get More Suggestions >